SYM-AM-18-033

Proceedings

of the

Fifteenth Annual Acquisition Research Symposium

Thursday Sessions
Volume II

Acquisition Research:
Creating Synergy for Informed Change

May 9–10, 2018

March 30, 2018

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

Acquisition Research Program
Graduate School of Business & Public Policy
Naval Postgraduate School

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

Table of Contents

Welcome: James B. Greene, RADM, USN (Ret.), Chair of Acquisition, Naval Postgraduate School ... viii

Keynote: Vice Admiral David H. Lewis, USN, Director, Defense Contract Management Agency ... x

Plenary Panel 17: State of the Defense Acquisition System ... 1

Defense Acquisition Trends 2017: A Preliminary Look ... 1

Panel 18. Rapid Product Development: Variations on the Theme ... 17

The Use of COTS in Defense Acquisition Programs: A Research Synthesis and Framework ... 18

Crossing the Valley of Death: The Case of the MDUSV ... 36

Seven Tips to Support Rapid Product Deployment: Lessons Learned ... 57

Panel 19. Augmenting the Acquisition Decision Processes With Data Analytics ... 73

Informing DoD Program Planning Through the Examination of the Causes of Delays in Acquisition Using Acquisition Data ... 75

Utilizing Public Data for Data Enhancement and Analysis of Federal Acquisition Data ... 94

Toward Cognitive Supremacy via Quantitative Augmentation ... 114

Panel 20. Program Management in an Age of Complexity ... 134

Acquisition and Development Programs Through the Lens of System Complexity ... 136

Developing a Sense of Reality Within Complex Program Management Environments ... 155

Fielding Better Combat Helmets to Deploying Warfighters ... 175

Panel 21. Considerations in Accelerating Technology Adoption in Defense Acquisitions ... 193

Applying a DEVelopment OPerationS (DevOps) Reference Architecture to Accelerate Delivery of Emerging Technologies in Data Analytics, Deep Learning, and Artificial Intelligence to the Afloat U.S. Navy ... 195

Conducting Viability Assessments for Acquisition Planning ... 210

Panel 22. Enhancing Small Business Participation in Defense Markets ... 224

The Impact of 8(a) Small Business Graduation ... 226

New Entrants and Small Business Graduation in the Market for Federal Contracts ... 235

Examining Small Business Set Asides: Evidence and Implications for Small and Mid-Sized Suppliers in Federal Procurement 254

Panel 23. Designing Cybersecurity Into Acquisition Programs........ 265

Applying Cause-Effect Mapping to Assess Cybersecurity Vulnerabilities in Model-Centric Acquisition Program Environments 266

Automated Methods for Cyber Test and Evaluation 280

Cybersecurity: Converting Shock Into Action (Part 1) 289

Panel 24. Implementing and Costing Rapid Prototyping 304

Maturing Cost Estimation in a Rapid Acquisition Environment 305

Set-Based Evaluation Tool (SET): A Software Analysis Tool to Support Set-Based Decision Methods 322

Application of Set-Based Decision Methods to Accelerate Acquisition Through Tactics and Technology Exploration and Experimentation (TnTE2) 335

Panel 25. Why Causal Learning Must Be Adopted Within Acquisition Research 362

Experience Searching for Causal Factors in Personal Process Student Data ... 364

Further Causal Search Analyses With UCC's Effort Estimation Data 382

Panel 26. Enhancing Supply Chain Effectiveness 400

Literature Review: Metrics for Naval Humanitarian Assistance and Disaster Relief (HADR) Operation 401

Navy Expeditionary Logistics 426

Additional Papers 444

Additional Papers Continued 445

A Systems Theory Based Examination of Failure in Acquisition System Reform 446

Industrial White Paper Briefing: Monterey Bay Regional Spaceport Lowest Priced Technically Acceptable COTS Launch System 469

In the Fullness of Time: Towards Realistic Acquisition Schedule Estimates 472

Searching Hidden Links: Inferring Undisclosed Subcontractors From Public Contract Records and Employment Data 491

A Method for Identification, Representation, and Assessment of Complex System Pathologies in Acquisition Programs 500

A Mathematical Framework to Apply Tradespace Exploration to the Design of Verification Strategies 525

A Review of Trusted Broker Architectures for Data Sharing 537

Towards Game Theoretic Models for Agile Acquisition 551

Optimizing Contract Modifications Under One Universal Mod 573

Assessing Vulnerabilities in Model-Centric Acquisition Programs Using Cause-Effect Mapping 592

SYM-AM-18-033

Proceedings
Of the
Fifteenth Annual Acquisition Research Symposium

Thursday Sessions
Volume II

Acquisition Research:
Creating Synergy for Informed Change

May 9–10, 2018

March 30, 2018

THIS PAGE INTENTIONALLY LEFT BLANK

Welcome: James B. Greene, RADM, USN (Ret.), Chair of Acquisition, Naval Postgraduate School

THIS PAGE INTENTIONALLY LEFT BLANK

Keynote: Vice Admiral David H. Lewis, USN, Director, Defense Contract Management Agency

Vice Admiral David H. Lewis, USN, a native of the state of Washington, graduated in 1979 from the University of Nebraska with a Bachelor of Science in Computer Science and was commissioned through the Naval Reserve Officers Training Corps program. He graduated from the Naval Postgraduate School in 1988 with a Master of Science in Computer Science.

At sea, Lewis served aboard USS *Spruance* (DD 963) as communications officer, where he earned his surface warfare qualification; USS *Biddle* (CG 34) as fire control officer and missile battery officer; and USS *Ticonderoga* (CG 47) as combat systems officer. His major command assignment was Aegis Shipbuilding Program manager in Program Executive Office Ships, where he led the delivery of seven DDG 51 class ships and procured another 10 ships.

At shore, Lewis' assignments included assistant chief of staff for maintenance and engineering; commander, Naval Surface Forces; the Navy secretariat staff; Naval Sea Systems Command staff; Aegis Shipbuilding Program Office; supervisor of shipbuilding, Bath, ME; and Readiness Support Group, San Diego.

Upon selection to flag rank in 2009, Lewis served as vice commander, Naval Sea Systems Command and then served four years as program executive officer, Ships, where he directed the delivery of 18 ships and procurement of another 51 ships. From 2014–2017 he served as commander, Space and Naval Warfare Systems Command, where he led a global workforce of 10,300 civilian and military personnel who design, develop, and deploy advanced communications and information capabilities. In May 2017 he assumed command as the director, Defense Contract Management Agency, Fort Lee, VA.

Lewis' personal awards include the Distinguished Service Medal (two awards), Legion of Merit (four awards), Meritorious Service Medal (three awards), Navy and Marine Corps Commendation Medal (two awards), Navy and Marine Corps Achievement Medal, and various service and unit awards.

Plenary Panel 17: State of the Defense Acquisition System

Thursday, May 10, 2018	
9:30 a.m. – 11:00 a.m.	**Chair: The Honorable David J. Berteau,** President and CEO, Professional Services Council **Discussant: Stan Soloway,** President & CEO, Celero Strategies, LLC **Discussant: The Honorable Katrina McFarland,** Sehlke Board of Advisors Member and President of Blue Oryx LLC, Former Assistant Secretary of Defense (Acquisition) ***Defense Acquisition Trends 2017: A Preliminary Look*** Rhys McCormick, Center for Strategic and International Studies Greg Sanders, Center for Strategic and International Studies Andrew Hunter, Center for Strategic and International Studies

David J. Berteau—became the President and Chief Executive Officer of the Professional Services Council (PSC) on March 28, 2016. With nearly 400 members, PSC is the premier advocate of and resource for the federal technology and professional services industry. As CEO, Berteau focuses on legislative and regulatory issues related to government acquisition, budgets, and requirements, helping to shape public policy, lead strategic coalitions, and work to improve communications between government and industry. PSC's member companies represent small, medium, and large businesses that provide federal agencies with services of all kinds, including engineering, logistics, operations and maintenance, information technology, facilities management, international development, scientific, and environmental services.

Prior to PSC, Berteau was confirmed in December 2014 as the Assistant Secretary of Defense for Logistics and Materiel Readiness. He managed logistics policy and processes to provide superior, cost effective, joint logistics support to the entire Department of Defense. He oversaw the management of the $170 billion in Department of Defense logistics operations.

Earlier, Berteau served as Senior Vice President and Director of the National Security Program on Industry and Resources at the Center for Strategic and International Studies (CSIS) in Washington, DC. His research and analysis covered national security, management, contracting, logistics, acquisition, and industrial base issues. Berteau is a Fellow of the National Academy of Public Administration and has previously served as an adjunct professor at Georgetown University and at the Lyndon B. Johnson School of Public Affairs, a Director of the Procurement Round Table, and an Associate at the Robert S. Strauss Center at the University of Texas.

Before he joined CSIS full time in 2008, he served as a CSIS non-resident Senior Associate for seven years. In addition, he was director of national defense and homeland security for Clark & Weinstock, director of Syracuse University's National Security Studies Program and a professor of practice at the Maxwell School of Citizenship and Public Affairs, and senior vice president at Science Applications International Corporation (SAIC). He served a total of 14 years at senior levels in the U.S. Defense Department under six defense secretaries.

Berteau graduated with a BA from Tulane University in 1971 and received his master's degree in 1981 from the LBJ School of Public Affairs at the University of Texas.

Stan Soloway—is President & CEO of Celero Strategies, LLC, a full-service strategic consultancy focused on the federal market. Celero Strategies is Soloway's latest step in a career during which he has become widely regarded as one of the nation's leading experts on the federal market, the factors and dynamics that drive it and how to translate that expertise into meaningful strategies and action. With Celero, Soloway's goal is to combine two core passions: helping good companies bring innovative solutions to government and helping government significantly improve its performance and delivery of service.

Prior to founding Celero Strategies in January, 2016, Stan served for 15 years as the President & CEO of the Professional Services Council, the largest and most influential national association of government technology and professional services firms. While at PSC, Soloway was the industry's leading voice, policy strategist and resource for both government and the private sector. He regularly testified before Congress, was a prolific writer, appeared often on radio and television; and was routinely sought out by both corporate and government organizations to discuss current market trends, dynamics and strategies. He has also been a contributing author for books published by Cambridge University, Harvard Law School, the University of Pennsylvania, and the IBM Center for the Business of Government.

Stan was the recipient of the 2016 Consumer Electronics Show (CES) Government Technology Leadership Award and, in 2015, was inducted into the Greater Washington Government Contractor Hall of Fame. He also was named the IT Industry Executive of the Year in 2013 by Government Computer News; has regularly been named one of the 100 most influential business leaders in Washington (Washington Business Journal) and one of the 100 most influential figures in national defense (Defense News and Gannett). He is a three-time winner of the Federal 100 Award for his leadership in federal information technology and is a Fellow of both the National Academy of Public Administration and the National Contract Management Association, where he also serves on the Executive Advisory Board. He is a principal at the Partnership for Public Service where he serves as a Senior Advisor to Government Executives (SAGE) and in 2016 was appointed to the Community Advisory Board of WAMU Radio, Washington, DC's National Public Radio outlet.

During the second half of the Clinton Administration, Stan served as the Deputy Undersecretary of Defense and was responsible for wide-ranging reforms to defense acquisition and technology policy and practices, and broader department-wide re-engineering. In recognition of his leadership in the department, Stan was awarded both the Secretary of Defense Medal for Exceptional Public Service and the Secretary of Defense Medal for Distinguished Public Service.

As passionate believer in the importance and value of public service, Stan also served from 2007 to 2013 as a Senate-confirmed member of the Board of Directors of the Corporation for National and Community Service, the federal agency that oversees AmeriCorps and other national service programs, and is a major source of funding and leadership for community service organizations across the nation. Earlier in his career he was a public policy and public affairs consultant for nearly 20 years. He also co-produced the acclaimed PBS series "Great Confrontations at the Oxford Union."

Katrina McFarland—is currently on the Sehlke Board of Advisors and President of Blue Oryx, LLC, and was formerly the Acting Assistant Secretary of the Army (Acquisition, Logistics, & Technology) following designation by President Barack Obama on February 1, 2016.

As Assistant Secretary of the Army (Acquisition, Logistics, & Technology) and Army Acquisition Executive, McFarland oversaw the execution of the Army's acquisition function, including life cycle management and sustainment of Army weapons systems and research and development programs, and manages the Army Acquisition Corps and greater Army Acquisition Workforce. McFarland also serves as the science advisor to the Secretary of the Army and as the Army's senior research and development official and senior procurement executive. In addition, McFarland held principal responsibility for all Department of the Army matters related to logistics.

Prior to joining the Department of the Army, McFarland served as the Assistant Secretary of Defense (Acquisition). In this role, she was the principal adviser to the Secretary of Defense and the Under Secretary of Defense for Acquisition, Technology, & Logistics on matters related to acquisition. Previously, she served as the President of the Defense Acquisition University (DAU). Under her leadership, DAU provided practitioner training, career management, and services to enable the acquisition, technology, logistics, and requirements community to make smart business decisions and deliver timely and affordable capabilities to the warfighter. Prior to joining DAU, McFarland was the Director for Acquisition for the Missile Defense Agency (MDA)—a position she held since May 2006. As MDA's principal acquisition executive, McFarland advised the Director of MDA on all acquisition, contracting, and small business decisions. Other core responsibilities included the development of process activities and program policy associated with the execution of the single integrated Ballistic Missile Defense System research, development, and test program, and the establishment of the Baseline Execution Review to ensure that an integrated program execution of the BMDS occurred across the baselines of schedule, cost, performance, contracting, test, and operational delivery.

McFarland began her civil service career in 1986 as a general engineer at Headquarters Marine Corps where she was accredited as a Materials, Mechanical, Civil, and Electronics Engineer. She has received an Honorary Doctoral of Engineering from the University of Cranfield, United Kingdom, the Presidential Meritorious Executive Rank Award, the Secretary of Defense Medal for Meritorious Civilian Service Award, the Department of the Navy Civilian Tester of the Year Award, and the Navy and United States Marine Corps Commendation Medal for Meritorious Civilian Service. She is DAWIA Level-III certified in Program Management, Engineering, and Testing as well as having a professional engineer license and having attained her PMP certification.

Defense Acquisition Trends 2017: A Preliminary Look

Rhys McCormick—is an Associate Fellow with the Defense-Industrial Initiatives Group (DIIG) at CSIS. His work focuses on unmanned systems, global defense industrial base issues, and U.S. federal and defense contracting trends. Prior to working at DIIG, he interned at the Abshire-Inamori Leadership Academy at CSIS and the Peacekeeping and Stability Operations Institute at the U.S. Army War College. He holds a bachelor's degree in security and risk analysis from Pennsylvania State University and a master's degree in security studies from Georgetown University.

Greg Sanders—is a Fellow in the International Security Program and Deputy Director of the Defense-Industrial Initiatives Group at CSIS, where he manages a research team that analyzes data on U.S. government contract spending and other budget and acquisition issues. In support of these goals, he employs SQL Server, as well as the statistical programming language R. Sanders holds a master's degree in international studies from the University of Denver, and he holds a bachelor's degree in government and politics and a bachelor's degree in computer science from the University of Maryland.

Andrew Hunter—is a Senior Fellow in the International Security Program and Director of the Defense-Industrial Initiatives Group at CSIS. From 2011 to 2014, he served as a senior executive in the Department of Defense, serving first as Chief of Staff to Under Secretaries of Defense (AT&L) Ashton B. Carter and Frank Kendall, before directing the Joint Rapid Acquisition Cell. From 2005 to 2011, Hunter served as a Professional Staff Member of the House Armed Services Committee. Hunter holds a master's degree in applied economics from Johns Hopkins University and a bachelor's degree in social studies from Harvard University.

Abstract

This paper presents a preliminary look at the Fiscal Year (FY) 2017 Department of Defense (DoD) contracting trends available in the Federal Procurement Data System (FPDS). This data provides important insights concerning the defense industrial base through analysis of contract characteristics such as defense component, area (products, services, R&D), component, level of competition, platform portfolio, and vendor size. These trends provide vital information that can inform and highlight critical issues in the defense industrial base, such as the historical trough in development pipeline for major weapon systems. Given that FY 2016 was the end of seven consecutive years of DoD contract obligation drawdown, the trends for FY 2017 are particularly interesting.

Introduction

This paper presents a preliminary look at the Fiscal Year (FY) 2017 Department of Defense (DoD) contracting trends available in the Federal Procurement Data System (FPDS). This data provides important insights concerning the defense industrial base through analysis of contract characteristics such as defense component, area (products, services, R&D), component, level of competition, platform portfolio, and vendor size. These trends provide vital information that can inform and highlight critical issues in the defense industrial base, such as the historical trough in the development pipeline for major weapon systems. Given that FY 2016 was the end of seven consecutive years of DoD contract obligation drawdown, the trends for FY 2017 are particularly interesting.

This report uses the methodology used in CSIS reports on federal contracting. For over a decade, the Defense-Industrial Initiatives Group (DIIG) has issued a series of analytical reports on federal contract spending for national security by the government. These reports are built on FPDS data, which is downloaded in bulk from USAspending.gov. DIIG now maintains its own database of federal spending, that includes data from 1990–2017. This database is a composite of FPDS and DD350 data. For this report, the study team relied on FY 2000–FY 2017 data. All dollar figures are in constant FY 2017 dollars, using the latest Treasury deflators. For additional information about the CSIS contracting data analysis methodology, see https://csis.org/program/methodology.

For this paper, CSIS focused on the following research questions identified in previous DIIG defense contracting reports:

- DoD Contract Spending in a Budgetary Context: How has the defense contracting topline responded to the recent increases in the defense budget?
- Area: Have the different areas (products, services, and research and development) responded differently to the defense contracting rebound?
- Vendor Size: How did the share of contract obligations change among vendors of differing sizes, particularly small vendors?
- Competition: Did the share of contract obligations awarded after effective competition change?[1]
- R&D: Has the seven-year trough in the development pipeline for major weapon systems continued in FY 2017?

DoD Contract Spending in a Budgetary Context

Figure 1 shows that overall DoD contract obligations continued to grow in FY 2017 as the overall defense budgete increased. Total DoD contract obligations increased from $304.1 billion in FY 2016 to $319.8 billion in FY 2017, a 5% increase. Since DoD contracting obligations bottomed out in FY 2015, overall DoD contract obligations have increased by 13% over the past two years. Overall DoD contract obligations have increased as a share of DoD Total Obligation Authority (TOA) over the past two years, going from 48% in FY 2015 to 51% in FY 2016 and 53% in FY 2017. With the defense budget set to increase in FY 2018 and FY 2019, defense contract obligations are likely to continue to grow in the near future.

[1] *Effective competition* is defined as competitively sourced contracts receiving at least two or more offers.

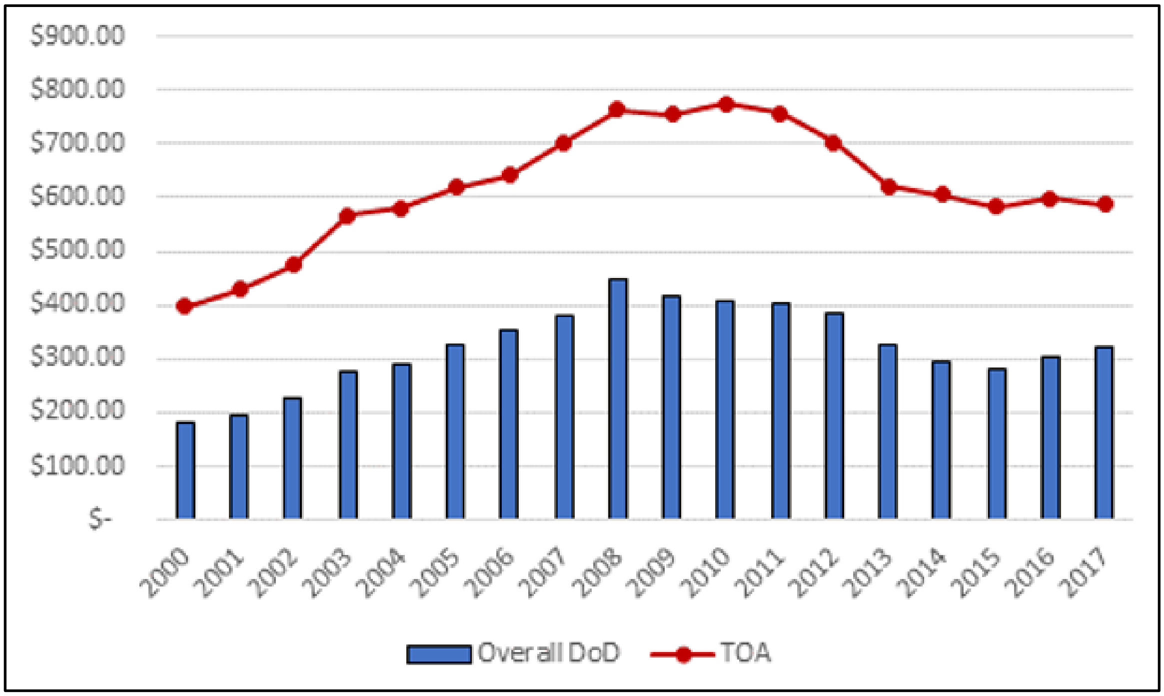


Figure 1. Defense Contract Obligations and Total Obligational Authority, 2000–2017
(FPDS; OUSD[C], 2017; CSIS analysis)

Defense Contracting Spent by Area

Within the overall DoD contracting porfolio, contract obligations for Products has increased faster than either Services or Research and Development (R&D). In 2017, overall DoD Products obligations increased by 8%, compared to the 3% growth in both Services and R&D. Since 2015, overall DoD Products contract obligations have increased by 22% compared to the 6% increase in overall DoD R&D contract obligations and the 5% increase in overall DoD Services contract obligations.

Over the past two years, there have been notable shifts in the overall DoD contract portfolio as a share of overall DoD contract obligations. Across all of the DoD, the share of average contract obligations going to Products increased to 50% in FY 2016 and 51% in FY 2017. Previously, Products had averaged 46% of overall DoD contract obligations since FY 2000. Meanwhile, the share of overall DoD contract obligations for Services declined from 44% in FY 2015 to 41% in FY 2017. Over the past two years, the share of overall DoD R&D contract obligations held steady at 8%.

Figure 2 shows defense contract obligations by area from FY 2000 to FY 2017.

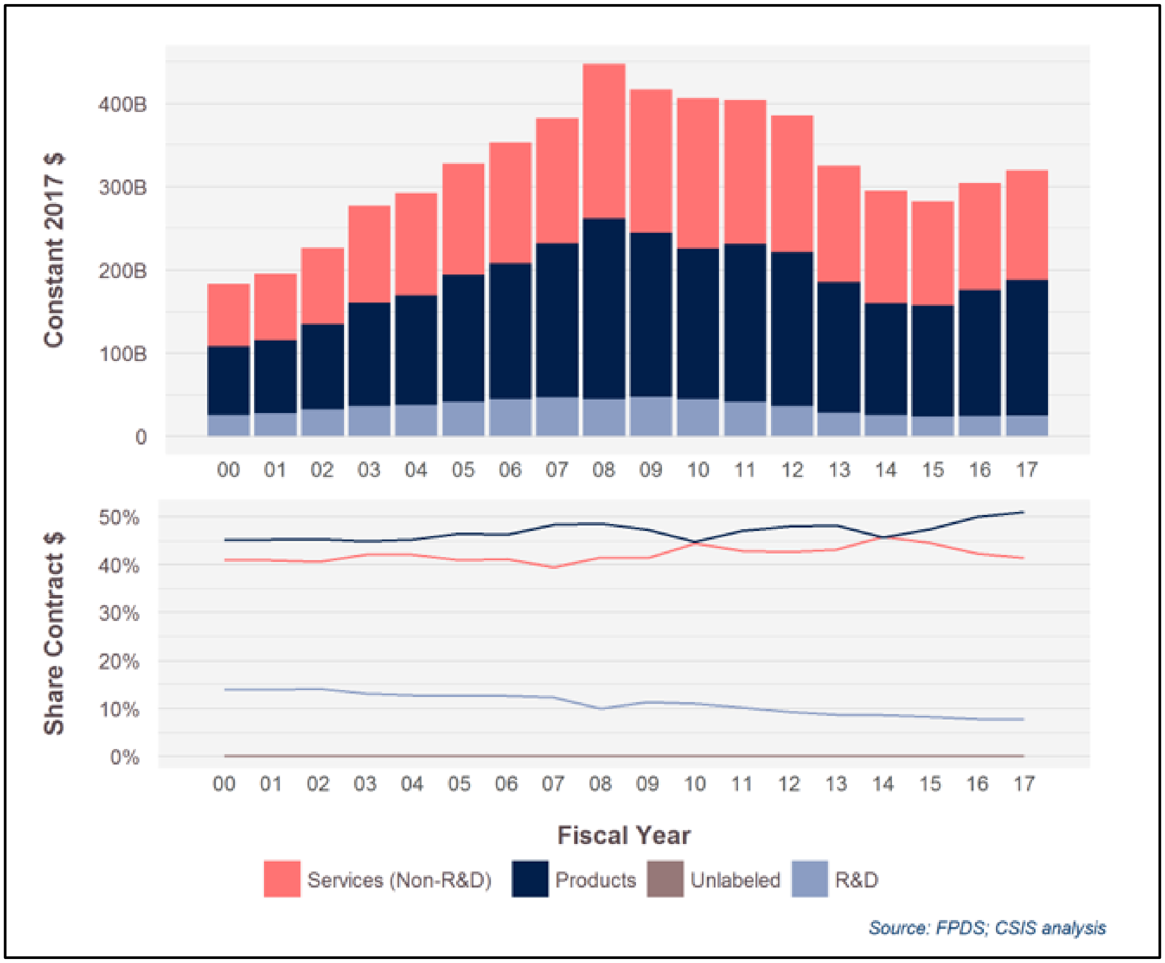


Figure 2. Defense Contract Obligations by Area, 2000–2017
(FPDS; CSIS analysis)

Overall DoD: Stage of R&D

Previous CSIS research showed a seven-year trough in the development pipeline for major weapon systems. From FY 2009 to FY 2015 overall DoD contract obligations for Advanced Technology Development (6.3) and System Development & Demonstration (6.5) declined by two-thirds as notable MDAPs were either canceled or matured into production (Hunter et al., 2017). The eight-year trough in the development pipeline for major weapon systems continued into FY 2017, but there are signs that the trough might have reached rock-bottom. For the first time since FY 2005, Defense System Development & Demonstration (6.5) contract obligations grew compared to the previous year. Defense System Development & Demonstration (6.5) contract obligations grew 11% in FY 2017, increasing from $3.8 billion in FY 2016 to $4.2 billion. Defense Advanced Technology Development (6.3) contract obligations increased 3% from $4.04 billion in FY 2016 to $4.17 billion in FY 2017.

Although Advanced Technology Development (6.3) and System Development & Demonstration (6.5) contract obligations are still at near-historic lows, the 3% and 11% growths respectively in FY 2017 are positive signs that the bleeding has stopped for now.

Advanced Component Development & Prototype (6.4) contract obligations in FY 2017 grew 25% from FY 2015. This rate constitutes a significantly higher rate of growth than the 6% overall growth of defense R&D between FY 2015 and FY 2017. Advanced Component Development & Prototype (6.4) grew 3% in FY 2017—a significantly lower rate of growth when compared to the 22% increase in FY 2016. As a share of the defense R&D portfolio, Advanced Component Development & Prototype (6.4) rose from 17% in FY 2015 to 21% in FY 2017 and are now the second largest R&D category after Applied Research (6.2).

Basic Research (6.1) contract obligations declined slightly in FY 2017 (-2%), but are still higher than in FY 2015. Applied Research (6.2) contract obligations, the largest share of the defense R&D portfolio (28%), grew 7% in FY 2016 and 1% in FY 2017.

Operational Systems Development (6.7) grew 1% in FY 2017 after declining 13% in FY 2016.

Figure 3 shows defense R&D contract obligations by stage of R&D from FY 2000 to FY 2017.

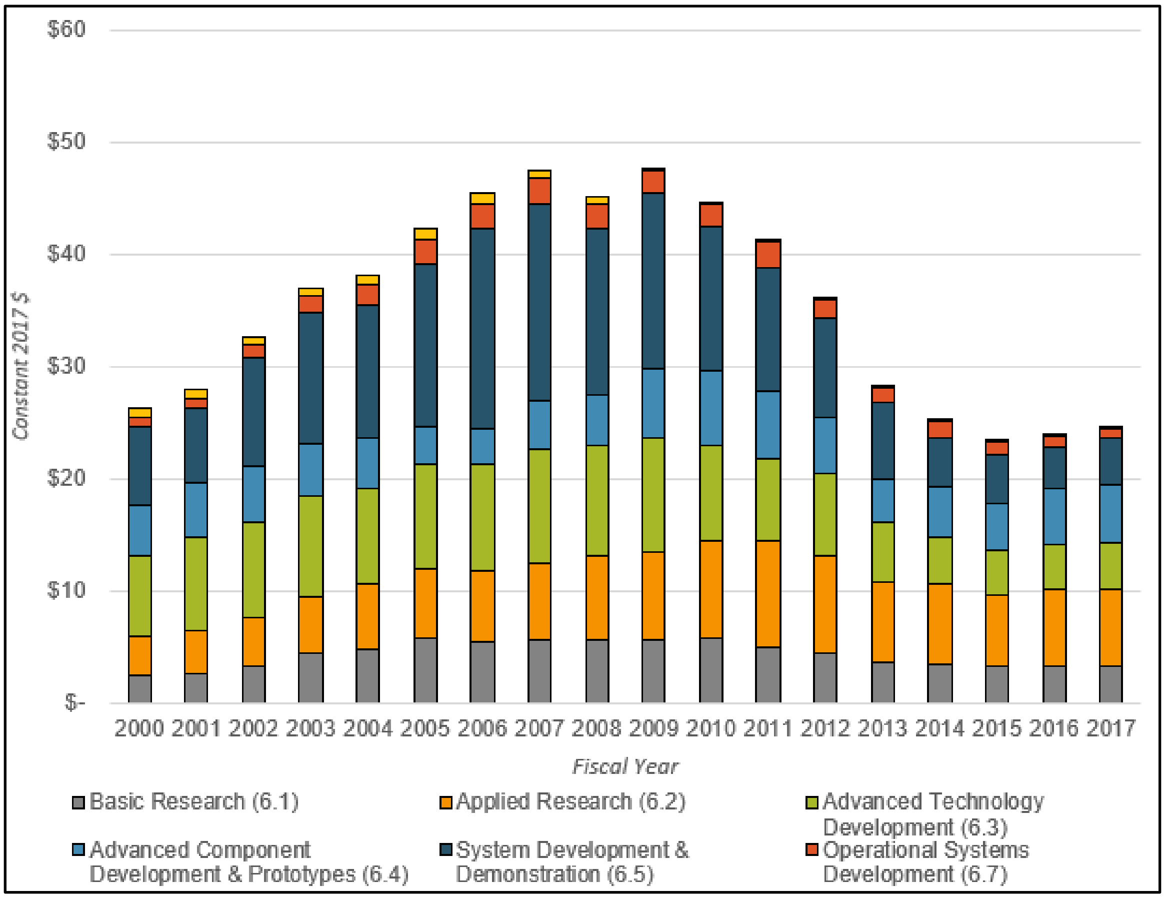


Figure 3. Defense R&D Contract Obligations by Stage of R&D, 2000–2017
(FPDS; CSIS analysis)

Overall DoD: Component

Over the past two years as defense contracting has rebounded, the trends between the defense components has varied significantly despite total contract obligations within each component rising since FY 2015. Figure 4 shows Defense Contract obligations by component from FY 2000 to FY 2017.

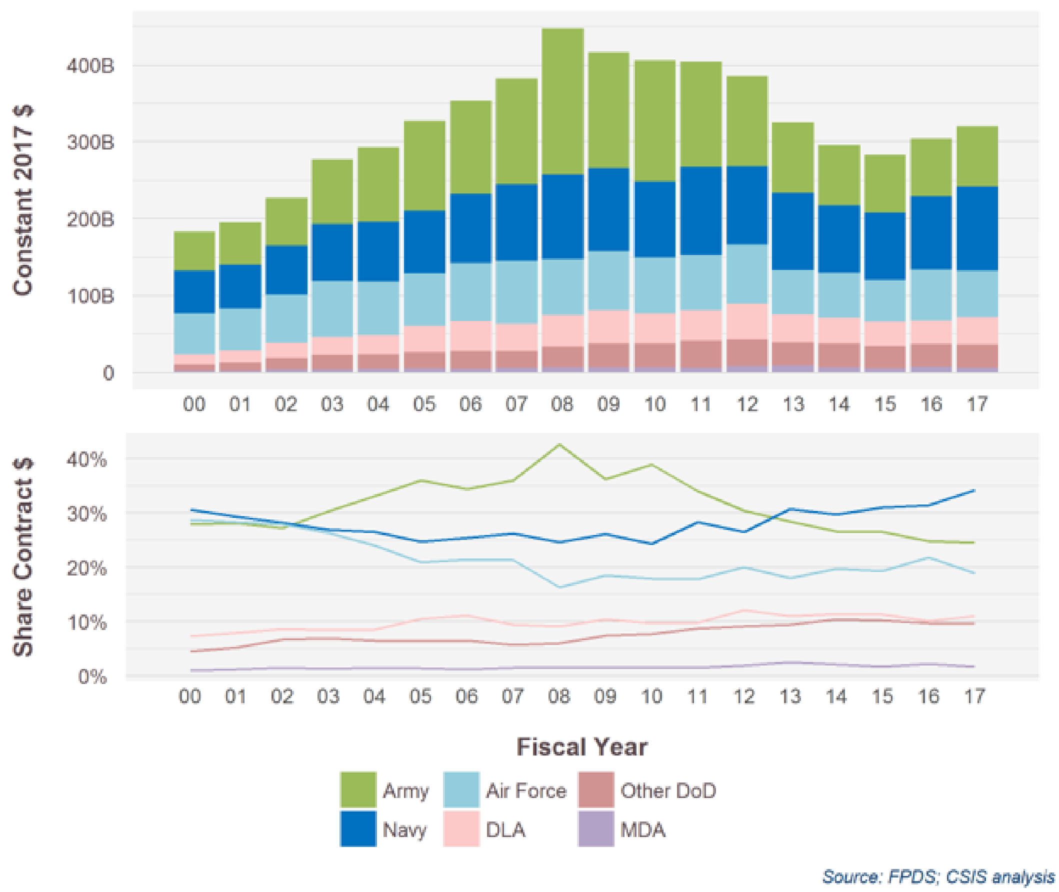


Figure 4. Defense Contract Obligations by Component, 2000–2017
(FPDS; CSIS analysis)

Air Force contract obligations have increased by 11% since FY 2015, but there has been a significant whipsaw effect over the past two years in the Air Force. In FY 2016, Air Force contract obligations increased by 21%, rising from $54.6 billion in FY 2015 to $66.3 billion in FY 2016. However, in FY 2017 Air Force contract obligations declined 9%, falling to $60.6 billion, a total slightly above drawdown levels, but still 21% lower than the Air Force's $77.1 billion in FY 2012.

The Army has seen a gradual increase in contract obligations over the past two years, but below the overall rate of growth experienced by the DoD as a whole. In FY 2016, Army contract obligations only grew 1%, compared to overall DoD contract obligations increasing by 8% that year. In FY 2017, Army contract obligations increased by 4%, a rate just below the 5% rate of overall growth. As the Army seeks to accelerate its modernization program that effort will require continued steady contracting growth in the near-term.

Navy contract obligations have grown 25% over the past two years and have rebounded to pre-drawdown levels. In FY 2016, Navy contract obligations grew 9%, a rate just above the overall rate of growth of contract obligations. In FY 2017, Navy contract obligations grew at a rate significantly higher than the overall rate of growth. Navy contract obligations grew 25% from $95.3 billion in FY 2016 to $109.4 billion in FY 2017. Of note, Navy contract obligations in FY 2017 were 7% higher than they were in the FY 2012, the only component of DoD to not only rebound to pre-drawdown levels but to exceed these levels.

MDA and DLA both experienced a whiplash effect between FY 2016 and FY 2017, but in opposite directions. In FY 2016, DLA contract obligations declined by 3% before increasing 15% in FY 2017. Meanwhile, MDA contract obligations increased 39% in FY 2016, before declining 19% in FY 2017.

Overall DoD: Platform Portfolio

Except for the Air & Missile Defense, Facilities and Construction, Other Products, and Space Systems, contract obligations are up across platforms portfolios since FY 2015.

Land Vehicles contract obligations increased 10% in FY 2017 after suffering "catastrophic" declines during sequestration and the defense drawdown (McCormick, Hunter, & Sanders, 2017). Land Vehicles contract obligations rose from $7.5 billion in FY 2016 to $8.2 billion in FY 2017. The 10% increase was slightly offset by the 3% decline in FY 2016, but Land Vehicles contract obligations are up 7% from their low point in FY 2015—still well below historical averages.

Ships & Submarines and Air & Missile Defense saw the smallest decline in contract obligations during sequestration and the defense drawdown but have faced very different trajectories since. Over the past two years, Ships & Submarines have grown at a steady rate, increasing by 13% in FY 2016 and 8% in FY 2017. Since FY 2015, Ships & Submarines contract obligations increased from $24.2 billion to $27.2 billion in FY 2017, a 22% increase. Comparatively, Air & Missile Defense contract obligations grew 5% in FY 2016 before declining 15% in FY 2017. Total Air & Missile Defense contract obligations fell 11% from $9.7 billion in FY 2015 to $8.6 billion.

The Aircraft and Ordnance & Missiles platform portfolios have both grown at a significantly higher rate than overall DoD topline growth. Aircraft contract obligations increased to $77.2 billion in FY 2016 from $63.2 billion in FY 2015, a 22% growth. Aircraft contract obligations then grew an additional 10% in FY 2017 to $85.3 billion, a historic high. Ordnance & Missiles contract obligations increased 23% in FY 2016 and then an additional 7% in FY 2017. In total, Aircraft and Ordnance & Missiles contract obligations have grown 34% and 32% respectively since FY 2015.

Space Systems and Facilities and Construction have seen slight declines even as overall defense contract obligations grew. After increasing by 1% in FY 2016, Space Systems contract obligations declined 2% in FY 2017. In total, Space Systems contract obligations have fallen from $6.1 in FY 2016 to $6.0 billion in FY 2017, a 1% decline. Facilities and Construction contract obligations remained relatively steady in FY 2016 (-0.3% decline), before falling 2% in FY 2017.

Electronics, Comms, and Sensors grew at nearly the same rate as the overall defense rate of growth over the past two years. In FY 2016, both Electronics, Comms, and Sensors and overall defense contract obligations increased by 8%. In FY 2017, Electronics, Comms, and Sensors, increased 4%, just slightly less than the 5% overall growth.

Figure 5 shows defense contract obligations by platform portfolio from FY 2000 to FY 2017.

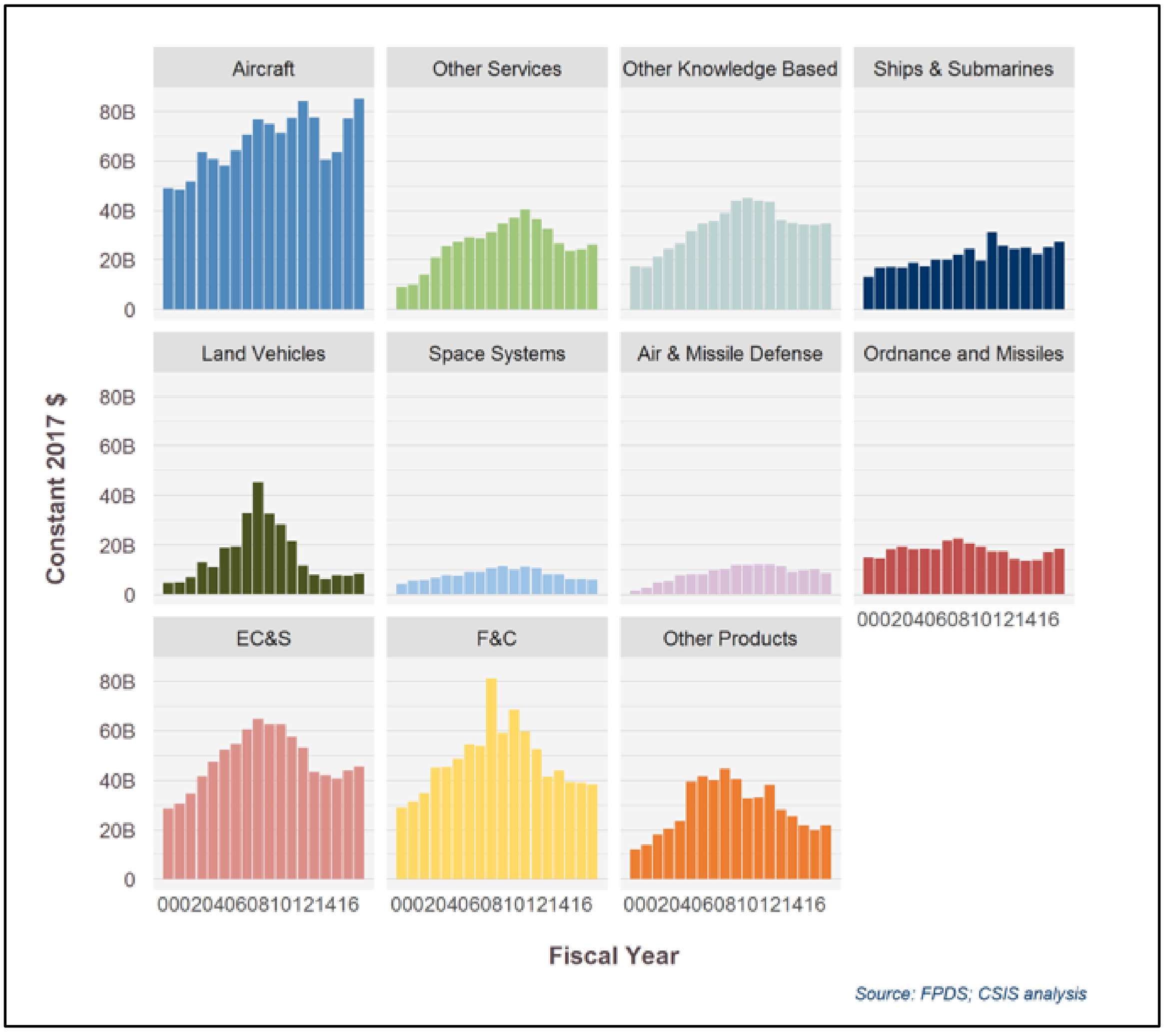


Figure 5. Defense Contract Obligations by Platform Portfolio, 2000–2017
(FPDS; CSIS analysis)

Overall DoD: Vendor Size

Figure 6 shows defense contract obligations by size of vendor from FY 2000 to FY 2017.

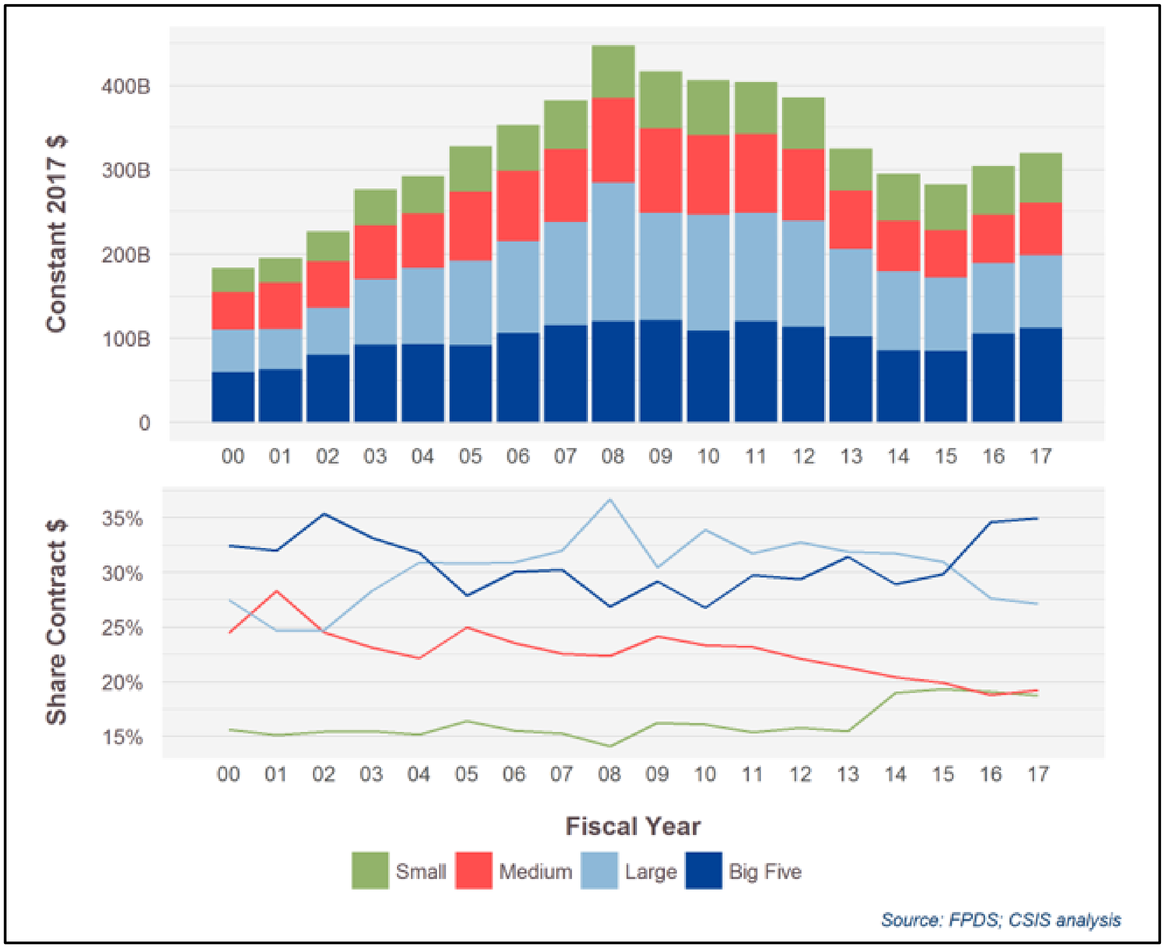


Figure 6. Defense Contract Obligations by Size of Vendor, 2000–2017
(FPDS; CSIS analysis)

The Big Five have benefited the most from the rebound of defense contracting. Since FY 2015, Big 5 contract obligations have increased by 33%.[2] Big 5 contract obligations have grown 25% from $84.3 billion in FY 2015 to $105.2 billion in FY 2016. In FY 2017, Big 5 contract obligations grew to $111.8 billion, a 6% increase from FY 2017. As a share of defense contract obligations, the Big 5 have risen from 30% in FY 2015 to 35% in FY 2016 and FY 2017.

Small vendors have been the second largest beneficiary of the defense contracting rebound growing 10% since FY 2015. Defense contract obligations going to Small vendors

[2] The Big 5 are the largest defense contractors: Lockheed Martin, Boeing, Raytheon, Northrop Grumman, and General Dynamics.

rose to $60 billion in FY 2016 from $54.6 billion in FY 2015, a 6% increase. In FY 2017, Small vendors' contract obligations totaled $59.8 billion, a 3% increase. However, despite absolute growth in the amount of contract obligation held by Small vendors, as a share of total defense contract obligations, Small vendors have remained steady at 19%.

Medium vendors' contract obligations have increased by 9% since FY 2015. Medium vendors only grew 1% in FY 2016, before increasing sharply in FY 2017. Last year, Medium vendors' contract obligations grew 8% from FY 2016, the largest percentage growth amongst vendors of all sizes. As a share of defense contract obligations, Medium vendors fell slightly from 20% in FY 2015 to 19% in FY 2016 and FY 2017.

Large vendors have seen the least benefit from the defense contracting rebound. Since FY 2015, contract obligations held by Large vendors declined 1%. Large vendors continued their decline, which started in FY 2011, in FY 2016, falling by 4%. Large vendors fared better in FY 2017, as contract obligations awarded to Large vendors rose from $83.9 billion to $86.7 billion, a 3% increase. As a share of defense contract obligations, Large vendors fell from 31% in FY 2015 to 28% in FY 2016 and 27% in FY 2017.

Overall DoD: Vendor Size by Area

Previous CSIS research has shown that beyond the topline vendor size trends, sequestraiton and the defense drawdown impacted "vendors of differing sizes depending on what area (products, services, or R&D) vendors are contracted for." For example, Big 5 R&D contract obligations fell nearly three and a half times faster than Small, Medium, and Large vendors R&D contract obligations (McCormick et al., 2017).

Big 5 contract obligations have increased for Products, Services, and R&D since FY 2015, but Products has significantly outpaced the other two categories. Since FY 2015, Big 5 Products contract obligations have increased by 43% compared to 15% growth in R&D and 10% growth in Services. Big 5 Products contract obligations increased by 32% in FY 2016 and 8% in FY 2017. Big 5 R&D contract obligations increased 2% in FY 2016 and 12% in FY 2017 but remain well below historical averages. Big 5 Service contract obligations declined 1% in FY 2017 after having grown 11% in FY 2016.

Contract obligations increases have been closer among all three categories for Small vendors since FY 2015. Small vendors' Products and Services contract obligations have both grown 9%, while R&D contract obligations have increased slightly faster, growing 14%. Small vendors' Products contract obligations increased 3% in FY 2016 and 6% in FY 2017. Small vendors R&D contract obligations grew 10% in FY 2016 before slowing to a 3% growth in FY 2017. Small vendors Services contract obligations grew 7% in FY 2016 and 2% in FY 2017. Of note, Small vendors' $35.7 billion in defense services contract obligations is 2% higher than the $34.9 billion obligated in FY 2012.

Medium vendors' trends were comparable to those seen by Small vendors. Since FY 2015, Medium vendors' contract obligations for Products grew 8%, R&D grew 7%, and Services grew 11%. Medium vendors' Products obligations declined 2% in FY 2016 but grew 11% in FY 2017. Medium vendors' R&D contracts have grown steadily over the past two years, increasing 3% in FY 2016 and 4% in FY 2017. Finally, Medium vendors' Services contract obligations increased 3% in FY 2016 and 7% in FY 2017.

Finally, trends within Large vendors' portfolios varied significantly as contract obligations declined 1% overall. Since FY 2015, Large vendors' R&D contract obligations have declined by 16% compared to the 4% decline in Services, and 4% growth in Products. Large vendors' R&D contract obligations declined 4% in FY 2016 before declining 12% in FY 2017. Large vendors' Services contract obligations declined 6% in FY 2016 but increased 4% in FY 2017. Finally, Large vendors' Products contract obligations declined 2% in FY 2017, but increased 6% in FY 2017.

Figure 7 shows defense contract obligations by size of vendor by area from FY 2000 to FY 2017.

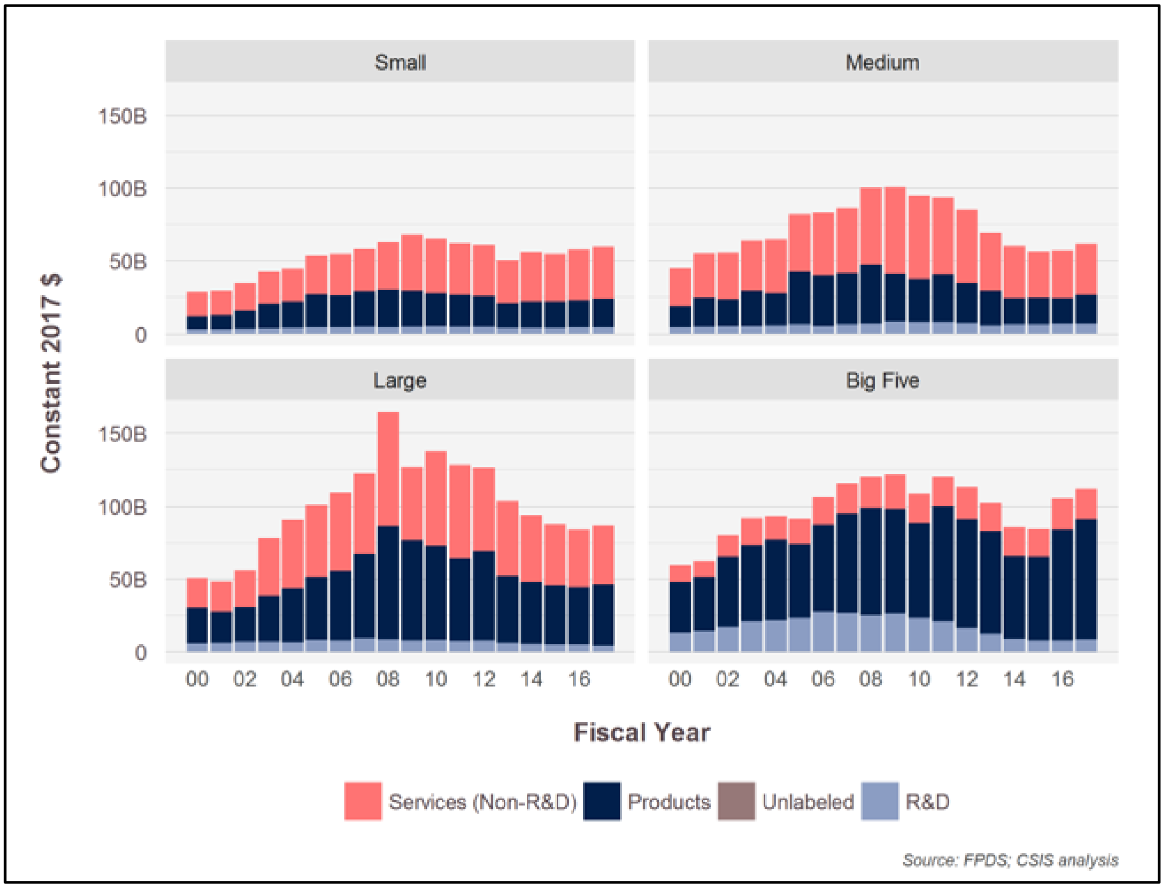


Figure 7. Defense Contract Obligations by Size of Vendor by Area, 2000–2017
(FPDS; CSIS analysis)

Overall DoD: Competition

Previous CSIS research has shown that the rate of effective competition has remained relatively steady since 2000 despite policy guidance favoring increased competition (Ellman et al., 2016; McCormick et al., 2015). Figure 8 shows the rate of effective competition for defense contract obligations from FY 2000 to FY 2017.

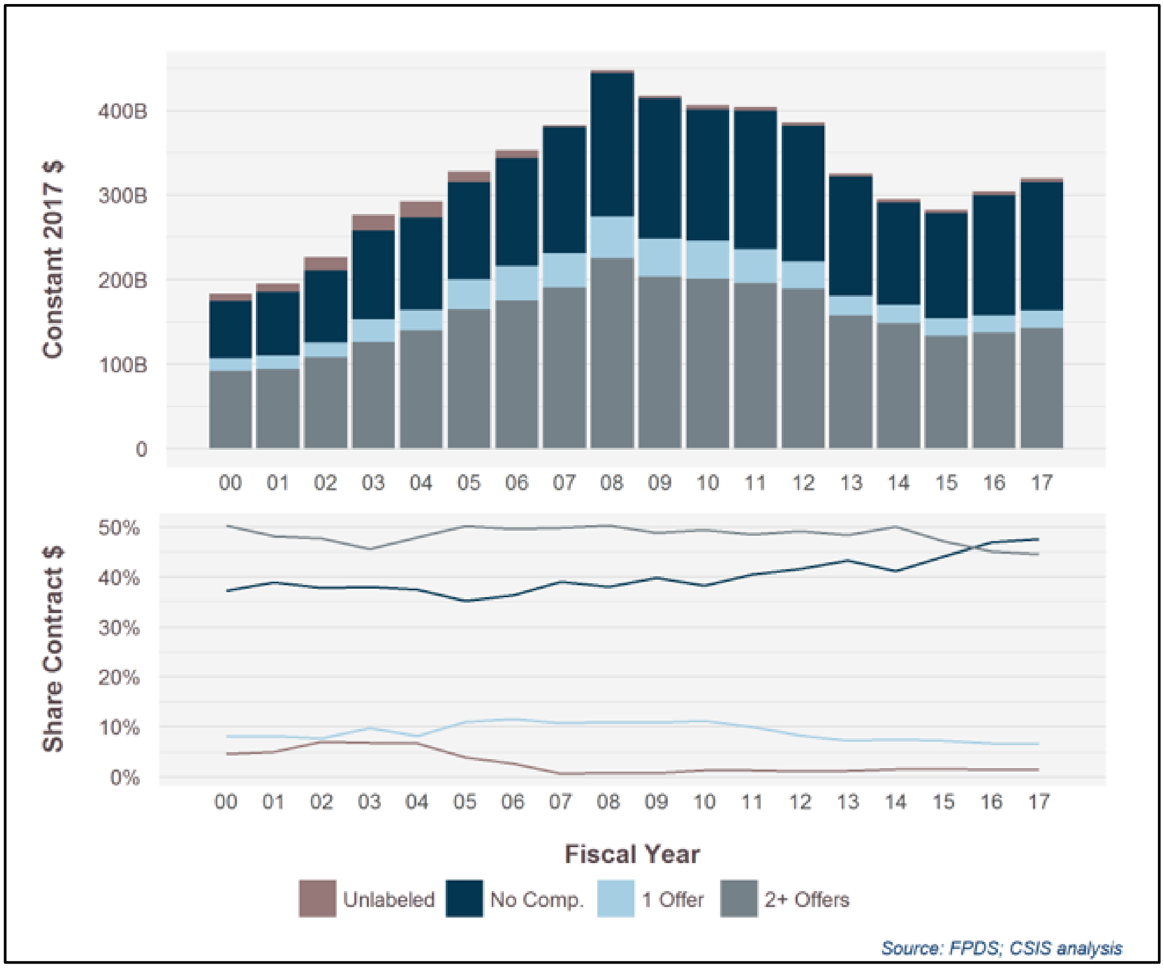


Figure 8. Defense Contract Obligations by Level of Competition, 2000–2017
(FPDS; CSIS analysis)

The data show that the rate of effective competition has fallen slightly over the past two years. In FY 2015, 47% of contract obligations were awarded after effective competition compared to 51% awarded without effective competition. In FY 2016, the share of contract obligations awarded after effective competition fell to 45%. The effective competition rate continued its decline in FY 2017, falling to 44%.

The declining effective competition rate in the rate of effective competition has been driven by significant increases in the total sum of contract obligations awarded without effective competition. Since FY 2015, contract obligations awarded with no competition has grown from $124.4 billion to $152 billion, a 22% increase. Comparatively, contract obligations awarded after effective competition has grown from $133.2 billion to $142.2 billion, a 7% increase. Of note, policy guidance issued to reduce the number of contracts awarded after receiving only one offers seems to be working. Over the past four years, contract obligations awarded after receiving only one offer has held relatively steady in terms of both raw dollars and share of defense dollars.

Conclusion

Products contract obligations growth has significantly outpaced R&D and Services.

Over the past two years, defense Products contract obligations have grown 22% compared to R&D and Services increasing by 6% and 5% respectively. Whereas Services and R&D have grown between 2 to 3% annually, Products contract obligations increased 13% in FY 2016 and 8% in FY 2017. It is likely that Products contract obligations growth continues to outpace Services and R&D given the considerable number of new legacy weapon system platforms purchases in the recent budget deal.

Navy's anchors aweigh; Air Force does a barrel roll; and the Army goes rolling along.

There were notable differences in the contracting trends between the military components.

The Navy fared best amongst all DoD components, growing 25% since FY 2015. Navy contract obligations increased 9% in FY 2015 and 15% in FY 2016. In FY 2017, the Navy accounted for 34% of Defense contract obligations, 10% higher than the next closest component and a high-water mark for this century.

Although Air Force contract obligations are up 11% since FY 2015, there has been a significant whipsaw effect over the past two years. The Air Force was the biggest beneficiary of the FY 2016 defense contracting rebound in absolute dollar terms, increasing 21% from FY 2015 totals. However, in FY 2017 Air Force contract obligations declined 9%. It remains to be seen whether the Air Force's long-term trajectory will resemble FY 2016 or FY 2017 trends or somewhere between the two.

The Army, the largest bill-payer during sequestration and the defense drawdown, has rolled along these past two years seeing slow, but steady growth. Over the past two years, Army contract obligations grew 1% in FY 2016 and 4% in FY 2017. Continued steady growth is critical as the Army seeks to recover from its modernization triple whammy (McCormick & Hunter, 2017).

Weapon system development pipeline trough might have bottomed out.

The seven-year trough in the weapon systems development pipeline appears to have hit its lowermost point. For the first time in years, contract obligations for System Development & Demonstration (6.5) and Advanced Component Development & Prototypes (6.4) increased from the previous year. Although System Development & Demonstration (6.5) and Advanced Component Development & Prototypes (6.4) contract obligations increased 11% and 3% respectively, it is too early to declare that the trough in the weapon systems development pipeline is over. Even after seeing positive news for the first time in years, System Development & Demonstration (6.5) contract obligations are still just above historic lows. It will likely take a few years of growth before it is possible to declare the end of the weapon system development pipeline trough.

Land Vehicles starts bounce back; Aircraft and Ordnance & Missiles up; Air & Missile Defense down.

The Land Vehicles platform portfolio started to bounce back in FY 2017 after suffering catastrophic cuts during sequestration and the budget drawdown. In FY 2017, Land Vehicles contract obligations increased from $7.5 billion in FY 2016 to $8.2 billion in FY 2017, a 10% increase.

Aircraft and Ordnance & Missiles were the two platforms that experienced the greatest growth during the defense contracting rebound. Aircraft contract obligations increased 34% since FY 2015, while Ordnance & Missiles increased 32%. Of note, Aircraft accounted for 27% of defense contract obligations (In FY 2017?), Aircraft's highest share of the defense budget since FY 2000.

Four platform portfolios experienced declines over the past two years: Air & Missile Defense; Facilities and Construction; Other Products; and Space Systems. Amongst those four platform portfolios, Air & Missile Defense experienced the greatest declines, falling by 11%. Interestingly, Air & Missile had been amongst the platform portfolios that fared best during sequestration and the drawdown.

Big 5 winner, but all up except Large.

The Big 5 were the big winners from the defense contracting rebound, while Large vendors have fared the worst. Big 5 defense contract obligations have grown 33% since FY 2015. This has largely been driven by the 43% increase in Products contract obligations going to the growth, but the Big 5 have also seen increases in Services (10%) and R&D (15%). Additionally, the Big 5 increased their overall share of defense contract obligations from 30% to 35%, largely at the expense of Large vendors.

Large vendors were the only vendor size category to decline since FY 2015, falling 1%. However, the trends suggest that Large vendors could fare better in future years as Large contract obligations increased 3% in FY 2017 compared to FY 2016's 4% decline.

Small (10%) and Medium (9%) vendors have grown at roughly equivalent rates since FY 2015. Small vendors' greatest increase came in R&D, which was up 14% compared to Products and Services, which both increased 9%. For Medium vendors, Services were the greatest source of growth, increasing 14%, compared to 7% growth for R&D and 8% growth for Products.

Rate of effective competition is down across the DoD.

Worryingly, over the past two years, the rate of effective competition for DoD contract obligations has declined. Whereas the rate of effective competition had held steady at around 50% over the past decade, FY 2016 and FY 2017 have departed from the trend. The share of contract obligations awarded after effective competition fell to 45% in FY 2016 and then 44% in FY 2017. This trend is troublesome given the importance of competition and given the previous imperviousness of the rate of effective competition to previous policy guidance. CSIS will explore potential reasons for these declining competition rates in a future report.

On a positive note, the share of contract obligations awarded after receiving one offer has continued to remain steady.

Final Thoughts

Defense contract obligations continued to grow in FY 2017 after rebounding in FY 2016, albeit at a slower pace than last year. With the defense budget set to continue rising for at least the next two years, defense contract obligations are poised to continue growing for the near-future. Beyond the next two years, the long-term forecast for the defense budget is unclear, making the DoD's decisions about where to spend that money critically important especially given the 2018 *National Defense Strategy's* focus on great power competition. The most recent budget and recent contracting trends show the prioritization of procurement over RDT&E, but the DoD's greatest challenge in the coming years will be finding the proper balance its investment portfolio. The DoD will need to balance procurement of upgraded versions of systems already in production that help tackle the current readiness challenge with RDT&E investments in future capabilities like artificial intelligence, hypersonics, and autonomy. Overinvestment in either direction could be detrimental to the DoD as over-emphasis on current platforms could increase existing readiness at the expense of the future fighting force, while overinvestment in future capabilities could create a death spiral for parts of the force like the F/A-18 Super Hornet fleet that are facing potential breaking points.

These investment dynamics present a critical follow-on challenge: resourcing and accessing innovation from nontraditional defense suppliers and the broader research community. The advances being made in the critical warfighting capabilities of the future are not being driven by the DoD or the traditional defense industrial base, but instead by commercial firms, universities, and other research entities globally. If the DoD hopes to gain access to these firms, it will need to create clear resourcing opportunities in the budget, yet the latest trends in both procurement and RDT&E have heavily favored the traditional defense industrial base. As policymakers tackle the difficult challenge of balancing current readiness and future capabilities, it must be careful not to crowd out resourcing for sources of innovation outside the traditional defense industrial base if the DoD is to succeed at accomplishing the *National Defense Strategy's* goal of refocusing on great power competition. Understanding the trends of what, how, and from whom the DoD has been buying can provide important insights into how the acquisition system responds to these and other challenges.

This paper presents only the preliminary findings of CSIS's analysis of the FY 2017 defense contracting trends. CSIS will further analyze the trends discussed in this paper and more in future reports.

References

Ellman, J., McCormick, R., Hunter, A. P., & Sanders, G. (2016). *Defense acquisition trends, 2015*. Washington, DC: Center for Strategic and International Studies. Retrieved from https://csis-prod.s3.amazonaws.com/s3fs-public/legacy_files/files/publication/160126_Ellman_DefenseAcquisitionTrends_Web.pdf

Hunter, A., McCormick, R., Ellman, J., Cohen, S., Johnson, K., & Sanders, G. (2017). Defense acquisition trends, 2016: *The end of the contracting drawdown*. Washington, DC: Center for Strategic and International Studies. Retrieved from https://www.csis.org/analysis/defense-acquisition-trends-2016

McCormick, R., & Hunter, A. (2017). *The Army modernization imperative: A new big five for the twenty-first century*. Washington, DC: Center for Strategic and International Studies. Retrieved from https://csis-prod.s3.amazonaws.com/s3fs-public/publication/170530_Hunter_ArmyModernization_Web.pdf?230oluRM4PwJBi4XRunDpVRMndOnunc

McCormick, R., Hunter, A. P., & Sanders, G. (2017). *Measuring the impact of sequestration and the drawdown on the defense industrial base.* Washington, DC: Center for Strategic and International Studies. Retrieved from https://csis-prod.s3.amazonaws.com/s3fs-public/publication/180111_McCormick_ImpactOfSequestration_Web.pdf?A10C65W9Qkx07VaJqYcJguCH.7EL3O7W

McCormick, R., Hunter, A. P., Sanders, G., Cohen, S., & McQuade, M. R. (2015). *Measuring the outcomes of acquisition reform by major DoD component.* Washington, DC: Center for Strategic and International Studies. Retrieved from http://csis.org/publication/measuring-outcomes-acquisition-reform-major-dod-components

Disclaimer

The Center for Strategic and International Studies (CSIS) does not take specific policy positions; accordingly, all views expressed in this presentation should be understood to be solely those of the author(s).

Panel 18. Rapid Product Development: Variations on the Theme

Thursday, May 10, 2018	
11:15 a.m. – 12:45 p.m.	**Chair: John Birkler,** Senior Fellow, RAND Corporation ***The Use of COTS in Defense Acquisition Programs: A Research Synthesis and Framework*** Timothy G. Hawkins, Lt Col, USAF (Ret.), Western Kentucky University Michael J. Gravier, C.T.L., Bryant University ***Crossing the Technology Valley of Death: The Case of the MDUSV*** David N. Ford, Texas A&M University John T. Dillard, COL, USA (Ret.), Naval Postgraduate School ***Seven Tips to Support Rapid Product Deployment: Lessons Learned*** Bruce Nagy, NAVAIR, China Lake

John Birkler—is a senior fellow at the RAND Corporation. He has held a variety of research and management positions since joining RAND in 1977: He has managed RAND's Maritime Program, overseeing research for the U.S. Navy, Office of the Secretary of Defense, U.S. Special Operations Command (SOCOM), U.S. Coast Guard, the Australian DoD, and the UK Ministry of Defence, and mentors U.S. Navy, Marine Corps, and Coast Guard executive fellows at RAND.

Birkler's research spans RDT&E strategies and planning, industrial base, acquisition, management, and organization issues. In addition to the above maritime clients, his research has covered a wide range of aircraft systems (including the Joint Strike Fighter, F-15, F-14, B-1, B-2, A-12, C-5, C-17, F-117, F/A-18 E/F), missiles and munitions (including the advanced cruise missile, the Tomahawk cruise missile, and precision conventional munitions), and surface and subsurface combatants. He also has led studies on the links between the health of the defense industrial base and levels of innovation and competition. His most recent work has involved managing or leading multiple Analyses of Alternatives (AoAs) for the Navy, USMC and Army, and SOCOM, Australian DoD, and leading a high-profile RAND analysis of Australia's Naval Shipbuilding Enterprise.

Birkler received his MS in nuclear and solid state physics from the University of South Carolina and completed the UCLA Executive Program in Management. After completing his third Command tour, he retired from the Navy Reserve with the rank of Captain**.**

The Use of COTS in Defense Acquisition Programs: A Research Synthesis and Framework

Timothy G. Hawkins, Lt Col, USAF (Ret.)—is an associate professor in the Department of Marketing, Western Kentucky University, and a National Contract Management Association Fellow. He researches and teaches in the realms of supply chain management, marketing, government contracting, and strategic sourcing. He has 20 years of sourcing experience in industry and government. Hawkins has published articles on opportunism in buyer–supplier relationships, source selection, services procurement, performance-based logistics, collaborative pricing, and electronic reverse auctions in scholarly publications such as the *Journal of Supply Chain Management, Journal of Business Logistics, Journal of Purchasing and Supply Management, Journal of Business Research, International Journal of Logistics Management, Journal of Defense Analytics and Logistics, Defense Acquisition Research Journal, Industrial Marketing Management, Journal of Business Ethics, Supply Chain Management: An International Journal, Journal of Marketing Channels, Air Force Journal of Logistics, Journal of Contract Management, International Journal of Procurement Management, Journal of Product and Brand Management*, and *Journal of Public Procurement*. His current research interests include procurement ethics, buyer–supplier relationships, strategic sourcing, services procurement, and supplier performance management. [timothy.hawkins@wku.edu]

Michael J. Gravier—C.T.L., is an Associate Professor of Marketing and Global Supply Chain Management at Bryant University. He received his PhD in marketing and logistics from the University of North Texas, an MS in Logistics Management (with a specialization in transportation management) from the Air Force Institute of Technology, and BA with majors in Spanish and anthropology from Washington University in St. Louis. Prior to his academic career, he spent 12 years as an active duty logistics readiness officer in the U.S. Air Force and worked as a research assistant in the Washington University School of Medicine. He has research interests in supply chain ethics, procurement, logistics pedagogy, logistics in emerging countries, and the evolution of supply chain networks in response to risk factors like obsolescence and changing information needs. [mgravier@bryant.edu]

Abstract

The DoD faces pressure to sustain its competitive advantages in national security. Enduring budget pressures, a record-long high operations tempo, the blitzing pace of technology, and adversaries that are leveraging commercial technology compound the challenge. The adoption of COTS products into defense acquisitions has been offered to help meet these challenges. A literature review of 62 sources was conducted with the objectives of better understanding COTS product implementation performance. It explored (1) characteristics of the research, (2) policies, laws, regulations, and directives that govern the use of COTS, (3) the known barriers to COTS implementations, (4) the known success factors to COTS implementations, (5) the recommendations have previously been made with respect to COTS implementations, and (6) recommendations for more timely and more effective COTS implementations. From the literature emerged a framework of COTS product usage and a scale to measure COTS product appropriateness that should help to guide COTS product adoption decisions and to help manage COTS product implementations ex post.

Introduction

The United States positions itself as the global leader in national defense, power projection, and the defense of its allies. To attain that vision, the U.S. must stay on the leading edge of technology; that is, it must maintain a competitive advantage in each domain—land, sea, air, space, and cyberspace. However, the U.S. Department of Defense (DoD) is not unbound by its resources. There are ceilings on the number of ships, soldiers, sailors, airmen, and fighter squadrons—to name a few. And the annual allocation of dollars—the ability to acquire resources—is constrained. The provisions of the Budget Control Act of 2011 linger as a reminder of the exploded national deficit, the need for a balanced budget, and sequestration. The estimated budget deficit for fiscal year (FY) 2017 is $577 billion (Amadeo, 2017), while the cumulative national debt is $19.968 trillion, or $61,554 per citizen (USDebtClock.org, 2017).

According to the GAO (2017),

> the Department of Defense faces five key challenges that significantly affect the department's ability to accomplish its mission. These include the need to (1) rebalance forces and rebuild readiness; (2) mitigate threats to cyberspace and expand cyber capabilities; (3) control the escalating costs of programs, such as certain weapon systems acquisitions and military health care, and better manage its finances; (4) strategically manage its human capital; and (5) achieve greater efficiencies in defense business operations. (p. 8)

These challenges are not expected to wane any time soon. Hence, the DoD must continue to innovate in a way other than just technology and weapons—it must figure out how to do even more with less.

Notwithstanding, technology is advancing at a breakneck pace. New developments in autonomous units, light-bending hyper stealth, electromagnetic rail guns, hypersonic missiles, 3D printing, artificial intelligence, big data, lasers, and social media—to name a few—cost money to develop and to harness. Hence, it is very expensive to remain on the leading edge versus current and potential adversaries and against different types of adversaries—conventional and asymmetric. Coupled with the demand on funds is the demand for faster response time. Yet time is no friend to a defense acquisition system that consumes, on average, 8.25 years to field a system from program initiation to initial operating capability (Riposo et al., 2014). Drastic change is needed in the DoD (Garber et al., 2011).

Additionally, adversaries and potential adversaries have expanded into unconventional domains posing threats via space and cyberspace. Even adversaries such as ISIS and Hezbollah have figured out the benefits of commercial technology and have adopted them (Hambling, 2017). They have also expanded into some of the most complicated domains by leveraging commercial technology. This is not surprising since many developments no longer originate in government-owned or contracted laboratories. Rapidly advancing commercial capabilities are deteriorating the United States' advantage (Tucker, 2017).

The use of commercial off-the-shelf (COTS) products is one strategy to help the DoD overcome its challenges. The implementation of COTS products offers faster development time, reduced cost, and higher quality compared to custom development (Torchiano et al., 2002). Yet in some settings, actually achieving those desired outcomes has been fleeting. COTS usage is no panacea (Carney & Oberndorf, n.d.), and is fraught with complexity, difficulty, and risk. According to Ben FitzGerald, a senior fellow at the Center for a New American Security, the DoD consistently struggles with the insertion of commercial technology (Erwin, 2016). Based on a review of approximately 40 programs, defense acquisitions continue to be plagued by immature architectures, COTS integration, interoperability, and obsolescence (Baldwin, 2007).

While some attention was afforded buying commercial items as far back as five decades, the brunt of the thrust occurred in the mid-1990s with the Perry Memorandum and the Federal Acquisition Streamlining Act of 1994. Pockets of success implementing COTS products exist, as do spectacular failures. With greater attention recently on the budget resulting from the Budget Control Act of 2011, coupled with the realization of that the pace of technology is accelerating and that adversaries are leveraging commercial technology, there has been recent renewed attention on accelerating the infusion of COTS products into defense acquisition.

Though the use of COTS products has been widely researched, as apparent from the DoD's struggles to harness it, COTS product usage is not completely understood. The literature on the use of COTS across various contexts is fragmented. There are some DoD-specific case studies of COTS product usage and numerous non-DoD studies—albeit mostly concentrated in the COTS software realm. It has been 17 years since the last comprehensive synthesis of COTS implementations—then conducted by the Air Force Scientific Advisory Board (Grant, 2000). There is no known comprehensive synthesis of COTS usage research.

Scope and Objectives

The purpose of this research, therefore, is to review the literature surrounding the use of COTS technology to better understand COTS product implementation performance. Such a research synthesis seeks to bring together previously disparate streams of work (Webster & Watson, 2002), namely DoD system acquisition, software engineering, supply chain management, marketing (new product development), and knowledge management. The scope of this review includes hardware and software. The following research questions will be explored:

1. What are the known barriers to COTS implementations?
2. What are the known success factors to COTS implementations?
3. What policies, laws, regulations, and directives govern the use of COTS?
4. What recommendations have been made with respect to COTS implementations?
5. What are the typical research types, contexts, research methods, target markets, and foundational theories utilized in COTS-based research?
6. What is recommended for more timely and more effective COTS implementations?

The answers to these six questions are crucial; they should help reduce program risks of poor performance, failure, cost growth, and schedule slippage. The gained knowledge should also help the DoD acquisition community to more effectively and more efficiently leverage COTS products in order to meet its mission mandates and retain a competitive advantage against existing and potential foes.

The remainder of this paper is organized in the following manner. First, the study presents the review methodology. Following the synthesis of the literature, results are then presented. Lastly, discussion, limitations, implications, future research directions, and conclusions are offered.

Methodology

To address the research questions, this research employed a literature review,

> the selection of available documents (both published and unpublished) on the topic, which contain information, ideas, data and evidence written from a particular standpoint to fulfil (sic) certain aims or express certain views on the nature of the topic and how it is to be investigated, and the effective evaluation of these documents in relation to the research being proposed. (Hart, 1998, p. 13)

The process for a systematic literature review outlined by Tranfield et al. (2003) was followed. This process consists of three stages: planning the review, conducting the review, and reporting and dissemination. In the planning stage, the need for the review is identified and a review protocol is developed. In stage two, the relevant literature is searched, identified, and selected. Additionally, particular data is extracted and synthesized. In the final stage, a report is drafted that includes recommendations. It is then disseminated.

There exists a mountain of information surrounding the implementation of COTS technologies. A simple Google search of "commercial off-the-shelf" yielded 512,000 hits. Academic databases searched included ProQuest ABI/Inform Global, LexisNexis Academic, JSTOR, and EBSCOHost. Publications by the Acquisition Research Program (ARP) were reviewed. GAO reports were found on the GAO's website. Regulations were found from the Navy's repository found at: https://doni.daps.dla.mil/default.aspx; 1,140 regulations were scanned for COTS applicability. Academic courseware was obtained from the Defense Acquisition University (DAU). Sources were also traced backward from reference lists (Leedy and Ormrod, 2005). Sources searched included peer-reviewed journals, conference proceedings, Acquisition Research Program reports, case studies, GAO reports, DoD reports, search engine (Google and Google Scholar), DAU Acquisition Community Connection, GAO bid protests (on the basis of COTS), U.S. Court of Federal Claims bid protests (on the basis of COTS), books, trade press, white papers, guidebooks/handbooks, patents, and conferences/practitioner organizations.

The massive number of sources found was narrowed by inclusion and exclusion criteria (Table 1). The scope of the knowledge base was expanded beyond the DoD context since there is very little rigorous, peer-reviewed academic research examining only DoD acquisitions involving the use of COTS products. However, the exemplar case studies and the summary of prior recommendations were constrained to DoD COTS product implementations. The literature search terminated when no new viewpoints emerged (Leedy & Ormrod, 2005).

Table 1. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Defense acquisition context	COTS case studies published prior to 2000
Hardware	COTS implementations by non-U.S. entities
Software	COTS usage in scientific discovery in which COTS product usage is not the study's focus
Studies of for-profit sector COTS usage	Classified COTS product implementations/programs

Once the literature was accumulated, the data extraction form was used to construct concept matrices of barriers and success factors (Webster & Watson, 2002). These tabulations depict the most prevalent antecedents to COTS implementation performance—the key dependent variable in the emerged framework. Looking across sources, patterns and themes were sought (Webster & Watson, 2002). A pattern was considered to exist when a concept appeared in four or more sources as barriers and as enablers (i.e., success factors).

Each article was categorized according to its theory type using Gregor's (2006) typology. Gregor classified information systems theories according to their four objectives: analyzing, explaining, predicting, and prescribing. The resultant typology included five types: analyzing, explaining, predicting, explaining and predicting, and design and action. Analyzing theories simply describes *what is.* They sometimes take the form of classifications or taxonomies. The analyzing theory makes no causal inferences or predictions. Explaining theories do just that; they explain how, what, why, when, and where. Yet, the explaining theories do not posit testable hypotheses. Conceptual models and theory development fit this type. Many case studies fit this classification. Predicting theory says *what is* and *what will be* in the future. While the theory makes predictions and includes testable hypotheses, it does not very well explain why the hypotheses should be (or are) so. In contrast, explaining and predicting theories make predictions, offer testable hypotheses, and explain the causality. Finally, design and action theories explicate *how to* do something. They are prescriptive in nature.

Then, each article was classified by its stage in the knowledge management process per the framework of Beesley and Cooper (2008). Process stages include knowledge creation, dissemination, knowledge transfer, knowledge adoption, and innovation.

To assess the quality of each article, several methodological aspects were evaluated for academic rigor. In Appendix A, this assessment appears in the column labeled Scholarly Academic Evidence. Each article is coded as yes (Y) or no (N). Yes indicates that the article was published in a peer-reviewed source, provides sufficient evidence of validity and reliability, explains type of data, data source, and data collection method with confidence that error is mitigated, and describes an appropriate data analysis method. Otherwise, the article was coded no.

Results

Emerged Constructs and Relationships

From the literature, concepts were coded as individual barriers and enablers to COTS product usage. For the barriers, 86 concepts were identified. For the enablers, 89 concepts were identified. Looking across concepts for commonality and repetition, themes rose to the surface. The central theme seemed to address the fitness of COTS products to the situation, henceforth termed *COTS appropriateness*. The following discussion will explain COTS appropriateness and each of its antecedent factors. See Figure 1 for a depiction of the comprehensive COTS product usage framework.

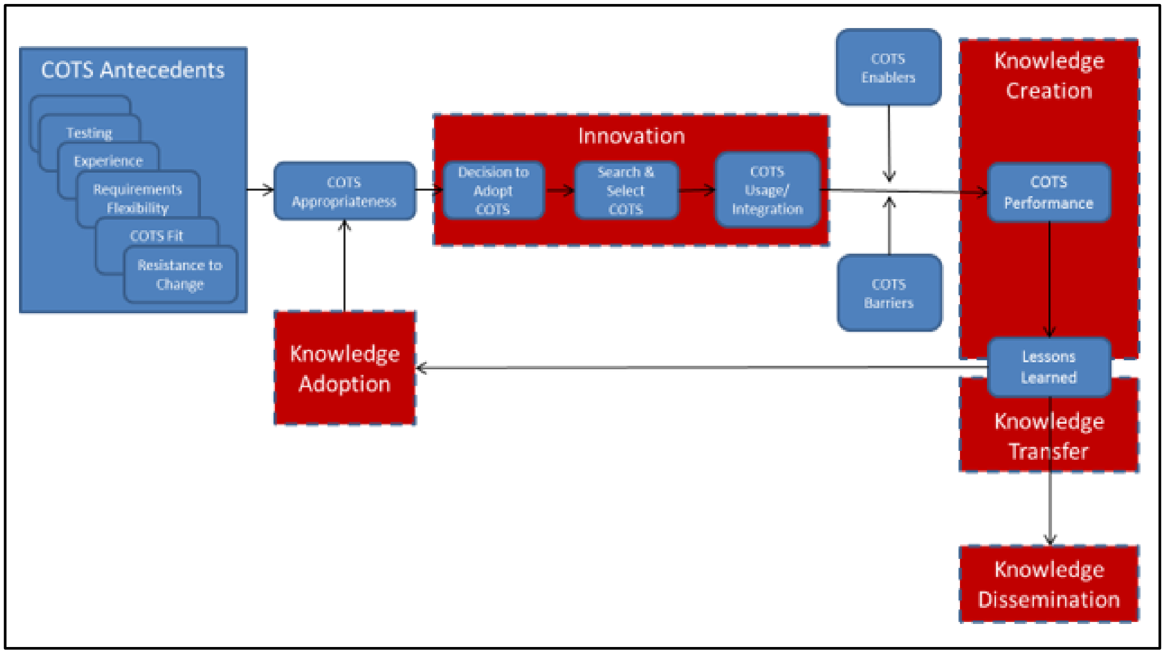


Figure 1. COTS Product Usage Framework

COTS Appropriateness

COTS appropriateness is the focal construct in the emerged COTS framework. Grant (2000) concluded, "Not enough emphasis has been placed on understanding and implementing the process to determine the applicability (that is, the appropriateness) of COTS" (p. 31). The DoDIG (2006a) mentions the inappropriateness of COTS implementation on numerous occasions by the Air Force, then links the inappropriateness to performance failures (e.g., excess costs). Academicians have also taken notice of the importance of COTS appropriateness. Jilani (2008) mentions the selection of inappropriate COTS components. Keil and Tiwana (2005) also mention the disastrous ramifications of selecting inappropriate COTS software. Couts and Gerdes (2010) question the appropriateness of COTS to meet the needs of some integrations. Cechich and Piattini (2007) offer a procedure for detecting the suitability of COTS candidates.

COTS appropriateness is herein defined as the extent to which a COTS product—adopted for use as-is or integrated into another product or system—can meet the program objectives with very little or no modification without introducing excess risk to cost, schedule, performance, safety, or security. It considers the fit between the COTS product functionality and that desired by the DoD user for a particular intended mission effect.

In order to assist researchers studying COTS implementations and to assist practitioners in assessing COTS usage opportunities, a scale is developed to measure COTS appropriateness (See full Technical Report.). This scale is intended to assess the degree of appropriateness, measured on an interval scale of 1 to 7, as determined by the presence (or absence) of the following antecedent conditions.

Antecedents to COTS Appropriateness

Certain situations lend themselves to COTS product usage while others do the opposite. From the DoD case studies and the at-large literature, the following attributes (i.e., factors) determine, at least partially, whether a COTS product should be adopted. These antecedent factors are listed in order of expected strength of the relationship, with the strongest predictors listed first. The Technical Report elaborates on rationales for inclusion with citations from the supporting literature.

RQ1: What are the known barriers to COTS implementations?

There are several antecedent factors that decrease COTS product use appropriateness, as discussed above. These factors, once a COTS product is adopted and as implementation is attempted, reappear as barriers to success. They include a “black box” design, organizational resistance to change, intellectual property constraints, short product life cycles, and complexity.

DoD Examples of Barriers to COTS Implementations

The following 15 DoD programs exemplify barriers to COTS product usage for various reasons. Each program is followed by a citation enabling the reader to trace back the details. As mentioned previously, there is substantial variance in the rigor and details provided for each case.

- Navy Littoral Combat Ship (DoD, 2009)
- USMC Presidential Helicopter Replacement (VH-71) (DoD, 2009)
- Army Armed Reconnaissance Helicopter (ARH) (DoD, 2009)
- Air Force F-22 (Grant, 2000)
- Air Force Depot Maintenance Management Information System (Grant, 2000)
- Air Force Expeditionary Combat Support System (ECSS) (Charette, 2013)
- Air Force KC-767A Tanker Lease (DoDIG, 2006a; DoDIG, 2004a; GAO, 2006)
- Air Force C-130J (DoDIG, 2006a; DoDIG, 2004b)
- Air Force and Navy T-6A Texan II, Joint Primary Aircraft Training System (JPATS) (DoDIG, 2006b)
- Air Force Wideband Gapfiller Satellites (DoDIG, 2006a)
- Navy and Air Force MV/CV-22 Osprey engines (DoDIG, 2006a)
- Air Force C-17A engines (DoDIG, 2006a)
- Army High Mobility Multipurpose Wheeled Vehicle (HMMWV) (DoDIG, 2006a)
- Air Force T-3 Firefly (Baker, 2002)
- Army DCGS-A (Brill, 2017)

RQ2: What are the known success factors to COTS implementations?

There are several antecedent factors that increase COTS product use appropriateness, as previously discussed above. These factors, once a COTS product is adopted and as implementation is initiated, reappear as enablers to success. They include the fit between requirements and COTS product capabilities, requirements flexibility, COTS product experience, open systems architecture, a robust COTS product evaluation and selection process, post-adoption COTS product change preparedness, COTS product training, communication, evaluating total cost of ownership, a priori and post hoc testing, marketplace knowledge, leadership, stakeholder buy-in, and contractual financial incentives.

DoD Examples of Effective COTS Implementations

The following 23 DoD programs exemplify enablers of COTS product usage. Like the aforementioned barriers, each program is followed by a citation enabling the reader to trace back the details.

- DoD Common Access Card (GlobalPlatform, 2003)
- Air Force Manufacturing Resources Planning (MRP) (Grant, 2000)
- Advanced Amphibious Assault Vehicle (AAAV) (Grant, 2000)
- New Attack Submarine and Acoustic Rapid COTS Insertion (ARC-I) (DoD, 2009; Grant, 2000; Boudreau, 2006; Ford & Dillard, 2009)
- Navy Sea Fighter (FSF-1) (DoD, 2009)
- Airborne Warning and Control System (AWACS) (Grant, 2000)
- Navy E-2 Hawkeye Early Warning Program (Gansler & Lucyshyn, 2008)
- Army Light Utility Helicopter (Gansler & Lucyshyn, 2008)
- DLA Business System Modernization (BSM) (Gansler & Lucyshyn, 2008)
- Army's General Fund Enterprise Business System (GFEBS) (Kendall, 2015)
- Lightweight Autonomous Underwater Vehicles (AUVs) (Incze, 2011)
- Navy P-8 Poseidon maritime patrol and reconnaissance aircraft (DoD, 2009; Naegle & Petross, 2010)
- Air Force C-5 Reliability Enhancement and Re-Engining Program (RERP) (Lorell et al., 2017)
- Air Force and Navy Joint Direct Attack Munition (JDAM) (Grant, 2000; Lorell et al., 2017)
- Air Force Small Diameter Bomb (SDB I) (Lorell et al., 2017)
- Air Force Wideband Global SATCOM (WGS) System (Lorell et al., 2017)
- Mine Resistant Ambush Protected (MRAP) vehicles (Morrow, 2010)
- Army M-ATV (Morrow, 2010)
- Aegis Ballistic Missile Defense System (Lockheed-Martin, 2017)
- Defense Healthcare Management Systems Modernization (DHMSM) (DoDIG, 2016b; Landi et al., 2017)
- Army's Single Stock Fund (SSF) program (Alcide, 2006)
- USMC utility task vehicle (UTV) program (Tadjdeh, 2017)
- Army Ka-Band Satellite Transmit and Receive System, AN-GSC-70(V) (Stein, 2006)

RQ3: What policies, laws, regulations, and directives govern the use of COTS?

Most of the attention to buying commercial items occurred in the mid-1990s with the Perry Memorandum and the Federal Acquisition Streamlining Act of 1994. There has been recent renewed activity in the amount of COTS-related law, policy, regulation, and directives. The Technical Report lists the relevant laws, policies, regulations, and directives.

RQ4: What recommendations have been made with respect to COTS implementations?

Over the years, several oversight authorities and researchers have made recommendations for practitioners in order to improve their management of COTS product implementations. A list of those recommendations is provided in the Technical Report.

RQ5: What are the typical research types, contexts, research methods, target markets, and foundational theories utilized in COTS-based research?

The Technical Report shows the data collection methods and data analysis methods employed. It lists the publications from which COTS product usage literature was found. The report also shows the types of research and the process steps of Beesley and Cooper's (2008) knowledge management framework in which each reviewed article fits.

Discussion

Managerial Implications

COTS product implementation is complex and difficult to successfully navigate. This is evident in simply the number of antecedent factors that affect COTS usage appropriateness that emerged from the literature. Additionally, some additional factors are likely to be significant actors, yet may not have risen to the top as a pattern due to the limited number of published case studies.

While there appears to be a desire to use COTS products (evidenced by statutory requirements and policy directives), the actual integration of COTS products into systems is easier said than done. It introduces one more risk to programs that is unlikely to be welcomed by program managers who spend their days anticipating and defending against risks. What has the DoD structurally infused to alleviate those perceived risks from program managers? The emerged framework, based on findings from academic studies and case studies of DoD COTS product implementations—coupled with knowledge management literature—clearly indicate the importance of monitoring the commercial marketplace. An organization must possess the ability to recognize the value of new external information (Grandinetti, 2016). In order to recognize the value, marketplace observers must know the technical and scientific details, know the DoD's existing infrastructure, and be familiar with user needs and desired effects. This not a novel idea; market intelligence cells were recommended in 2014 (Finkenstadt et al., 2014). The number of available organic personnel with these skills, experience, and education—that is, with the requisite knowledge—is scant. Thus, it is likely that, without intervention, the DoD will continue to rely on systems integrators to conduct the commercial marketplace monitoring. This outsourcing of sorts raises serious implications of agency theory. In whose interest is the monitor working, and how is knowledge transformation (aka, assimilation or transfer) being manipulated or withheld? Since the ability to take on new knowledge to some extent depends on the amount and type of knowledge already possessed, how is the integrator's knowledge being managed such that it is not lost?

Commercial off the shelf, as a topic, appears to be waning since the 2005–2009 timeframe. The quantity of source hits resulting from the search term "commercial off the shelf" in the ProQuest ABI/Inform Global database has modestly decreased recently. The quantities of hits are distinguished between peer-reviewed journals (PRJ) and all COTS-related articles. This decrease has not gone unnoticed (Maras et al., 2012). This trend is somewhat corroborated by examining the number of patents (Google, 2017) using the term commercial off the shelf. The quantity of COTS-related patents also seems to have peaked and is now waning. These trends could suggest that the practice of using COTS products is in decline, or it could simply mean that labeling COTS usages as such may be in retreat as the practices become rather standard (Maras et al., 2012). This reduction would be expected as the usage of COTS becomes ubiquitous; thus, perhaps authors perceive the term *COTS* to be implied and therefore, unnecessary to mention.

"There is a failure to assure correct, predictable, safe, secure execution of complex software in distributed environments" (Baldwin, 2007, p. 8). Research does not address the issue of security involved with adopting COTS products (Grant, 2000). The new DFARS clause 252.204.7012 requiring the protection of defense information and cyber incident reporting applies to systems that integrate COTS, but not to purely purchased commercial items (Cassidy & Stanton, 2017). Included within the realm of security is counterfeiting. Very little research addresses counterfeiting though it clearly poses a risk to system performance and to security. Supply chain risks with respect to IT may include insertion of counterfeits, unauthorized production, tampering, theft, insertion of malicious software and hardware, and poor manufacturing and development practices (Gump et al., 2015); thus, grey market products—those distributed beyond the manufacturer's intended channel—should be avoided. But gaining control of a free-market supply chain is daunting, as indicated by the Aerospace Industry Association's concern over recent DFARS changes (AIA, 2014). The security of COTS-based systems is and will continue to be a serious issue (DoDIG, 2016a). And the DFARS requirements for counterfeit electronic part detection and avoidance (DFARS 252.246-7007) that flow down to suppliers might repel viable COTS product sources.

Research hardly addressed the issue of intellectual property (IP) involved with adopting COTS products (Grant, 2000). However, the literature since 2000 suggests that intellectual property rights is a formidable barrier. This is logical particularly in systems that have to reconcile the IP rights of multiple pieces of hardware or multiple software components. One component repository, ComponentSource, currently makes available 1,933 components, 705 applications, and 384 add-ins to systems integrators and developers available from 343 publishers (ComponentSource, 2017). Imagine keeping track of the use restrictions, access rights, royalties, warranties, and liabilities of only 10 components. Then imagine that each of those sets of 10 terms and conditions is different.

Commercial firms rapidly update their products to keep pace with technology and in the pursuit of new avenues of differentiation and, thereby, competitive advantage. Short product life cycles and short time-to-market make design and acquisition time critical. Experimentations of new ways to quickly access new commercial technology will be important. One example is the DoD's pilot program called Commercial Solutions Opening (CSO) established by Section 879 of the National Defense Authorization Act (NDAA) for FY 2017 (Public Law 114-328) and implemented by DFARS Case 2017-D029. A CSO is a merit-based source selection strategy that utilizes Other Transaction Agreements (OTA) rather than contracts pursuant to the FAR. Under the Defense Innovation Unit Experimental (DIUx) program (https://diux.mil/), 25 OTAs have been awarded valued at $48.4 million (Defense Innovation Unit Experimental [DIUx], 2017). This program is drawing private investment from venture capitalists and participation from firms that normally do not transact with the DoD. Recent initiatives include autonomy, personal aerial vehicle, tactical autonomous indoor drone expansion, human cooling, digitally aided close air support platform, hardened network defense, knowledge management, multifactor authentication for network access, and advanced analytics from synthetic aperture radar imagery.

While COTS software has been researched extensively, COTS hardware receives very little scholarly attention. This could be attributed to the newness and magnitude of software issues. It could also be due to the expectation that the commercial sector will favor commercial hardware integration when it is cost effective.

From the literature, user satisfaction is a key measure of information systems COTS success (Kakar, 2013); however, user satisfaction did not appear from the DoD case studies as a key to successful COTS implementation. This could be attributed to a top-down paradigm that the user gets what the program office delivers. Hence, while system performance defines success, the literature shows ambivalence toward the user's perception. Nevertheless, the ubiquitous technology acceptance model (Davis, 1989) shows that IT system adoption is driven by perceived ease of use and by perceived usefulness.

The DoD struggles to use COTS products to create a military advantage (Erwin, 2016). Clearly, some commercial products are not designed and built to meet the rugged needs of military applications. Nonetheless, the DoD's struggle is perhaps most brightly illuminated by the Palantir case—a commercial analytics product which soldiers have lauded as life critical but which was refused by the Army somewhat arbitrarily (U.S. Court of Federal Claims, 2016). Thus, antecedent factors for COTS product adoption beyond functional capability deserve special attention.

The relevant literatures surrounding COTS implementations is severely lacking in theoretical grounding (Hall & Rapanotti, 2016). This void can stymie understanding and the pace of progress. Other business-oriented, applied disciplines have also struggled to find unique theoretical foundations explaining and predicting their phenomena, such as supply chain management (Defee et al., 2010) and information systems (Gregor, 2006). Few studies dig into causal relationships explaining or predicting phenomena. Yet, such studies yield the strongest evidence answering why things are the way they are and how things might be expected to be in the future. Hence, explaining and predicting is the essence of theory and discovery. Since knowledge is cumulative (Cohen & Levinthal, 1990), more research attention should explore causal relationships.

Few case studies of COTS product usage would qualify as scholarly contributions. Thus, it is difficult to discern between truth and conjecture, or more likely, to get beyond the visible symptoms and discover the underlying causes. Therefore, consumers of information in many of the existing case studies may be forming beliefs and making decisions based on anecdotal evidence and hasty conclusions. Most "case studies" lack methodological rigor and sufficient detail explaining how findings were determined, what type of data was collected, how data was collected, how data was analyzed, and how validity and reliability were assured. Very few case studies involving interviews mentioned the location of interviews or whether they were conducted face-to-face, over the phone, or online. Few cases mentioned recording the interviews, transcribing them, interview durations, transcript lengths, and sending transcripts to informants for validity. Few cases summarized the demographics of who was interviewed such as duty title, industry, organization, years of experience, nationality, location, etc. Likewise, few case studies mentioned triangulating data with other sources (e.g., archival records—how many and what type) to corroborate data and analyses. Few case studies mentioned the qualitative data analysis methodology such as coding qualitative text, seeking themes, the number of themes identified, identifying patterns, and unveiling associations among themes via constant comparison—a process of continuously returning back to all text once a new theme or pattern emerged and via code matrices (Miles & Huberman, 1994). Few case studies offered any information about validity and reliability such as using multiple coders of themes and measuring inter-rater reliability and conducting member checking sessions (Yin, 2009) to validate findings and analyses. Few cases reconciled the findings with the relevant literature as evidence of further validity.

Some have called for a new defense acquisition process tailored to COTS product usage. The literature reviewed herein, while offering a substantial number of considerations when adopting COTS products, does not compellingly suggest that the DoD's 5000 series cannot effectively integrate COTS products. Perhaps some changes could be made to provide guidance and consistency to the field to account for some of the nuances and complications of COTS product adoption. Horowitz and Lambert (2006) offer some insight:

> An assembly sequence (components to be assembled, corresponding dates and costs) has several risks including: 1) technical risk: successful (or not) function of assembled components by planned schedule milestones; 2) operational risk: achieving (or not) the desired business value by using the new system of assembled components; and 3) programmatic (schedule and cost) risks: accomplishing the assembly within time and budget constraints. (p. 286)

They thus presented a framework (called "learn as you go") for planning and adjusting milestone sequences in assembling off-the-shelf software components. Principles from this framework could be borrowed to tweak, or allow for special cases within, the DoD 5000 series of directives and instructions.

RQ6: What is recommended for more timely and more effective COTS implementations?

1. Apply the proposed COTS Product Appropriateness scale (see Technical Report) to prospective programs when contemplating integrating major COTS components. This scale captures the emerged antecedent factors (from barriers and enablers), and therefore, should serve as a helpful indicator of the prospect.
2. To facilitate knowledge management, DoD activities should record COTS product implementations in contract action reports. This will enable future program managers, technical authorities, and contract managers a single, reliable source from which to search for prior COTS implementations by similarity of COTS technology type (e.g., software components, avionics, land-based robots, etc.). This knowledge can rapidly inform decision-makers of where to go to gather additional detailed information on lessons learned, market research, and suppliers to facilitate knowledge dissemination.
3. Expounding on the previous recommendation, COTS product implementations should be catalogued in a central repository in order to make detailed lessons learned available to future acquisition teams. Since no single, optimal solution to knowledge management can be developed (Bjornson & Dingsoyr, 2008), this central repository could complement other knowledge management practices. For example, the deposited lessons learned could be pushed to educators and trainers at DAU, NPS, AFIT, ICAF, senior service shools, and interested university centers.
4. Since tacit knowledge resides with people, organizations should set, via policy, maximum program employee turnover rates. Turnover has repeatedly been found a culprit in failed and low performing programs (Charette, 2013).

5. Over the years, several oversight authorities and researchers have made recommendations for practitioners in order to improve their management of COTS product implementations such that desired, and in some cases mandated, outcomes are achieved. However, the extent to which all of these recommendations have been implemented is unknown. Therefore, an audit of the recommendations would be useful to reconcile the deficiencies and weaknesses of current practice with required and helpful practices (i.e., the recommendations). The audit results would provide a gauge of the extent that current processes and policies are sufficient and that COTS product usage is sufficiently managed.
6. It appears that, in the realm of software, the use of COTS products is such a pervasive commercial practice that products involving software nearly cannot be developed without at least some integration of COTS products. This is undoubtledly due to the significant savings in costs and time. Nevertheless, what is not as ubiquitous is the extent of reuse of physical COTS products (i.e., hardware). Thus, a study should be conducted to quantify the extent of COTS implementation, and quantitatively validate the positive and negative antecedents to COTS implementation performance.
7. The DoD should not establish quotas for COTS implementations. Quotas have, in the past, manifested in percentage goals (i.e., COTS products have to constitute a certain percentage of a system). Extrinsic forcing mechanisms could result in gaming and unnecessary risk-taking.
8. Set policy that requires a technical evaluation sub-factor in all source selections that: (1) requires offerors to submit their plan for making their deliverables (including components of them) open to competition during sustainment, and (2) allows for meaningful evaluation credit (i.e., ratings, strengths, and reduced risk ratings) for superior plans. These plans, in turn, should become part of the resultant contract.
9. In contracts involving award fees, consider making the extent of COTS implementation one of the criterion for award fee determination.
10. For all contracts requiring the use of COTS products, add an assessment of: (1) the extent of COTS product usage and (2) COTS product implementation effectiveness to the contractor performance assessment reporting (CPAR). This follows recommendations by Rendon (2007). It should motivate contractors to pursue the integration of COTS products since many suppliers place significant attention on achieving desired CPAR scores (Hawkins, 2016).
11. Expand the scope of the DoD's Strategic Capabilities Office (CSO) organized as a Janus-facing orgnization around desired effects and simultaneously around commercial industries. Within the CSO, technology expert councils (i.e., industry-facing organization) would need to matrix to the revolutionary effects council (i.e., warfighter-facing organization). A sufficient number of standing councils would be needed to adequately cover the various high-potential industries and the most-impactful effects.
12. The DoD should build structure to facilitate knowledge management and absorptive capacity. This means that resources such as people, time, and technology should be allocated to monitoring the marketplace for commercial products and new technology capabilities. There are pockets of excellence such as the CSO and DIUx; however, their scope and capacity is likely too

small to assist all current and yet-to-be-discovered needs. Those monitoring the marketplace must be technically adept so that they will be able to recognize valuable information when they see it. Additionally, the curb on travel should be lifted for the defense acquisition workforce. If anything, these technical and business professionals need more exposure to commercial knowledge, not less. Conferences are efficient forums to interact with numerous experts in a short amount of time. Finally, discovered knowledge should be codified (i.e., made explicit) and be available to future market monitors since absorptive capacity depends on the amount of knowledge previously acquired.

13. In developing the COTS implementation framework, a scale to measure the focal construct, COTS appropriateness, was developed. This scale, in its current form, should be considered exploratory. Hence, it should be empirically tested to ensure reliability and all types of validity (i.e., content, construct, discriminant, convergent, nomological, and external). Once validated, the scale should be used by practitioners to assist in their decisions whether to adopt COTS products. The scale can also be used by academicians to empricially study COTS implementations.
14. Researchers pursuing COTS-based inquiry should ground their research in relevant theory. Journals and academic conferences publishing COTS-based works should add to their requirements a review of the relevant literature and an explicit positioning of the work into that body of knowledge.
15. Case studies of COTS product usage should demonstrate greater methodological rigor and provide more detail explaining how findings were determined, how data was collected and analyzed, and how validity and reliability were assured. This will prevent the adoption of anecdotal evidence and hasty conclusions. A commonly-adopted method is provided in *Case Study Research: Design and Methods* by R. K. Yin (2009).
16. The DoD should leverage its commercial business internships, such as the Air Force's Education With Industry program and the Navy's Supply Corps Training With Industry program, to glean commercial practices with respect to new product design, development, manufacturing, and sustainment. A specific focus could be placed on gaining knowledge of COTS product insertion and accompanying intellectual property rights. These uniformed officer interns can then return to the DoD to help implement the practices.

Conclusion

This literature review was commissioned with the objectives of better understanding COTS product implementation performance. It explored (1) the typical research types, contexts, research methods, target markets, and foundational theories utilized in COTS-based research, (2) policies, laws, regulations, and directives that govern the use of COTS, (3) the known barriers to COTS implementations, (4) the known success factors to COTS implementations, (5) the recommendations have previously been made with respect to COTS implementations, and (6) recommendations for more timely and more effective COTS implementations. From the literature emerged a framework of COTS product usage that should help to guide COTS product adoption decisions and to help manage COTS product implementations ex post.

These six aspects of COTS product implementations are crucial; they should help reduce program risks of poor performance, failure, cost growth, and schedule slippage. The gained knowledge should also help the DoD acquisition community to more effectively and more efficiently leverage COTS products in order to meet its mission mandates and retain a competitive advantage against existing and potential foes.

References

Alcide, D. W. (2006). *Transformation of organizational legacy logistics systems and facilitating an integrated lean enterprise: A case study within the United States Army* (Doctoral dissertation). Capella University.

Amadeo, K. (2017). FY 2017 Federal Budget: Obama,Trump, and enacted budgets. *The Balance*. Retrieved from https://www.thebalance.com/how-trump-amended-obama-budget-4128986

Baker, B. (2002). The fall of the firefly: An assessment of a failed project strategy*. Project Management Journal, 33*(3), 53–57.

Baldwin, K. (2007). DoD software engineering and system assurance: New organization—New vision. Presentation by USD(AT&L), Defense Software Strategy Summit. Retrieved from http://csse.usc.edu/events/2007/arr/presentations/baldwin.ppt

Beesley, L. G., & Cooper, C. (2008). Defining knowledge management (KM) activities: Towards consensus. *Journal of Knowledge Management, 12*(3), 48–62.

Bjornson, F. O., & Dingsoyr, T. (2008). Knowledge management in software engineering: A systematic review of studied concepts, findings and research methods used. *Information and Software Technology, 50*(11), 1055–1068.

Boudreau, M. W. (2006). *Acoustic rapid COTS insertion—Case study* (NPS-PM-06-041). Monterey, CA: Acquisition Research Program.

Brill, S. (2017, April). Donald Trump, Palantir, and the crazy battle to clean up a multi-billion-dollar military procurement swamp. *Fortune,* 78–90.

Carney, D. J., & Oberndorf, P. A. (n.d.). The commandments of COTS: Still in search of the Promised Land. Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University. Retrieved from https://acc.dau.mil/CommunityBrowser.aspx?id=24403&lang=en-US

Cassidy, S., & Stanton, P. (2017). Cybersecurity requirements clarified. *National Defense*. Retrieved from http://www.nationaldefensemagazine.org/articles/2017/3/21/cybersecurity-requirements-clarified

Cechich, A., & Piattini, M. (2007). Early detection of COTS component functional suitability. *Information and Software Technology, 49*(2), 108.

Charette, R. N. (2013, December 6). The U.S. Air Force explains its $1 billion ECSS bonfire. *IEEE Spectrum*. Retrieved from http://spectrum.ieee.org/riskfactor/aerospace/military/the-us-air-force-explains-its-billion-ecss-bonfire

Cohen, W. L., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly, 35*(1), 128–152.

ComponentSource. (2017). Retrieved from https://www.componentsource.com/

Couts, C. T., & Gerdes, P. F. (2010). Integrating COTS software: Lessons from a large healthcare organization. *IT Professional Magazine, 12*(2), 50–58.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–339.

Defee, C. C., Williams, B., Randall, W. S., & Thomas, R. (2010). An inventory of theory in logistics and SCM research. *International Journal of Logistics Management, 21*(3), 404–489.

Defense Innovation Unit Experimental (DIUx). (2017). DIUx quarterly results—Q2 2017. April 20, 2017. Retrieved from https://diux.mil/library

DoD Inspector General (DoDIG). (2004a). *Acquisition of the Boeing KC-767A Tanker Aircraft* (D-2004-064). Arlington, VA: Author.

DoD Inspector General (DoDIG). (2004b). *Contracting for and performance of the C-130J Aircraft* (D-2004-102). Arlington, VA: Author.

DoD Inspector General (DoDIG). (2006a). *Commercial contracting for the acquisition of defense systems* (D-2006-115). Arlington, VA: Author.

DoD Inspector General (DoDIG). (2006b). *Acquisition of the Joint Primary Aircraft Training System* (D-2006-075). Arlington, VA: Author.

DoD Inspector General (DoDIG). (2016a). *DoD needs to require performance of software assurance countermeasures during major weapon system acquisitions* (DoDIG-2016-082). Arlington, VA: Author.

DoD Inspector General (DoDIG). (2016b). *Audit of the DoD Healthcare Management System Modernization program* (DoDIG-2016-094). Arlington, VA: Author.

DoD. (n.d.). Defense Federal Acquisition Regulation Supplement (DFARS). Washington, DC: Author.

DoD. (2009). *Report of the Defense Science Board Task Force on integrating commercial systems into the DoD, effectively and efficiently—Buying commercial: Gaining the cost/schedule benefits for defense systems*. Washington, DC: Author.

Erwin, S. I. (2016). Pentagon taking a more serious look at off-the-shelf technology. *National Defense*. Retrieved from http://www.nationaldefensemagazine.org/articles/2016/12/12/pentagon-taking-a-more-serious-look-at-off-the-shelf-technology

Federal Acquisition Regulation, 48 C.F.R. § 3.502-1 (2017).

Finkenstadt, D., Hawkins, T., & Hosey, W. (2014). Growing contracting competency: Market intelligence cells. *Contract Management* (November), 20–31.

Ford, D. N., & Dillard, J. T. (2009). Modeling open architecture and evolutionary acquisition: Implementation lessons from the ARCI Program for the rapid capability insertion process (NPS-AM-09-043). In *Proceedings of the Sixth Annual Acquisition Research Symposium* (Vol. 2). Monterey, CA.

Gansler, J. S., & Lucyshyn, W. (2008). *Commercial-off-the-shelf (COTS): Doing it right* (UMD-AM-08-129). Monterey, CA: Acquisition Research Program.

GAO. (2006). *Efforts needed to address Air Force commercial acquisition risk* (GAO-06-995). Washington, DC: Author.

GAO. (2007). *Defense acquisitions: An analysis of the Special Operations Command's management of weapon system programs* (GAO-07-620). Washington, DC: Author.

GAO. (2017). *Actions needed to address five key mission challenges* (GAO-17-369). Washington, DC: Author.

Garber, R., Willen, B., Heckler, A., & Skarda, T. (2011). *Affordability in the U.S. Defense Department: A better approach to dealing with declining DoD budgets.* Atlanta, GA: AT Kearney.

GlobalPlatform. (2003). *Case study: Common Access Cards—Expanding the functionality of ID cards for the U.S. Department of Defense*. Retrieved from http://www.globalplatform.org/fcs/DoD_Case_Study.pdf

Google Patents. (2017). Retrieved from https://patents.google.com

Grandinetti, R. (2016). Absorptive capacity and knowledge management in small and medium enterprises. *Knowledge Management Research & Practice, 14*, 159–168.

Grant, J. (2000). *Ensuring successful implementation of commercial off-the-shelf products (COTS) in Air Force systems.* Ft. Belvoir, VA: Defense Systems Management College Press.

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly, 30*(3), 611–642.

Hall, J. G., & Rapanotti, L. (2016). *A design theory for software engineering* (Technical Report TR2016/01). Milton Keynes, England: Department of Computing and Communications, The Open University.

Hambling, D. (2017). Islamic State's new weapon of choice: Off-the-shelf drones. *Aviation Week*. Retrieved from http://aviationweek.com/defense/islamic-state-s-new-weapon-choice-shelf-drones

Hart, C. (1998). *Doing a literature review: Releasing the social science research imagination.* London: Sage.

Hawkins, T. (2016). *Antecedents and consequences of supplier performance evaluation efficacy* (WKU-CM-16-150). Monterey, CA: Acquisition Research Program.

Horowitz, B. M., & Lambert, J. H. (2006). Assembling off-the-shelf components: "Learn as you go" systems engineering. *IEEE Transactions on Systems*, Man & Cybernetics: Part A, 36(2), 286–297.

Incze, M. (2011). Lightweight Autonomous Underwater Vehicles (AUVs) performing coastal survey operations in REP 10A. *Ocean Dynamics, 61*(11), 1955–1965.

Jilani, J. K. (2008). Using COTS components in software development. *AIP Conference Proceedings, 1052*(1), 203–208.

Kakar, A. K. (2013). *Feature selection for evolutionary commercial-off-the-shelf* (Doctoral dissertation). Tuscaloosa, AL: University of Alabama.

Keil, M., & Tiwana, A. (2005). Beyond cost: The drivers of COTS application value. *IEEE Software, 22*(3), 64–69.

Kendall, F. (2015). *Compendium of annual program manager assessments for 2015*. Washington, DC: DoD.

Leedy, P. D., & Ormrod, J. E. (2005). *Practical research: Planning and design* (8th ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Lockheed-Martin. (2017). Next generation Aegis ballistic missile defense system successfully engages medium range ballistic missile target. Retrieved from http://news.lockheedmartin.com/2017-02-06-Next-Generation-Aegis-Ballistic-Missile-Defense-System-Successfully-Engages-Medium-Range-Ballistic-Missile-Target

Lorell, M. A., Payne, L. A., & Mehta, K. R. (2017). *Program characteristics that contribute to cost growth: A comparison of Air Force Major Defense Acquisition Programs*. Santa Monica, CA: RAND.

Maras, J., Lednicki, L., & Crnkovic, I. (2012). 15 years of CBSE Symposium: Impact on the research community. In *Proceedings of the 15th ACM SIGSOFT Symposium on Component Based Software Engineering* (pp. 61–70). ACM.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.

Morrow, D. (2010). No. 20: Integrating COTS: Lessons from recent ground vehicle acquisitions. Center for Strategic and International Studies, *Current Issues*, 20. Retrieved from https://csis-prod.s3.amazonaws.com/s3fs.../100115_DIIG_Current_Issues_n20.pdf

Naegle, B., & Petross, D. (2010). *P-8A Poseidon Multi-mission Maritime Aircraft (MMA) software maintenance organization concept analysis* (NPS-LM-10-006). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Rendon, R. (2007). Analysis of Modular Open Systems Approach (MOSA) implementation in Navy acquisition programs (NPS-AM-07-016). In *Proceedings of the Fourth Annual Acquisition Research Symposium.* Monterey, CA: Naval Postraduate School, Acquisition Research Program.

Rentsch, J. R. (2014). Written testimony, Aerospace Industries Association, June 16, 2014. Arlington, VA: Aerospace Industries Association.

Riposo, J., McKernan, M., & Duran, C. K. (2014). *Prolonged cycle times and schedule growth in defense acquisition: A literature review*. Santa Monica, CA: RAND.

Stein, F. (2006). TSM-Satcom update Ka-Band Satellite Transmit And Receive System AN-GSC-70(V). *Army Communicator, 31*(1), 31–33.

Tadjdeh, Y. (2017). Marine Corps investing in light vehicles to take the load off troops' shoulders. *National Defense*. Retrieved from http://www.nationaldefensemagazine.org/articles/2017/4/3/marine-corps-investing-in-light-vehicles-to-take-the-load-off-troops-shoulders

Torchiano, M., Jaccheri, L., Sørensen, C. F., & Wang, A. I. (2002). COTS products characterization. In *Proceedings of the 14th International Conference on Software Engineering and Knowledge Engineering* (pp. 335–338). ACM.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management, 14*, 207–222.

Tucker, P. (2017). The Army seeks internet-of-battlefield-things, distributed bot swarms. *Defense One*. Retrieved from http://www.govexec.com/defense/2017/07/army-seeks-internet-battlefield-things-distributed-bot-swarms/139551/?oref=govexec_today_pm_nl

United States Congress. (2016). National Defense Authorization Act of 2017.

U.S. Court of Federal Claims (USCOFC). (2016). PALANTIR USG, INC., Protestor, v. UNITED STATES, Defendant. Filed 3 Nov 16. No. 16-784C. 129 Fed. Cl. 218; 2016 U.S. Claims LEXIS 1716.

USDebtClock.org. (2017). Retrieved from http://www.usdebtclock.org/

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), 13–23.

Yin, R. K. (2009). *Case study research: Design and methods.* Los Angeles, CA: Sage.

Crossing the Valley of Death: The Case of the MDUSV

David N. Ford—PE, Graduate School of Business & Public Policy, Naval Postgraduate School, Monterey, CA. [davidford@tamu.edu]

John T. Dillard, COL, USA (Ret.)—Graduate School of Business & Public Policy, Naval Postgraduate School, Monterey, CA [jtdillar@nps.edu]

Abstract

Technology transition from research to "programs of record" (a.k.a. crossing the valley of death) has often been challenging, especially when new capabilities emerge that weren't originally envisioned, such as next-generation aircraft, fighting vehicles, and so forth. The recent evolution of unmanned aerial systems (UASs) is a good example of extemporaneous proliferation of new capabilities. These technology-driven advances may not fit into conventional paradigms of warfighting concepts and may have organizational and infrastructure impacts. The Anti-Submarine Warfare Continuous Trail Unmanned Vessel (ACTUV) project by DARPA built a prototype surface ship, christened *Sea Hunter*, that was tested in San Diego and then transitioned to the Office of Naval Research (ONR) at the end of 2017. It endeavored 70-day missions of up to 7000nm without a manned crew aboard. To cross the valley of death and transition to a Program of Record, a validated requirement must exist, along with funding for development/procurement across the Future Years Defense Program. The current research proposes and applies a framework for planning successful crossing of the valley of death to the current version of the ACTUV program, Medium Displacement Unmanned Surface Vessel (MDUSV). Results include important specific challenges, behaviors, methods, recommendations, and impacts on practice and research.

Context

Innovation is required to remain competitive in many domains, including commercial enterprises and national defense. Innovations are often classified as either incremental (e.g., increasing computer speed or sonar offset distance) or disruptive (e.g., smartphones, aircraft carriers). Disruptive innovations are distinguished from incremental innovations by their causing changes in the fundamental behavior of communities. Innovation of technologies is a knowledge development and technology application process that typically moves from understanding concepts and causal relationships in basic research through a series of discovery and development phases to a useful application of the technology.

Technology innovation is critical to the DoD fulfilling its mission "to provide the military forces needed to deter war and to protect the security of our country" (DoD, n.d.) by keeping American warfighters armed with materiel solutions that maintain competitive advantage over adversaries. Maintaining a steady stream of innovative materiel solutions requires the effective and efficient design and management of the technology innovation process. The technology transition "valley of death" (a.k.a. herein as "the valley") describes a particularly difficult part of the innovation process that lies near the middle of the journey from basic research to application.

The innovation process can be pulled forward by unmet needs, pushed forward by new technologies and capabilities, or both. In their study of innovation failure in the acquisition of the Unmanned Carrier-Launched Airborne Surveillance and Strike (UCLASS) system, Turner and Wickert (2016) describe the three DoD offset strategies based on their needs and development of technologies:

- The First Offset Strategy (Eisenhower's New Look) was driven by technologies being pulled forward to meet nuclear deterrence needs. Requirements preceded and defined innovation.
- The Second Offset Strategy (Cold War era) was driven by technology push as technologies, e.g., in stealth and precision strike, were developed independently and then integrated into a strategy. Innovation preceded requirements.
- The current Third Offset Strategy (autonomy[1] and artificial intelligence) reflects both the need to address current emerging threats from near-peer adversaries and also the fast evolution of new technologies.

The remainder of this paper is organized as follows. The valley of death is introduced, followed by a description of the challenges it creates and relevant extant theories and recommendations for crossing the valley successfully in the form of a framework that will be used to analyze the MDUSV program. The MDUSV program is then described as it relates to crossing the valley. The framework is applied to the program by specifying program challenges, behaviors, and methods. This leads to the recommendation of a new and unique organization which can address the MDUSV needs for crossing the valley. Finally, implications of the formation and use of the recommended organization are discussed.

Background

The Valley of Death

The *valley of death* is a metaphor for the difficulty experienced by innovators in transitioning technologies that have been successfully researched and initially developed into successful applications. The valley most often includes a lack of funding and other forms of development support to progress from late research and pre-materiel decision-making, through technology development, to application (Pusateri et al., 2015). The metaphor is applied to the experiences of a wide range of products, including both incremental improvements and disruptive innovations, in many industrial and public settings. As a major developer and user of new technologies, the DoD suffers greatly from the valley of death (National Research Council, 2004).

Graphical descriptions of the valley of death abound. One example of a simple depiction is from a Canadian natural resources agency (Figure 1).

[1] As defined by the DoD (2017, p. 15), autonomous vehicles and remotely controlled vehicles are mutually exclusive categories.

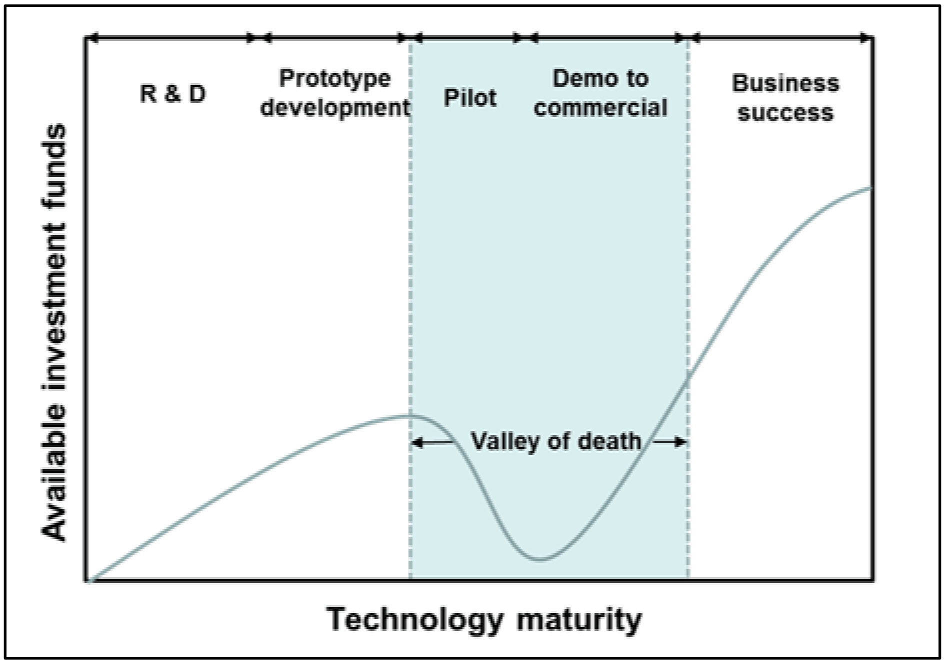


Figure 1. The Valley of Death in Technology Development
(Natural Resources Canada, n.d.)

Gunderson (2014) suggests that the “Hype Curves” developed by Gartner, Inc. also depict the valley. Generically, hype curves describe innovation life cycles with five phases: Innovation Trigger, Peak of Inflated Expectations, the Trough of Disillusionment (the valley of death), Slope of Enlightenment, and Plateau of Productivity. Figure 2 illustrates Gartner, Inc.’s hype cycle for emerging technologies.

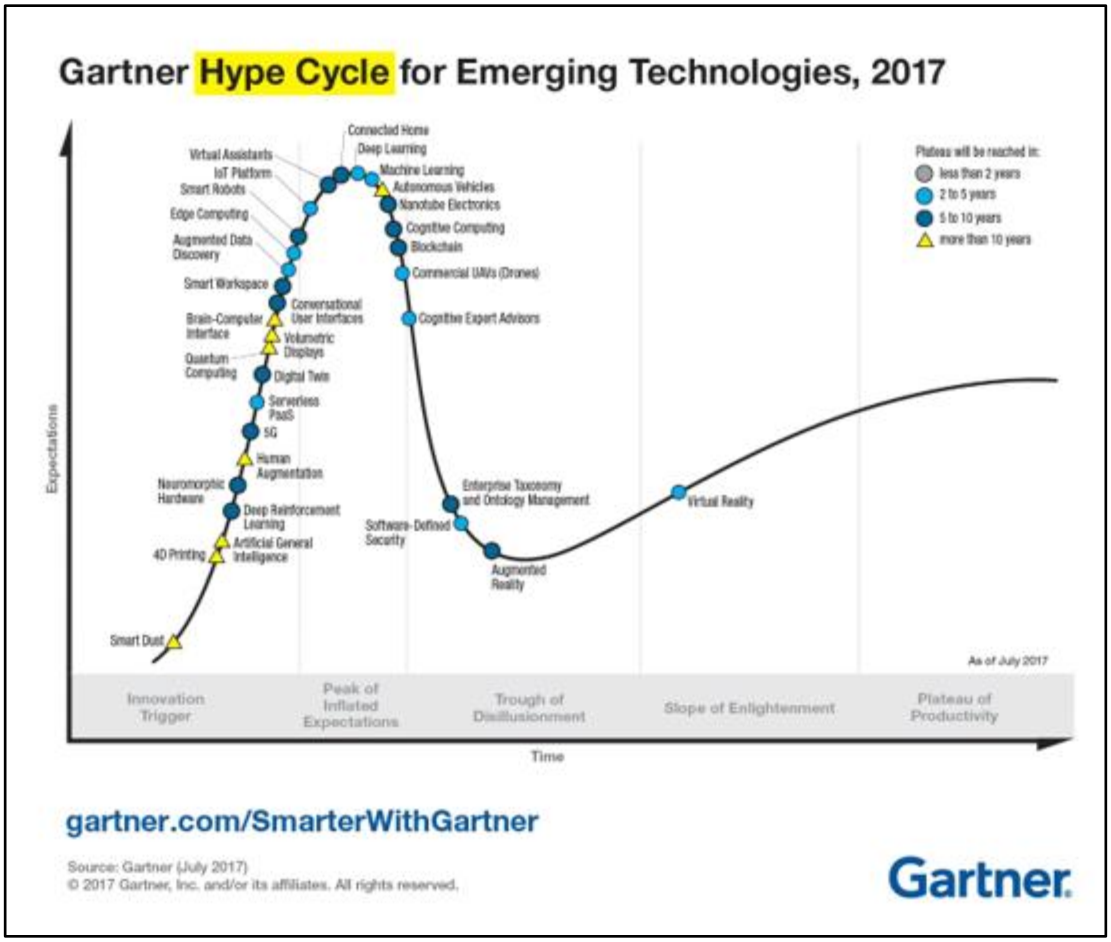


Figure 2. The Gartner Hype Curve for Emerging Technologies
(Panetta, 2017)

Much has also been written about the difficulties of technology transition in public (DoD) sectors, where successful transition is often elusive. The concept of a valley of death applies commercially as well, wherein products fail to be fully commercialized or adopted sufficiently by the market place. Products sometimes fail to successfully cross the valley of death to meet user needs. Newman (2018) notes that the iPad succeeded in 2010 only after the failure or much more limited success of the Microsoft Tablet PC in 2002, Microsoft Pocket PC 2000 in 2000, Intel Web Tablet in 1999, NewsPad in 1997, Palmpilot in 1996, Fujitsu Stylistic 1000 tablet in 1996, Apple Newton MessagePad in 1993, Compaq Concerto in 1993, EO Personal Communicator in 1991, GRIDPad tablet in 1989, Letterbug in1986, and others. Being able to describe the valley of death and efforts to facilitate crossing it is critical for Navy and DoD materiel acquisition. Recent operations in Iraq and Afghanistan have brought many capabilities to U.S. forces without going through the slow, disciplined, and burdened processes necessitated in peacetime. Developed and purchased with Overseas Contingency Operations funding, some of these discrete line items that are already distributed in the force are now having challenges getting fully authorized and resourced for sustainment as the DoD returns to normal peacetime operations. Similar challenges can face the development of new technologies and reuse of existing technologies in projects such as ACTUV/MDSUV.

The Medium Displacement Unmanned Surface Vehicle (MDUSV) Program

Program History

Medium Displacement Unmanned Surface Vessel (MDUSV) is the current moniker for the U.S. Navy's effort in autonomous technology demonstration, as it evolves toward a more fully defined set of capabilities. Formerly called the Anti-Submarine Warfare Continuous Trail Unmanned Vessel (ACTUV), implying early recognition for potential primary missions, it has completed its first phase of prototyping and experimentation, as an outgrowth of a Defense Advanced Research Projects Agency (DARPA) project.

In 2016–2017, the Defense Advanced Research Projects Agency (DARPA) demonstrated autonomous operation of a naval surface vessel in the Anti-Submarine Warfare Continuous Trail Unmanned Vessel (ACTUV) project. This project had three primary goals (Littlefield, 2017):

- Explore the performance potential of a surface platform conceived from concept to field demonstration under the premise that a human is never intended to step aboard at any point in its operating mission cycle.
- Advance unmanned maritime system autonomy to enable independently deploying systems capable of missions spanning thousands of kilometers of range and months of endurance under a sparse remote supervisory control model.
- Demonstrate the capability of the ACTUV system to use its unique characteristics to employ nonconventional sensor technologies that achieve robust continuous track of the quietest submarine targets over their entire operating envelope.

The project was also structured to explore and advance the potential of autonomous vessel performance, to include independent multi-mission operations over time and distance, with varying payloads. The prototype vessel DARPA produced, named *Sea Hunter*, was focused initially upon the anti-submarine (ASW) mission (thus its name). However, a wider range of missions and configurations came into view and are already envisioned for future experimentation and exploitation. Key questions to be answered going forward are not only within the business and technical realm of acquisition, but also the operational framework of future surface combatant operations covering a myriad of missions and concepts of operations (CONOPS). Not at all (to date) deemed an "orphan technology" in search of utility, the capability and cost savings perceived as apparent from autonomy, both in the near and far term, have already given rise to strong OPNAV advocacy and resource sponsorship, suggesting that MDUSV's successful crossing of the valley of death can make a significant contribution to naval surface warfare capabilities. In addition to the benefits of human life risk reduction and obvious life-cycle cost savings, perhaps the most compelling aspects of the autonomy concepts that are at the heart of the MDUSV are the opportunities to contribute to the yet-to-be-fully-defined Third Offset Strategy. A shrinking U.S. military force structure with declining technological superiority faces a current era of near-peer threats and military power competition. How can autonomy, and the MDUSV specifically, help to deliver a large quantity of relatively inexpensive, though technologically advanced, surface vessels to better augment and distribute U.S. forces and maximize survivability?

The first vessel to emerge from the MDUSV program was designated as the *Sea Hunter*, a Class III vessel (displacement of approximately 145 tons), of several displacement size classes, launched in 2016. *Sea Hunter* has since been undergoing sea trials and experiments along the western U.S. coast and throughout areas of the Pacific Ocean.

Transitioning this year to the Office of Naval Research for two years of further tests, Sea Hunter had completed its demonstration of over-water speed and stability, with system reliability during extended operations throughout 2017. Perhaps chief among these was compliance with maritime collision regulations (COLREGS). Prior to and during this period, Office of the Chief of Naval Operations (OPNAV) staff began their analysis of required capabilities that might be performed by multiple classes of autonomous surface vessels.

MDUSV Approaches the Valley of Death

Like the recent history of unmanned aerial vehicles, there are organizational, cultural, and doctrinal, as well as business and technological barriers to the acceptance and employment of new technologies like *autonomy*. Autonomy actually represents a spectrum of unmanned systems spanning those under human remote control to systems with sparse or no human supervision. Having one prototype built for testing and experimentation, the ACTUV working group has successfully brought together the appropriate stakeholders to ensure a successful crossing of the technology transition “chasm,” or valley of death. However, substantial uncertainty lies ahead for them and the larger naval force it seeks to serve.

In response to this need, the Naval Postgraduate School (NPS) hosted an ACTUV Workshop on February 14–15, 2017, to host representatives from practically every community that might constitute stakeholders. Organizations represented were the Office of Naval Research, DARPA, Leidos Corporation (DARPA’s ACTUV prime contractor), PEO Littoral Combat Ships, OPNAV N96, N9I, SPAWAR, and multiple departments of NPS. The workgroup was fortunate to have actual “owners” participate, as is necessary for successful technology transition. The resource sponsor, N96, is perhaps the most important of these. But also important are members of the Science and Technology (S&T) community, who must transition the outputs of the DARPA project seamlessly into an extended phase of testing, and then their eventual transfer to an existing program office structure such as PMS 406, Unmanned Maritime Systems. An author of this research served, along with N96, as co-leads of the Acquisition Breakout Session, whose task it was to brief the 2017 workshop on progress made with identification of challenges for technology transition (i.e., in crossing the valley of death). The acquisition strategy breakout session of the two-day work group sought to identify technical and business challenges associated with the technology transition of ACTUV from its current status as a DARPA project to become an official “Program of Record.” (This term refers to a program with its own line of funding in the Future Years Defense Plan, a database of programmed funds, and denotes also that it has been formally initiated with a Milestone B decision (thus necessarily having a validated requirement in the form of a Capability Development Document).) History is replete with examples of promising technologies not being able to cross a mythical valley of death or technology transition or commercial marketing “chasm.” Over a dozen of these typical challenges were pulled from existing literature about DoD tech transfer and used for a group discussion on a later breakout session.

The first output of this breakout session’s work was recognition that *a substantial number of accomplishments had occurred to date*. Not only has technology been demonstrated with obvious revolutionary capabilities, but opportunities are easily envisioned for cost savings as well within a mixed fleet of manned and unmanned naval service vessels. Significant is the amount of user interest and support already evident, extending to the highest levels of the Navy. Depending upon outputs from other working groups to further develop operational concepts, roles, and missions, it is already apparent that autonomous vessels such as ACTUV, now MDSUV, can become a force multiplier.

Developing capabilities is a critical part of crossing the valley of death. The initial configuration of ACTUV already allowed for multi-mission payloads. Besides anti-submarine warfare, other "dull and dirty" missions are emerging, such as mine/countermine operations, long-haul resupply, etc. During the 2017 workshop, an N96 representative revealed that the Navy staff had conducted Capabilities Based Analyses (CBA) to verify the need for unmanned and autonomous vessels, with follow-on analyses of alternatives (AoA) proceeding through FY19, and then development of an overarching Initial Capabilities Document (ICD) to follow. This formalization of a validated requirement document is key in establishing service needs, which drive the acquisition process, and help establish programs of record. He also described plans for a "development squadron" (DEVRON) to be in place by FY20 to further demonstrate technologies for basic missions to at least levels of Technology Readiness Levels (TRL) 4–5.

Lastly, the resource sponsor assured that continued funding will be reflected in the Future Years Defense Plan (FYDP), Future Service Combatant (FSC) line. From an acquisition process point of view, intellectual property should not be a substantial issue when later development efforts are ready for competitive procurement, as the ACTUV source code and other technical data are believed to be either nonproprietary or otherwise in-house within the government.

Important results from this 2017 workshop were the introduction of key players and cementing of their partnerships and respective responsibilities for the near future. However, while technology transition was discussed, much remained, and still remains, to fully develop a technology transition plan. Challenges include the need for a validated Joint Capabilities Integration and Development System (JCIDS) requirement and full funding for development and procurement across the Future Years Defense Program (FYDP). Entering advanced development and low/full rate production will necessarily include potential paradigm shifts regarding system autonomy in the Navy as it ascertains missions and operational concepts for integration into a mixed fleet of future surface combatant vessels.

A Framework for Crossing the Valley of Death

The proposed framework herein structures the relevant literature on the crossing of the valley of death into three perspectives: (1) challenges faced in crossing the valley, (2) behavior modes in crossing the valley, and (3) methods for crossing the valley. Specifying the components of the framework for individual technology transition programs can facilitate crossing the valley.

Challenges in Crossing the Valley of Death

Many challenges make crossing the valley of death difficult. In addition to the development of the underlying technologies, Newman (2018) identifies the development of the technologies, manufacturing readiness, which addresses the feasibility and affordability of producing the technology at the required scale and rate, and the integration of the technology into other, larger systems. Within the DoD, the *Manager's Guide to Technology Transition in an Evolutionary Acquisition Environment* (DoD, 2005) describes technology transition challenges, many that apply to crossing the valley of death. These challenges are organized around the three types of problems: (1) technology transition, (2) cultural barriers, and (3) knowledge management. More specific challenges identified in the *Manager's Guide* include the following:

- The technology may not develop rapidly enough to be ready when it is needed (p. 4-23)
- A focus on a preferred solution may prevent the adoption of better solutions

- Designs may not adequately incorporate needed future upgrades (p. 4-6)
- A suboptimal technology may be chosen (pp. 4-11–4-12)
- Teaming is critical (p. 4-22)

Resistance from within innovating organizations can make also crossing the valley difficult. Established organizations and systems often tend to support retaining the status quo by opposing the development and adoption of disruptive technologies. For example, Turner and Wickert (2016) describe the erosion of requirements of the Navy's Unmanned Carrier-Launched Airborne Surveillance and Strike (UCLASS) system, a potentially-disruptive next-generation unmanned combat system, until the system was diluted into a tanker for traditional manned missions. Other challenges can include users not being ready for a mature-enough technology (e.g., inline skates were available in 1972, but Rollerblades did not become popular until 15 years later; Newman, 2018) or regulatory approvals lagging technology and demand (e.g., UAV pizza delivery).

The proposed framework structures these challenges of crossing the valley of death into three potential bottlenecks:

- **Technology development**, including performance, speed of development and phasing of development, and costs.
- **Application development**, including identifying and developing needs and uses and capabilities, individually and with other systems, and matching technologies to uses.
- **Overcoming social resistance**, including addressing inertia and support of the status quo in users, sponsors, and regulators, providing adequate knowledge transfer communication and partnering to no constrain progress, and maintaining forward momentum.

Identifying which of three types of challenges best describes a specific issue or need can assist in identifying what organization or persons has the knowledge, skills, capabilities, and capacity to best address the issue or need.

Behavior Modes of Crossing the Valley of Death

Gulbrandsen (2009) described two behavior modes of crossing the valley, a linear process and a social process. They refer to the linear approach as "Mode 1" and the social approach as "Mode 2."

Crossing the Valley as a Linear Process

The linear behavior model of crossing the valley of death (Mode 1) is process-based and objective. In this behavior mode, the transition from research to application moves through a sequence of phases, evolving from basic science to applied science to "development" to production (Mirowski and Sent, as cited in Gulbrandsen, 2009, p. 20). Gulbrandsen quotes Gibbons et al.'s (1994) description of this behavior mode:

> Mode 1 is discipline-based and carries a distinction between what is fundamental and what is applied; this implies an operational distinction between a theoretical core and other areas of knowledge such as the engineering sciences, where the theoretical insights are translated into applications. (as cited in Gulbrandsen, 2009, p. 4)

Although the linear behavior mode includes negotiations about the evolution of solutions, participants are (in theory) objective, and all participants are guided by meeting the goals of the project.

The successful use of the linear process for crossing the valley of death requires clear, specific, shared, and enforced methods and measures of how technologies will be developed and how other aspects of innovation will be managed.

Crossing the Valley as a Social Process

In contrast to the linear behavior mode, the social behavior mode (Mode 2) is a highly interactive process in which a trans-disciplinary team negotiates and renegotiates the technology. For example, an inter-disciplinary project team (IPT) may re-conceptualize a solution, move to a different technology if progress on the first technology chosen stalls, or change how the technology will be used, thereby redefining the nature of the materiel solution. In the social process, knowledge production is characterized by a diffuse trans-disciplinarity that can blur the lines between disciplines and the traditional stages of acquisition. Progress within the social process can interrupt progress within the linear process. An example is if discussions with users were to identify a previously unidentified but very valuable potential use (the social process) that requires changes in the performance targets for technology development (the linear process).

Successful use of the social process for crossing the valley of death requires the socializing of ideas across the various and diverse organizational participants in the innovation effort. This requires the establishment and maintenance of linkages and relationships across organizational boundaries and between participants with differing local objectives (e.g., cost control vs. speed of innovation vs. risk reduction), and methods. Social processes are notoriously challenging and the failure to manage them can slow and stop momentum in innovation. Therefore, crossing the valley successfully using a social process is based on relationships within an IPT and others (e.g., contractors, research organizations) and collaboration among stakeholders who hold varied interests. According to Doheny-Farina (1992),

> At their core these processes involve individuals and groups negotiating their visions of technologies and applications, markets and users in what they all hope is a common enterprise. This means that the reality of a transfer does not exist apart from the perceptions of the participants. Instead, the reality—what the transfer means to the participants—is the result of continual conceptualizing, negotiating, and reconceptualizing. (as cited in Gulbrandsen, 2009)

Posen (1984) supports the need for a social process in military innovation by suggesting that it requires internal champions and pressure from commercial stakeholders. The *Manager's Guide* (p. 4-5) says that crossing the valley requires a partnership among communities such as S&T, R&D, PM, capability needs, T&E, sustainment, and financial.

A superficial understanding of innovation reveals the need for a combination of linear and social processes to successfully cross the valley of death. The proposed framework describes the following aspects of behavior modes for crossing the valley of death:

- **The linear processes used for crossing the valley**, including identifying and describing actual practice vs. espoused processes and gaps between (resource constrained) practice and needed practices.
- **The social processes used for crossing the valley**, including identifying and describing practice vs. espoused processes and gaps between social practices within the IPT and practices needed, and places where the social process is likely to interfere with linear process.

- **Interactions of the linear and social processes in crossing the valley**, including identifying and describing places where the linear process is likely to interfere with social processes, identifying and describing places where the social process is likely to interfere with linear process, and means of managing those interfaces.

Identifying which of three types of behavior modes best describes the actual or desired process for addressing a specific issue or need can assist in identifying what organization or persons has the knowledge, skills, capabilities, and capacity to best address the issue or need.

Methods for Crossing the Valley of Death

Many commercial and military innovation accomplishments demonstrate that innovative organizations can successfully cross the valley of death. See the DoD's *Manager's Guide* (DoD, 2005), Pusateri et al. (2015), and "The latest unmanned drone" (2017) for military examples. The literature also recommends how to do so.

Christensen (2003) recommends three strategies for crossing the valley in the case of disruptive technologies:

- Targeting underserved early adopters who are less committed to legacy systems, thereby gaining adoption without threatening the status quo. Williams and Gibson (1990) refer to this approach as "dissemination."
- Provide solutions that are superior to the status quo (the "better mousetrap" approach). Williams and Gibson (1990) refer to this approach as "appropriability."
- Introduce the innovation gradually, first through familiar methods and settings to accelerate adoption and reduce resistance from those defending the status quo. In addition, Williams and Gibson (1990) observed facilitation of the interfaces among stakeholders through communication as a means of crossing the valley.

Tippens (2004) contrasted "high-velocity" technology firms that successfully cross the valley of death with those that hold onto technologies into obsolescence. The former had the following:

- Short, iterative processes
- Collaborative concurrent development
- A passionate focus on user needs
- A willingness to take risks
- Early and rapid prototyping

Within the DoD, Pusateri et al. (2015) developed a Joint Transition Planning Process with supporting meetings and a working group for crossing the valley in DoD medical development. Their process positions products in late-stage S&T for successful transition to AD, thereby facilitating, without replacing, current processes. Meetings structure and improve IPT communication, particularly awareness of progress and technology transition issues. The working group is like a temporary IPT that focuses on technology transition. Its activities can include assessments of status, analysis of alternatives, and program management. They emphasize communication across parts of the IPT and processes and document multiple successes using this Joint Transition Planning.

Lewis, in his analysis of the DoD's Third Offset Strategy (Lewis, 2017) recommends that DoD be a "fast follower" (of commercial efforts) instead of a first mover in acquiring autonomy and artificial intelligence (AI) technologies. First movers are organizations that initially invest in and develop a new technology. History has shown that first movers are often overtaken by fast-followers, organizations that refine a technology based on the work of the first mover and quickly adapts them for application. Lewis provides examples of information technology (IT) products that became dominated by fast followers, including Google (fast follower) superseding AltaVista in search engines and Excel superseding Lotus 123 in spreadsheets. Fast followers are particularly likely to move past first movers in environments characterized by rapid innovation, such as autonomy and AI. Lewis suggests how the DoD can be an effective fast follower of autonomy and AI technology to accelerate acquisition, including crossing the valley of death. Critical acquisition capabilities for doing this include deep learning about specific technologies (effective following) and judiciously increasing government risk-taking to accelerate acquisition processes (move faster). Lewis's specific recommendations that can help in crossing the valley include the following:

- Develop internal autonomy and AI expertise
- Track and use specific commercial technologies
- Track technology develop by others
- Learn from other related DoD efforts
- Build interoperability into autonomous and AI systems

Although the need to be a leader in the application of autonomy and AI in the MDUSV program may preclude the adoption of a fast follower strategy, some of Lewis's recommendations may be effectively applied to MDUSV crossing the valley of death.

The proposed framework structures the recommended methods for successfully crossing the valley of death described above according to how they address the three types of challenges described above:

Technology Development

- Provide better solutions
- Collaborate and facilitate stakeholder interfaces
- Iterate early and fast
- Be willing to take risks
- Hold and keep deep knowledge of technologies

Application Development

- Target underserved users and needs
- Collaborate and facilitate stakeholder interfaces
- Iterate early and fast
- Focus on user needs
- Be willing to take risks

Overcoming social resistance

- Introduce innovations gradually
- Focus on user needs

Identifying which type of method can best address a specific issue or need can assist in identifying what organization or person has the knowledge, skills, capabilities, and capacity to best address the issue or need. In addition, recommendations can be further disaggregated for analysis into those that apply the linear behavior mode, the social behavior mode, or are at an interface between the linear and social behavior modes, thereby integrating the first and second part of the framework.

Summary of Proposed Framework for Crossing the Valley of Death

Challenges of crossing the valley of death

- Technology development
- Application development
- Overcoming social resistance

Behavior modes for crossing the valley of death

- The linear processes used
- The social processes used
- Interactions of the linear and social processes in crossing the valley

Methods for crossing the valley of death

Technology Development

- Provide better solutions
- Collaborate and facilitate stakeholder interfaces
- Iterate early and fast
- Be willing to take risks
- Hold and keep deep knowledge of technologies

Application Development

- Target underserved users and needs
- Collaborate and facilitate stakeholder interfaces
- Iterate early and fast
- Focus on user needs
- Be willing to take risks

Overcoming Social Resistance

- Introduce innovations gradually
- Focus on user needs

Disaggregate recommended actions into those aspects that apply the linear behavior mode, social behavior mode, or interface between those modes.

Application of the Framework to the MDUSV Program

MDUSV Challenges in Crossing the Valley of Death

The previously described 2017 ACTUV workshop and the successive interdisciplinary MDSUV working group series of sessions at NPS in 2018 continues to identify what these authors see as eight primary groups of challenges, incomplete work, or simply important things that needed to be accomplished. They are summarized and further disaggregated into 14 more challenges in our framework of categorization.

Technology Development

Processes

- Four different milestone decision documents need to be produced along with a Navy roadmap for future service combatants.
- Upon completion of S&T activities, the PMO will construct a full acquisition strategy for what will probably be a traditional acquisition approach to development. If technology enablers have not at that point been demonstrated to TRL 6-7, a Technology Maturation and Risk Reduction phase may be needed before Engineering and Manufacturing Development.
- Moving to a common test and evaluation "scorecard" is a challenge with regards to safety etc.
- Traditional acquisition strategies contain a myriad of elements including life-cycle cost estimate, contractual competition, cyber security, etc. This all feeds into the contractual scope and type of transaction vehicle for continued industry efforts.
- Quantities of initial and final on buys for full operational capability must be planned and programmed.
- Specific feature sets for a completely configured system are needed to drive technical specifications and requirements.
- There was no way to ascertain even a rough order of magnitude for the cost of a follow-on prototypes. But it could be presumed that at least initial buys would be analogous to the first vessel of $20 million, with an added payload of $3 million, spanning 24 months of time to produce.
- Future costs will rise with complexity, but production quantities and production schedules should certainly achieve some economies of scale commensurate with what we see across other systems/platforms.

Products

- Autonomous tactics and behaviors are still conceptual and not fully mature.
- Endurance and reliability of autonomous vessels, amounts of corrective maintenance actions, etc. are yet to be proven
- Several more years of development are needed for the maturation of autonomous technologies, especially for more complex missions.
- The resource sponsor should avoid a hiatus or loss of momentum by providing continuous funding for FY 18 and beyond.

Application Development

- Four different milestone decision documents need to be produced along with a Navy roadmap for future service combatants.
- A documented and validated requirement must be developed with missions and operational concepts fully identified.
- The need for interoperability with other systems and platforms demand that some top level requirements emerge for common command and control.
- The resource sponsor should avoid a hiatus or loss of momentum by providing continuous funding for FY 18 and beyond.

- Moving to a common test and evaluation “scorecard” is a challenge with regards to safety etc.

Overcoming Social Resistance

- Key to the development of the four different milestone decision documents is the maintenance of dialogue amongst key players previously mentioned. The idea of a Tech Transfer Agreement (TTA) might facilitate this as a formal memorandum of sorts.
- The need for interoperability with other systems and platforms demand that some top level requirements emerge for common command and control.

MDUSV Innovation Behavior Modes

The MDUSV program is next described based the three portions of the “behavior modes for crossing the valley of death” portion of the framework: (1) a linear innovation process, (2) a social innovation process, and (3) interactions between linear and social innovation processes.

The Linear Innovation Behavior Mode Used in the MDUSV Program

The DoD acquisition process (Figure 3) is an example of a linear behavior mode of innovation. In this process work is done to add knowledge about materiel solutions to move those solutions from S&T and Major Decision A, through Technology Maturation & Risk Reduction, Major Decision B, Engineering & Manufacturing Development, and into Production.

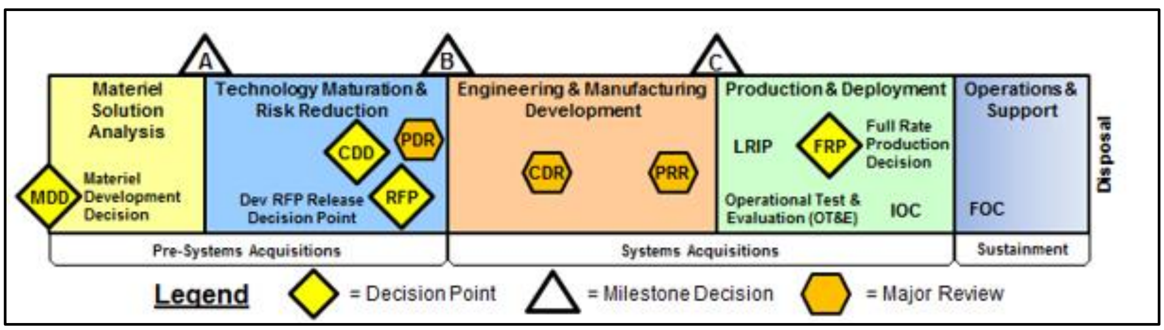


Figure 3. The DoD Acquisition Process

The MDUSV program will implement the DoD acquisition process and thereby be deeply embedded in a linear behavior mode. Comments at the 2017 workshop and the 2018 series of working group sessions that have continued at NPS also strongly support the use of a linear process in MDUSV innovation, primarily in discussions of the design and management of the DoD acquisition process for the MDUSV program. These addressed milestone decision-making, the development of an acquisition strategy and requirements, metrics (the “scorecard”), and purchase quantities.

The Social Innovation Behavior Mode Used in MDUSV Program

The NPS working group sessions have provided an important means of exploiting social innovation in the MDUSV by introducing the key players to each other and providing a setting within which they could openly discuss program issues, including crossing the valley of death. 2017 Workshop products described the need for and role of a social innovation process in the MDUSV program. For example, the considerations included the following questions:

- Any joint or Interoperability/interdependency aspects of the system?
- Who is the likely Navy sponsor of ACTUV?
- Where should the JCIDS CDD/CPD be initiated?
- Risk Areas (Programmatic/Technical, Need, Funding)?

These are areas where decisions will be based largely on discussions among program stakeholders, making the social processes critical to program success. Some of these have emergent resolutions underway already.

Interactions of the Linear and Social Innovation Behavior Modes in the MDUSV Program

Interactions between the linear and social innovation processes will likely be one of the most critical requirements for MDUSV successfully crossing the valley of death. Participants have noted that "Key to the development of the four different milestone decision documents is the maintenance of dialogue amongst key players previously mentioned," and suggested a solution in "the idea of a Tech Transfer Agreement (TTA) might facilitate this as a formal memorandum of sorts." This illustrates that participating stakeholders understand and appreciate the importance and challenges of designing and managing the interfaces between *linear* and social innovation in MDUSV. Notwithstanding the necessary adherence to *linear* processes inherent in the DoD acquisition management structure, it will be the *social* communication and coordination among stakeholders that, if maintained, will ensure safe passage across the valley of death."

MDUSV Methods for Crossing the Valley of Death

The 2017 workshop and follow-on sessions have also generated recommendations, including the following:

- Leverage the S&T community's time and efforts as much as possible to flesh out requirements, doctrinal concepts and to perhaps resolve other legal and ethical concerns.
- Use this time to explore innovative contracting methods (such as the use of development/production options) and anything else that will alleviate bureaucracy and allow development to continue.
- Manufacturing Readiness Levels (MRL) are also an aspect to be considered, given the relative certainty about the type of vessel this will be.
- Prepare for the use of prudent business practices, commensurate with an investment of this size, will be required under the DoD 5000 series instructions and Federal Acquisition Regulation, etc. due to the program becoming a Major Defense Acquisition Program (MDAP).
- Address where such vessels will fit within the operational architecture of doctrinal war fighting. These vessels have great potential for lower unit cost, huge savings during their operation and support phase, and the saving of lives that aren't placed into harm's way for missions performed by them.

- Legislative changes in acquisition reform over the past year afford several new areas of relief to challenges identified.
    - **Investment Decision Authority**—Lowest level Milestone Decision Authority is another recent legislative change that can speed the attainment of autonomous surface vessel capability. Service-level Decision Reviews can be minimized, along with the costly multi-level and adjacent agency preparatory briefings that have added off-core scope activities to program managers and hindered timely progress in the past.
    - **Abbreviated Documentation**—Along with a lowered threshold for decision-making, the dozens of bureaucratic documents traditionally required for milestone review should be consolidated and abbreviated where possible to fulfill the steps necessary for sensible but pragmatic satisfaction of information needs for decision-making.
    - **Simplified Contracting**—Recent legislation now allows Other Transaction Authority (OTA) in lieu of Federal Acquisition Regulation contracting instruments from prototyping through production phases of acquisition. However, care must be taken in the exercise of prudence for legal compliance regarding competition, rewards and incentives structuring, scope of work specification, performance measurement, etc. to avoid pitfalls already being seen in acquisitions attempting to exploit this method of shortening transaction timelines. Cautions are in the areas of proprietary hardware and software from the selected industry partner. Modular Open Systems Architecture should be emphasized in both business as well as technological functions.
    - **Tailored Acquisition Strategy**—Leveraging of the DARPA project results, along with ONR's experimentation and sea trials should alleviate the necessity for a Technology Maturation and Risk Reduction Phase in the traditional model of acquisition. With validated need statement (CDD) and FYDP funding programmed, transition directly into Engineering and Manufacturing Development Phase, with concurrent Low Rate Initial Production should be approved by Milestone Decision Authority, given match exists between requirements and resources, and finalized designs giving assurance of capability attainment.
    - **Streamlined Test and Evaluation—**As with contracting instruments, a balance must be struck among elements of good prudence and due diligence versus testing to the point of unnecessary extremes. That MDUSV is unmanned, except for occasional maintenance and back-up functions that may become necessary, justifies a lesser expenditure of resources for suitability factors such as safety and survivability, while nonetheless stressing system performance and reliability.

Notwithstanding these changes that could facilitate crossing the valley, a few acquisition imperatives remain. The recommendations for activities to be conducted in parallel, as the remaining months of sea trials and experimentation continue, under the auspices of ONR, before the hand-off to PMS 406 are as follows:

- **Requirements Capture and Refinement**—Requirements definition should be better informed from experimentation efforts, with evolutionary growth of capabilities planned for the technology enablers that are identified as not yet fully mature, especially when such are defined along a range or spectrum (versus binary attainment) of performance. The *Sea Hunter* now afloat will give insights into multiple capability payload packages for various missions and concepts of operations (CONOPS). Prototyping and sea trials that are now moving along in parallel with JCIDS efforts for Initial Capability Development documentation and formal validation already constitute a large advantage in the transitioning of technology across the valley of death.

  - **Maximize Modeling and Simulation**—For early requirements and product realization ranging from Force-on-Force simulations to computerized design and platform integration, M&S efforts will pay dividends along the entire path of MDUSV development for operational employment utility, anomaly discovery and test scope or sample size reduction.
  - **Disciplined Systems Engineering**—There are seldom shortcuts with regard to the necessarily disciplined engineering efforts at system and sub-system level. And systems engineering processes have proven their value for issue discovery venues, risk management, configuration control and technical performance measurement along the iterative development path that attacks complexity and resolves uncertainty. However, such need not impede progress in technology transition.

To apply the framework, these recommendations and other characteristics of the program and acquisition process were organized into a two dimensional matrix that aggregates recommendations from the literature according to the type of challenge addressed and identifies which behavior mode is used to apply specific recommendations for MDUSV (Table 1). The result facilitates analysis of the program plan for crossing the valley and the identification of methods and behavior modes that may not be being applied but could facilitate crossing the valley.

Table 1. Application of Crossing the Valley Framework to MDUSV Recommendations

Methods for Crossing the Valley of Death	Behavior Modes for Crossing the Valley of Death: Linear Behavior Mode	Social Behavior Mode	Linear and Social Behavior Modes
Technology Development			
Provide better solutions	Analysis of Alternatives	Maintain dialogue among key players	Tech Transfer Agreement among key players, Maintain dialogue
Collaborate and facilitate stakeholder interfaces	**Explore innovative contracting methods**	***Use S&T time to develop requirements, concepts, etc.***	Tech Transfer Agreement among key players, Maintain dialogue
Iterate early and fast	**Use S&T time to develop requirements etc. Prepare for MDAP**	Introduce users (DEVRON) to prototypes	
Be willing to take risks	Develop full acquisition strategy Analysis of Alternatives	Maintain dialogue among key players	Tech Transfer Agreement among key players, Maintain dialogue
Hold and keep deep knowledge of technologies	Rapid and iterative prototyping; Fast follower of commercial efforts	Collaborate with industry to become "Fast follower" of commercial efforts	
Application Development			
Target underserved users and needs	Develop requirements for interoperability, command, and control.	***Consider operations with other vessels***	
Collaborate and facilitate stakeholder interfaces		Maintain dialogue among key players	Tech Transfer Agreement among key players, Maintain dialogue
Iterate early and fast	**"Fast follower" of commercial efforts**	Common Test and evaluation scorecard	
Focus on user needs	Develop CONOPS, requirements, & features	Maintain dialogue among key players	Tech Transfer Agreement among key players, Maintain dialogue
Be willing to take risks	Develop full acquisition strategy Analysis of Alternatives	Maintain dialogue among key players	Tech Transfer Agreement among key players, Maintain dialogue
Overcoming Social Resistance			
Introduce innovations gradually	Evolutionary acquisition as increments of desired capability emerge	Introduce users (DEVRON) to prototypes	Modeling and simulation efforts for operational scenarios as well as technical parameters
Focus on user needs	Develop requirements for interoperability, command, and control.	Maintain dialogue among key players	Tech Transfer Agreement among key players, Maintain dialogue

The matrix provides a starting point for the analysis and design of the MDUSV program's preparation for crossing the valley of death. Through review and revision, key players can improve the description by adding information. Blank cells can be used to identify methods (rows) and means (columns) that are not currently being used to consider additional efforts to accelerate innovation. Descriptions within specific cells can be the basis of discussions among relevant program participants about challenges, behavior modes, and methods of crossing the valley of death.

Recommendations for MDUSV Crossing the Valley of Death

The application of the framework to the MDUSV program suggests several aspects of the program's crossing of the valley of death that may be improved.

These improvements could take many forms. Typically they will cross the process categories of linear, social, and interfaces between linear and social innovation processes. As an example, Gallup, Trask, MacKinnon, Wood, and Dillard (2018) proposed a specific method for managing the critical interfaces during innovation of MDUSV under the title "Coordinating a Multi-Organization Research and Development Program to Enable MDUSV Acquisition." After describing the program and the roles of its primary organizations in general, they describe a coordination challenge that threatens to prevent the program from crossing the valley of death and provide an illustration:

> All {primary organizations} agree with the need to incorporate unmanned systems in the future naval force but no one office is in charge of putting all the pieces together to provide a solution at a known point in time. Because the operational community has not documented and validated specific mission requirements for the design parameters of MDUSV, the acquisition community is not yet able to initiate a program to acquire MDUSV. The overall effort lacks organization, strategic alignment and an understanding of the inherent roles each organization must play to bring the MDUSV concept to fruition.
>
> These complexities are illustrated by the recent investment of $120M by the Special Capabilities Office (SCO) in Project Overlord, with the intention of creating one ship that will demonstrate some autonomy. This objective has already been proven and is being tested through the DARPA ACTUV/ONR MDUSV program which is being further enhanced by the commitment to build a second hull for testing and development. Expending resources on Project Overlord provides the illusion of progress while treading ground already covered. More could have been accomplished if SCO had invested these funds in the MDUSV program.

Gallup et al. (2018) then propose a realistic means of overcoming this challenge:

> The solution proposed is to create a SECNAV-approved consortium of organizations, cross-functionally responsible for conducting research and development activities so that each is solving an essential element necessary to make MDUSV operational at the earliest possible date. The organizational structure should be headed by a SECNAV-level office with the following organizations participating:
>
> - N96, N2/N6, ONR, SPAWAR, NPS, Naval War College, NRL, SCO, and universities funded to pursue technical, operational, and acquisition research as directed.

- The coordinating office at SECNAV would grant authority to member organizations to use “other transaction authority” to secure contracts with commercial vendors such as Boeing, Leidos, IBM, and others.[2]
- SECNAV office would take responsibility for coordinating the effort and protecting/adding funds as necessary to achieve goals and stay on schedule.

Tool and methods such as the one proposed by Gallup et al. (2018) can greatly facilitate crossing the valley of death by creating and maintaining linkages across diverse parts of the innovation effort (users, developers, funders; challenges, behaviors, and solutions; technology development, application development, social resistance). Such a method would facilitate the purposeful and planned incorporation of platform flexibility that would allow fast adoption of existing technologies, near adoption (e.g., in 10–15 years) of developing technologies, and the adoption of currently-unknown technologies in out years. Doing so would provide the justification for continued development and realistic bases for forecasted cost savings and operational improvements in the future.

Conclusions

The current work describes the technology transition valley of death and the challenges in crossing it based on the literature and background on the ACTUV/MDUSV program as relative to same. A three-part framework for the analysis and design of crossing the valley is proposed and then applied to the current MDUSV program. Potential uses of the framework products are described. A specific example of a recommendation, as viewed through the lens of the framework, is provided, and how it can facilitate the program crossing the valley. Additional development of the framework for describing, analyzing, and designing program’s crossing of the valley of death is recommended.

The current work can impact practice through the MDUSV program. It provides an initial evaluation of the MDUSV plan for crossing the valley of death. This predicts where the program may encounter challenges and suggests underlying causes such as coordination across linear and social innovation behavior modes. Those challenges and underlying causes can be used by program leaders to identify, design, and implement solutions, thereby speeding the crossing of the valley.

The current work impacts research on the crossing of the valley of death by proposing and initially testing a framework for analyzing and designing a DoN program’s crossing of the valley of death. This framework can be expanded and improved based on other programs and tested through application to other programs. By doing so, a valuable tool for acquisition can be developed and applied.

---

[2] “*Other transactions”* is the term commonly used to refer to the 10 U.S.C. 2371 authority to enter into transactions other than contracts, grants, or cooperative agreements. The Department currently has temporary authority to award ‘other transactions’ (OTs) in certain circumstances for prototype projects that are directly relevant to weapons or weapon systems proposed to be acquired or developed by the Department” (Under Secretary of Defense [AT&L], 2000, p. 7). OT is used by DARPA to speed contracting necessary for rapid prototyping.

References

Christensen, C. (2003). *The innovator's dilemma: When new technologies cause great firms to fail*. New York, NY: Harper Collins.

DoD. (n.d.). About the Department of Defense (DOD). Retrieved from https://www.defense.gov/About/

DoD. (2005). *Manager's guide to technology transition in an evolutionary acquisition environment.* Fort Belvoir, VA: Defense Acquisition University Press. Retrieved from https://www.dau.mil/cop/stm/DAU%20Sponsored%20Documents/Managers%20Guide%20to%20Technology%20Transition%202005.pdf

DoD. (2017). *Unmanned systems integrated roadmap, FY2013–2038* (Reference number 14-S-0553). Retrieved from https://www.defense.gov/Portals/1/Documents/pubs/DOD-USRM-2013.pdf

Gulbrandsen, K. E. (2009). *Bridging the valley of death: The rhetoric of technology transfer* (Doctoral dissertation). Retrieved from http://lib.dr.iastate.edu/etd/10740

The latest unmanned drone is a version of an existing manned one: Back to the unicopter. (2017, November 2). *The Economist*. Retrieved from https://www.economist.com/news/science-and-technology/21730865-back-unicopter-latest-unmanned-drone-version-existing-manned

Lewis, L. (2017, September). *Insights for the third offset: Addressing challenges of autonomy and artificial intelligence in military operations*. Retrieved from https://www.cna.org/CNA_files/PDF/DRM-2017-U-016281-Final.pdf

National Research Council. (2004). *Accelerating technology transition: Bridging the valley of death for materials and processes in defense systems*. Washington, DC: National Academy of Sciences.

Natural Resources Canada. (n.d.). [Figure]. Retrieved from https://cfs.nrcan.gc.ca/assets/image/2825

Newman, D. (2018). Technology readiness and the valley of death. Retrieved from http://www.boeing.com/features/innovation-quarterly/may2017/feature-thought-leadership-newman.page

Panetta, K. (2017). Top trends in the Gartner Hype Cycle for emerging technologies, 2017. Retrieved from https://blogs.gartner.com/smarterwithgartner/files/2017/08/Emerging-Technology-Hype-Cycle-for-2017_Infographic_R6A.jpg

Posen, B. R. (1984). *The sources of military doctrine: France, Britain, and Germany between the World Wars.* Ithaca, NY: Cornell University Press.

Pusateri, A. E., Macdonald, V. W., Given, M. B., Walter, S. F., & Prusaczyk, W. K. (2015, November–December). Bridging the technology valley of death in joint medical development. *Defense AT&L,* 40–45.

Turner, M. D., & Wickert, D. P. (2016). *Innovation lost: The tragedy of UCLASS.* Fort McNair, Washington, DC: Eisenhower School for National Security and Resource Strategy, National Defense University. Retrieved from http://www.jcs.mil/Portals/36/Documents/Doctrine/Education/jpme_papers/turner_m.pdf?ver=2017-12-29-142159-597

Under Secretary of Defense (AT&L). (2000, December 21). *Other transactions (OT) guide for prototype projects*. Retrieved from https://www.darpa.mil/attachments/otguideDec15.pdf

Seven Tips to Support Rapid Product Deployment: Lessons Learned

Bruce Nagy—holds the position of Chief Architect for advanced development activities within NAWCWD. He leads forward thinking research efforts involving machine intelligence, simulation/regression techniques, software architecture, and battle engagement strategies. He specializes in proactive management using predictive metrics. Using predictive metrics, Nagy has shortened the development schedule for advanced R&D DoD and NATO-based projects. He has been noted to recover a Navy satellite program from excessive schedule delays. While as a Naval Officer, he specialized in troubleshooting critical path issues for high visibility programs. Nagy has degrees in mathematics, biology, and electrical engineering from The Citadel and Naval Postgraduate School. [bruce.nagy@navy.mil]

Abstract

Navy leadership has tasked its workforce to respond to an urgent need of identifying good ideas that fulfill immediate warfighter gaps in offensive and defensive capability. This paper describes how following seven tips gave a 3.5-person team the ability to develop a working prototype demonstrating the framework of a cutting-edge, counter Unmanned Aerial System (cUAS) technology, supporting a complete kill chain solution. When good ideas are accepted by Navy leadership for feasibility investigation, budgets are tight, teams are small, and expectations are high. This necessitates an environment that introduces a need for rapid product prototyping of a working system/framework that will inspire additional monies to support rapid deployment. Seven tips describe detailed lessons learned involving project management, resource management, system engineering, and architectural analysis, including the use of Open System Architecture (OSA). Specific best practice techniques applied to the rapidly prototyped cUAS technology example, including generic discussions of Program of Record technologies, are used as case studies emphasizing the benefits of each tip described. By considering the seven tips, the workforce is given guidance, examples, and food-for-thought as to how to meet Navy leadership's urgent need for rapid deployment of cutting-edge technology.

Background

"Months not years" is now the leadership's call for urgency in getting products out to the fleet in a shorter period of time. This call has become a mantra regarding the need for new ideas taking advantage of existing technology in creative ways.

"Easier said than done" is one phrase heard by frustrated acquisition professionals. For example, a reader's response as annotated in a blog of the AirTALKs' recipe (recited in the previous introduction) was, "what bureaucratic barriers were eliminated?"

The current "recipe for success" described by AIRTalks on August 15, 2017, to deliver products more rapidly to the fleet is as follows:

- Leadership set a *clear* and *urgent goal*
- Focused on *schedule* and *outcomes*; *tailored in* only what was critical
- *Empowered* the team to manage risk and make decisions; and
- *Eliminated* bureaucratic *barriers* to speed

Another initiative supported by the DoD's MD5 National Security Technology Accelerator focuses on teaching employees how to deliver a well-crafted elevator pitch and interview stakeholders in support of future adaptation.

The above points are valuable to heed and the related training is important, but the recipe misses a key ingredient: educating the workforce on how to deliver products more rapidly to the fleet. This paper does not attempt to reinvent the acquisition cycle, but to more effectively work with it to provide an approach to support rapid prototyping by using more effective Open Systems Architectures (OSAs) and standards supported by industry best practices associated with project management, system engineering, architectural analysis, and resource management. The lessons learned described in this document also include creating a constructive workforce environment—using a common-sense best practice, as will be described.

The need to begin the birth of a program by rapid prototyping that uses more effective OSAs and proper application of best practices ensures that the system of systems architecture will support a rapid product deployment. This document describes examples and related approaches associated with project management, system engineering, architectural analysis, and resource management to ensure the proper selection of OSA and application of best practices. Using the tips provided, a team can have a better chance to take a "good idea" and get it funded, proving out concepts through the use of rapid prototyping. Rapid prototyping, in the counter Unmanned Aerial System (cUAS) case provided as the primary example within this document, means delivering a working system of systems that demonstrates the concept and validates the selection of OSA that enables cost reduction, productivity improvement, and product reliability.

In today's urgent need to maintain technical superiority within a theater of operations, rapidly deployable good ideas, using the proper OSAs through the use of best practices, are more likely to gain attention and get funded. Most ideas are supported with limited funding, therefore rapid prototyping, minimizing cost but emphasizing capability, becomes a necessity.

The Necessity to Rapidly Prototype a Good Idea

It is logical to any acquisition professional to be concerned about rapid deployment and its potentially adverse effects on the quality and reliability of product. Will the product suffer if pushed into the hands of the warfighter too soon? That is certainly one of the reasons why the acquisition cycle is so structured and rigorous. What emphasizes this concern of a poorly developed product from skipping steps is that warfighters need highly reliable, quality products because their lives might be in jeopardy and they are dependent on the technology they are using. In most commercial products, this level of rigor or dependability isn't required. For example, if a smartphone breaks, the user is without a phone until he or she takes it to the store for a replacement. If a warfighter's communication device stops working within a battle engagement, his lifeline for fire support may not be available, creating a potentially life-threatening situation.

The main challenge is that the acquisition turnaround time for a new technology consists of many approval cycles and gates, sometimes consuming a decade of reviews and meetings before the proposed product is seen by the fleet. Within this same period of time, the commercial market may have produced hundreds of components that might have affected the performance, cost, and quality of the product under development.

The obvious answer is to focus on OSA and use of industry best practices that make going through the acquisition cycle faster, but still allow for highly reliable, quality results. A popular analogy is to be able to create a system of systems architecture in a similar manner as putting together Legos, where the OSA associated with best practice interface standards represent the nubs on each Lego block. Therefore, in the simplest of explanations, this paper describes lessons learned in terms of using best practices based on specific, proven tips and sound engineering practices supporting the proper use of OSA.

The goal of the paper is to offer seven tips for consideration in helping acquisition professionals more easily create an effective Lego-based system of systems prototype products.

The tips are centered on rapid prototyping to ensure that the selected OSA through use of best practices is tested to be practical, economically feasible, and reliable in terms of meeting goals. Examples in this document will be provided in which the wrong OSA selection created increased acquisition issues. Specific solutions to this dilemma will be described in terms of best practice approaches.

If rapid prototyping is done properly using the "right" OSA and following effective best practices, then the benefit will more likely be a rapidly developed prototype, using minimal cost over a short period of time. With regard to the cUAS example used to support tip development, the technology was developed within two months using a team of 3.5 developers.

An additional benefit is that as new commercial technologies are created during the acquisition cycle, the prototyped architecture, following the tips offered in this document, will have proven to allow for a plug-and-play development environment. This will offer the product, under the acquisition cycle, greater opportunities to keep in pace with technology advancements, naturally reducing the acquisition cycle, increasing the reliability of the product, and significantly decreasing overall program cost.

Seven tips will be discussed in the form of lessons learned to share how to practically use OSA and industry best practices to answer Navy leadership's "months not years" call for urgency.

Seven Tips

Tip 1: "A Picture Is Worth a Thousand Words!"—Use a Storyboard to Clarify the Problem and Solution

This tip supports the first step of gaining clarity regarding the problem and solution. To gain this clarity, four questions need to be answered:

1. What specific problem does the Navy need to solve?
2. What assumptions are being made about the problem domain, and do those assumptions still support Navy needs?
3. Is someone already solving this problem using the same assumptions? If so, was this group contacted and solutions compared?
4. With regard to the proposed solution, is a complete kill chain scenario described?

This first tip has to do with how the problem can be qualified to be a viable candidate for funding. It avoids spending time solving a problem that lacks interest from potential stakeholders willing to provide funds for the proposed solution. As stated, a cUAS project is offered as an example to illustrate the value this tip offers in clearly defining a problem and describing a complete solution.

A short background regarding cUAS: An effective cUAS solution is becoming an important goal for all armed services, especially because of recent events in the news. From a Navy perspective, the issue is potential threats to naval facilities from small, hard-to-spot drones.

Because of this Navy focus, the problem defined was limited to an attack from sea. The imagined attack involved a small boat, a small Unmanned Aerial Vehicle (UAV), and a sinister intent. The problem definition naturally eliminated many solutions developed by the Army or other branches. The Navy needs to be worried about sea attacks, more than other government agencies. In this project, the assumptions were that the UAV could be launched miles away from an ocean launch site. Depending on the distance, land-based sensors may have difficulty identifying small objects from long distances away. The small object would be a group 1 UAV with autonomous capability, designed as a fixed wing drone. The drone's performance would be typical of any fixed wing autonomous vehicle purchased on some popular website that could fly long distances without operator intervention.

Again, to adequately address the third question regarding this tip, it was important to investigate if there was already a solution to identify a group 1 UAV from far distances, launched from the sea platform, and then eliminate this threat, if needed. It seemed that a solution to support this water-based attack using the approach described would add to an arsenal of Navy capabilities. Although, the Army supported a detailed cUAS solution and because of this, a Technical Interchange Meeting was held to ensure the Navy cUAS solution related to its unique problem was not already solved.

Once it was determined that the solution was unique, where no other groups were providing similar solutions, the next step was defining a complete kill chain or mission scenario. The goal was not to just define a piece of the solution, but to support the entire kill chain scenario. The solution ranged from how the UAV was identified when approaching to its elimination.

To capture these ideas, Department of Defense Architecture Framework (DoDAF) Operational View-1 (OV-1) view was used in a unique way. Two OV-1s were combined to make up a storyboard. OV-1 views are commonly used. A two-frame storyboard, as shown in Figure 1, is a creative use of OV-1s. The point is that DoDAF views allow for creative license. In Figure 1, the cUAS example is described from supporting identification using video to a kamikaze defense solution.

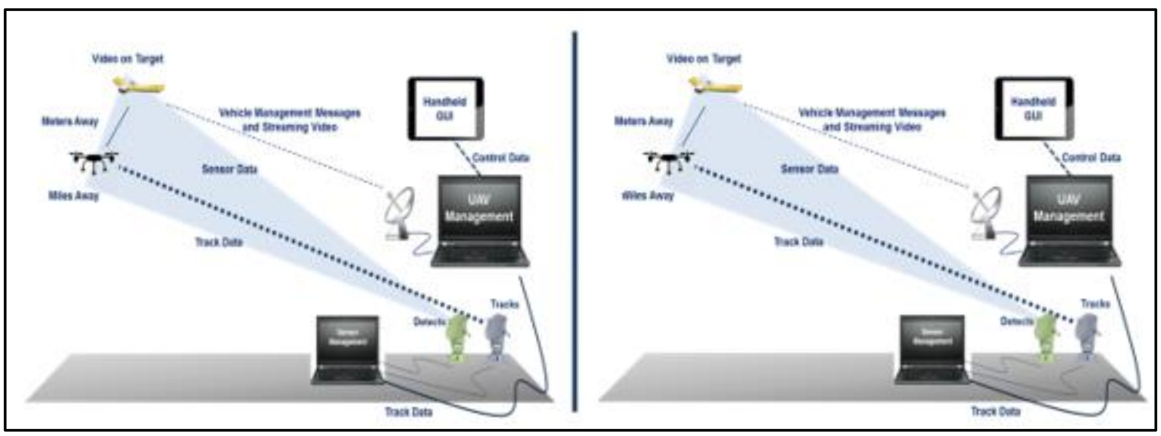


Figure 1. A Simplified Version of Two Related DoDAF OV-1s Used as a Two-Frame Storyboard

Figure 1 represents assumptions regarding both the problem and solution, which allowed stakeholders and the development team to understand what type of subsystem elements might be needed. "A picture is worth a thousand words," but it is still recommended that words provide the details behind each OV-1 views. With one sheet of paper, the views along with dialogue can now communicate both solution and assumptions to verify if both still support Navy needs. The goal is to use the OV-1 to ensure that there is a clear definition of the problem and a complete solution. "Complete" refers to including all parts of the kill chain scenario, from identification to elimination, as described in Figure 1.

Therefore, in following the approach described in this tip using OV-1s, all four questions were more easily answered and explained to others.

Tip 2: Apply Assembly Line Thinking to Make "Months Not Years" a Possibility

This tip illustrates how DoDAF views can greatly enhance the ability to meet the "months not years" goal without hindering productivity.

For review, an assembly line consists of a series of already prepared parts that are integrated over a set period of time, normally defined by the speed of the assembly line process. Using this analogy, consider platforms or subsystem elements to be the already prepared parts. The schedule defines the assembly line time period. Each of these parts are welded together via software interfaces or hardware constructs, like Ethernet cable. Note: Even if the assembly line's purpose is to only develop one item, the assembly line process is still valid.

In the project mentioned above involving cUAS, the assembly line subsystem elements/platforms were as follows:

- Kinetic Integrated Low-Cost Software Integrated Tactical Combat Handheld (KILSWITCH), which has been used by the Marine Corps to support situational awareness and other combat-related goals.
- Navy Unmanned Common Control System (CCS) Science and Technology (S&T) version using the Office of Naval Research's (ONR's) Topside. Topside is a multi-dimensional ground control station. The combination of using CCS and Topside has been successfully demonstrated over the last three years with a variety of demonstrations, including Large Displacement Unmanned Undersea Vehicle (LDUUV), Autonomous Aerial Cargo/Utility System (AACUS), and Common Mission Command Center (CMCC) with the K-MAX Unmanned Aerial System (UAS).
- Maneuver Aviation Fires Integrated Application (MAFIA), Joint Multi-Platform Advanced Combat (JMAC), and MAFIA Association System (MAS), which is currently fielded as cUAS technology.
- Maneuver Aviation Fires Integrated Application (MFOCS), which is currently fielded.
- Standardized Payload Management Systems (SPMS), which has been demonstrated to manage a variety of weapons and payload systems on Navy UAVs and is currently being enhanced using an Unmanned Aerial Vehicle Control Segment (UCS) service interface.

The main interface, a common bridge, to connect these elements together was determined to be UCS, a Navy supported standard. For the subsystem elements selected, the bridges/interfaces to the UCS standard would therefore be

- Cursor on Target (COT) supporting KILSWITCH communication to UCS
- Tactical Counter-Unmanned Technologies (TCUT) supporting MFOCS communication to UCS

This common bridge ensured interoperability while keeping communication under a single bridge standard. The need for a common standard connecting all system elements is described in Tip 3.

To implement an assembly line paradigm, a series of sequence diagrams were chosen. The first diagram type started with operational sequence diagrams (DoDAF OV-6c; see Figure 2) and then moved to system sequence diagrams supporting a system view (SV). Figure 3 describes a DoDAF SV-10c which translated platforms in OV-6c to sub-elements within those platforms. Figure 2 describes the use of various types of radars to support various needs in identifying small targets. The Joint Integrated Fire Control System (JIFCS) was the technology developed within months using the tips described in this paper. Figure 2 describes how JIFCS needed to interface with sensor data, handheld devices, and a group 1 UAV to implement the desired cUAS solution. The goal of both Figures 2 and 3 is not to describe the JIFCS technology, but to show how assembly line thinking was applied to rapidly develop a new product.

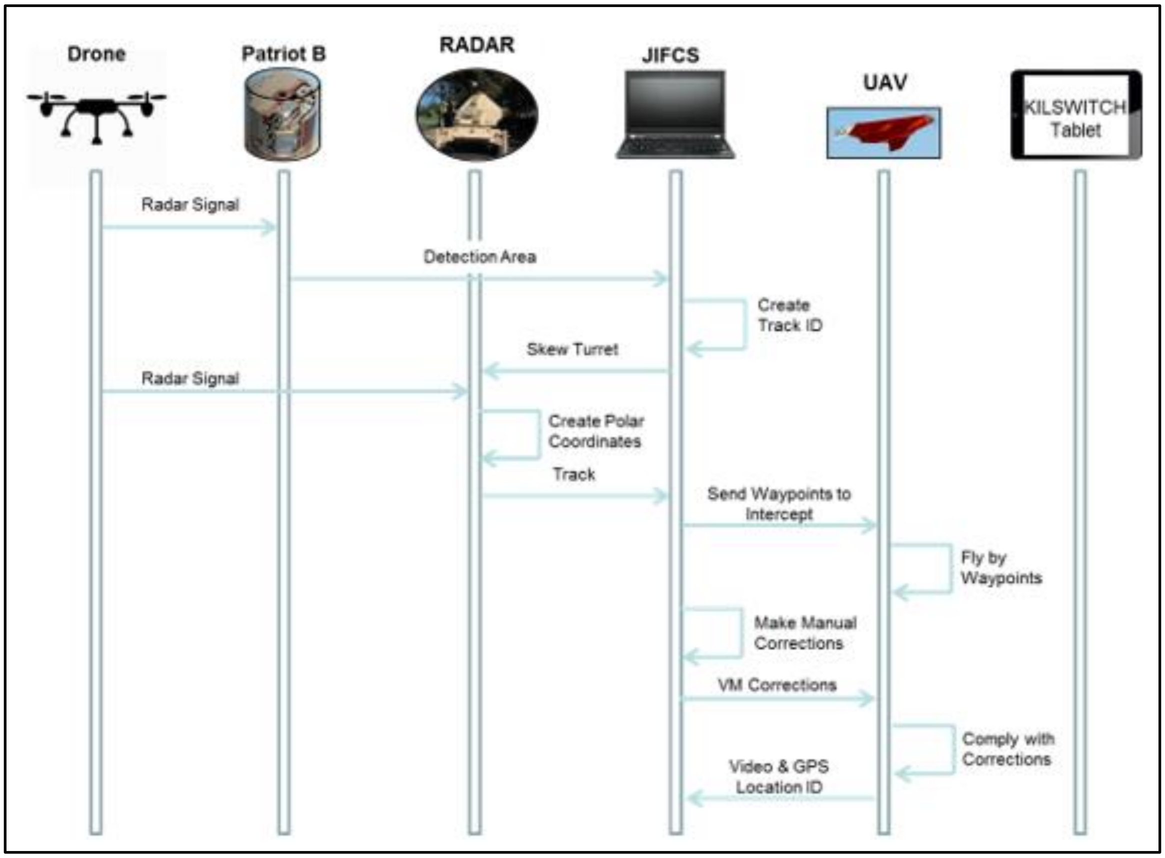


Figure 2. OV-6c View Regarding One Part of the cUAS Solution

The final suggestion related to Tip 2 is make sure the pieces to be assembled are mature enough to be used within the project's schedule. This pitfall was realized in a rocket development project, where one of the main elements relied on a rocket engine that was still in the experimental phase. The rocket engine needed several years to mature before it could be considered to be realistically used in a combat environment. In this case, because this element made the assembly line process take years, and not months, the project was rejected. Redesigns were eventually investigated, but the project's focus was lost. This type of rocket development project became a very low priority for funding and eventually was forgotten. Be careful: If an element takes several years to mature, it obviously should not be used as part of the assembly line process when attempting to support leadership's "months not years" call for urgency.

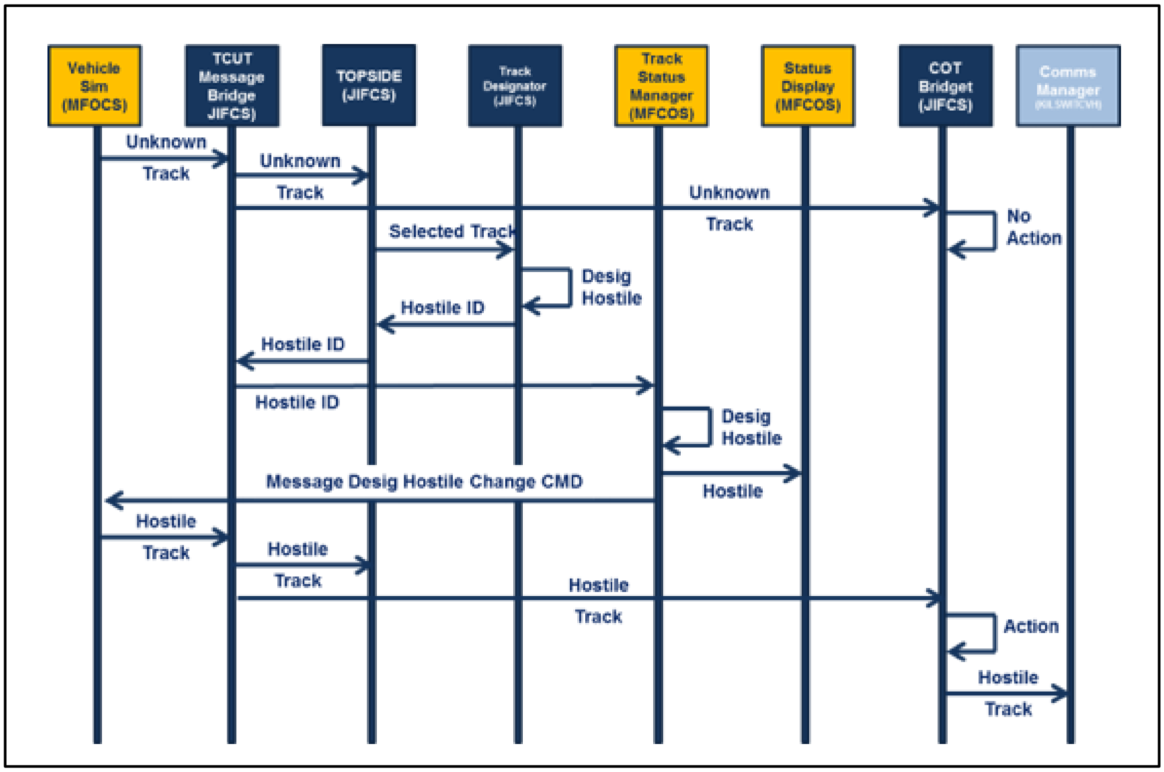


Figure 3. SV-10c View Regarding One Part of the cUAS Solution

Tip 3: "To Be or Not to Be?"—Analyze the Layers of Interfaces to Determine What Should or Should Not Be Used

Tip 3 provides a "rule of thumb" on how to determine whether it makes architectural engineering sense to proceed with regard to interfaces between subsystem elements.

In the cUAS example, there were subsystem elements that required the use of three to five protocols to exchange needed data. This caused an over-complication of the architecture and implementation, resulting in integration and performance issues. This came about because each of the original subsystems were developed using different interface standards. To create even more complications, some of the elements were developed in a Linux Operating System (OS) and others in Windows OS, causing additional enterprise gluing issues.

Although the protocol associated with each subsystem involved well-known and popular standards, the need to use many standards to support straightforward communication caused installation and performance issues.

As a tip, if a straightforward bridge cannot be used between standards of different subsystem elements, then consider reevaluating the subsystem elements. If there are no other choices, then consider new development instead of forcing a square peg to fit a round hole.

In the cUAS example, the UCS standard was determined to be able to bridge all subsystem elements. The other elements were assessed as to whether straightforward bridges could be produced, and the assessment came out positive. If it didn't, then a different architecture would have been investigated.

Figure 4 is an example of multiple protocols used through a thread of information in the cUAS project. The upper thread represents four translator services that would potentially be needed for communication between subsystem elements and a system that shouldn't be put together. The lower thread in Figure 4 represents an ability to architecturally reduce subsystem communication to one translator service, creating a good system. The architecture introduced an Air Force standard named Unmanned Command and Control Initiative (UCI); however, the ongoing challenge was to ensure UCI and UCS efficiently talked to each other.

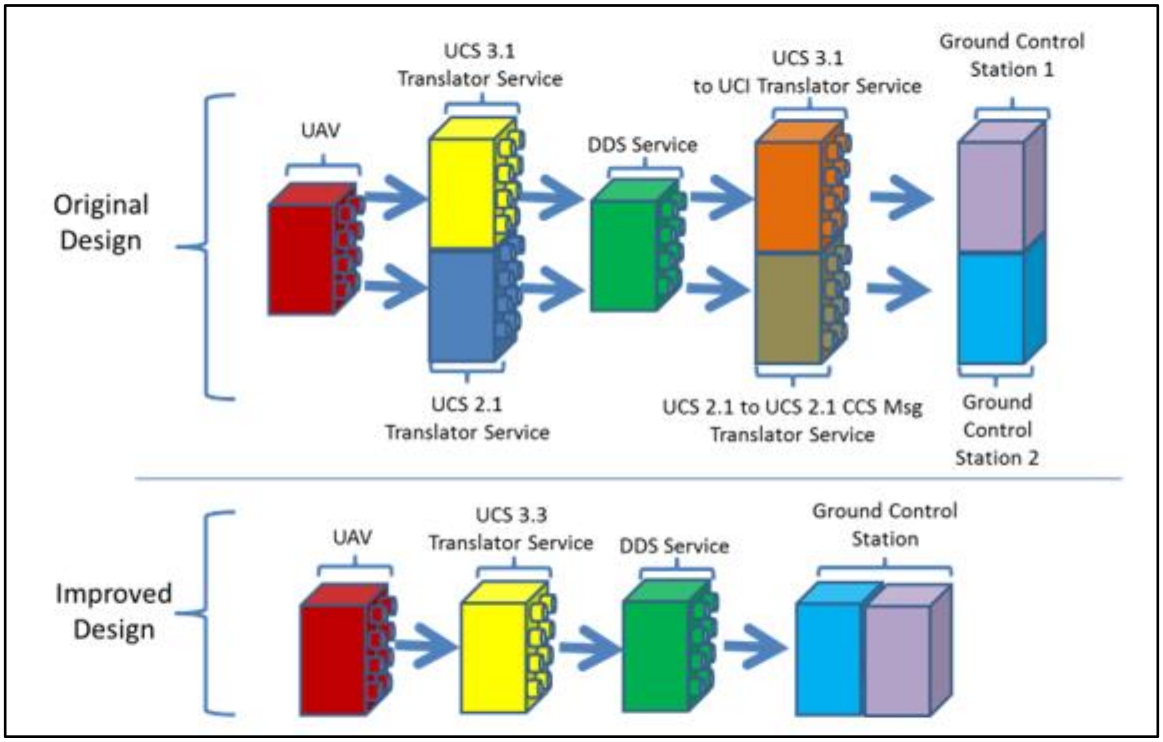


Figure 4. Translator Service Designs Used for Communication Between Subsystem Elements

Tip 4: Focus More on Integration, Less on New Development, to Create a "Months Not Years" Project Plan

Once it has been agreed to proceed with the project, Tip 4 provides suggestions on how to quickly create the first demo within the "months not years" timeframe. It provides examples of system engineering techniques that support rapid development.

In the case of the cUAS project, using Tip 4's technique showed that the demonstration could be done within four two-week sprints, which included architectural analysis, implementation, and test.

If the first three tips are already completed, then this tip will naturally follow and be easier to achieve. The Tip 4 technique suggests using the time it takes to do the integration/assembly of each primary element, including data usage, to determine the shortest timeline to release a demo. If the "right" primary elements aren't correctly identified from your assembly line, it's difficult to successfully use this tip. Primary elements are identified by understanding how the key platforms are used. Sometimes the best way to identify primary elements is to identify non-primary elements. Non-primary elements are subsystems that use primary elements as their hosted platforms.

In the cUAS example, the primary elements were the MFOCS emulator, JIFCS emulator, and KILSWITCH. Using the technique just described, the timeline for connecting the primary elements first involved the development of two bridges/interfaces:

1. From the COT message standard to the UCS message standard between KILSWITCH and JFICS
2. From the TCUT message standard to the UCS message standard between MFOCS and JIFCS

It should also be noted that the timeline also included the time it took to develop the software necessary to display the related data. Therefore, the completion of the demo was principally based on two time related factors:

1. The time it took to develop the bridges that connected the standards (and therefore connected these primary elements)
2. The time it took to display the related data on non-primary elements (e.g., for JIFCS, non-primary elements would be the Topside display)

Additional features were determined based on whether they could be done in parallel to bridge development or support bridge development, while keeping the main focus of the project on bridge development or data that used the messages translated by the bridge. Notice that the UCS message standard was the common connectivity between the primary elements. The reason for having a common standard for connectivity was described in the previous tip. If this technique is followed properly, there should be little to no lag time within the assembly line timeline associated with connecting these primary elements.

After identifying this primary element assembly line timeline, the next key question was what non-primary element features could be integrated without adding any time to the schedule? Understandably, sometimes it is necessary to require additional time, above and beyond the primary element timeline.

Figure 5, the cUAS project schedule, emphasizes that the timeline focused on bridge development or relate data management/display. This meant that most times, these tasks were on the critical path, meaning the smallest (to no) lag time between tasking. The purpose of putting emphasis on the primary element assembly line timeline means that the tasking is mainly focused on assembling primary and non-primary elements instead of creating technology to integrate.

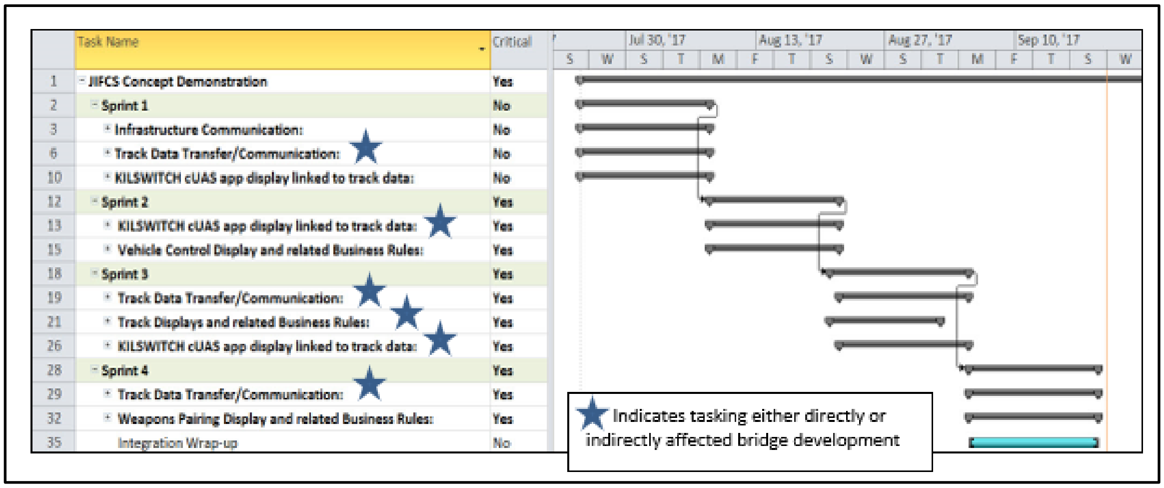


Figure 5. Schedule to Ensure Bridge Work Is on Critical Path and Not Added Features

The tasking "Track Data Transfer/Communication" described in Figure 5 was the incremental development of the bridges. Most of the tasks in Figure 5 containing the words "track data" and "track displays" involved bridge development for primary elements to share data, data management, or data display.

Again, Tip 4's technique allowed the cUAS project team to maintain the goal of reducing the schedule as much as possible to support a rapid demonstration. This technique also supported a scheduling discipline when assessing how much added feature development the timeline could permit without taking the focus away from bridge development or related use of the messages translated.

In general, once a project's original concept is demonstrated, then the next phase of the project would be to add features to the existing framework, knowing the bridge development and message use were working properly. This technique allows the project to lay a foundation, like a solid chassis in a car assembly line. The only difference is that in this assembly line, only one product is being created. This technique also allows the project to identify the "ideal" shortest period of time. In following this technique, future development risk is reduced, and potential customers are given a better understanding of value.

Tip 5: Constantly Remember—It Takes a Village to Raise a … Product

During implementation of a project, Tip 5 can help ensure that the support network is adequately defined and that everyone remains "willingly helpful" during the development cycle. This tip also provides a suggestion on how to deal with folks who may not have time to be in a support network, but need to be available and willing in order for the project to succeed.

Although subtle, this tip should be followed, maybe before all other tips suggested in this paper: Be nice and help everyone as much as possible, because one day a challenging task may need a helping hand and the money to pay for it may not be available. With regard to the cUAS project being discussed, from video folks to weapons pairing experts, 15 to 20 minutes of conversational help about key problems became invaluable.

If individuals on a project team can sincerely promote another group's work or help out, even in small ways, great dividends are received. So, when the time comes that a project team member needs help, someone will probably show up, without a need for a charge object.

In the case of the cUAS project, a kind videographer took a few minutes to do some editing. Yet, those few minutes made a significant difference in the realism of the demo. Figure 6 is a snapshot of the video that took minutes to edit. Another example was a weapons pairing expert who provided a check and balance for related algorithms that were being developed regarding automated weapons pairing. These and other folks provided invaluable support for the success of the cUAS project.

Figure 6. Snapshot of Picture Related to cUAS Project

It's the old saying, "what goes around, comes around." Tip 5: Help when possible, and one day the rewards will appear.

Tip 6: Consider the 80/20 Rule to Support New Talent Growth and Challenging Schedule Goals

There is a temptation to have only experienced people do the work when a short timeframe is involved. Tip 6 describes how to use a version of the 80/20 rule regarding developers and resource management.

In the cUAS example, the project team size was fairly small. Everyone had to pull their own weight. The team consisted of 80% (or four) seasoned developers, and allowed for 20% (one) highly energetic, highly motivated person to work on the project. This meant that the new developer would be allowed more scheduled time to accomplish the assigned tasks. It was important that the new developer was not put under the pressure of being on the critical path. Additionally, workarounds needed to be readily available. "Months not years" is a short-term philosophy, but developing people within the organization provides long-lasting results.

There's no substitute for experience, but there is no greater reward than giving a new developer a chance to shine. As stated, the cUAS development team consisted of five people, represented on the left side of Figure 7. There were senior people on and off of the project and "in the wings" ready to help all the developers, which goes back to following Tip 5: Be sincerely helpful to others and good things will become available.

Figure 7. 80/20 Resource Mix to Support Professional Development

As an important note, in addition to providing a successful demonstration of a needed technology, the project also had the privilege to support the development of one who became a much more capable and seasoned software developer—a significant win for the organization.

Again to reemphasize per the cUAS project example, the development of junior personnel is still possible within a rapid development project in which timelines are tight.

Tip 7: Share Lessons Learned—Know That Sharing Experience Creates a Village of "Smart People"

Along with the "months not years" call for urgency, there is another popular leadership phrase offered: "Take risks!" In some people's mind, that means potential failure. Tip 7 is about sharing knowledge so others learn how to overcome risks and succeed. In other words, it describes the best way to make "lemonade out of lemons." This paper was written to share lessons learned and hopefully help some other project team create an urgently needed product in a shorter period of time for the warfighter.

Lessons learned described via the previous tips came in two popular categories: (1) what worked and (2) what didn't work. Table 1 represents how the previous tips were categorized in terms of what worked and didn't work. During the previous description of each tip, greater detail was included as to why the suggestion worked or why it didn't work.

If one of the seven tips proves to support the ability of someone launching a good idea to support warfighter supremacy, then the cUAS project described has had a significant additional success, beyond cUAS. In other words, this tip suggests that success also includes sharing lessons learned for others to have successes.

Table 1. Lessons Learned: What Worked and What Didn't Work

What Worked	What Didn't Work
"A Picture Is Worth a Thousand Words!" —Use a Storyboard to Clarify the Problem and Solution. (Related Tip 1: "A Picture Is Worth a Thousand Words!"—Use a Storyboard to Clarify the Problem and Solution)	Forcing standards to work together by needed to use an excessive amount of bridges in order for subsystems and platforms to communicate with each other. (Related Tip 3: "To Be or Not to Be?"—Analyze the Layers of Interfaces to Determine What Should or Should Not Be Used)
Using OV-6c and SV-10c views to describe an assembly line approach that connects existing technology. (Related Tip 2: Apply Assembly Line Thinking to Make "Months Not Years" a Possibility)	
Identifying primary and non-primary elements and focusing the schedule on integrating primary elements. (Related Tip 4: Focus More on Integration, Less on New Development, to Create a "Months Not Years" Project Plan)	
Supporting other people and their projects unconditionally to create an environment of mutual support, independent of available charge objects. (Related Tip 5: Constantly Remember—It Takes a Village to Raise a … Product)	
Growing talent without risking schedule by keeping less experienced implementers off of the project plan's critical path and giving them time to learn. (Related Tip 6: Consider the 80/20 Rule to Support New Talent Growth and Challenging Schedule Goals)	
Sharing what worked and didn't work with others—and making it as easy as possible for people to understand and be interested. (Related Tip 7: Share Lessons Learned—Know That Sharing Experience Creates a Village of "Smart People")	

Conclusions

These are exciting times because the U.S. Navy is looking for solutions to warfighter needs, and the need to rapidly deploy a good idea is vital. Good ideas that can be rapidly deployed are more likely to receive funding. The goal of sharing these tips is to provide guidance to help navigate the challenges of assessing, designing, and planning a successful project using OSA and best practices. “Months not years” is a hard mantra to follow. But also consider the phrase, “Take risks!” With the myriad of technology currently developed, putting these elements together to create a new system of systems can be an exciting adventure. Yet, in an ironic way, the excitement is also associated with and sometimes driven by the risk. Prototyping through the use of the proper OSA by applying the appropriate best practices ensures greater rapid deployment success.

In the cUAS project, lessons learned showed the need to focus on sequence diagrams to help determine whether a demonstration was possible within months. The analysis showed how different elements could be integrated rather easily to fit within a months’ timeframe. It showed the need to get as many people involved (within the village) as possible. It doesn’t mean everyone needs to be on the payroll either. Fifteen minutes here and there from various experts regarding key areas associated with your project pays off in big dividends. These lessons learned also describe why a new programmer in the learning phase using the “right” resource management schema can add value, both short and long term, to an organization. Finally, whether a project is successful or not, sharing lessons learned is an important aspect to the rapid development process.

Consider the seven tips described when a good idea pops up and the assessment as to how to begin becomes an inviting next step. Table 2 describes a checklist based on the seven previous tips suggested to help determine if any good idea qualifies to be a “months not years” candidate. Answer the questions in this table using a product prototype.

Table 2. Checklist to Support “Months Not Years” Development Effort

Tip 1: “A Picture Is Worth a Thousand Words!”—Use a Storyboard to Clarify the Problem and Solution		
Tip 1	Was a specific problem defined that the Navy needs to have solved? (Ref: DoDAF Views, OV-1s to start in storyboard fashion)	Yes/No
Tip 1	Were assumptions made about the problem domain and did those assumptions still support Navy needs? (Ref: DoDAF Views, OV-1s to start in storyboard fashion)	Yes/No
Tip 1	Is someone already solving this problem using the same assumptions? If so, was this group contacted and solutions compared? (Ref: DoDAF Views, OV-1s to start in storyboard fashion)	Yes/No
Tip 1	With regard to the solution, is a complete kill chain scenario described? (Ref: DoDAF Views, OV-1 to start in storyboard fashion)	Yes/No
Tip 2: Apply Assembly Line Thinking to Make “Months Not Years” a Possibility		
Tip 2	Have the platforms been selected and their timing relationships been defined? (Ref: DoDAF Views, OV-6c and SV-10c in assembly line fashion)	Yes/No
Tip 2	Have the subsystem elements regarding the platforms been selected and their timing relationships been defined? (Ref: DoDAF Views, OV-6c and SV-10c in assembly line fashion)	Yes/No
Tip 3: “To Be or Not to Be?”—Analyze the Layers of Interfaces to Determine What Should or Should Not Be Used		
Tip 3	Are there straightforward bridge/interface connections between standards of different subsystem elements?	Yes/No
Tip 3	Is there one common communication standard used to connect the different subsystems elements and platforms?	Yes/No
Tip 4: Focus More on Integration, Less on New Development, to Create a “Months Not Years” Project Plan		
Tip 4	Is developing bridge work/interfaces between elements the main driver to the length of schedule?	Yes/No
Tip 4	Is the development of the bridge work/interfaces between elements most times on the critical path?	Yes/No
Tip 5: Constantly Remember—It Takes a Village to Raise a … Product		
Tip 5	Is a formal (paid) support network, including experts in various fields related to the proposed solution, available to help?	Yes/No
Tip 5	Is an informal (not needing to be paid) support network, including experts in various fields related to the proposed solution, available to help?	Yes/No
Tip 6: Consider the 80/20 Rule to Support New Talent Growth and Challenging Schedule Goals		
Tip 6	Is around 80% of the development team experienced regarding the work that needs to be performed?	Yes/No
Tip 6	Are the experienced developers assigned work on the critical path?	Yes/No
Tip 6	Are the inexperienced developers not assigned work on the critical path?	Yes/No
Tip 6	Are the inexperienced developers highly motivated?	Yes/No
Tip 6	Are the inexperienced developers provided more time to complete tasks to account for their learning process?	Yes/No
Tip 7: Share Lessons Learned—Know That Sharing Experience Creates a Village of “Smart People		
Tip 7	Are you identifying what worked and why?	Yes/No
Tip 7	Are you identifying what didn't work and why?	Yes/No
Tip 7	Are you making suggestions to improve the process?	Yes/No

If the answers to the questions in the checklist in Table 2 are predominately “yes” based on a prototype, not just a paper exercise, then those good ideas are more likely to be a good candidate for rapid product deployment. And rapidly deployable good ideas are more likely to get funded. And if funded, those good ideas are more likely to be deployed in time, proven through your prototype, to ensure U.S. combat superiority is maintained, saving lives and eliminating enemy threats.

References

AirTALKS [Blog post]. (2017, August 15). Retrieved from https://myteam.navair.navy.mil/corpapps/NAVAIRComm/NAVAIRBlog

Dam, S. H. (2006). DoD architecture framework: *A guide to applying system engineering to develop integrated, executable architectures*. Marshall, VA: SPEC.

DoD. (2010). The DoDAF Architecture Framework Version 2.02. Retrieved September 1, 2017, from http://dodcio.defense.gov/Library/DoD-Architecture-Framework/

Grady, J. O. (2010). *System synthesis: Product and process design*. Boca Raton, FL: CRC Press.

Kerzner, H. (2013). *Project management: A systems approach to planning, scheduling, and controlling.* Hoboken, NJ: John Wiley & Son.

Langford, G. O. (2012). *Engineering systems integration, theory, metrics and methods*. Boca Raton, FL: CRC Press.

Maier, M. W., & Rechtin, E. (2009). *The art of systems architecting*. Boca Raton, FL: CRC Press.

Solomon, B. (n.d.). Crafting your “elevator pitch”; Building a tech transition plan.

Distribution Statement

Distribution Statement A. Approved for public release. Distribution is unlimited.

Panel 19. Augmenting the Acquisition Decision Processes With Data Analytics

Thursday, May 10, 2018	
11:15 a.m. – 12:45 p.m.	**Chair: Mark Krzysko,** Deputy Director, Enterprise Information, Acquisition Resources and Analysis, OUSD (AT&L) **Discussant: Gary Bliss,** Director, Performance Assessments and Root Cause Analyses, Office of the Assistant Secretary of Defense ***Informing DoD Program Planning Through the Examination of the Causes of Delays in Acquisition Using Acquisition Data*** Charles K. Pickar, Naval Postgraduate School ***Utilizing Public Data for Data Enhancement and Analysis of Federal Acquisition Data*** Ningning Wu, University of Arkansas at Little Rock Richard Wang, Massachusetts Institute of Technology M. Eduard Tudoreanu, University of Arkansas at Little Rock ***Toward Cognitive Supremacy via Quantitative Augmentation*** Col Clark Quinn, USAF, Air Force Institute of Technology Maj David M. Smalenberger, USAF, Air Force Institute of Technology

Mark Kryzsko—serves as the Deputy Director, Enterprise Information. In this senior leadership position, Krzysko directs data governance, technical transformation, and shared services efforts to make timely, authoritative acquisition information available to support oversight of the Department of Defense's major programs, a portfolio totaling more than $1.6 trillion of investment funds over the lifecycle of the programs.

Preceding his current position, Krzysko served as Assistant Deputy Under Secretary of Defense (ADUSD) for Business Transformation, providing strategic guidance for re-engineering the Department's business system investment decision-making processes. He also served as ADUSD for Strategic Sourcing and Acquisition Processes and as Director of the Supply Chain Systems Transformation Directorate, championing innovative uses of information technologies to improve and streamline the supply chain process for the Department. As the focal point for supply chain systems, Krzysko was responsible for transformation, implementation, and oversight of enterprise capabilities for the acquisition, logistics, and procurement communities. In addition, Krzysko served as advisor to the Deputy Under Secretary of Defense for Business Transformation on supply chain matters and as the functional process proponent to the Department's Business Transformation efforts, resulting in the establishment of the Business Transformation Agency.

In March 2002, Krzysko joined the Defense Procurement and Acquisition Policy office as Deputy Director of e-Business. As the focal point for the Acquisition Domain, he was responsible for oversight and transformation of the acquisition community into a strategic business enterprise. This included driving the adoption of e-business practices across the Department, leading the move to modernize processes and systems, and managing the investment review process and portfolio of business systems. Krzysko served as the Division Director of Electronic Commerce Solutions for the Naval Air Systems Command from June 2000 to March 2002. From April 1991 until March 2000, Krzysko served in various senior-level acquisition positions at the Naval Air Systems Command, including Contracting Officer of F/A-18 Foreign Military Sales, F/A-18 Developmental Programs, and the F-14. In addition, he served as Program Manager of Partnering, the Acquisition Business Process Re-Engineering Effort, and as Acquisition Program Manager for the Program Executive Office for Tactical Aircraft.

Krzysko began his career in the private sector in various executive and managerial positions including Assistant Managing Director for Lord & Taylor Department Stores and Operations Administrator for Woodward & Lothrop Department Stores. Krzysko holds a Bachelor of Science Degree in Finance from the University of Maryland, University College, College Park, MD, and a Master of General Administration, Financial Management, from the same institution.

Gary Bliss—is Director, Performance Assessments and Root Cause Analyses (PARCA), in the Office of the Assistant Secretary of Defense for Acquisition. PARCA carries out performance assessments of Major Defense Acquisitions Programs (MDAPs) and conducts root cause analyses for MDAPs with Nunn-McCurdy breach status or when requested by senior Department of Defense (DoD) officials.

Bliss previously held the position of Deputy Director, Enterprise Information and Office of the Secretary of Defense (OSD) Studies in the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (AT&L). His responsibilities included oversight of the five OSD-funded Federally Funded Research and Development Centers and the OSD's university research program, as well as review and development of innovations to overhaul the AT&L enterprise management systems.

Earlier in his career, Bliss served 13 years as the Director of Office of the Director, Program Analysis & Evaluation Weapon System Cost Analysis Division (WSCAD). WSCAD's 10 staff members constitute one of the two offices dedicated to OSD Cost Analysis Improvement Group (CAIG) functions, and is responsible for the preparation of independent development and procurement cost estimates for major systems that range from munitions (e.g., tactical missiles) through platforms (e.g., helicopters, submarines, fighter aircraft, tanks, etc.). As such, Bliss has been a key player in the DoD's most important system decisions by the Services, OSD, and Congress.

Generally recognized in both industry and government as a leading authority on the economics of defense procurement, Bliss has an established track record in institutional reform and enterprise reengineering. He is an experienced lecturer, often speaking to varied audiences on such topics as

- Management information system governance and reengineering
- Manufacturing enterprise reengineering
- Acquisition institutional reform

Owing to this expertise, Bliss has been hosted by the governments of Australia, Taiwan, Japan, and the United Kingdom to lecture their staffs on matters of defense acquisition.

Bliss has a BA, Mathematics and Economics (Highest Honors in Economics), from College of William and Mary.

Informing DoD Program Planning Through the Examination of the Causes of Delays in Acquisition Using Acquisition Data

Charles K. Pickar—DBA, is a member of the NPS faculty where he teaches project management, defense acquisition and systems engineering. Before joining NPS, he led the Applied Systems Engineering Program Area at the Johns Hopkins University Applied Physics Laboratory. He is a retired Army officer with extensive experience in the U.S. defense industry, including director and VP levels at Lockheed Martin, Northrop Grumman, and SAIC. He is the current Chair of the Systems Education Technical Committee of the IEEE Systems Council. His research and published work focuses on applying systems engineering and system dynamics analytical approaches to defense acquisition problems. [ckpickar@nps.edu]

Abstract

This research has two objectives. First, using DoD acquisition data, the study details the process developed to mine, convert, and use DoD acquisition schedule data, including a discussion on how the data was refined. Part of this effort was an identification of the factors that caused delays. This data is used to describe a method for project managers to use in their project planning process.

Introduction

Department of Defense (DoD) project management is focused on awareness, anticipation, and decision-making. In order to address these three imperatives, project and program managers in the DoD must plan in detail the expected path and duration of development projects as senior leadership requires reports on progress relative to a plan. However, since the nature of the weapons development process (R&D) is so uncertain and the scheduling tools provided are, at best, stochastic, there is a need for better understanding of the many factors that influence activity/task planning, network development and project execution. This understanding includes ways to estimate schedule beyond the stochastic methods of today. This research has two objectives. First, using the OSD acquisition information databases, determine and develop ways to extract and make that information on scheduling available to DoD project managers. The second goal is to identify important delay factors, so those factors can be considered in project planning.

This research uses DoD acquisition data to inform the schedule planning process. Specifically, it identifies the many factors that have historically led to schedule delays and provides a methodology for PMOs to use when they plan and schedule their weapons system program. This study has three parts. First is an examination of the literature on the current state of schedule estimating. The second part describes the process developed for this study to mine, convert, and use DoD acquisition data, including a discussion on how the data was refined. The last section presents some initial findings from this research and proposes some uses for the information.

The widely used definition of a project includes the assumption that each project is something unique: “[a] project is a temporary endeavor undertaken to create a unique product, service or result” (Project Management Institute [PMI], 2017). Perhaps it is not. Instead, perhaps a project is not unique, and perhaps we can use the experience the DoD has in project management to our benefit. That is the value of using data in defense acquisition.

As long ago as 1988, Morris and Hough were critical of the practice of project management:

> Curiously, despite the enormous attention project management and analysis have received over the years, the track record of projects is fundamentally poor, particularly for the larger and more difficult ones. Overruns are common. Many projects appear as failures, particularly in the public view. Projects are often completed late or over budget, do not perform in the way expected, involve severe strain on participating institutions or are cancelled prior to their completion after the expenditure of considerable sums of money. (Morris & Hough, 1988)

Instead, the basic premise of this study is to address this criticism head-on and suggest instead, that maybe … when problems persist, practitioners and scholars are getting something wrong (Christensen & Bartman, 2016). Therefore, this study explores how to find and use data to help PMs understand the dynamic nature of weapon system development.

Managing defense acquisition schedules has become even more important in recent years for many reasons including the following:

- Longer "cycle times" for defense acquisition programs, especially for high-priority combat systems—in both absolute and relative terms
- The rise of competitor nations with greatly increased capabilities, sophistication, and agility—threatening U.S. national interests (getting inside our development cycle)
- Significantly limited resources available for defense modernization programs, which makes management of funding profiles especially important

This research explores one of the available sources of acquisition data, the Selected Acquisition Report (SAR). SAR data is collected and stored in the Defense Acquisition Management Information Retrieval (DAMIR) database, a repository for, *inter alia*, the DoD Selected Acquisition Reports (SAR). The SAR is a summary of the acquisition data of selected Major Defense Acquisition Programs (MDAP). SARs are required by Title 10 USC § 2432 to be submitted to Congress periodically.

Weapons system development projects are infamous for exceeding time and cost constraints. Study of this time phenomena however, generally focuses on the resultant time it takes to develop a weapon system, not the front-end planning necessary to address schedule overruns. We examine this topic first by reviewing the basics of project scheduling, then examining the project planning process and how scheduling is currently done, what is considered in the development of project schedules, as well as what should be considered.

Project Scheduling

The concept of time in project management can be divided into two major categories: task duration estimation, and task sequencing and project scheduling. First, the technical process of estimating the duration of the project task must be determined. Once duration is established, the management process of project sequencing and scheduling must be defined.

Broad review of the literature on project scheduling reveals research roughly divided into three areas. First, the bulk of literature on scheduling is devoted to the networking and probabilistic techniques which have dominated schedule estimation since the 1960s. This focus is logical in that the "science" of scheduling originated with the almost simultaneous development of the critical path method (CPM), and the Program Evaluation and Review Technique (PERT). CPM or critical path method places activities in a logical network sequence. When completed, this sequence is expressed as a network and provides the total time necessary to accomplish the project, as well as the total time of the individual activities which is expressed as the critical path. PERT also used in building the network provides a probabilistic assessment of the actual schedule time. PERT (also known as three-point estimation) uses the weighted average of three measures of task duration, the most likely duration (M), the pessimistic duration (P), and the optimistic duration (O).

The strength of CPM and PERT (apart from the fact they are used almost exclusively in schedule development, and in most enterprise project management software packages) is it allows management focus by identifying the critical path, thus, the key activities that must be monitored and controlled. Monitoring provides a means to oversee costs including, among others, anticipating personnel changes. Controlling allows the PM to determine whether the project is on schedule, as well as ensuring the defined length of the project is met.

Disadvantages include project management being unable to react to instability and changes, as well as managing resources to "feed" the critical path and not being able to "see" and comprehend the overall effort. This is because CPM and PERT take a static view of project activities—which fails to account for the relationships and interdependencies inherent in complex projects (Balaji & James, 2005).

Second is the basic assumption that work proceeds as planned in the network—that there is a direct flow from work to be done, to work accomplished. That is, every task has a discrete start and end—work is either started or not, finished or not. More importantly, there is no accommodation for work that might not be done correctly or to the required quality. Further, the subjective nature of defining not only the most likely time duration, but also the optimistic and pessimistic durations potentially magnifies schedule uncertainty especially in large, complex projects (Franck et al., 2017). A last disadvantage of the current scheduling method is that it does not recognize management decisions and the feedback from those decisions.

The next major area in the schedule literature examines project schedule from the perspective of the time it takes to develop weapons systems. This research focus assesses schedules by asking the question, "Why does weapon system development take so long?" Central to this line of research is the idea of "cycle time," also referred to as "schedule interval." The issue examined is how to provide weapons systems to the operational force as soon as possible. Research questions ask, "Has the time to develop weapons systems increased?" (Van Atta et al., 2015).

The final area of research interest is that of software project estimation. This area represents the focus of the most recent research. Some suggest that because of the complexity of software, as well as the degree of software in most modern weapons systems, software schedule estimation most closely resembles weapons system development scheduling.

Estimating Activity Duration

Surprisingly, little information is available in the literature on the "how" to estimate the measures of a schedule—the task duration. While the major defense contractors have formal in-company processes, little formal literature is available on the specifics of task estimation. Further, most available information on estimating task duration is found in project management textbooks, but even then, the specifics are scarce.

There are however, similarities between cost estimating and activity duration estimating. This is because accurate cost estimates require the insight into scope and schedule that only duration estimating can provide. Both processes use similar techniques. Both depend on expert judgment, both use parametric methods, and both employ a bottoms-up methodology as one of the techniques is estimated at the task level, then rolled up. Central, however, to schedule estimation is the idea of sequencing. The network is a central element of determining duration.

The PMBOK (Project Management Body of Knowledge) lists five methods for estimating project activity duration. These methods include the following (PMI, 2017):

- Expert Judgment
- Alternatives Analysis
- Published Estimating Data
- Project Management Software
- Bottoms-Up Estimating

Expert judgment acknowledges that technical and engineering experts should be able to estimate the effort necessary to accomplish tasks and translate those estimates to duration. This assumes the chosen experts have significant experience in the execution of those tasks and are therefore competent to judge time required (Hughes, 1996).

Alternatives analysis recognizes that activities or tasks can be accomplished in different ways—alternatives. These different ways include defining different techniques, differing levels of resources, and using different machines.

Published estimates are databanks that gather resources measures. These measures include hourly rates by skill level, acknowledged production rates for various development and manufacturing activities. In most cases, this data is available internal to the organization. However, there are data companies that track and report this data. An example is the IEEE-USA Salary & Benefits Survey. This data is often available for different locations in the United States, as well as worldwide.

Project management software is not really an estimation method. Instead, it provides a means to identify and organize information necessary for resource estimates.

Finally, an engineering or bottoms-up estimate is a comprehensive schedule (and cost) process that starts at the work package level and aggregates costs to build a complete estimate. Bottoms-up estimates are necessary when schedule activities cannot be accurately estimated using another technique. As the name implies, bottoms-up estimates start at a level of activity or task that can be confidently estimated. The activities are then rolled-up to the required level. These estimates are extremely work-intensive but are also the most accurate.

Other recognized methods include parametric techniques. A parametric or top-down estimate builds an activity estimate for the development project from historical data comparing variables through a statistical relationship. All the methods listed are used to estimate the length of time each of the activities or Work Breakdown Structure tasks lists. “Simply stated, the duration of an activity is the scope of the work (quantity) divided by a measure of productivity” (Hendrickson et al., 1987, p. 278).

Thus, activity duration estimation establishes the actual time required to complete discrete tasks in an overall project, while project scheduling fixes the start and end dates, as well as execution approaches of the project. Once the overall schedule is established, management activities driven by either time or resource constraints will determine the actual execution of the project (Schwindt & Zimmerman, 2015). The analogy that comes to mind is that of an orchestra. The individual instruments (and of course, the musicians) are the discrete tasks of the project. The orchestra leader is the project manager, and the music score is the “plan” the orchestra leader uses to execute the “project.” Building on this information, the next step in this effort is to identify schedule data that can be used to augment these estimating activities.

Data Methodology

While there is significant information available on DoD procurements, the overwhelming majority of that information is on cost. Cost is tracked and reported in detail at both the service as well as DoD level, and there is significant numerical-type data available on cost. Cost is also reported in a format that lends itself to analysis (spreadsheets). In fact, both in government and industry, cost is significantly more frequently reviewed than schedule (Smith & Friedman, 1980).

Schedule information, on the other hand, is reported by DoD program managers, but normally in prose or tables in reports such as the SARs and others. The challenge for this effort was to identify schedule data and render it into a form that can be mathematically compared and examined. This section discusses the process developed to convert schedule information into schedule data.

Data for this research was obtained from the Defense Acquisition Management Information Retrieval (DAMIR) database, a repository for, *inter alia*, the DoD Selected Acquisition Reports (SAR). The SAR is a summary of the acquisition data of selected Major Defense Acquisition Programs (MDAP). SARs are required by Title 10 USC § 2432 which states,

> The Secretary of Defense shall submit to Congress at the end of each fiscal-year quarter a report on current major defense acquisition programs. Except as provided in paragraphs (2) and (3), each such report shall include a status report on each defense acquisition program that at the end of such quarter is a major defense acquisition program. Reports under this section shall be known as Selected Acquisition Reports.

The available DAMIR database includes SARs from 1997 to 2017. The schedule section of the report consists of a Gantt chart and table showing the major milestones and current estimates. Figure 1 is an example of the schedule data found in a SAR. The section titled Change Explanations (CE) provides a description of the schedule changes. Both the graphic and the change explanation sections are rendered as unformatted text.

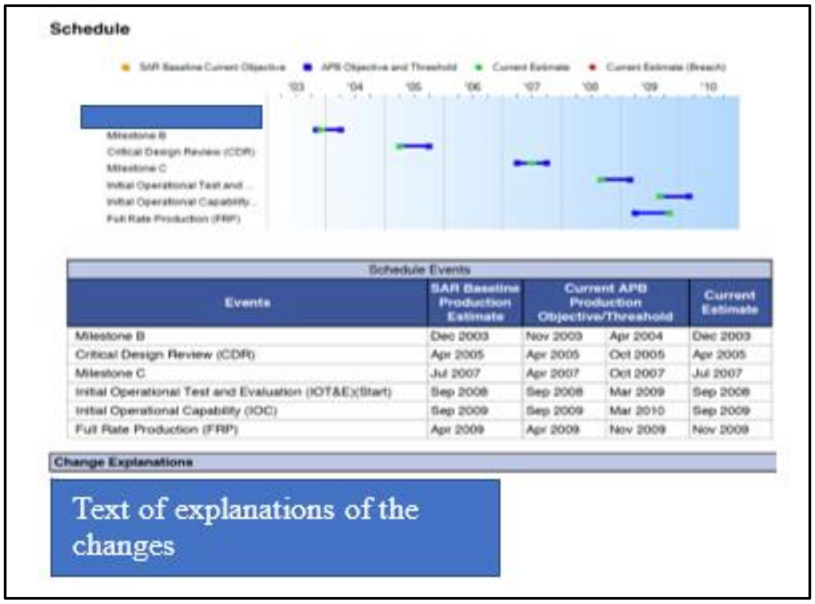


Figure 1. SAR Schedule of Data Example

The information reported by the program managers in the SAR consists of an executive summary, a brief description of the overall program with separate sections for major subprograms and identification of threshold breaches and discussions on cost, schedule, and/or performance issues. The database can be searched by program and year. Data accessibility work by OSD now provides the possibility of extracting SAR data from PDF forms into a spreadsheet. Specific queries allow an analyst to mine particular sections of the SAR, to include that used for this effort—schedule. The extracted data for schedule includes program information, key dates (milestones) and other identifying information. However, the data describing what changed and by how much is provided as text. Thus, a process needed to be developed to convert textual explanations to normalized, measurable data.

The total number of schedule records in the available SAR database was 3,969. The data used in this study are a subset of the SAR reports of 1,224 programs from 1997 to 2017. Each program potentially had between one and 20 entries (corresponding to the 20-years period and depending on when the program was initiated, and whether any schedule changes were reported). Of those 1224 programs, the available SAR schedule data consisted of 1,948 entries in the "change explanation" (CE) field of the database. In this preliminary study, those systems with no entries in the CE field were not examined. Table 1 details the overall data.

Table 1. Overall SAR Data Information, 1997–2017

Time period covered	1997–2017
Total number of records in the obtained data	3,969
Total number of programs in the database	1,224
Number of Programs/ Subprograms with Schedule Change Explanations	165
Total number of change explanations	1,994

Central to an understanding of weapon system scheduling (and as a way of converting change explanation text to data) is an examination of those factors that historically have led to increases in weapon system development times. The major studies of the past two decades have identified a number of factors that have contributed to increased duration. Thus, the next step was to identify these factors. A literature review revealed several studies that have classified weapon system development delay factors. An example of the explanations includes budget, funding, complexity, technical difficulty, and requirements stability (Drezner & Smith, 1990; Smith & Friedman, 1980; Van Atta et al., 2015). A list of these project delay factors is at Table 2. Not all these previously identified delay factors were evident in this study; however these factors provided a starting point for this analysis.

Table 2. Identified Generic Factors Causing Delays in Weapons System Development

Factors
Competition at the prime contractor level
Concurrency, overlap in time and effort between the development and production phases of a program
Funding adequacy/ stability
Existence of prototyping
Separate contracts for each phase of the program
Priority of the program to the service relative to other ongoing programs
External guidance such as OSD or congressional direction, reviews, restrictions, and designations
Joint management with other agencies
Program complexity, or interactions with agencies external to the program
Technical difficulty
Concept stability, or stability in mission, operational concepts, and doctrine
Contractor performance changes/Contract changes
External events such as inflation, earthquakes, labor strikes, etc.
Major requirements stability
Program manager turnover
Rework
Design Freeze

The classification of the change explanation (CE) entries was a two-step process. First, each change explanation was examined and a determination on causality made. Using the abovementioned factors as a classification mechanism, in the first pass, the project office change explanations were examined, and an initial determination of the schedule factor(s) was identified. It became clear in this preliminary analysis that in many cases, there was more than one cause of the delay. Those explanations with more than one cause were initially classified, then further refined. Those entries that required further analysis were flagged in order to return and further refine the classification. Some data were not assigned a code because of either duplicative information, or because of what appeared to be arbitrary schedule updates that appeared to be no real changes in schedule activities. Table 3 shows the delay factors identified/determined in this analysis, and the number of identified cases of each factor. Numbers do not total because of more than one factor identified on some of the programs.

Table 3. Minimum and Maximum Schedule Delays (Months)

Delay Factor	# instances	Maximum Delay (months)	Minimum Delay (months)
Administrative changes to schedule including updates to APB, ADM changes as well as changes resulting from Nunn-McCurdy processes and program restructuring	460	168	5
Technical	291	60	4
Testing delays	283	66	1
Delay in availability of key capabilities/facilities (launch vehicle/testing facilities/IOT&E units)	3	13	6
Budget/Funding Delays	52	43	1
Delays attributed to the Contractor	50		
Delays because of Rework	16	4	1
External events such as inflation, earthquakes, labor strikes, etc. *(Force Majeure)*	4	4	1
Delays due to Contracting/Contract Negotiation/Award delays	29	27	1
Actuals (updating previously reported dates to actual occurrence)	172	13	-39

The last step in the analytical process was to record the actual reported delays. The delays were listed in the SAR as dates. That required conversion from dates into a uniform format (months). Delays ranged from one month at the low end to 168 months on the high end. The delays were tracked to the identified factors. There were 1,216 instances of increases in time, and 150 instances of decreases in time. The delay factors, minimum and maximum delays, are shown in Table 3. Note that in some cases, dates were brought forward. In this case, those dates were noted as minus (-) numbers, representing a possible decrease in the schedule. In practice, however, a decrease in schedule captured in this manner is misleading since it is taken out of the context of the overall project. Over 95% of deceases noted were administrative in nature and either were corrections to mistakes recorded in previous SARs, or a reflection of actual dates versus planned dates. For purposes of this initial analysis, decreases were not examined.

Delay Factor Explanations

- Administrative changes include schedule updates because of APB and ADM changes, as well as changes including program restructuring as a function of decisions driven by Nunn-McCurdy results and program restructuring.
- Schedule changes identified those changes reported as a result of acknowledgement of the actual date of occurrence. These changes are also the result of receipt of approval documents from Milestone Decision Authorities to change specific dates.
- Technical schedule changes are a result of specific setbacks in technical development.
- Testing delays include both the ability to meet scheduled test dates, as well as technical issues discovered in the conduct of testing. When the testing discovered a technical issue, that technical issue was also counted as a technical problem.
- Explanations that produced no apparent changes in the schedule data reflect comments in the change explanation, but do not produce an actual change in the schedule. Examples include cases of achievement of IOC/FOC, as well as re-designations of milestones driven by ADM decisions.
- Delay in availability of key capabilities/facilities are a result of weather delays including satellite launches.
- Budget/Funding Delays are tied to specific notes on lack of budget, decrease in budget or changes by Congress to the specific program.
- Delays attributed to the Contractor result from construction and delivery delays as well as delays attributed to delivery of subcontractor materials.
- Delays because of Rework reflect both quality issues where the budgeted work must be redone in order to make it functional, as well as the feedback/follow-on problems caused throughout the development.
- *Force Majeure* are external events such as inflation, earthquakes, labor strikes, etc.
- Delays due to Contracting/Contract Negotiation stem from either problems in negotiation, delays in approvals for RFP releases, modification to contracts, or delays in awarding contracts.
- Actuals are the language used to describe simple updates to previously reported dates.

The CE section included the table described above; however the information was unformatted. In some cases, numerous dates for different events had changed and were reported. For purposes of this initial research effort, the date captured was the longest duration activity shown. Future efforts will identify and report all events.

Analysis

One of the objectives of this research was to identify, analyze, and provide those schedule factors causing delays in weapons system development. The final aspect of this study is to explore how the data extracted from this SAR analysis can be used to assist in schedule planning and development. The tools of scheduling (currently based on the CPM and PERT techniques discussed above) apply a network approach to define critical activities, slack, and the overall time required to complete the development. The network approach also provides the basis for cost estimation, resource allocation, management focus and risk assessment, and provides a visual flow of the effort. However, notwithstanding decades of study and countless man-years of experience, we are still missing something. One of those things we are missing is an acknowledgement of the dynamic nature of projects. Our current static view starting with planning has to change. A first step is to review the delay factors evident in the past 20 years of DoD Major Defense Acquisition Programs (MDAP). As a minimum, incorporating factors identified in both the planning and execution process could be a start.

The delay factors suggest PMs should attempt to plan for the time necessary to deal with oversight, information reporting and both the time takes, as well as the impacts of decisions—internal and external to the program. As the GAO pointed out in a 2015 study, the program office overheads associated with administrative activities added on an average of two years to complete:

> Programs we surveyed spent on average over 2 years completing the steps necessary to document up to 49 information requirements for their most recent acquisition milestone. This includes the time for the program office to develop the documentation and for various stakeholders to review and approve the documentation. (GAO, 2015)

For example, the GAO found that the F-22 Increment $3.2 billion Modernization spent 3,800 staff days to prepare 33 milestone documents and present 74 briefings for the Milestone B process (GAO, 2015). This work had a cost of some $10 million. Those 3,800 staff days obviously would also have impacts on the schedule, potentially more significant than on cost. This is not to argue the necessity for the program office to gather this information. Instead, it is a factor that should be accounted for in the program scheduling plan.

Figure 2 shows the cumulative schedule overrun hours for all programs analyzed in the period of 1997 to 2017. Of note are years 2010 and 2011 where the year-to-year increase in time is an order of magnitude larger than that any other year. This particular jump in delay was caused by the CVN-78 program.

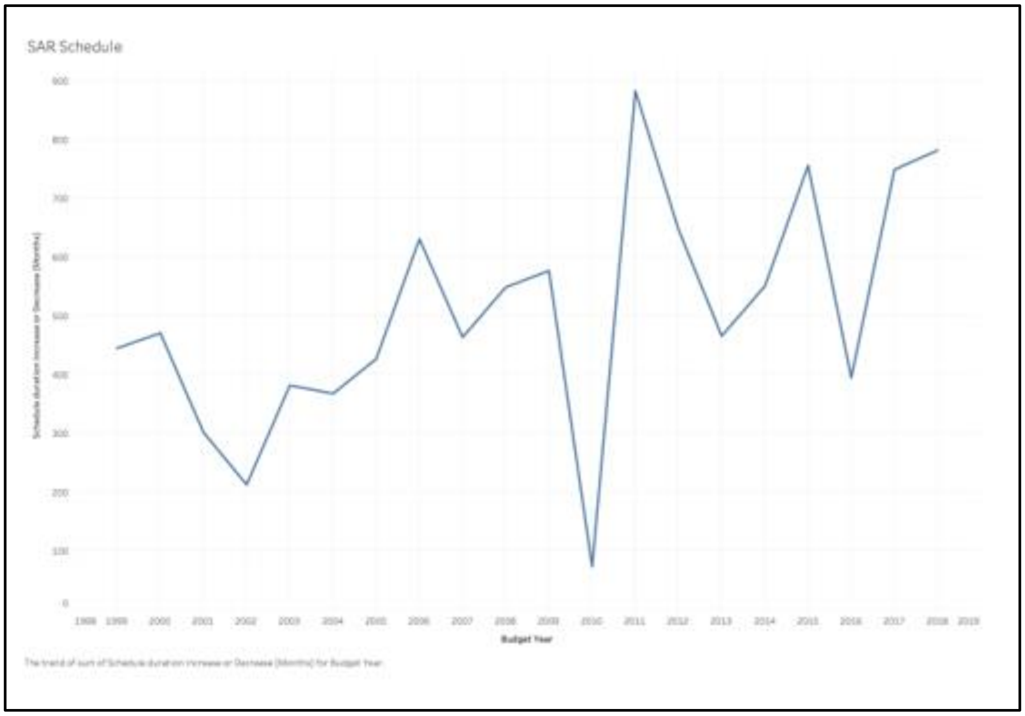


Figure 2. Sum of Schedule Delays by Year, 1998–2017

Figure 3 provides a trend line and forecast of the delay. Using this data, the forecast total delay hours across all programs in 2019 equals 712 hours, and in 2020 that forecast increases to 729 hours.

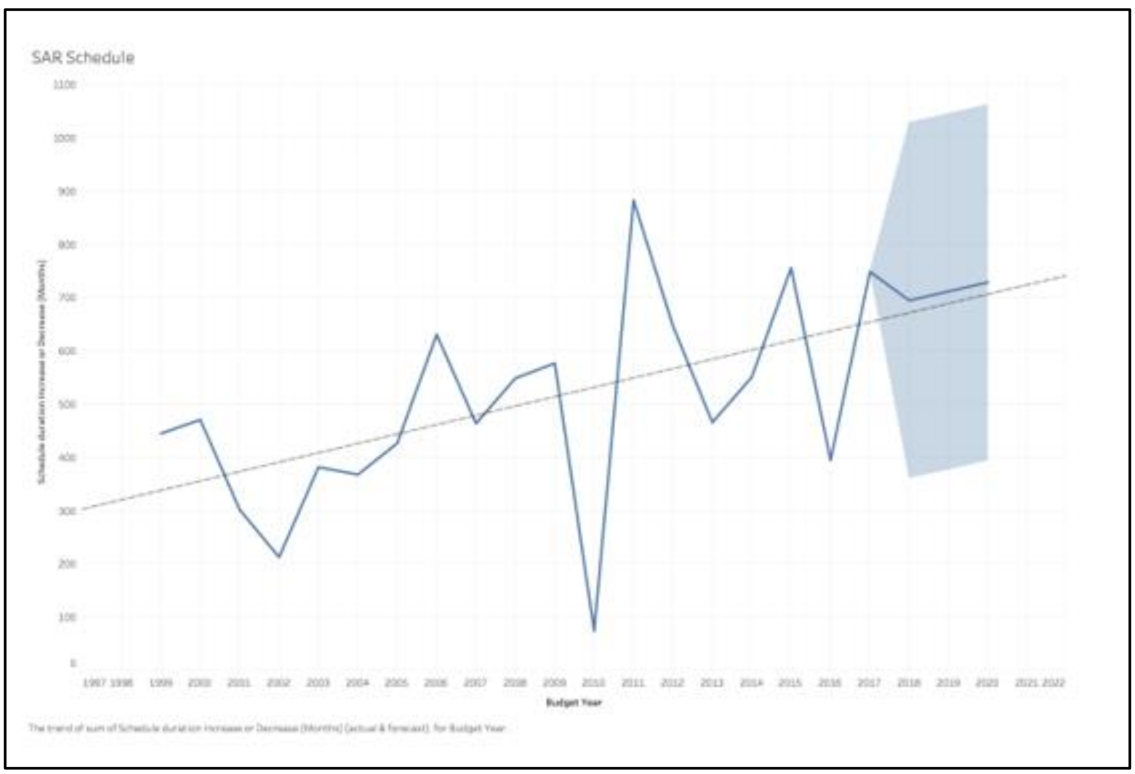


Figure 3. Forecast Total Hours for 2019 and 2020

Figure 4 shows the hours overrun by program for the same period. While there is no real proven correlation between schedule and cost overruns, the fact that the overruns are significant in this representative group of programs emphasizes the scope of the problem. And cost is not the only aspect involved. The delays also reflect added time before the systems are in the hands of the warfighters.

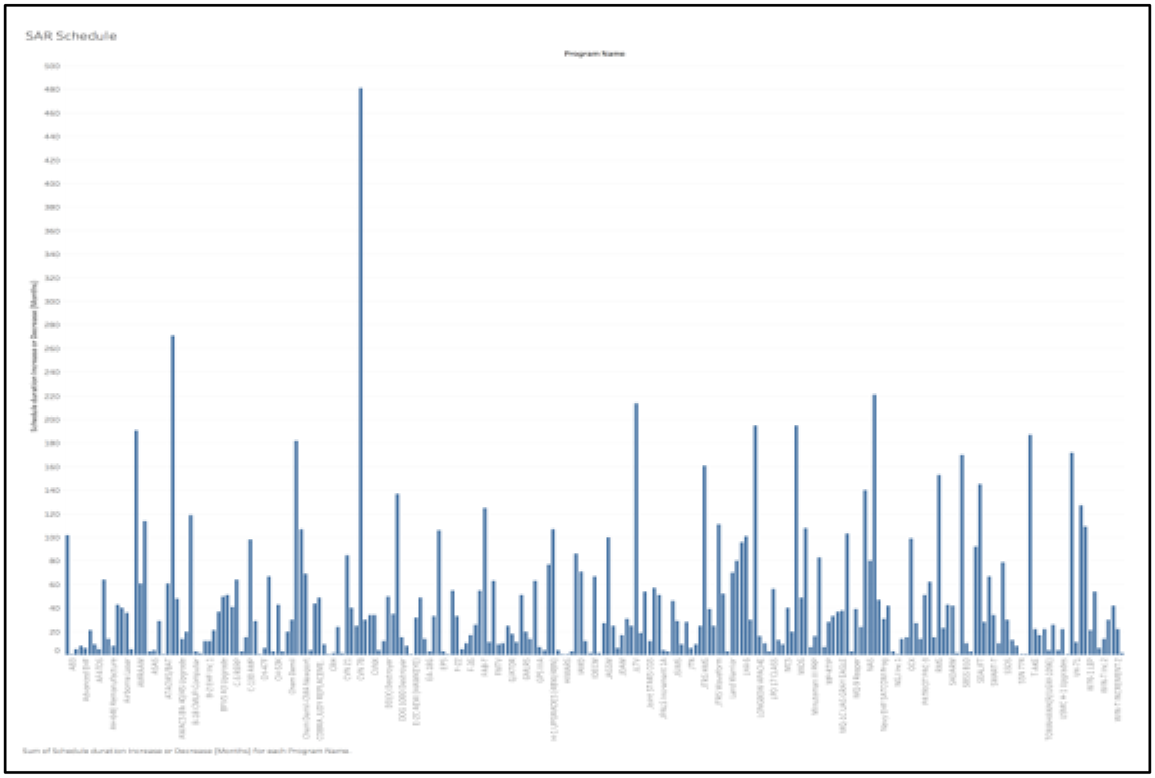


Figure 4. Total Hours Overrun by Program, 1997–2017

Using the Data

The last step in this research is to suggest a way for DoD program managers to use the data that comes from this examination of delay factors. As noted above, the CPM/ PERT approach to scheduling precludes the use of historical data at the program schedule level. And, while some companies track task estimation data, that data is often proprietary, and more focused on technical process estimation.

The basic assumption that work proceeds as planned in the network from start to finish is naïve at best (Franck et al., 2017). This static view provided by traditional project scheduling ignores the reality of project management that the work might not be done correctly or to the necessary quality (Cooper, 1993c). This same view also fails to consider that the results of decisions, whether good or bad, cause reactions in the project, much as inputs results in outputs in any system. Weapon system development reality using classical network analysis cannot delineate the progress of a project (Williams et al., 1994). Therefore, we should consider alternate ways of examining these problems.

System dynamics “deals with the time-dependent behavior of managed systems with the aim of describing the system and understanding through quantitative and qualitative models” (Coyle, 1996). Instead of the static view we are used to seeing in defense acquisition projects, system dynamics considers weapon systems development as a system, with inputs and outputs. Further, the rework cycle proposed by Cooper, also helps explain one of the project dynamics present in every development (Cooper, 1993a). Figure 5 shows the rework cycle. The concept is simple: Not all work attempted is completed correctly the first time. And, that work not completed correctly is not recognized. That work, initially undiscovered is at some point discovered and then moves into the "known rework" block. That known rework must be redone, both delaying completion of the overall project, and costing more. In practice, this effectively represents an increase in the work to be done.

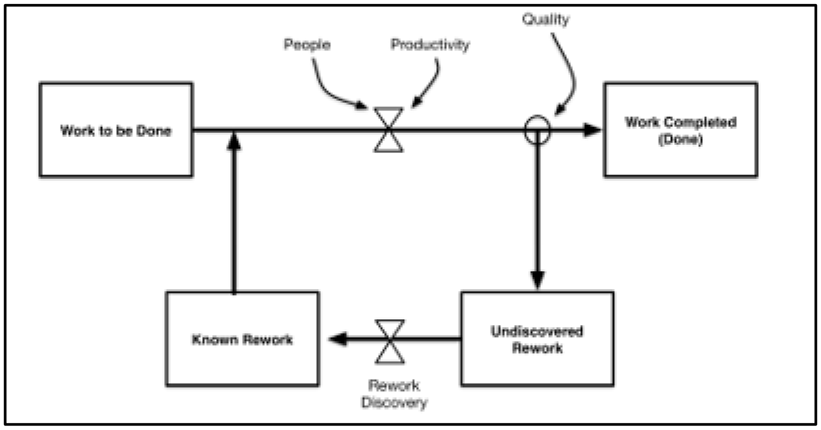


Figure 5. The Rework Cycle

The notations "people" and "productivity" flow into a valve that further controls the flow of work from that needing to be done to that work completed. People are the number of workers, and their level of training and expertise. Productivity is a measure of their efficiency. Simple scheduling in this instance takes the number of people times their efficiency and applies that to the number of tasks in the Work to be Done stock.

The rework cycle is a fundamental system dynamics concept first articulated by Cooper (Cooper, 1993a, 1993b, 1993c). The basic flow of work in a development is from work to be done (tasks or work packages) to work completed. Connecting that flow is a "valve" that regulates the flow. In the rework cycle, that flow is determined by people (numbers, skills, availability) and productivity. People times productivity provides a flow rate, for example: tasks per month. Quality is another modulator of the flow of work. Quality is simply a measure of whether the task was accomplished correctly and completely. Given the exploratory nature of research and development efforts, it is entirely possible that a planned development task fails to accomplish the task goal, and the task must be redone. Similarly, people may be operating at a high level of productivity, but not producing quality work.

There are two types of rework, known and undiscovered. These categories are integral to the nature of weapons system development. Developmental test does identify some of the work that needs to be redone, and that work flows to the known rework stock. However, there is work that may pass developmental test, but is later found to be deficient (software "bugs" are a good example). Those deficiencies may not be discovered for significant amounts of time and may also cause follow-on developmental efforts to slow or fail until they are finally discovered. Rework is generally a known issue for experienced project managers and was reported in some of the SAR data used in this study. Understanding the impact of the rework cycle coupled with the effects of other delay factors can provide project managers a tool to develop better schedule estimates.

Figure 6 shows a simplified, generic project with 1,000 tasks, executed by 10 people at a notional 90% productivity rate. The X-axis shows months, and the Y-axis shows the number of tasks. The graph shows both a steady reduction in work to be done, and an equally steady increase in the work completed. The graph shows completion of these 1,000 tasks at month 117.5. This representation of a CPM type schedule represents a deterministic view of a project that doesn't allow for delays or changes. This is one of the limitations of CPM and PERT and is recognized. Adding probability calculations to these schedules attempt to make them more realistic, but the root problem remains (Kerzner, 2013; Moder et al., 1983).

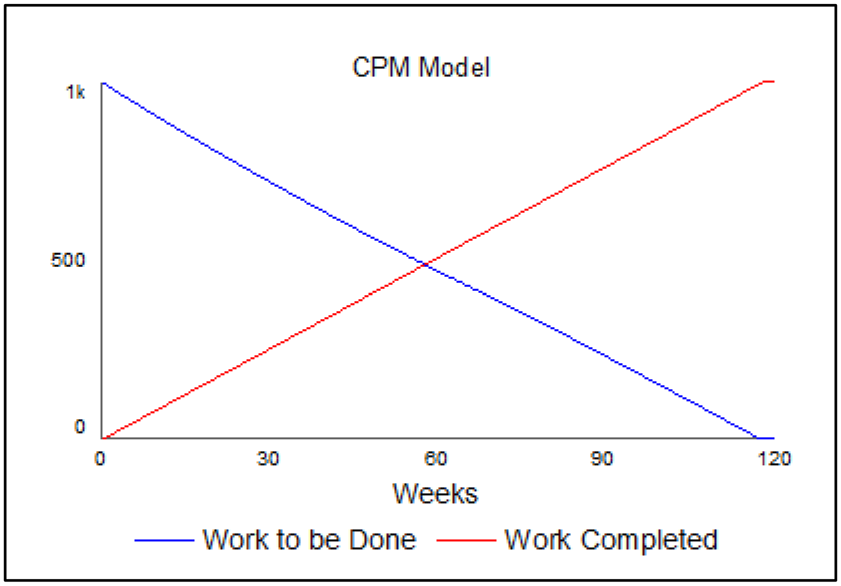


Figure 6. Model Showing the Effects of Rework

Figure 7 shows the results from the same generic model used in Figure 6, but this time incorporates the impact of the rework cycle. The X-axis shows time, and the Y-axis indicates number of tasks. Line A shows the Work Completed, line B shows Work to be Done, and line C shows a generic calculation of Rework. Comparing line A in this graph to that of work completed plot in Figure 6, demonstrates the effects of rework. In this case rework peaks at week 48 (line C) and is estimated at 75%. This means three of every four tasks must be redone, by some measures a conservative estimate especially when considering software development projects (Cooper & Mullen, 1993). Similarly, line B (Work to be Done) shows a much longer completion time than that of Figure 6. Completion time in this model run is 229 weeks, an increase of 111.5 weeks over the generic model in Figure 1, an almost 100% increase in schedule. Another way of considering the impact of rework is that instead of the 1000 tasks originally required, the number of tasks completed was 1,437—a significant increase in work requiring more time and money.

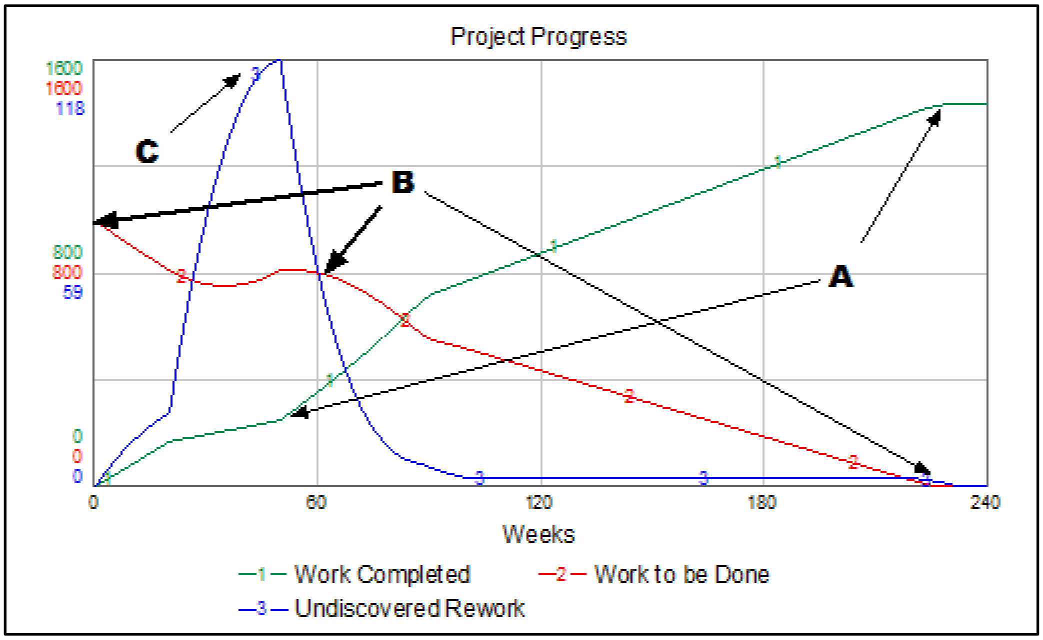


Figure 7. Model Showing the Effects of Rework

While this is an elementary model, it demonstrates that something as simple as rework can have a significant effect on project schedules.

A tool used in system dynamics to capture cause and effect is a causal map. The causal map becomes a tool used for the development of a model of the delay factors identified. Figure 8 is an initial causal loop diagram capturing some of the identified factors in weapons system program schedule delays. The factors shown are a subset of those identified for brevity in this paper.

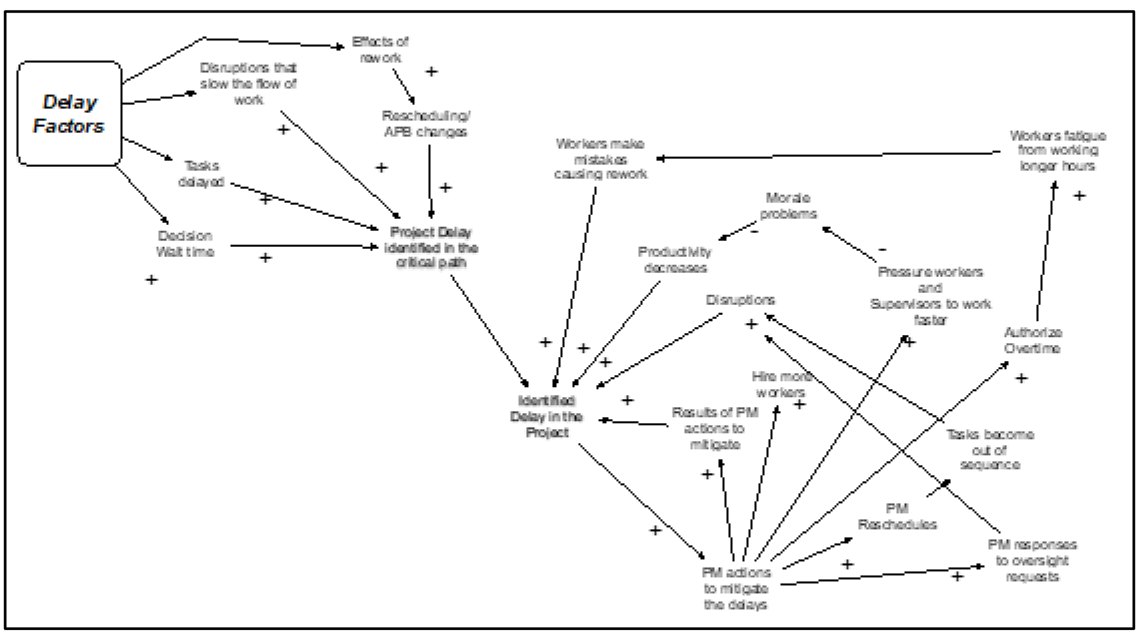


Figure 8. Delay Factors Triggers for Project Delays
(Howick, 2003)

Delay factors plus the effects of rework, decision wait time, tasks start delay, and other disruptions result in the PM (or PMO) recognizing a schedule problem (delay in the critical path). Invariably, the PM must take action to attempt to return the project to the equilibrium expressed as being on schedule. This, the PM could approve overtime, reschedule, or take some other mitigation. The pressure to get back on schedule is driven by many factors including cost considerations, pressure from the oversight organizations, and in weapons systems development, the necessity of delivering capability to the warfighter in the most efficient time. Regardless of the reason, the PM "does something." The plus and minus signs indicate the effect of the actions taken.

The project is a dynamic system with feedback loops, and invariably decisions taken to address one problem have impact on or create new problems. For example, approving overtime does initially address schedule issues as more work is being done in shorter periods. However, a recognized problem of overtime is fatigue. Fatigue causes workers to make mistakes, and those mistakes result in having to redo the work, thus perpetuating problems that were thought solved.

Similarly, hiring more workers causes more problems. Assuming the new workers have the requisite skills, they need to be trained/ acclimated to the actual project situation. In the *Mythical Man Month*, Brooks (1995) explained how this concept works in software development. In reality, it is universal (Brooks, 1995).

Finally, while many of the delay factors identified from the SAR analysis can be explained in Figure 4, others require further examination. One of the biggest challenges is the area of decisions, both internal and external. The internal decisions drive many of the actors discussed above. However, the PM must also deal with external decisions that can eventually impact the development.

Figure 9 is a notional graphic that represents a generic decision cycle in the context of the rework cycle. While the results of this data analysis included rework, the majority of the identified delay factors were decision-focused. Those decision centric factors included represent this decision cycle. The notation is shown between the work to be done and work completed boxes because many of the decisions identified occur outside the project manager's purview. The exogenous factors identified cause either reactions to those factors, or force other internal decisions. While not normally a part of the rework cycle, we suggest that a formal appreciation of a decision cycle, and the time it takes for decisions to be made both internal as well as external to the program management cycle must be considered.

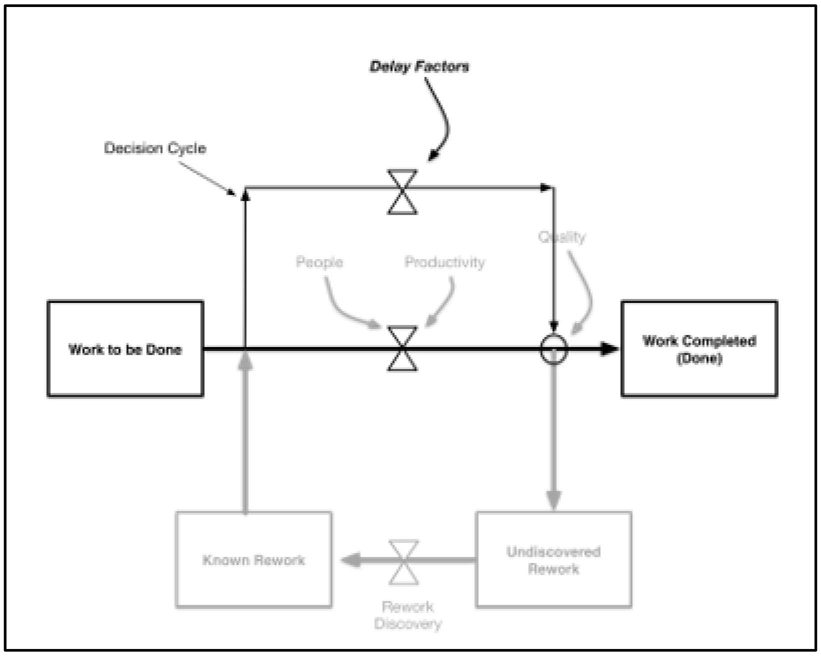


Figure 9. Notional Decision Cycle Added to Rework Cycle Diagram

Conclusion

No program manager sets out to overrun a schedule. "However, clients increasingly value not only cost and schedule control but cost and schedule certainty" (Godlewski et al., 2012, p. 18). Those clients for defense acquisition products seek certainty as well, both in cost and schedule. It is no secret that current methods for estimating and executing schedule are insufficient. In fact, certainty is one of the potential benefits of this examination of schedule factors. Project certainty starts in effective schedule planning.

This study presented a methodology to extract and identify schedule information from Selected Acquisition Reports, as well as a process for identifying classifying delay factors in weapons system acquisition programs. Finally, the study presented a suggested adjunct to the current scheduling methods that would allow project managers to use historically accurate delay factors to augment their decision processes.

This exploration of the big data aspects of defense acquisition is the first step in a continuing effort to explore not only details of schedule, but broader details and insights on the way we manage defense acquisition programs. The next step is to link the insights gained from this look at the scheduling part of the SAR to the Acquisition Program baseline (APB). The APB, oft referenced in this effort, is the vehicle to explore the entire scheduling history of an acquisition. This, we believe, could provide a better understanding of the causes of the delays by establishing a trace between results reported in a SAR, to the factor that caused the delay.

References

Balaji, J., & James, R. B. (2005). *Project dynamics with applications to change management and earned value tracking.* Presented at the System Dynamics Society International Conference.

Brooks, F. P., Jr. (1995). *The mythical man-month: Essays on software engineering* (Anniversary Ed., 2/E). Addison-Wesley.

Christensen, C. M., & Bartman, T. (2016). The hard truth about business model innovation. *MIT Sloan Management Review*.

Cooper, K. G. (1993a). The rework cycle: Benchmarks for the project manager. *Project Management Journal, 24*(1), 17–22.

Cooper, K. G. (1993b). The rework cycle: How it really works … and reworks. *PM Network*, 25–28.

Cooper, K. G. (1993c). The rework cycle: Why projects are mismanaged. Newtown Square, PA: Project Management Institute.

Cooper, K. G., & Mullen, T. W. (1993). Swords and plowshares: The rework cycles of defense and commercial software development projects. *American Programmer*.

Coyle, R. G. (1996). *System dynamics modelling*. Boca Raton, FL: CRC Press.

Drezner, J. A., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules*. Santa Monica, CA: RAND.

Franck, R., Hildebrandt, G., Pickar, C., & Udis, B. (2017). Realistic acquisition schedule estimates: A follow-on inquiry (pp. 1–20). In *Proceedings of the 14th Annual Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School.

GAO. (2015). *Acquisition reform: DoD should streamline its decision-making process for weapon systems to reduce inefficiencies* (GAO-15-192). Washington, DC: Author.

Godlewski, E., Lee, G., & Cooper, K. (2012). System dynamics transforms Fluor project and change management. *Interfaces, 42*(1), 17–32. http://doi.org/10.1287/inte.1110.0595

Hendrickson, C., Martinelli, D., & Rehak, D. (1987). Hierarchical rule-based activity duration estimation. *Journal of Construction Engineering and Management, 113*(2), 288–301. http://doi.org/10.1061/(ASCE)0733-9364(1987)113:2(288)

Howick, S. (2003). Using system dynamics to analyse disruption and delay in complex projects for litigation: Can the modelling purposes be met? *Journal of the Operational Research Society, 54*(3), 222–229. http://doi.org/10.1057/palgrave.jors.2601502

Hughes, R. T. (1996). Expert judgement as an estimating method. *Information and Software Technology, 38*(2), 67–75. http://doi.org/10.1016/0950-5849(95)01045-9

Kerzner, H. R. (2013). *Project management: A systems approach to planning, scheduling, and controlling* (11th ed.). Hoboken, NJ: Wiley.

Moder, J. J., Phillips, C. R., & Davis, E. W. (1983). *Project management with CPM, PERT and precedence diagramming* (3rd ed.). New York, NY: Van Nostrand Reinhold.

Morris, P., & Hough, G. H. (1988). The anatomy of major projects. Hoboken, NJ: John Wiley & Sons.

Project Management Institute (PMI). (2017). *A guide to the project management body of knowledge (PMBOK® Guide)—Sixth Edition and Agile Practice Guide* (ENGLISH). Newtown Square, PA: Author.

Schwindt, C., & Zimmerman, J. (2015). *Handbook on project management and scheduling* (Vol. 1). Springer.

Smith, G. K., & Friedman, E. T. (1980). An analysis of weapon system acquisition intervals, past and present (No. R-2605-DR&E/AF). Santa Monica, CA: RAND.

Van Atta, R. H., Kneece, R., Jr., Patterson, C. M., & Hermes, A. C. (2015). *Assessing weapon system acquisition cycle times: Setting program schedules*. Institute for Defense Analysis.

Williams, T., Eden, C., Ackermann, F., & Tait, A. (1994). The vicious circles of parallelism. *International Journal of Project Management, 13*(3), 151–155.

Utilizing Public Data for Data Enhancement and Analysis of Federal Acquisition Data

Ningning Wu—is Professor of Information Science at the University of Arkansas at Little Rock. She received a BS and an MS degree in Electrical Engineering from the University of Science and Technology of China and a PhD in Information Technology from George Mason University. Dr. Wu's research interests are data mining, network and information security, and information quality. She holds certificates of the IAIDQ Information Quality Certified Professional (IQCP) and the SANS GIAC Security Essentials Certified Professional. [nxwu@ualr.edu]

Richard Wang—is Director of the MIT Chief Data Officer and Information Quality Program. He is also the Executive Director of the Institute for Chief Data Officers (iCDO) and Professor at the University of Arkansas at Little Rock. From 2009 to 2011, Wang served as the Deputy Chief Data Officer and Chief Data Quality Officer of the U.S. Army. He received his PhD in information Technology from the MIT Sloan School of Management in 1985. [rwang@mit.edu]

M. Eduard Tudoreanu—is Professor of Information Science at University of Arkansas Little Rock. Professor Tudoreanu has expertise in human-computer interaction, information quality, advanced visualization of complex data, and virtual reality. He worked on visual data analysis, and has extensive experience in software development and user interface design. Professor Tudoreanu was the founding Technical Director of the Emerging Analytics Center. He has been the keynote speaker at ABSEL 2010, and served as a panelist for the National Science Foundation and Missouri EPSCoR. He earned his Doctor of Science degree in Computer Science in 2002 from the Washington University in St. Louis. [metudoreanu@ualr.edu]

Abstract

It is challenging to standardize data; yet, the capabilities to draw upon data across information systems hold huge potential for improving defense acquisition and procurement. Acquisition planning and management involves many decision-making and action-taking processes that cover a complex environment including actual acquisition, contracting, fiscal, legal, personnel, and regulatory requirements. A sound decision-making process has to rely on data—high quality data. Often the available data is dirty, outdated, incomplete, or insufficient for the expert to make a decision. On the other hand, there are enormous amounts of data on the web that can be utilized to crystalize the needed information. These data repositories are often publicly accessible and from a variety of sources including websites, government reports, news, wikis, blogs, online forums, and social media. This paper investigates how to leverage the information in public data sources to complement the internal data in order to support effective acquisition planning and management. This research is based on publicly available government acquisition databases at usaspending.gov and fpds.gov. It takes a data science approach for analyzing acquisition databases and focuses on two major tasks: (1) research on leveraging the web data for quality assessment and improvement of federal acquisition data and (2) research on appropriate data analytic techniques to discover useful information that can potentially help federal acquisition management and planning process.

Introduction

Military agencies collect, store, and integrate data from various sources in their acquisition and procurement decisions and management processes. However, data complexity is profound. Often, data are publicly available, but can be dirty, and become even dirtier due to biases during collection. Furthermore, acquisition, procurement, and contract data have varying data quality problems and can thus be difficult or even impossible to integrate.

Across the Department of Defense (DoD), there are hundreds of information systems that are drawn upon for defense acquisition and procurement tasks. It is challenging to standardize data across all of these information systems; yet, the capabilities to draw upon data from these systems not only are essential, but also hold huge potential for improving acquisition and procurement and reducing substantial costs across various acquisition and procurement programs. A critical challenge facing the DoD, and federal agencies in general, is how to develop data visibility capabilities to support various acquisition and procurement tasks without enforcing a single data standard across these hundreds of systems.

On the other hand, the vast quantity of online information provides great opportunities for us to harvest and enrich our data and knowledge. There are a variety of sources for the data including company and government websites and reports, news outlets, wikis, blogs, online forums, and social media. These sources contain rich information about almost everything and any subject we can think of. Indeed, searching for the needed information on the web has become a common practice for Internet users nowadays, thanks to the advancement of search engines and web technologies. If properly utilized, online information may help us assess and even improve the quality of the data we have. For instance, if a record contains a contractor's name but the address information is missing, we can fill the missing address by googling the contractor's address on the Internet. Similarly, if a contractor's DUNS number is found incorrect, then we may be able to find the right DUNS number by querying the websites that host DUNS number database.

A recent study by the Rand Corporation titled *Issues With Access Acquisition Data and Information in the Department of Defense* recommends several options for improving the DoD's acquisition data (McKernan et al., 2016). One option is to improve the quality and analytic value of acquisition data. It stated that according to information managers, **data verification and validation are top priorities** and the practice of building both manual and automated checks should be continued and expanded to other systems. Another option is to improve data analytic capabilities by continuing to collect both structured and unstructured data. It recommends that the DoD should try to come up with better ways of utilizing the unstructured data it collects.

This research aims to investigate appropriate data science approaches to improving the quality of federal acquisition data as well as discovering useful patterns that can further acquisition research. It will examine the feasibility of leveraging the information on the Internet for verification and validation of acquisition data. Utilizing online information faces several challenges. One of the key challenges is how to find the information that is credible and accurate from often an enormous amount of unstructured documents returned by a search. For instance, the information of an entity may spread out on various websites that have collected data from different sources and at different times. When searching the entity on the web, we may end up with thousands of if not millions of hits. Some of them may be incorrect and some out-of-date. Thus identifying the hits that contain both accurate and current information becomes a challenge. To make the problem even worse, the majority of online information is non-structured and textual. Thus, the question of how to extract the needed information from non-structured text becomes another challenge.

Research Issues

The web has greatly changed our ways of sharing and seeking information. At the same time, it has also altered traditional notions of trust due to the fact that the information can be published anywhere by anyone for any purpose, and there is no authority to certify the correctness of the information. It is up to the information consumers to make their own judgement about the credibility and accuracy of information they encountered online. To utilize online information effectively, this research needs to investigate appropriate methods to acquire valuable and reliable information online. Reliability of information can be measured from different aspects such as accuracy, timeliness, authority of information, trustworthiness of websites, and so forth. Consistency is another common issue with web data because the data on same subject might be different, or represented in different formats, scales, or metrics. Thus, resolving inconsistency of data from different sources and identifying the most accurate information become other key topics of this research.

As the majority of data on the web are unstructured text documents, it is challenging to identify, retrieve, and integrate the needed information from the web documents. Retrieval of desired information is not a trivial task and involves natural language processing, computational linguistics, text analysis, and entity identification and resolution. Other challenges of text analysis include complex and subtle relationships between concepts in text as well as ambiguity and context sensitivity of terms in text. The research will examine ways to identify and collectively integrate the needed information from both public and internal sources, and to leverage them for further acquisition research.

Research Methodology

This study is based on Federal Acquisition databases at USAspending.gov, which contains spending information of all U.S. departments between the years 2000 and 2018 for a selected state or all states. The data can be downloaded in different formats, such as CSV, TSV, and XML. Spending data are further categorized under prime award and sub-award. The types of spending include contracts, grants, loans, and other financial assistance. Our downloaded data contains 47GB data in total, covers the DoD budget between 2000 and 2017 including each type of spending data for both prime award and sub-award. We set up a database system to host the data.

Figure 1 shows the framework of the proposed Data Enhancement and Analytics System. The system has four major components, namely Quality Assessment engine (QA), Data Cleaning engine (DC), Data Enhancement and Analytics engine (DAE), and Text Retrieval and Analysis engine (TRA). The key component is Text Retrieval and Analysis engine as it supports the functionalities of the rest three components. TRA is responsible for four tasks: (1) performing searches on the Internet, (2) identifying the websites that contain the most reliable data, (3) extracting the desired information from the text, and (4) information fusion by collectively integrating information from multiple sources. When information needed for quality assessment and data cleaning is not available, TRA will search and extract the needed information online. Data Analytics and Enhancement engine aims to enhance our knowledge about data by discovering hidden and interesting patterns in the data as well as complementing the internal data with the information that is not found in the database but is potentially useful for advanced data analytics.

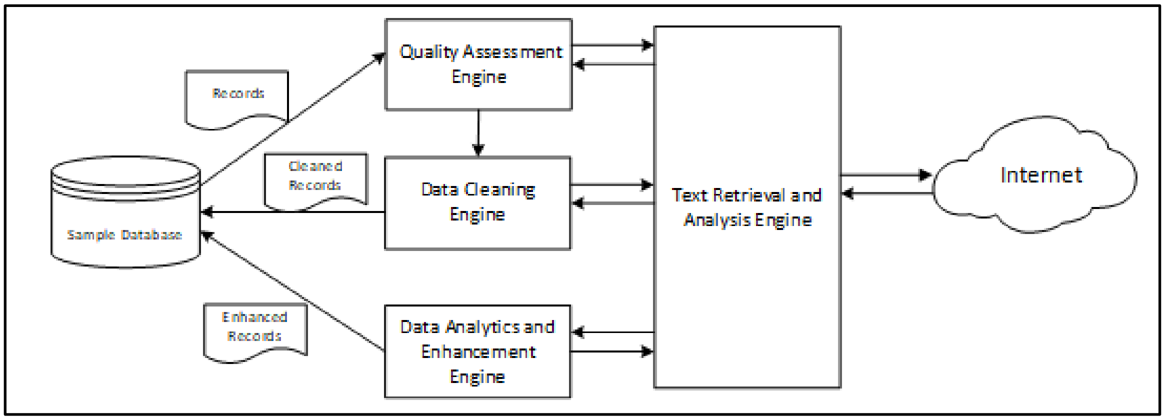


Figure 1. Framework of Data Enhancement and Analytics System

Our research methodology contains following steps:

- Assess the quality of the sample acquisition database in terms of accuracy, consistency, and completeness. Assessment on completeness is rather straightforward; however, assessment on accuracy and consistency is not, as it requires the extra knowledge about data and their semantics. For instance, to decide whether a value is accurate or not, we need to know what the expected correct value is. To evaluate whether the two values are consistent, we need to know their semantics and relationship. If they are not consistent, then we need to know which value is wrong and causing the inconsistency issue. Unfortunately, we do not always have the information we need for the quality assessment.
- Based on the quality assessment findings obtained in step 1, identify the fields for quality improvement. The key task of this step is to investigate the feasibility of leveraging the information online for both quality assessment and improvement. It will research on effective ways to evaluate the credibility of websites and to extract reliable information from a large amount of web pages.
- Apply appropriate data analytics methods to discover useful patterns from the data.
- Utilize online data to complement the information of the sample database. The primary objective of this step is to research appropriate text mining methods to retrieve the information for the purpose of advanced data analytics. Examples of information may include a business's product/service information, location, business type, business size, business relationship networks. This information can help us estimate the uniqueness of a business as well as the level of risk it might potentially pose to a project if it fails. The findings of this step can further acquisition research by identifying the room for improvement of a project.

Preliminary Research Findings

Quality Assessment

The data in sample database are organized into four tables: primeContracts, subContracts, PrimeGrants, and SubGrants. Table 1 shows general information about the tables, where *RecCnt* and *ColCnt* represent the number of records and number of columns in a table respectively; *CompleteCols* and *SingleValCols* represent the number of columns with no missing values and number of columns with only a single value across all records; and *EmptyCols* and *IncompleteCols* represent the number of empty columns and the number of columns with missing values respectively.

Table 1. Table Information of the Sample Database

Table Name	RecCnt	ColCnt	CompleteCols/ SingleValCols	EmtpyCols	IncompleteCols
PrimeContracts	23,677,787	212	50/1	0	162
SubContracts	395,569	101	41/0	3	57
PrimGrants	202,166	67	32/5	2	33
SubAGrants	11,115	101	29/4	25	47

For the quality assessment purpose, attributes are classified into two categories: identity attributes and non-identity attributes. Identity attributes provide identifier information for a contract or a contractor including project identifiers, contractor identifiers, address, telephone, and so forth. The rest attributes are non-identity attributes that do not provide identifier information. This study focuses on the quality assessment of only identity attributes on three dimensions: column completeness, accuracy, and field length consistency. Only PrimeContracts and SubContracts tables are used in this study as they have relatively more quality issues.

Column Completeness

Completeness can be measured in different aspects including column completeness, schema completeness, and population completeness. Column completeness measures the degree to which there exist missing values in a column of a table. Schema completeness measures the degree to which entities and attributes are missing from the schema. Population completeness measures the degree to which members of the population that should be present but are not present. Since there is not enough information for assessing schema and population completeness, the study will focus only on column completeness, which is measured by the percentage of non-missing values in the column.

Figures 2 and 3 show the completeness measures for identity attributes of PrimeContracts and SubContracts tables respectively. PrimeContracts table has perfect or near perfect completeness on three attributes. SubContracts table has 100% completeness on prime_award_piid and subawardee_dunsnumber, but it has missing values on both prime awardee's and subawardee's parent dunsnumbers. A possible reason might be some contractors may not have a parent company.

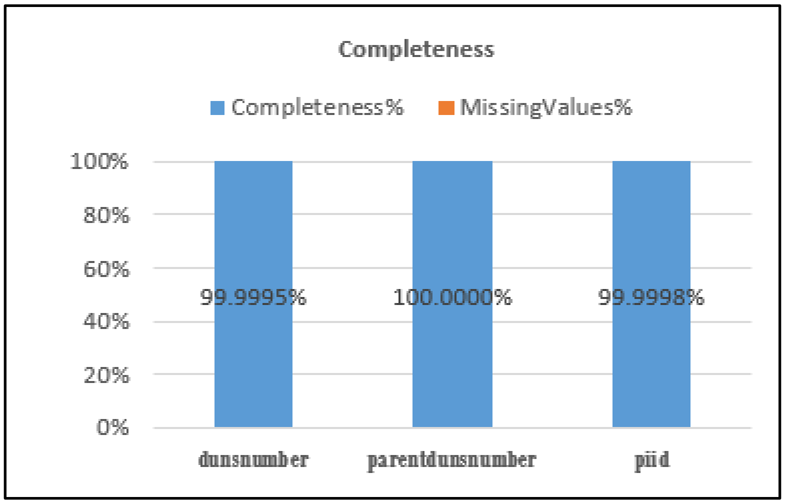


Figure 2. Completeness Measure of Identity Attributes for PrimeContracts Table

Figure 3. Completeness Measures of Identity Attributes for SubContracts Table

Attribute Length Consistency

Attribute length consistency measures how consistent lengths of an attribute's values are. Each identity attribute of the PrimeContrats table is supposed to have fixed-length values, as are the identity attributes of the SubContracts table. For example, a DUNS number, provided by Dun & Bradstreet (D&B), is a unique nine-digit identification number for each physical location of a business. Thus a DUNs number of other than nine digits is problematic. Figures 4 and 5 show the assessment of attribute length consistency of both tables. DUNS numbers in PrimeContracts table have a variety lengths; while DUNS numbers in SubContracts are consistently of nine digits.

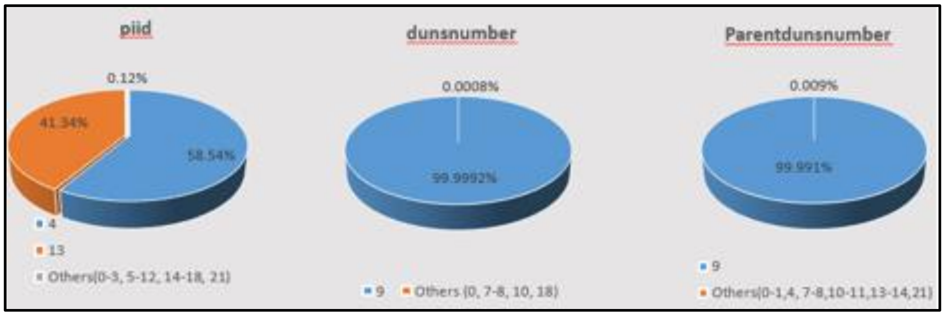


Figure 4. Field Length Consistency Assessment of PrimeContracts Table

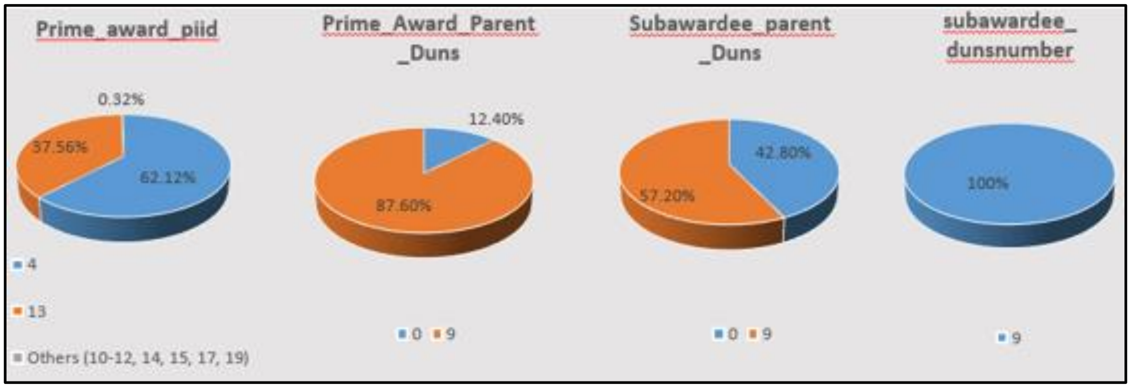


Figure 5. Field Length Consistency Assessment of SubContracts Table

Accuracy

Data is accurate if it is free of error and conforms to gold standards of data. Accessing accuracy is not an easy task as it requires the knowledge of correct data and needs to compare each data item with the known correct value, which is often not available. The accuracy assessment of this study will focus on the DUNS numbers first, because we can use the free service provided by Duns & Bradstreet database to search a business's DUNS number based on its name.

Quality Improvement

For proof of concept, the first phase of this research will focus on fixing incorrect DUNS numbers, and the online Duns & Bradstreet database will be used for this purpose. Duns & Bradstreet database contains DUNS numbers of 285 million commercial entities and 100 million associated contacts. It provides the services that allow users to search a company's information by name, telephone number, or DUNS number. When searching for a DUNS number, the database doesn't return the query results immediately, instead the results are sent through email. This somehow discourages the use of a script program to automatically submit a query and retrieve the results.

Given a business's DUNS number, the Duns & Bradstreet database can be queried for the corresponding business name, address, and telephone number. It is a bit tricky when querying the DUNS number based on the company name, as a company may have more than one DUNs number with one for each branch. To identify the right DUNS number for a branch, extra information such as the street address, zip code, and telephone number is needed.

The research started with the *dunsnumber* attribute of the PrimeContracts table. There are a total of 26 records in the table, all with a 10-digit *dusnumber*. A closer study on those DUNS numbers revealed that most of them are actually the phone numbers. Among those records, only three have information on *vendorname* attribute. The rest of the records misplace the vendor names into other fields. Since a company may have multiple branches with each located at a different address, searching the DUNs database by a business name may result in multiple matches. Thus address information is critical for finding the best match. Unfortunately, address information is misplaced in all 26 records. Figure 6 shows a few columns of the 26 records with misplaced values for vendor name and address.

ID	dunsnumber	multipleorsingleawardidc	vendorname	vendorenabled	vendorlocationdisableflag	streetaddress	streetaddress2	streetaddress3	city	state	zipcode	vendoralternatesitecode
1	2022611902	CB&I FEDERAL SERVICES LLC			2370 TOWNE CENTER BLVD		BATON ROUGE	LA	708068172	UNITED ST	LA	386491765
2	2082332929	SUNDANCE-TLI			275 S 5TH AVE_ STE 215		POCATELLO	ID	832016429	UNITED ST	ID	967391967
3	2164515588	LUCAS PRECISION_ LIMITED	LUCAS PRECISION_ LIMITED		13020 SAINT CLAIR AVE		CLEVELAND	OH	441082033	UNITED ST	OH	622006591
4	2536803243	U.S. OIL TRADING LLC			3001 MARSHALL AVE		TACOMA	WA	984213116	UNITED ST	WA	784187226
5	3053925669	WORLD FUEL SERVICES FL		9800 NW 41ST ST STE 400		MIAMI	FL	331782980	USA: UNITEI	FL	25	3054288000
6	3256738838	AVFUEL CORPORATION			47 W ELLSWORTH RD		ANN ARBOR	MI	481082206	UNITED ST	MI	131423808
7	4103792800	MABEY BRIDGE & SHORE_ INC			6770 DORSEY ROAD		ELKRIDGE	MD	210756205	UNITED ST	MD	296358146
8	4106942749			1580A W NURSERY RD		LINTHICUM	MD	210902202	USA: UNITEI	MD	2	8004439219
9	4155675899	UNITED BLOOD SERVICES		6210 E OAK ST		SCOTTSDALI	AZ	852571101	USA: UNITEI	AZ	9	4159010740
10	4794524727	GREENSCAPES TOTAL LAWNCAR			3118 S 64TH CIR		FORT SMITH	AR	729034971	UNITED ST	AR	799080705
11	4806000407			2423 W MINERAL RD		PHOENIX	AZ	850419559	USA: UNITEI	AZ	7	4806000407
12	6034311331	HECKLER & KOCH DEFENSE IN			19980 HIGHLAND VISTA DR S		ASHBURN	VA	201474189	UNITED ST	VA	166031588
13	6103858200	STV INCORPORATED			205 W WELSH DR		DOUGLASSVILLE	PA	195188713	UNITED ST	PA	106768252
14	6192388341			3589 DALBERGIA ST		SAN DIEGO	CA	921132123	USA: UNITEI	CA	51	6192388338
15	7038491000			6200 GUARDIAN GATEWAY DR		ABERDEEN F	MD	210051327	USA: UNITEI	MD	2	7036413735
16	7578734959	AH-BC NAVY JV_ A JOINT VE			804 OMNI BLVD STE 201		NEWPORT NEWS	VA	236064422	UNITED ST	VA	830647272
17	7736371666	DEHLER MANUFACTURING CO._			5801 W DICKENS AVE		CHICAGO	IL	606394030	UNITED ST	IL	5069661
18	7818717449	NOBLE SUPPLY AND LOGISTIC		302 WEYMOUTH ST		ROCKLAND	MA	23701171	USA: UNITEI	MA	9	7818711911
19	8044847840			12650 E ARAPAHOE RD BLDG		ENGLEWOO	CO	801123901	USA: UNITEI	CO	6	8044847839
20	8082451911	SENTER PETROLEUM INC			3011 AUKELE ST STE C		LIHUE	HI	967661465	UNITED ST	HI	606679520
21	8085229712	ALOHA PETROLEUM LTD	ALOHA PETROLEUM LTD		1132 BISHOP STREET STE 17		HONOLULU	HI	968132807	UNITED ST	HI	81909046
22	8088335825			707 KAKOI ST		HONOLULU	HI	968192017	USA: UNITEI	HI	1	8088361957
23	8177636775	LOCKHEED MARTIN CORPOR/	LOCKHEED MARTIN CORPORATI		1 LOCKHEED BLVD		FORT WORTH	TX	761083619	UNITED ST	TX	834951691
24	9036650030	TIGER LAWN AND LANDSCAPE			1389 MCR 3222		JEFFERSON	TX	756575878	UNITED ST	TX	969073886
25	9184266191			209 SW 7TH ST		ANTLERS	OK	745233834	USA: UNITEI	OK	2	9184262871
26	9519409686	PERRIS WIND TUNNEL		2093 GOETZ RD		PERRIS	CA	925709315	USA: UNITEI	CA	41	9518057728

Figure 6. Partial View of Records With Misplaced Values

Figure 7 shows an example of using online database to retrieve the DUNS number for the business with a single DUNS number. Figure 8 shows an example of extracting the DUNS numbers of companies that have multiple DUNS numbers. The vendor name, address, zip code, and telephone numbers are used to identify the highlighted best match, then the corresponding DUNS number is retrieved.

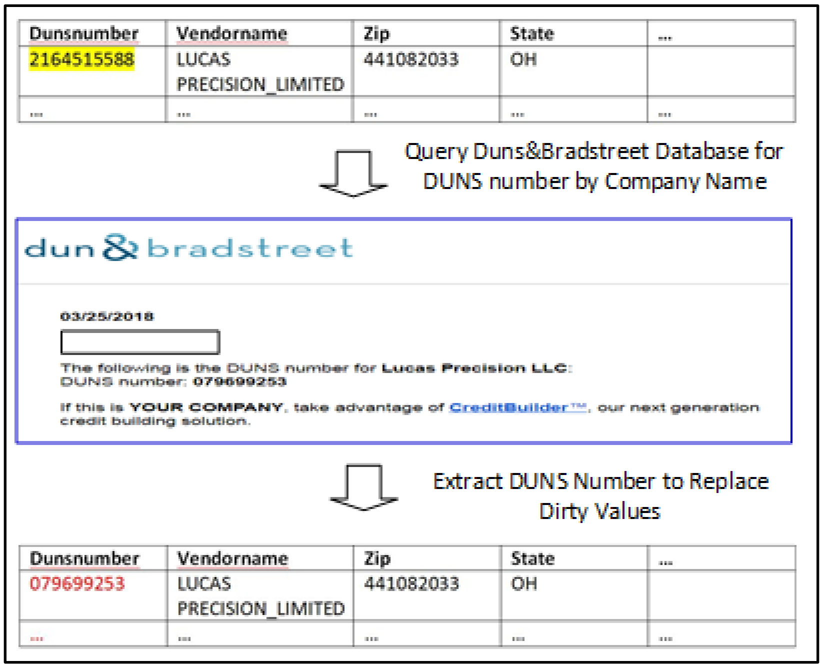


Figure 7. Examples of DUNS Number Extraction Using Online DUNS Database

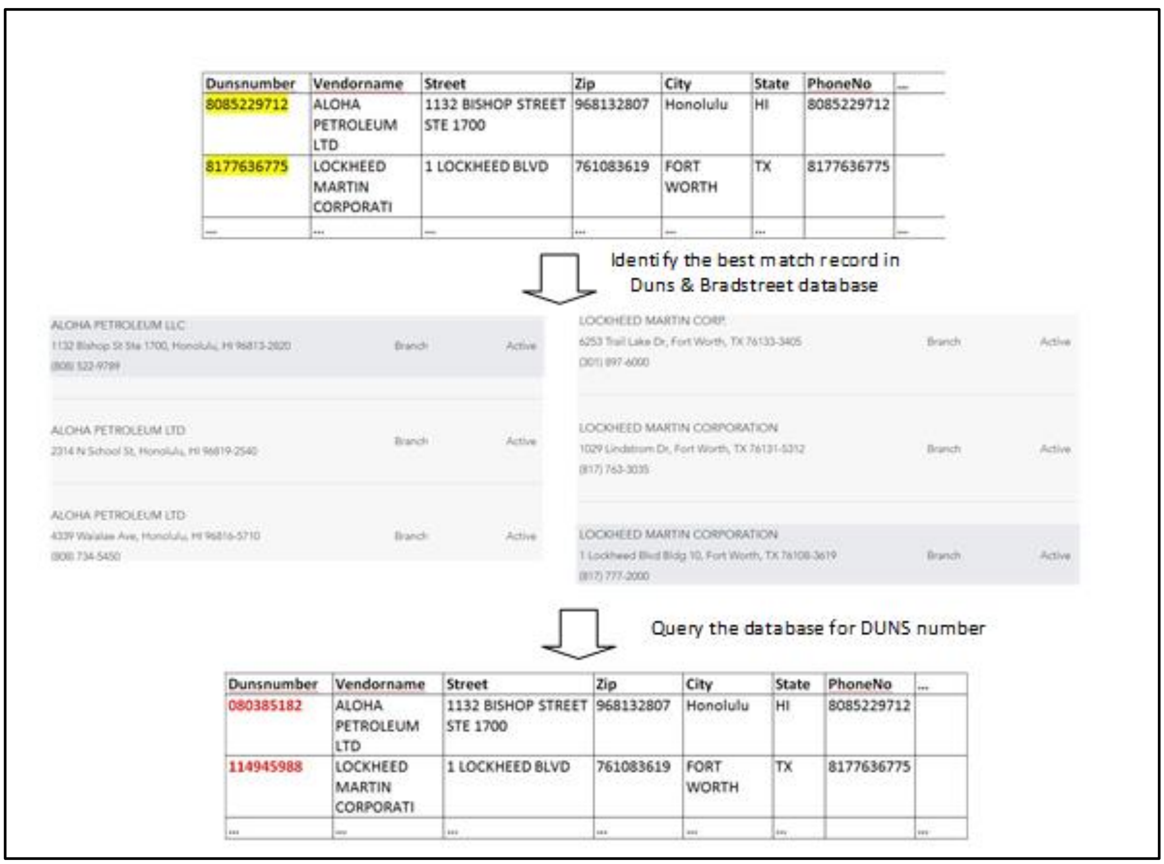


Figure 8. Examples of Identifying the Best Match in Online DUNS Database

For records with both incorrect DUNS number and missing vendor name, it is still possible to find the correct DUNS information if those records have correct and current phone numbers and address information.

Figure 9 shows the DUNS numbers, addresses, and telephone numbers that are retrieved from Duns & Bradstreet database for all 26 records. Both old and new DUNS numbers are displayed. For records with IDs 2 and 16, Duns & Bradstreet returns only a single match based on the business name; however, the returned street address is different from the one in the acquisition database. For records with IDs 8 and 11, Duns & Bradstreet returns multiple matches, all with the identical address and business name. So the DUNs number of the headquarters branch is chosen. Overall, 22 out of 26 records have *dunsnumber* filled by using the online DUNS database. The next phase of this study will be the validation and verification of the DUNS numbers and address information.

We are in the process of automating the DUNS number retrieval process, which is also a component of the Text Retrieval and Analysis engine. The Duns & Bradstreet database will be used due to its authority and reputation. The main challenge of DUNS number retrieval is to identify the best match for a company in the online database. This is indeed an entity resolution problem. Due to possible data quality problems in the sample database such as misplaced fields, typos, and missing and dirty values, identification of an exact match in DUNS database might not always be possible. The probabilistic matching methods appears promising as they can handle fuzzy matches more effectively. Another challenge is to identify the misplaced attribute values. The research will explore the methods for automatically sensing the semantics of a field based on its syntactic features and relationship with other fields and records.

ID	dunsnumber(old)	dunsnumber(new)	vendorname	Street	city	state	zip	tel
1	2022611902	079811713	CB*I Mahan LLC	2370 Towne Center Blvd	Baton Rouge	LA	70806-8172	(225)932-6500
2	2082332929	967391967*	**sundance-tli**	**305 N 3rd Ave Ste B**	**pocatello**	ID	83201-6306	(208) 233-2929
3	2164515588	079699253	LUCAS PRECISION_ LIMITED	13020 Saint Clair Ave	Cleveland	OH	44108-2033	(216) 451-5588
4	2536803243	784187226	U.S. OIL TRADING LLC	3001 Marshall Ave	Tacoma	WA	98421-3116	(253) 383-1651
5	3053925669	131504342	WORLD FUEL SERVICES FL	9800 Nw 41st St Ste 400	Miami	FL	33178-2980	(305) 428-8000
6	3256738838	020829396	AVFUEL CORPORATION	47 W Ellsworth Rd	Ann Arbor	MI	48108-2206	(734) 663-6466
7	4103792800	171902265	MABEY INC	6770 Dorsey Rd	Elkridge	MD	MD	(410) 379-2800
8	4106942749	005128988*	**Northrop Grumman Systems Corporation**	**1580a W Nursery Rd**	**Linthicum Heights**	**MD**	**21090-2202**	**(410) 765-1000**
9	4155675899	006902498	BLOOD SYSTEMS INC	6210 E Oak St	Scottsdale	AZ	85257-1101	(480) 946-4201
10	4794524727	799080705	GREENSCAPES TOTAL LAWNCARE& LANDSCAPING SI	3118 S 64th Cir	Fort Smith	AR	72903-4971	(479) 452-4727
11	4806000407	363821294*	**ORBIT INDUSTRIAL SERVICE & MAINTENANCE**	**5316 W Missouri Ave**	**Glendale**	**AZ**	85301-6006	(480) 704-4849
12	6034311331	134466999	HECKLER & KOCH DEFENSE INC.	19980 Highland Vista Dr Ste 190	Ashburn	VA	20147-4189	(703) 450-1900
13	6103858200	059946819	STV INCORPORATED	205 W Welsh Dr	Douglassville	PA	19518-8713	(610) 385-8200
14	6192388341	080090672	SOUTHBAY SANDBLASTING & TANK CLEANING INC	3589 Dalbergia St	San Diego	CA	92113-3810	(619) 238-8338
15	7038491000	827488979	ABERDEEN PROVING GROUND US ARM	6200 Guardian Gtwy	Aberdeen Proving Ground	MD	21005-1327	(410) 273-2640
16	7578734959	830647272*	**AH/BC NAVY JV LLC**	**11837 Rock Landing Dr Ste 300**	**Newport News**	**VA**	**23606-4493**	**(757) 873-4959**
17	7736371666	059530805	DEHLER MANUFACTURING COM	5801 W Dickens Ave	Chicago	IL	60639-4030	(773) 637-0615
18	7818717449	107910259	NOBLE SALES CO. INC	302 Weymouth St	Rockland	MA	02370-1171	(781) 871-1911
19	8044847840	079427300	PERFORMANCE FOOD GROUP	12650 E Arapahoe Rd	Englewood	CO	80112-3901	(303) 662-7100
20	8082451911	069891588	SENTER PETROLEUM INC	3011 Aukele St Ste C	Lihue	HI	96766-1430	(808) 245-1911
21	8085229712	080385182	ALOHA PETROLEUM LTD	1132 Bishop St Ste 1700	Honolulu	HI	96813-2820	(808) 522-9789
22	8088335825	007050586	GARLOW PETROLEUM	707 Kakoi St	Honolulu	HI	96819-2017	(808) 836-1957
23	8177636775	008016958	Lockheed Martin Corporation	1 Lockheed Blvd Bldg 10	Fort Worth	TX	76108-3619	(817) 777-2000
24	9036650030	969073886	CHRIS GIBBONS	1389 Mcr 3222	Jefferson	TX	75657	(903) 665-8190
25	9184266191	619338010	CHOCTAW MANUFACTURING DEFENSE CONTRACTC	209 Sw 7th St	Antlers	OK	74523-3834	(580) 298-2203
26	9519409686	141883517	PERRIS SKYVENTURE	2093 Goetz Rd	Perris	CA	92570-9315	(951) 940-4290

Figure 9. DUNs Numbers and Addresses Retrieved From Duns & Bradstreet

Data Mining

Data mining is the process of examining large data sets to uncover hidden but interesting patterns such as unknown correlations, market trends, customer preferences, and other useful business information. The analytical findings can shed significant insights to help add perspective to use the data and to lead to more effective decision makings. Some major data mining techniques include association discovery, classification, clustering, regression, sequence or path analysis, and structure and network analysis.

Association discovery aims to find frequent patterns that represent the inherent regularities in the datasets. Applications of association discovery include association, correlation, and causality analysis; basket data analysis; and cross-marketing, and so forth. Classification, also called supervised learning, is the task of inferring a function from labeled training dataset. The function can then be used to classify new data instances. Decision tree, Bayesian networks, support vector machine, and neural networks are some of the commonly used models for classification. Clustering, also called non-supervised learning, group a collection of data objects into groups according a predefined distance function. Clustering can be employed as a stand-alone tool to get insights about data or as a preprocessing tool for other algorithms. Sequence analysis discovers patterns among sequences of ordered events or elements. Application of sequence patterns include customer shopping sequence, DNA sequences and gene structures, sequences of stock market changes, and so forth. Graph and network analysis aims to discover frequent subgraphs, trees, or substructures. It has been used for social networks analysis and web mining.

Cluster and Network Analysis

As the first phase of the research, network analysis is performed on prime contractors and their subcontractors of the sample database in a hope to discover the business networks among contractors. Figure 10 visualizes some findings from the network analysis. It shows the top three big contractors that have the largest number of subcontractors, and top three highly demanded subcontractors who are working for the largest number of different contractors.

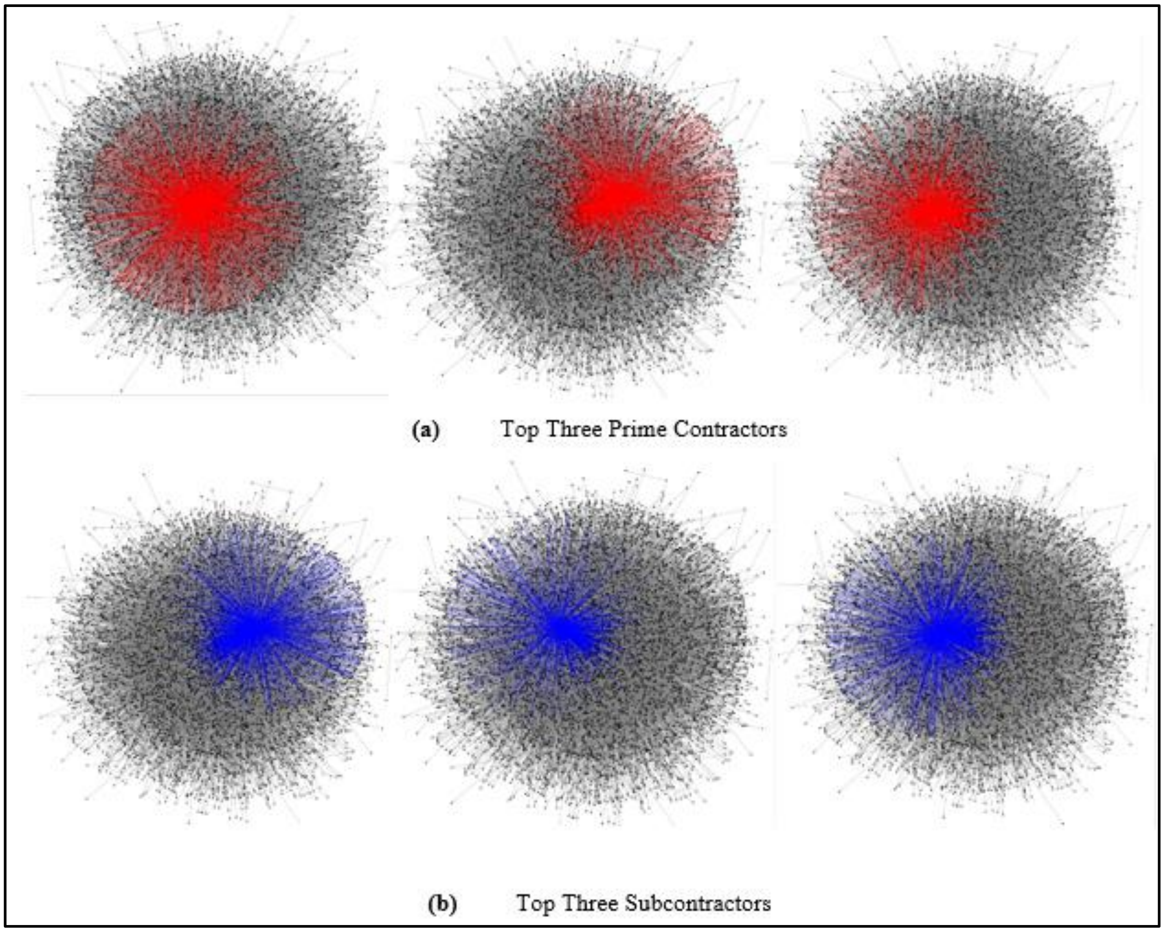


Figure 10. Cluster Analysis of Contractors

Figure 11 shows the clustering results of only contractors that worked with at least five subcontractors. Figure 11(a) shows overall clustering result, where each dot represents a primary contractor. The dots in orange are “big” primary contractors with many subcontractors. The dots in purple are relatively “small” primary contractors. Figure 11(b) shows zoomed-in clusters for two big prime contractors.

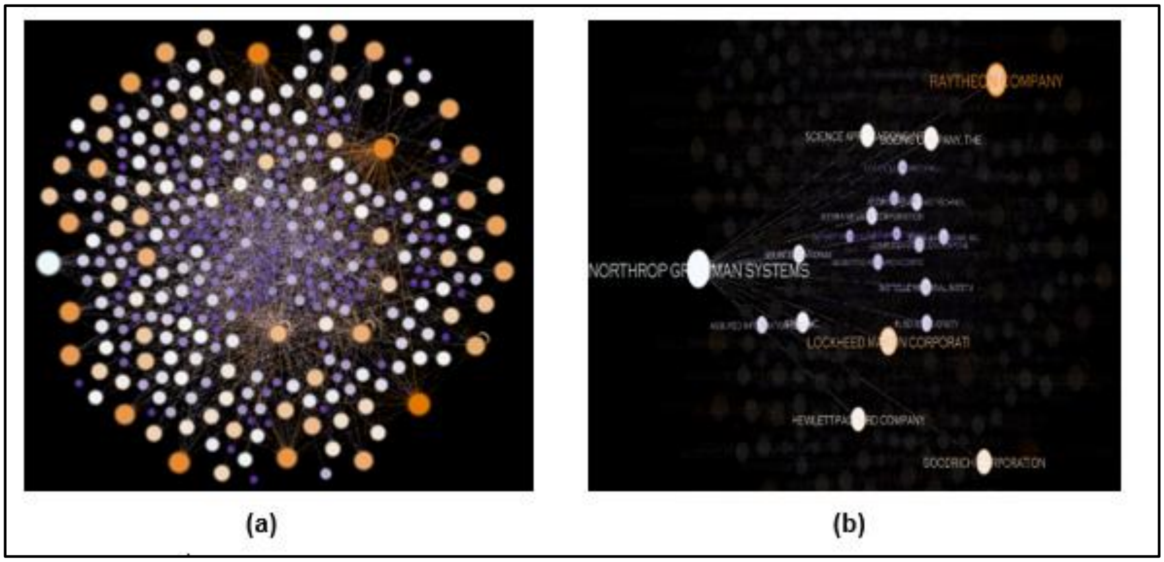


Figure 11. Clustering Results for Contractors Involved in More Than Five Projects

Figure 12 shows the analysis of the relationship between contractors and subcontractors by state. Each dot represents a state. The size of a dot is determined by the number of contracts awarded to a state. A directed edge represents the relationship between primary contractors and their subcontractors (pointed by an arrow). The thicker an edge is, the more contracts there are between the primary contractors and their subcontractors. The figure shows that some states, like California, get more contracts than others, and some states, like Illinois, tend to subcontract their projects to the other states.

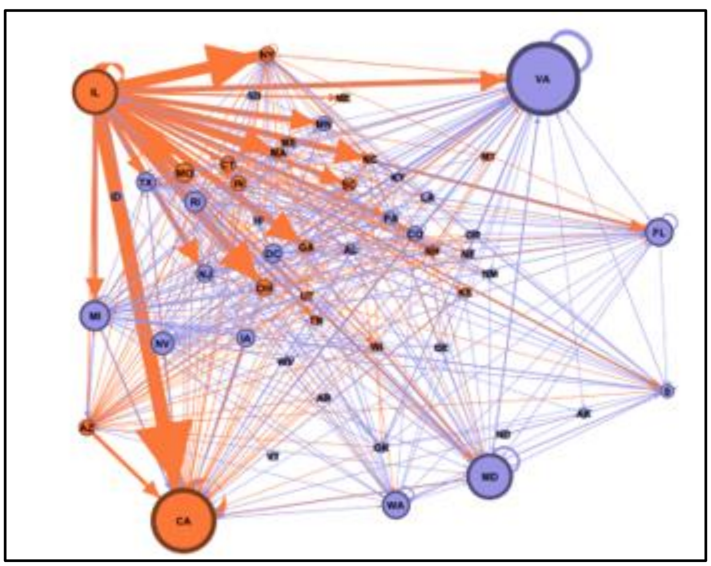


Figure 12. Clustering Results by State

Figure 13 shows the relationship of contractors of different business types. Each dot represents contractors of the same business type. A line between two dots indicates companies of two different business types are related by a contract. The figure reveals that companies tend to give subcontracts to the companies of the same business type. There are only two outliers that relate companies of different types.

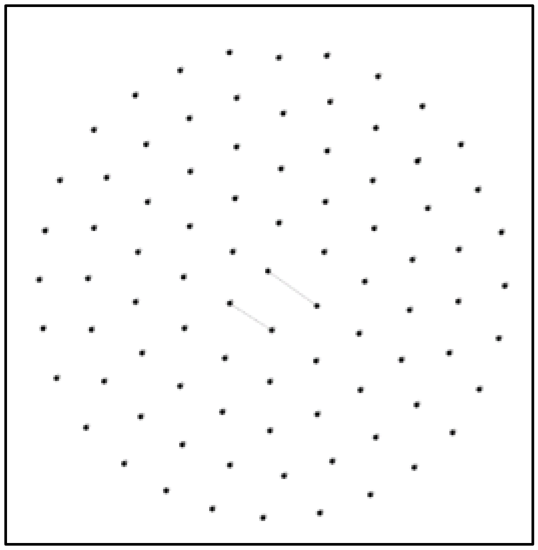


Figure 13. Relationship of Companies of Different Business Types

Pattern Discovery

A preliminary data analysis was performed and aimed to discovery patterns that may shed insights into possible areas for improvement in acquisition projects. For instance, small contractors are usually less robust and easier to fail compared to large contractors when facing natural or man-made disasters. Projects with subcontractors located in places that have a high risk of natural disasters such as earthquakes may have risks of a potential delay in delivery time. By taking into account of the risk factors in planning a project can help identify the room for improvement to ensure the successful and prompt delivery of the project.

The first round of analysis focused on finding the following patterns in the existing projects: (1) small-business subcontractors that are involved in different projects led by some key primary contractors and (2) projects that have multiple subcontractors located in a place that has a high risk of natural disasters such as earthquakes, hurricanes, flooding, or wild fires. The following section discusses two examples of our findings.

Finding 1

PTB is a small, single-location company with less than 200 employees. It was involved in six different projects led by some key primary contractors including Boeing, Lockheed Martin, and L-3 Communications. The average award amount is about $5,400. A close study of the company's website, shown as Figure 14, revealed it may provide some important services to its primary contractors. Since company websites and the acquisition database are all publicly accessible, they might be used by enemies for inferring sensitive information on a project or planning attacks to make the project fail.

Figure 14. Snapshot of PTB Webpages

Finding 2

We retrieved locations of 7.0-magnitude quake epicenters in United States from the U.S. Geological Survey website, www.usgs.gov, and identified 118 subcontractors located nearby an epicenter of 7.0-magnitude quake. In *SubContracts* table, 984 awards have at least one subcontractor located in the high-risk earthquake areas; 41 of them have at least two subcontractors located in the high-risk areas. Table 3 shows the top five contracts with the most number of subcontractors in high-risk earthquake areas.

Table 2. Top Five Contracts With the Most Number of Subcontractors in High Risk Earthquake Areas

Project ID	#Subcontractors
1	15
2	15
3	11
4	10
5	8

Finding 3

We did a preliminary exploration of whether a correlation exists between procurement data and employment information. The result shows that large reductions in federal contracts are correlated in a majority of cases (66% or 75%, depending on the metric used) to drops in employment in a given region and industry. This finding shows that it is possible to determine the location of an undisclosed contractor by examining public employment data at the times when large contracts are reduced or simply reach the end of their period. Such undisclosed contractors are typically employed by larger government contractors to achieve confidentiality, security, or a competitive advantage. Depending on the situation, acquisition experts may need additional planning to protect such hidden contractors if security is desired, or may rely on data science to identify these contractors and avoid them becoming a weak link in the acquisition process.

Conclusion and Future Work

The paper proposed a Data Enhancement and Analytics framework that is designed to use public data for improving quality and data analytic capabilities of the acquisition data. A proof of concept analysis was conducted to show the feasibility of using web information for quality assessment and improvement of the sample database. Still more needs to be done to implement the framework.

The future work will focus on the following two directions. First, we will research effective text analysis and trust evaluation techniques to identify credible and valuable information from the web. Second, we will research appropriate big data analytic techniques that can help enhance decision-making capabilities for acquisition management and planning.

Literature Review

This section summarizes some related work in the fields of federal acquisition data analysis, trust in web information, and Defense Acquisition Visibility Environment (DAVE).

Federal Acquisition Data Analysis

Apte, Rendon, and Dixon (2015) explored the use of big data analytic techniques to explore and analyze large dataset that are used to capture information about DoD services acquisitions. The paper described how big data analytics could potentially be used in acquisition research. As the proof of concept, the paper tested the application of big data analytic techniques by applying them to a dataset of Contractor Performance Assessment Report System (CPARS) ratings of 715 acquired services. It also created predictive models to explore the causes of failed services contracts. Since the dataset used in the research was rather small and far from the scope of big data, the techniques explored by the paper mainly focus on traditional data mining techniques without taking into account of big data properties.

Black, Henley, and Clute (2014) studied the quality of narratives in Contract Performance Assessment Reporting System (CPARS) and their value to the acquisition process. The research used statistical analysis to examine 715 Army service contractor performance reports in CPARS in order to understand three major questions: (1) To what degree are government contracting professionals submitting to CPARS contractor performance narratives in accordance with the guidelines provided in the CPARS user's manual? (2) What is the added value of the contractor performance narratives beyond the value of the objective scores for performance? (3) What is the statistical relationship between the sentiment contained in the narratives and the objective scores for contractor evaluations?

Our proposed research focuses on a much broader scope of acquisition projects. The research starts with cleaning and enhancing the acquisition data first. Once data is clean enough and has sufficient information, then advanced data analytic techniques will be applied in hopes of discovering interesting patterns that can be used to further acquisition management and planning research.

Trust in Web Information

Extensive research has been conducted on evaluating credibility and trust of online information, primarily textual information. The research by Corritore, Kracher, and Wiedenbeck (2003) identified three factors that impact trust in online environment: perception of credibility, ease of use, and risk. Cheskin (1999) identified six major features that encouraged trust in websites. The features are brand (the reputation of a company), seals of approval (icons from companies that certify a site as following security measures), ease of navigation, fulfillment (trust or distrust developed in using the web), presentation, and technology. In addition to the six features, the trust is expected to develop over time. The more interaction between the user and a website, the more information the user would gain to decide how much to trust it. Fogg et al. (2001) studied what makes a website credible. It defined credibility as believability and considered trustworthiness a major component of credibility. The paper also identified four factors as contributing to trustworthiness, namely linking (where the user was linked from and where the site links), policy statement, social recommendations, and business interest.

The commonly recommended approaches to online information evaluation include five criteria, including checking the accuracy, authority, objectivity, currency, and coverage or scope of the information and/or its source (Metzger, 2007). Accuracy refers to the degree to which a website is error free. The authority can be about a website or an author of information. The authority of a website is usually measured by its reputation and authority. The authority of an author is measured by the author's credentials and qualifications on the specific subject of information. Objectivity measures whether the information is fact or opinion. Currency refers to how up-to-date the information is, and coverage refers to the comprehensiveness or depth of the information provided.

Recent research shows that people tend to use verification strategies that require the least effort to perform. For instance, instead of using the recommended five criteria in evaluation, they opted to base decisions on factors like website design and navigability (Fogg et al., 2003). These findings are consistent with some recent credibility studies (Hilligoss & Rieh, 2008) and with theories from information processing and cognitive science (Sundar, 2008; Taraborelli, 2008). These theories stipulate that people have constraints on their ability to process information, and they tend to use cognitive resource that is just enough for a sufficiently optimal outcome for the evaluation task (Lang, 2000; Fogg et al., 2003).

Defense Acquisition Visibility Environment (DAVE)

Acquisition Visibility (AV) is the concept of providing the DoD with data and analysis support capabilities to inform the acquisition community. DAVE establishes a framework for improved and expanded support to the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]). As shown in Figure 15, DAVE employs a three-tiered architecture that contains the *DAVE portal, DAVE Platform*, and AV Data Framework.

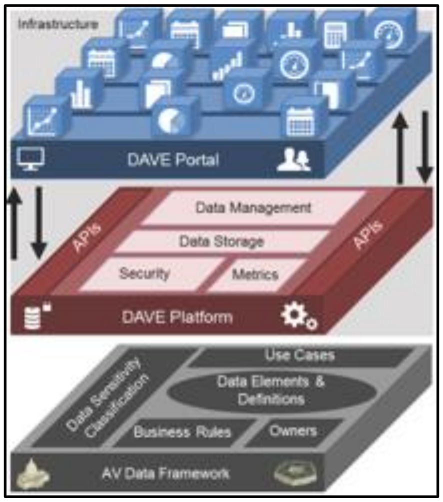


Figure 15. Defense Acquisition Visibility Environment (DAVE)

The *DAVE portal* is a synthesis of interactive infrastructure including data visualizations, calendars, and project management tools that are set to continue to grow in scope and capability as DAVE expands. These diverse tools with analysis capabilities will help users answer such questions as, “Are we solving a business problem by assessing the efficiency and effectiveness of the project?” and “What value does the project add to acquisitions in the Department of Defense?”

The DAVE platform includes the Application Programming Interfaces (APIs) for data management, data storage, metrics, and security. The DAVE platform determines the APIs for facilitating data access, and determines to which party the information can be shared. The APIs are the building blocks that allow for the integration of features or data, and the platform itself processes the data to get it to the state users require, as well as coordinating internal processes.

The AV Data Framework is the foundation on which the portal and platform are built and provides a number of essential elements including use cases, data elements and definitions, business rules, guidelines and markers regarding ownership of data, and data sensitivity classifications.

The proposed Data Cleansing and Enhancement System in this research can be used to support part of the functionalities of the DAVE platform as it prepares the data to the state that is suitable for user consumption or further analysis by the data leaning and enhance processes.

References

Apte, U., Rendon, R., & Dixon, M. (2016). Big data analysis of contractor performance information for service acquisition in DoD: A proof of concept. In *Proceedings of the 13th Annual Acquisition Research Symposium.* Monterey, CA: Naval Postgraduate School.

Augustine, N. R. (1997). *Augustine's laws*. AIAA.

Black, S., Henley, J., & Clute, M. (2014). *Determining the value of Contractor Performance Assessment Reporting System (CPARS) narratives for the acquisition process* (NPS-CM-14-022). Monterey, CA: Naval Postgraduate School.

Brown, B. (2010). *Introduction to defense acquisitions management* (10th ed.). Defense Acquisition University. Retrieved from http://www.dau.mil/publications/publicationsDocs/Intro%20to%20Def%20Acq%20Mgmt%2010%20ed.pdf

Cheskin, S. (1999). *Ecommerce trust: Building trust in digital environments*. Archetype/Sapient.

Cilli, M. Parnell, G. S., Cloutier, R., & Zigh, T. (2015). A systems engineering perspective on the revised defense acquisition system. *Systems Engineering, 18*(6), 584–603. doi:10.1002/sys.21329

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: Concepts, evolving themes, a model. *International Journal of Human-Computer Studies, 58*(6), 737–758.

Defense Acquisition University (DAU). (2016). *DAU Center for Defense Acquisition: Research agenda 2016–2017*. Retrieved from http://dau.dodlive.mil/files/2016/01/ARJ-76_ONLINE-FULL.pdf

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., … Treinen, M. (2001). What makes web sites credible?: A report on a large quantitative study. In CHI '01: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 61–68). New York, NY: ACM.

Gaither, C. C. (2014). Incorporating market based decision making processes in defense acquisitions. *International Journal of Defense Acquisition Management, 6*, 38–50.

Gallup et al. (2015, May). *Lexical link analysis (LLA) application: Improving web service to defense acquisition visibility environment*. Distributed Information Systems Experimentation.

Golbeck, J. (2008). Trust on the world wide web: A survey. *Foundations and Trends® in Web Science, 1*(2), 131–197.

Hagan, G. (1998). *Glossary: Defense acquisition acronyms and terms*. Fort Belvoir, VA: DoD, Defense Systems Management College, Acquisition Policy Department.

Krzysko, M. (2012, February). The need for acquisition visibility. *Journal of Software Technology*, 4–9.

Krzysko, M. (2016). *Acquisition decision making through information and data management.* Retrieved from http://www.digitalgovernment.com/media/Downloads/asset_upload_file917_5737.pdf

McKernan, M., Moore, N. Y., Connor, K., Chenoweth, M. E., Drezner, J. A., Dryden, J., … Szafran, A. (2016). *Issues with access to acquisition and information in the Department of Defense*. Santa Monica, CA: RAND.

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics, 59*(2013), 210–220.

Miller, A., & Ray, J. (2015, January 1). Moving from standard practices to best practices in defense acquisition. *Defense ARJ, 22*(1), 64–83.

Pennock, M. J. (2008). *Defense acquisition: A tragedy of the commons*.

Under Secretary of Defense. (2007, November). *Operation of the defense acquisition system* (DoDI 5000.01). Washington, DC: Author.

Under Secretary of Defense. (2015). *Operation of the defense acquisition system* (DoDI 5000.02). Washington, DC: Author.

Acknowledgment

We would like to thank Jamoris Miller and Leonardo Vieira for their help in visualization of some results.

Toward Cognitive Supremacy via Quantitative Augmentation

Col Clark Quinn, USAF—is the Chief of the Strategic Planning Integration Division, Headquarters U.S. Air Force. He leads a team of personnel tasked with integrating resources and capabilities into the Air Force's overarching $140 billion annual plan. Quinn graduated from the University of Florida with a degree in aerospace engineering and holds master's degrees in business administration and national security studies. Quinn has commanded a pilot training wing, served as a fighter wing Vice Commander, commanded at the squadron level, and held staff positions at multiple levels. He has deployed in support of operations in southwest Asia and northern Africa.

Maj D. M. Smalenberger, USAF—is the Quantitative Augmentation Branch Chief on the Air Staff and is the Creator and Chief Architect of Project QuANTUM and Laniakea. He holds a BS in Operations Research from the USAF Academy, an MS in Operations Research from the Florida Institute of Technology, and a PhD in Operations Research from the Air Force Institute of Technology. He holds certifications as an Advanced Lean Six Sigma Master Black Belt and as a Project Management Professional. His research includes PAC-learnable systems, financial warfare, mathematical statistics, information theory, quantum-hardened encryption, and multi-valued logic.

Abstract

The impetus of this work is to develop a complete and ubiquitous approach to augment the planning, programming, budgeting, and execution (PPBE) decision process quantitatively. The proposed solution provides a methodology that emphasizes the ontology of senior leadership questions, highlights those data sets that are relevant, and leverages a suite of methodologies that meet the ontological requirements. We propose a data democratization and utilization platform (Laniakea) as well as a four-stage analytical engine (QuANTUM) to achieve this objective. We posit that tethering these approaches to the PPBE problem set will enhance the derived solutions through a technologically-scalable, mathematically-flexible, and factually-rigid solution and that this will enhance leadership's overall awareness, and to understand and chart a way toward cognitive supremacy.

Introduction

To achieve victory, an adversary with fewer resources must neutralize their disadvantage. This economic pressure will likely require an evolution in their strategy which maneuvers around these disadvantages. As the U.S. military retains its dominance over certain domains while pursuing others, it causes evolutionary pressures to its adversaries to adapt accordingly. Given the U.S. military's current trajectory, these evolutionary pressures are showing beyond traditional battle spaces, such as with Unrestricted Warfare (UW)[1], hybrid warfare, *sub rosa*,[2] or *casus fortuitus*[3] operations. To prepare for these types of

---

[1] From the Chinese phrase 超限战 and directly translates to "warfare beyond bounds." Its core tenets include financial and economic warfare, lawfare, information warfare, cyber warfare, and terrorism (Xiangsui & Qiao, 1999).

[2] Translated from Latin and means "under the rose." It is used to describe clandestine types of operations (Libicki, n.d.).

challenges, we must modify our approach to plan, program, budget, and execute with these in mind. By augmenting our decision processes quantitatively and with appropriate analytical techniques and technologies, we can attempt to achieve praxeological optimization toward cognitive supremacy to help counter these emerging and exotic types of threats as well as continue to hedge against classical sorts as well.

The intent of this initiative is to integrate scientific and analytical precision in the DoD's Planning, Programming, Budgeting, and Execution (PPBE) process while appreciating the ontological differences that the questions leadership propose may exhibit, demonstrating the power of authoritative data sources (ADS)[4] under correct conditions, and fully leveraging the full suite of analytical methodologies that exist. The PPBE process spans each domain and functional area, involves multiple stakeholders, and is constrained to limit objective information toward logical and feasible solutions. There is a critical need to provide pertinent facts, identify interdependencies, and understand historical rationale to ensure the highest probability of success while making investment and divestiture decisions for DoD force structure. A cause of these challenges listed above is the lack of adequate access to pertinent data to substantiate their decision process quantitatively due to the federated style of data management that currently exists, consequently causing inefficiencies in the "Data-to-Decisions" paradigm. The DoD requires a change to address this problem toward a technologically-scalable, mathematically-flexible, and factually-rigid methodology to replace the current monopoly via subjective interpretation.

We define the following four concepts specifically as they provide the necessary terminology for this paper: Cognitive Supremacy (CS), Quantitative Augmentation (QuA), Ubiquitous Modeling (UM), and Question Class (QC).

- *Cognitive Supremacy* is the ability to hold complete control in the decision space over an adversary.
- *Quantitative Augmentation* is a process which provides technologically-scalable, mathematically-flexible, and factually-rigid methodologies to enhance the decision process scientifically.
- *Ubiquitous Modeling* is the ability to describe any system using appropriate ontological methodologies.
- A *Question Class* is a set of questions which are ontologically[5] equivalent.

---

[3] Translated from Latin and means "an inevitable accident." It is used to describe non-reputable operations which may be carried out on an adversary. An example would be financial market manipulation and then attributing those effects to normal market fluctuations.
[4] "[A] recognized or official data production source with a designated mission statement or source/product to publish reliable and accurate data for subsequent use by customers. An authoritative data source may be the functional combination of multiple, separate data sources" (DoD, 2007).
[5] An ontology is the compartmentalization of problem solving techniques based on complexity, approach, computation, and variables that are required (O'Connor & Wong, 2012).

Question Classes

A typical methodology to problem-solving is to dissect it into its causes. Once we identify these causes, then we are in a position to influence the aggregate behavior by affecting one or many of these causes. Several process improvement initiatives over the years, such as Total Quality Management (TQM) and Air Force Smart Operations for the 21st Century (AFSO21), have provided helpful tools and mechanisms to do so. However, using their standard techniques implies that the problems we wish to understand in more detail and dissect into their causes lend themselves to a deconstructable approach. As we will illustrate, this assumption does not hold true for many PPBE-level types of problems, and so this strategy has conditioned the community to persist in an ill-fated strategy.

We propose an alternative to the typical process improvement problem solving by categorizing the types of questions based on their ontology. Once we establish a question's ontology, then we may apply those methodologies which meet the ontological assumptions. Figure 1 illustrates the alternative view of problem-solving based on the concept of an ontology.

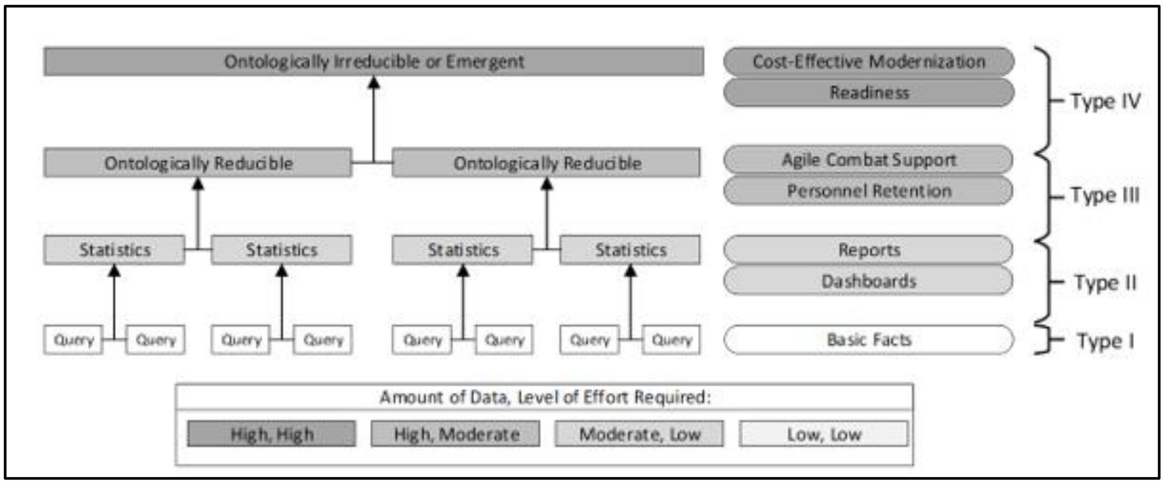


Figure 1. Illustration of the Hierarchical Structure for the Four Proposed QC With Pertinent USAF Examples to Provide the Reader Context: Personnel Recruitment and Retention (PR&R),[6] Agile Combat Support (ACS),[7] Readiness,[8] and Cost-Effective Modernization (CEM)[9]

---

[6] Represents the process that selects the personnel for the most appropriate Air Force Specialty Codes (AFSCs) based on their aptitude and preferences (Recruitment) and to optimize their dwelling within the enterprise (Retention).
[7] Represents the support tail of the USAF enterprise and comprises approximately 40% of Total Obligation Authority (TOA).
[8] The measure associated with efficiently allocating personnel and equipment to combat packets to meet operational objectives.
[9] A USAF priority to revamp aging aircraft, satellite, and nuclear fleets in a manner to maximize lethality and utility for a given cost structure.

Type I: Query

A query is a precise request for information retrieval from a database (DB) or information system (IS). The emphasis is on the content of the DB or IS and not on any ontology that may exist within the DB or IS. Queries require only the ADS on which we wish to run the query and require the least amount of technical expertise to do so. An example would be to list all females with the first name "Trista" from MilPDS.[10]

Type II: Statistics

A statistic condenses relationships which are stored in a DB or IS into parameters of a known statistical technique. There is an information transfer that occurs here from content to context. An example would be to test whether or not a normal distribution can accurately represent USAF entry test scores for enlisted members.

Type III: Ontologically Reducible Behavior

Deduced behavior based on the content of one or many DB or IS under consideration. We can dissect these relationships via one or several multivariate techniques and apply the typical cause analysis processes. An example would be how the fuel prices may constrain the number of sorties flown and may cause currency problems for fighter pilots.

Type IV: Ontologically Irreducible or Emergent Behavior

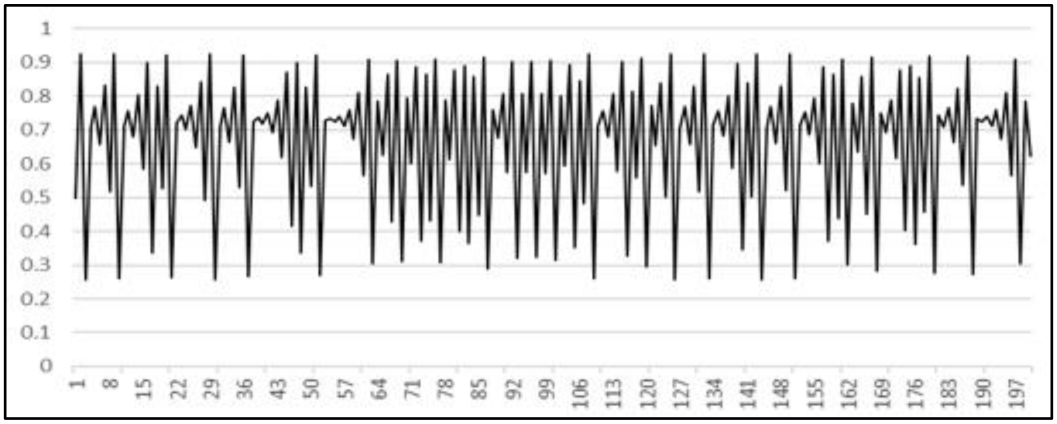
The resultant phenomenon where defined systems with clearly defined relationships may produce unpredictable and even unanticipated behaviors that are also more complex than any subsystem could produce in isolation. Therefore, we may not understand the etiological path as a set of few distinct and manageable factors. Furthermore, nearly all emergent-types of systems are ontologically irreducible to the lower levels (O'Connor & Wong, 2015). These systems include, but are not limited to, complex adaptive systems (CAS)[11] which many DoD systems and processes seem to be. This class will likely apply to all senior level questions that include "effectiveness," "agility," and "lethality."

We propose a very simple example to illustrate how simple systems can become intractable to deconstruct. The output of a system we wish to model follows the logistic map in Equation 1, where x_n is a number between null and unity and λ represents a value between null and 4. This is a relatively simple system, and the future value is dependent on the prior value. However, we obtain very different behavior from this system depending on the value of λ that we choose. For some values of λ, this system will reach a steady-state; for other values, it will enter into oscillations, and for some values of λ, the system will enter a state with no clear pattern.

$$x_{n+1} = \lambda \, x_n(1 - x_n) \qquad (1)$$

[10] "MilPDS is the single integrated 'Total Force' AF Human Resource system and authoritative data source for Total Force military records supporting all Active Duty, Guard, Reserve and retired AF members. MilPDS is the system of record that manages every aspect of an Airman's career, including accessions, assignments, career management, separation and retirement. MilPDS was the selected platform to realize the AF/A1 AF Integrated Personnel and Pay capability" (U.S. Air Force, 2017).
[11] A system in which a perfect information about each system subcomponent does not imply perfect information about the dynamics of all subcomponents in aggregate (Miller & Page, 2007).

As the analysts, we do not assume we know this relationship *a priori*, but wish to deduce this behavior from an ADS which captures observations from this system. Figure 2 illustrates an observation subset.



Note. We notice the erratic and unpredictable behavior that the system presents and which is difficult to predict or to deduce the underlying relationship analytically.

Figure 2. Illustration of the First 200 Observations From the Logistic Map With $x_0 = 0.5$ and $\lambda = 3.7$

Even with relatively simple systems such as in Equation 1, it may be difficult or impossible to deduce that these observations originate from a logistic map. A further complication is that the more complex the system is, the more observations you will likely require to validate a deduced relationship. However, there are likely insufficient observations necessary in a combination of ADS to sufficiently represent USAF questions of Type IV, and so we require a different approach which we will discuss later in this paper.

We posit that understanding and appreciating the class in which a question resides is a necessary but not sufficient step to articulating the question quantitatively. It provides a glimpse into the level of effort and the types of methodologies that you will likely require to answer it in a scientifically-rigorous way, and this is a focus later in this article. By understanding which class in which a question resides will help the analyst understand which types of approaches they will have to use to formulate the question in the Specific, Measurable, Attainable, Results-based, and Time-bound (SMART) framework (O'Neil & Conzemius, 2006). When taking the QC into account, then several Lean Six Sigma[12] techniques such as Ishikawa diagrams, 5 Whys, and so forth can assist an analyst in dissecting a question into pertinent and quantifiable sub-questions.

The query logic for Type I questions are a codified representation of the original question, and so as long as the data is available, this class is easy to answer. The answers we seek are only in the content of the data. For Type II questions, we seek the relationships, or the context, of specific data elements. Since the data content is available, we may run descriptive analytics against it and so arrive at specific statistics. These two QCs are nearly

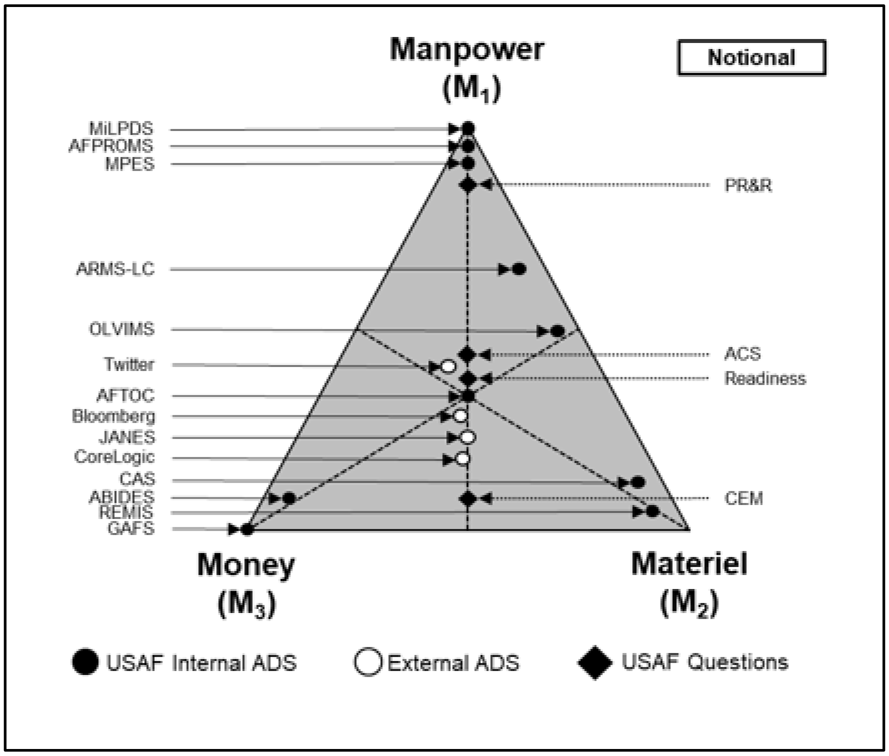
[12] A process improvement methodology which leverages collaborative team efforts to improve a system's or process's performance by removing waste and variation (Mills, Carnell, & Wheat, 2001).

trivial due to an analyst's ability to obtain an answer directly from the data. However, with Type III questions, an analyst is now interested in the interrelationships between disparate ADS. There may be some hidden behavior that an analyst may not understand or observe, and this may lead to missing ADS that we should have included or the introduction of an ADS into the solution that provides no additional utility. Type III questions will likely require an iterative process with question formulation and ADS selection. Lastly, there is no guarantee that any combination of ADS may sufficiently address Type III and IV questions. It is very easy to construct a scenario in which an infinite amount of data may not provide an adequate understanding of the underlying system behavior. For Type IV questions, we may be able to understand broad behavior, but it remains an open question if we may achieve a more granular level of understanding.

M^3-Space

The second component is to determine the fields in which ADS pertain to the question of interest. How to standardize and automate this approach remains an open question since there may be certain underlying behaviors present in the data that are not readily apparent and may involve fields that an analyst may overlook.

We propose a first step in developing a solution strategy. By analyzing many senior-level types of questions, a common theme becomes apparent. Many questions involve a combination of three core tenets: *Manpower, Materiel, and Money*, or M^3. Therefore, it seems advantageous to map each ADS (and then later map each field within each ADS) to the region between these core tenets to gain insight to which ADS may provide utility to certain types of questions once they, too, are mapped to the M^3-space. Figure 3 provides a notional example of several USAF core and external ADS and certain questions mentioned previously which are currently of interest to senior USAF leadership.



Note. The dashed lines represent the altitudes and represent locations where an ADS or question has equal representation for those core tenants on which it lies. Notionally, OLVIMS lies between M_1 and M_2 core tenants for example.

Figure 3. Illustration of the Proposed M^3-Space With Several USAF-Specific ADSs Such as GAFS,[13] MiLPDS, and REMIS;[14] Several External ADSs to the DoD Such as Bloomberg,[15] CoreLogic,[16] and Janes;[17] as Well as Several USAF Enterprise-Wide Questions

With this paradigm, we may chart the ADS locations between these three core tenets. This location will give us an idea of how likely an ADS may support a particular question of interest. Furthermore, if we map ADS which we have yet to obtain, then such an

---

[13] "General Accounting & Finance System (GAFS) is owned and functionally managed by DFAS-Columbus. It is used to process more than 3.2M accounting transactions totaling $3.4B monthly. GAFS-DTS processes more than$4.4M traveler payments annually; more than $4.5B in DoD travel payments" (U.S. Air Force, 2017).

[14] Reliability and Maintainability Information System (REMIS) is the primary USAF data system for collecting, validating, editing, processing, integrating, standardizing, and reporting equipment maintenance data, including reliability and maintainability data, on a global, world-wide basis (U.S. Air Force, 2017).

[15] One of the leading financial data vendors and brokers worldwide.

[16] CoreLogic is a data broker which specializes in U.S. economic, housing, and personnel data.

[17] Jane's Information Group was a British publishing company which specialized in military, aerospace, and transportation topics. It was acquired in 2007 by IHS Inc. which continues to sell their data products.

approach may assist in determining which ADS we should prioritize next. We will likely wish to have a balanced mix of ADS that may contribute to all three core tenets at any given time, and so this approach may assist in that objective.

Although the mapping of ADS is conceptually helpful, it only provides limited utility. Preferably, we require a granularity that is at the attribute level. The granularity will enable us to map each attribute to each core tenet or a combination of them. Unlike with the ADS mapping, an attribute mapping provides an added benefit that any attribute which maps to any altitude within the M^3 is a key ID across those core tenets. Furthermore, an attribute that maps to the centroid is a key ID for all three core tenets and would be of high value to tie multiple ADSs together.

Data Landscape

Data Policy

Standard practice for data analytics is to amass the necessary ADS locally or in the locally-owned and regulated environment. Each party signs some agreement to acquire one or more ADS to gain access or obtain a copy of one or more ADS. For DoD personnel and DoD organizations, this is usually via a Memorandum of Agreement (MoA) or via a System Authorization Access Request (SAAR) or DD2875. The MoA is a binding agreement between the parties involved and establishes which ADS the parties will access, how the customer will access, store, and use this data; what data protection measures they will enforce; and what their procedures will be to eliminate the data once the agreement has expired. Parties typically use an MoA when one or more customers request more than one ADS. Customers typically submit the variation DD2875 via their chain of command and then on to the specific ADS owner to gain approval to access or obtain a copy of a particular ADS.

Data Situation

A general observation is that ADS are heterogeneous over multiple operating systems (primarily Windows, Linux, and Solaris) and utilize several DB environments including but not limited to Actian, DB2, Ingres,[18] MS SQL Server, and Oracle as well as a variety of IS dating as early as the 1970s. In some cases, non-U.S. government third parties maintain these ADS outside of any DoD installation and are geographically disparate.

Observations and Challenges

The lack of a consistent DoD data acquisition process requires a considerable level of time investment to ascertain and comply with the nuances of any particular ADS. Typically, there is a low probability of establishing a relationship successfully with the owning organization if there is no potential for a symbiotic relationship or no higher authority mandates cooperation. A compounding factor to this challenge is that the DoD has little or no situational awareness about the number or location of DB or IS it possesses. Therefore, it can be difficult to request an ADS if it is arduous to locate the proper authority to request access. Furthermore, system administrators or database administrators (DBA) often are extremely specialized and therefore are no longer experts on an entire IS or DB. Therefore,

[18] The INteractive Graphics REtrieval System is an open-source SQL relational database which was developed at the University of California, Berkeley, in the 1970s and is still in use in the DoD today.

we may require several teams of individuals for a single ADS. A particularly concerning observation is the compliance with outdated data management policies. Some data owners retain data for a couple of years and then purge or overwrite them with new data. This practice is particularly troublesome since we are unable to provide meaningful analytics within a problem set that requires this data once this data no longer exists. The rationale for these policies was a cost savings measure when electronic storage costs were considerably higher than only a few years ago. However, it is not sufficient to gain access or to obtain a copy of an ADS to conduct a meaningful analysis. One potentially overlooked, yet critical, item is a legend or explanation of the attributes and values contained in any ADS—or data dictionary. There are many instances where data dictionaries are either outdated, incomplete, or in some cases nonexistent. Lastly, there is a systemic lack of continuity to explain the changes a DB, IS, or ADS have undergone since their inception. The lack of continuity provides difficulties in interpreting changes within these systems or data sources and which may have significant implications on the analytical quality which an analyst can produce.

Proposed Solution

We offer a potential strategy to address these challenges by bifurcating, yet tethering, the problem into a data approach and analytics approach, as shown in Figure 4. To provide high-quality analytics that are useful to the larger community, we require a standardized process and framework to access and leverage many ADS. Since ADS within the DoD are heterogeneous, we require a considerable level of effort to Extract, Transform, and Load (ETL) them into a common environment. To leverage a large number of feasible tools, a cloud environment in the private sector is likely an improved solution over purchasing a suite of licenses for different ETL tools. A cloud solution would enable the ETL process to utilize tools that may also not possess the requisite approval for installation on machines connected to a U.S. government network. The Enterprise Information Model (EIM) provides the aggregation of knowledge for all data dictionaries, their commonalities, and combinations of cross-ADS attributes which analysts have used in the past. The environment that enables the ETL process, storing of ADS, and provides an overarching cloud solution for the AF and possibly DoD, is Laniakea.[19]

We also suggest a common analytical base for the larger DoD community. The original intent of this analytical base was to have the ability to access and utilize ADS within Laniakea easily and so provide a more useful analytical toolset to the communities which wish to conduct analytics on a wide span of ADS. As previously argued, any analytical approach must provide sufficient flexibility to address each of the four types of QC. The platform that provides accessibility to Laniakea while providing a suite of analytical tools to the DoD community is the Quantitative Augmentation via Neuro-evolutionary Technologies toward Ubiquitous Modeling or QuANTUM.

---

[19] The word *Laniakea* is Hawaiian and means ``immeasurable heaven.'' It is the galaxy supercluster which contains the Milky Way and an estimated 100,000 other galaxies and spans approximately 520 million light years. In this context, Laniakea refers to the USAF's galactic data supercluster which seems appropriate due to the vastness of the USAF's and DoD's ADS.

The solution connects to the DoD Information Network (DoDIN)[20] via an Internet Access Point (IAP), internet gateways, and Virtual Private Clouds (VPC). Each component, the web application firewalls, the bastion hosts, Laniakea, and QuANTUM, are physically and virtually distinct entities which connect with VPC peering which would enable added levels of security and inhibit transitivity of access to each isolated component in the chance an adversary exploited a vulnerability. Also, both Laniakea and QuANTUM have persistent monitoring to protect against threats or anomalous activity actively.

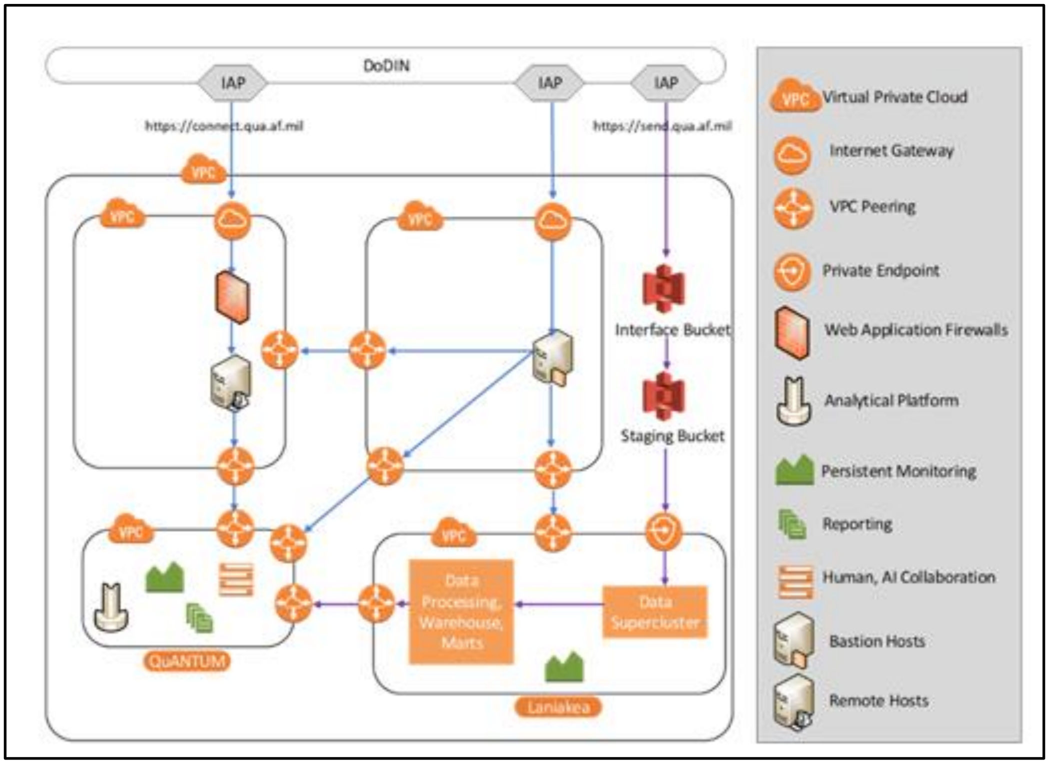


Figure 4. Illustration of the QuA Ecosystem as Conceived by the Authors

[20] "The set of information capabilities, and associated processes for collecting, processing, storing, disseminating, and managing information on-demand to warfighters, policy makers, and support personnel, whether interconnected or stand-alone, including owned and leased communications and computing systems and services, software (including applications), data, security services, other associated services, and national security systems" (Committee on National Security Systems, 2003; Joint Chiefs of Staff, 2010).

The Analytical Engine—QuANTUM

We require an analytical approach which would enable us to leverage the scientific process and, concerning the questions above, classes and ADS since there are multiple types of QCs and these classes require different strategies to address them adequately. Providing a common platform to access ADS and analytics is preferable since it would provide the opportunity to develop a common terminology throughout many communities. We posit that one of the difficulties in explaining a viewpoint is that the terminology is defined in disparate colloquialisms and human language and not on mathematics.

We illustrate the example of Fully-Burdened Cost (FBC). The first challenge to using a definition is to locate one that is accepted by the larger DoD community. The definition for FBC seems to vary based on Service, organization, and experience. Typically, the Government Accountability Office (GAO) publishes lists periodically of general terms for the PPBE and other processes (GAO, 2005). There are three types of FBC listed in the Financial Management Regulation (FMR), namely cost for fuel, the unit cost of contractor-acquired property, and composite rate of pay, allowances, taxes, and accruals (Under Secretary of Defense, Comptroller, n.d.). Each of these definitions is not quantitatively well-defined, and so the DoD guides estimate these types of fully-burdened costs (DoD, 2013). Since the DoD only provides not completely quantifiable guidance on how to interpret these definitions, it leaves the opportunity for variability in the calculation of these costs. This variability may provide confusion and inconsistencies in planning, programming, budgeting, and execution, and these inconsistencies may compound and result in suboptimal choices. An alternative approach would be to define these terms using a query language, such as Structured Query Language (SQL) that can access many records within ADS and so calculate a value that we can tether to fact. This proposed approach would incentivize many DoD communities which use similar terms such as FBC to develop common definitions so they may communicate with one another more intelligibly. This unprecedented clarity would be of additional benefit to higher levels of leadership, such as the Office of the Secretary of Defense (OSD) since it would no longer need to interpret many inconsistent definitions of the same terminology.

The next logical step after creating a standard data repository, such as Laniakea, and providing quantitative definitions for common terminology, would be to provide an analytical environment that may augment the decision process quantitatively and for varying QCs. The typical approach to conducting analysis is to amass those necessary ADS which are most pertinent to answering a question locally and then to devise an approach with those analytical tools which an analyst has at their disposal. Often the analyst is limited to the types of analytics they can perform since they have very few tools, and so they must create a toolset. This limitation is typical since many organizations cannot afford analytical suites of tools that would allow them to answer their questions more swiftly and rigorously. Also, the authorities for DoD IS have not yet cleared many formidable tools for installation on computers connected to government networks. This primary approach to conduct analytics is no longer feasible given financial, security, and other constraints. Therefore, we posit that an alternative solution exists that provides a higher level of accommodation for these concerns. We observe that the primary approach involves aggregating ADS to where the analytics reside. The alternative approach, or dual, would push the analytics to where the ADS reside. This alternative would enable an analyst to generate a virtual machine (VM)[21] in the cloud solution we mentioned and to virtually install an instance of many types of analytical products that would not be allowed otherwise. Although this provides an improved solution, the difficulty remains with having a sufficient number of licenses or proprietary products for the communities.

An improvement would be an analytical platform which functions as a Platform-as-a-Service (PaaS) and utilizes the ADS stored in Laniakea as a Data-as-a-Service (DaaS) approach. Therefore, this platform would function as a utility such as water, electricity, natural gas, or internet. This framework leverages economies of scale in that the cost per user declines as the number of users rises. Therefore, the incentive is for the developmental teams to maintain a viable and demonstrably superior platform for the communities to use. Therefore, this paper proposes a Government-Off-The-Shelf (GOTS)-owned platform that leverages a suite of ADS, provides common terminology via a query language, and provides a suite of analytical methodologies from which to choose. The platform takes advantage of computing at multiple classification levels and the ability to scale with its compute infrastructure.

QuANTUM is an analytical engine that is comprised of four stages and is in parallel development with Laniakea as its data source. Each stage addresses and enhances the efforts of the prior stages and addresses their potential limitations. Stage I, Argos,[22] is the analytical platform that provides a suite of tools ranging from simple queries from ADS to multivariate and advanced approaches such as topological data analytics (TDA), machine learning (ML), forecasting, and optimization. However, Argos can only address Types I through III. The reason is that we require historical observations to address these questions. With historical data, Argos is incapable of adequately addressing Type IV questions due to those reasons we discussed.

To address the Type IV QC, we require an ontologically distinct methodology from that of Type I through III, and this difference is the focus of Stage II, Krishna.[23] Since this is an approach to address the challenges as discussed briefly with the logistic map, we require a methodology that would provide some level of insight that lies beyond both internal and external ADS but can still leverage these ADS when appropriate. One possible methodology would be to utilize modeling and simulation (M&S) and so attempt to gain insights into the emergent behavior that may exist as simpler systems interact with one another. We may wish to compute many simulations with different initial conditions to get a general sense of how the system under consideration may behave. If there are significant deviations from the baseline system, then this might provide evidence to suggest that the system may require additional modifications to align both real and simulated systems more closely. The M&S environment that the Type IV QC concerning the DoD would require is a physics-based, continuous-time, imperfect information multi-agent framework at the campaign or multi-campaign level. A physics-based approach is necessary to provide realism to the

[21] An emulator of an IS. It is a software substitute for a physical machine.

[22] Also known as Argus, Panoptes is a many-eyed, all-seeing giant in Greek mythology

[23] The eighth avatar of Vishnu. He is the central figure in the Mahabharata, Bhagavata Puana, and the Bhagavad Gita. The impetus of the name is due to verse 32 in the Bhagavad Gita which also drew Robert Oppenheimer's attention: "Now I am become death, the destroyer of worlds" (Mascaro, 1962).

environment. The model includes the application of physical laws and constraints within the environment such as gravity for munitions or aircraft and Line of Sight (LoS) for communication links. The feature of continuous time is important since complex systems may behave very differently when discretized. Also, providing the correct approach to the flow of time will provide an improved model concerning other types of flows (e.g., information) within the framework. A common feature many wargames fail to appreciate fully is the limitation of timely and perfect information about the situation on the battlefield, and is analogous to von Clausewitz's (1832) "fog of war." By limiting the fidelity and flow of information, we introduce risk into the decision space and thereby provide the framework to learn how to operate in these degraded conditions. The intent is to introduce these features in an environment that remains scale-invariant and therefore provides utility to leadership at any level. Lastly, and perhaps most importantly, is the introduction of allied and adversarial adaptive systems. We may begin to develop this environment by utilizing Reinforcement Learning (RL)[24] and then exploring the utility of Generative Adversarial Networks (GAN).[25] These types of approaches provide evolving solution strategies given a complex environment and with minimal rules provided *a priori*. Also, these types of approaches may find novel ways to interact with a complex environment that someone may not anticipate. Those types of strategies which exploit indirect and cascading effects are those which the authors feel may be of importance in future conflicts and require further research to explore adequately.[26] If we provide this system flexibility with determining which assets to locate where given a certain set of campaigns, we may also gain insights into how our asset portfolio should be allocated to achieve the highest probability of victory given those campaigns (or any other objective function we choose). Finally, we wish that this framework exploit advances in massively parallel computational platforms and that these methodologies may learn from many environments simultaneously, thereby providing a richer awareness of the solution space to leadership in less time.

Krishna is no panacea. One of its major limitations is that it provides feasible strategies given the current weapon systems and infrastructure that exist. Allied and adversarial weapons portfolios are dynamic, and systems with enhanced, even new, capabilities arrive with some regularity. Stage III, or Oracle,[27] provides the framework necessary to address Krishna's limitations. It introduces degrees of freedom (DoF) to those weapon system specifications (e.g., maximum range, maximum altitude, maximum munitions capacity, etc. for aircraft, and so forth) that may change with new weapon systems or modifications of older systems. However, any analysis which provides modifications to

---

[24] One of the three main types of learning: supervised, unsupervised, and reinforcement learning. Over successive iterations, an RL model will learn based on the environment in which it is introduced. If the aggregate behavior of the RL system is desirable, then it is rewarded and so reinforced otherwise it is not (Russell & Norvig, 2015).
[25] Two unsupervised artificial neural networks which contest in a zero-sum test space (Goodfellow et al., 2014).
[26] A possible scenario may involve targeting several seemingly ancillary and unguarded power substations simultaneously. Their cumulative effect may introduce a rolling blackout which may degrade an adversarial military installation communications network and so degrade the adversary's command and control.
[27] Named after the Oracle of Delphi and not to be confused with the company that bears the same name.

specifications alone is inherently flawed, as we would also need to consider the trade-offs which we incur with these modifications and how these new or modified systems should interact with the rest of the portfolio in a given a set of campaigns. One of the main trade-offs is that we incur Research, Development, Test & Evaluation (RDT&E) costs for any given modification. These need to be captured into the specification changes as well.

We may observe that in this framework there may arise common strategies and themes within the behavior of the systems that may confirm or repudiate a commonly-held belief. These observations would provide a level of substantiation not yet reached via alternative methods. The intent would be to augment policies and strategies at the Service, Joint, Department, and coalition levels where and when appropriate. Furthermore, the advantage of this approach is that it does not rely on the analysis of the levels above to obtain a solution. This approach might be advantageous since each of these levels attempts to achieve optimality within their sphere of influence, and this effect might result in many local optima rather than an optimum solution for all actors at the same time.[28] Referring to Type IV types of questions, models of subsystems may also not provide an accurate description of aggregate behavior, and therefore an overarching model may also be more appropriate here.

Unlike Argos and Krishna, Oracle is a composite stage, and therefore Oracle draws on specific capabilities that the prior stages deliver. We may leverage certain methodologies incorporated into Argos to develop a set of relationships between the cost of a certain specification set and attempt to infer an overall relationship between cost and specifications. We define this as the COst-to-Specifications Manifold (COSM). Although we may apply multivariate techniques such as Response Surface Methodologies (RSM)[29] to capture this relationship accurately, there is a growing trend to leverage alternative techniques such as Artificial Neural Networks (ANN)[30] over RSM (Himmel & May, 1991; Carpenter & Barthelemy, 1993; Hussain, Xuanqiang, & Johnson, 1991). This realization enables us to recycle techniques present in Argos and utilize them for Oracle. Figure 5 provides an overview of the process discussed in this section.

---

[28] We can never guarantee reaching a global optimum when utilizing heuristics, and so the argument is that we need to traverse the solution space given all actor constraints simultaneously.

[29] The approach to utilize mechanistic models, empirical, and response surface models to identify and fit factors and experimental data to an appropriate mathematical representation (Myers, Montgomerey, & Anderson-Cook, 2009).

[30] A synthetic analog to biological learning which is also robust against errors and incomplete input (Mitchell, 1997).

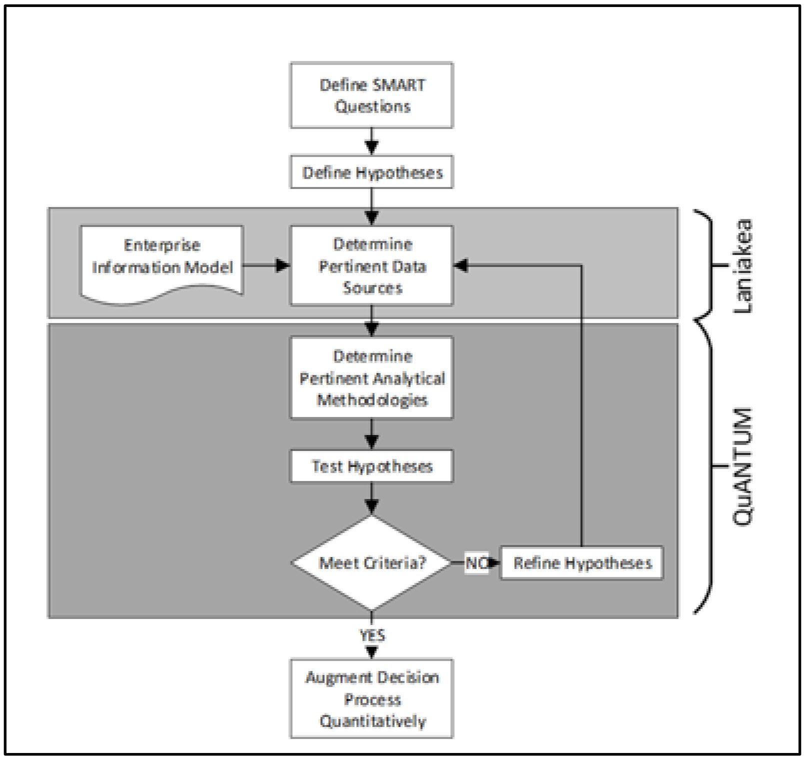


Figure 5. Illustration of the Scientific Process Utilizing the QC to Help Frame SMART Questions and to Shape Hypotheses While Leveraging the Galactic Data Supercluster, Laniakea, and the Analytical Engine, QuANTUM

Once we have established an acceptable COSM model, we can then provide some insight into the trade space between these two factors. The COSM functional approximation now provides the necessary cost penalty to any proposed improvement by an M&S solution, such as Krishna. As the agents within Krishna hypothesize what improvements to make to existing systems, they will incur the necessary cost penalties of making certain specification modifications. Therefore, their resultant strategies will account for these nuances as they try to achieve their overall objectives.

No change can happen instantaneously to a weapons plan, strategy, or weapons portfolio. These changes will require appropriate planning for these modifications, reprogramming funds, budgeting resources within future years, and executing these resources in a manner that remains feasible within political, policy, fiscal, and other types of constraints. We therefore require feasible scheduling trajectories for these improvements given their constraint set. This problem is a classic resource-constrained weighted scheduling optimization problem (RCWSOP) for which Genetic Algorithms (GA),[31] among

[31] A synthetic analog to Darwinian evolution at the genetic level. Information is encoded within artificial chromosomes and scored based on a supplied objective. As in biology, these chromosomes undergo crossover and mutations to provide variety and so traverse the solution space more completely. The chromosomes which provide the best fit are likely to carry that information on into successive generations (Mitchell, 1998).

other heuristics, have been used successfully to provide acceptable solutions. These approaches are highly parallelizable, and so if we supply additional computational resources, we may traverse a larger portion of the solution space and so attempt to find and improve on those feasible solutions we have obtained. Since the RCSOP is Type III ontologically, we may define Kronos as a derivative stage based on methodologies within Argos. Figure 6 provides a detailed overview of the QuANTUM analytical engine.

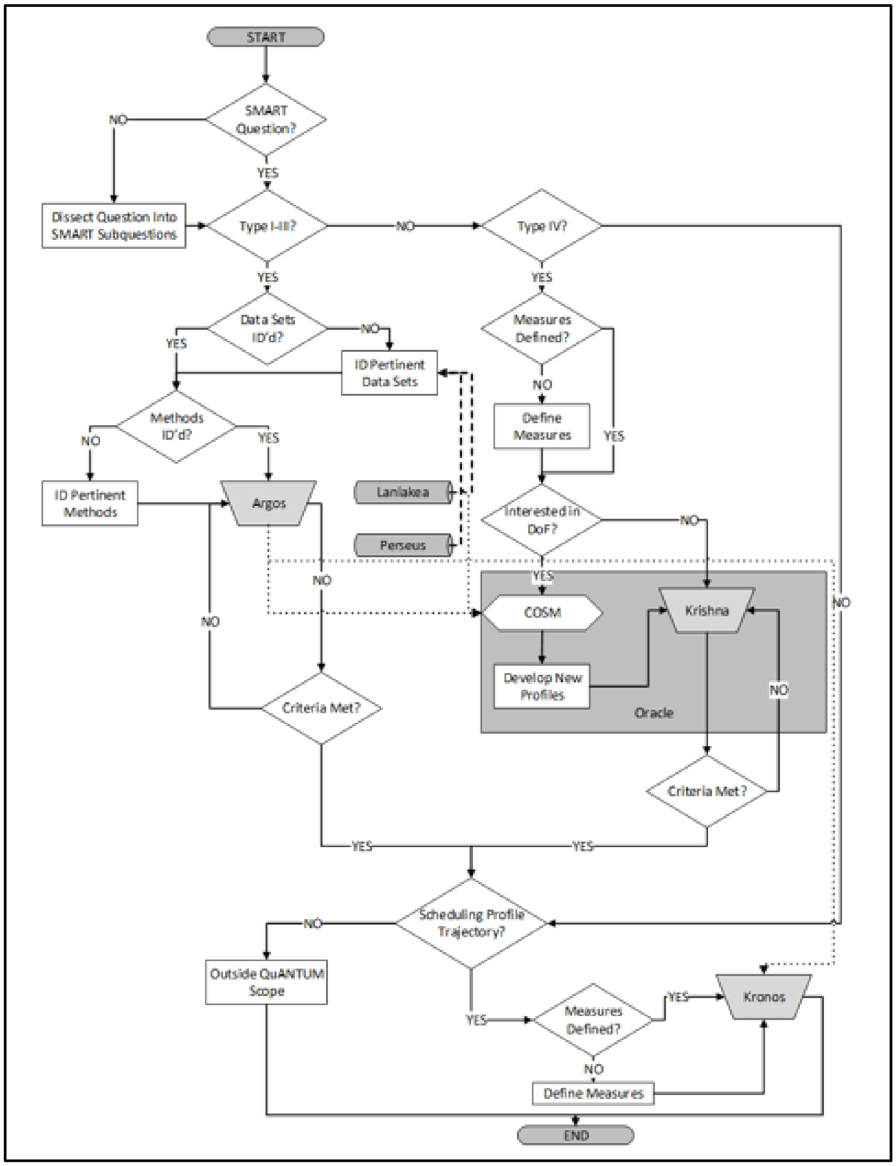


Figure 6. Proposed QuANTUM Analytical Engine Process

The Galactic Synthetic Data Supercluster—Perseus[32]

Currently, most DoD Data Democratization (DoD[3]) and analysis efforts face limitations due to the nature of data (or observation) classification. The impetus of these classification guidelines is to protect data specific to an individual, such as Personally Identifiable Information (PII) and Personal Health Information (PHI), as well as data about national security which exists at levels such as secret and top secret. Proper automated and consistent classification methods of aggregated ADS remain an open question and an active field of research. The lower bound of a classification process is seemingly trivial, as any aggregation would inherit the highest classification level of the individually aggregated ADS. However, it is unclear how to determine the upper bound at aggregation, as well as after the implementation of specific analytical methods on those ADS. Therefore, the analytical communities require a feasible way forward as an interim solution as the defense community continues to tackle this question.

The classification of those ADS post-aggregation depends on their information. DB and IS store information in the content and the encoded in the relationships within and between those ADS. Ergo, if we can regulate the information of the ADS, then we may adjust their classification levels as necessary. Different classification levels require different levels of emphasis within these two core tenets. Specifically, about PII and PHI-types of data, any method must eliminate the possibility of direct targeting of an individual if the data set remains intact. However, it is not necessary to eliminate relationships within the data as these amongst individuals are not PII or PHI. About operational data, there are circumstances where the content of the ADS, as well as the relationships within the ADS may be at the same or higher classification level, and so both tenets require consideration. We assume the following:

1. The utilized ADS are sufficiently complete and of sufficient dimension to imitate given some methodology.
2. The underlying ontology is Probably Approximately Correct (PAC)[33]–learnable.

About Assumption 1, to effectively use real observations to derive synthetic observations, we require observations that are sufficiently complete to do so. We assume that the data is usable and does not require considerable preparation, interpolation, or any other approach for its use. Many of the ADS within the DoD have not yet reached a sufficient level of completeness to derive synthetic observations. Furthermore, and based on prior data extraction observations, many DoD data sets do not necessarily possess sufficient depth for the level of width. Therefore, there will likely be challenges with the generation of synthetic data if there are a sufficient number of observations available for their ontology. We posit that most systems are likely to exhibit highly nonlinear behavior and so would require many more observations to imitate the underlying behavior better. About

---

[32] The Perseus-Pisces supercluster (SCI 40) stretches approximately 250 million lightyears and is one of the largest known structures in the nearby universe, while the other is Laniakea, in which the Milky Way resides.
[33] Proposed by Leslie Valiant in 1984 and is an aspect in machine learning which provides a framework for mathematical analysis. The objective is to define a system that has low generalization error of the selected generalization function (or hypothesis) which is bounded in polynomial time.

Assumption 2, we assume that there is sufficient ontological information encoded within the real observations under consideration. This assumption would imply that some methodology exists which can approximate the content and context within the data sets under consideration.

This capability is exceedingly important, as aggregation will likely increase the overall classification level and so will likely diminish the number of organizations which may be privy to using this data. In many cases, the DoD relies on academic institutions, government contractors, or other government organizations or agencies to enhance their understanding and so increase their potential capability. Providing a mechanism to pre-select how closely synthetic observations may imitate real observations would enable an organization to retain these relationships without having to sacrifice aggregation. The underlying mechanism to generate synthetic observations in a meaningful way and the necessary halting criteria can be found in other works (Smalenberger, 2018).

Data Security—A Flawed Concept

Systems that contain or transmit data (and so transmit information) have always been prone to attack. We may attribute attacks to the fourth general principle of economics: "People respond to incentives" (Miller, Benjamin, & North, 2002). The typical approach has always been to focus on enhanced security features, including physical security, firewalls, limiting access to the data, and more exotic encryption techniques. Each of these approaches ultimately fails, and it is only a function of time until they do. It seems likely that a better version of the same old approaches is insufficient to diminish the incentive structure for the adversary.

These older strategies originated when data storage and processing on that data was at a premium. Therefore, the only data you wished to store was that data which was meaningful. Any user would delete the rest. Today, however, data storage and IS are orders of magnitude cheaper than they were previously. We propose a strategy to take advantage of that.

Assume that the synthetic observations stored in Perseus are not degraded in any form to provide them to academia, etc., but imitate the observations stored in Laniakea in a way that is no longer distinguishable between the two. If we derive the entire Perseus data supercluster in this way, we could generate a fully synthetic version of Laniakea within Perseus. This tactic would mean that given both data superclusters, it would not be possible to distinguish between the two which one contains real observations and which one contains synthetic observations. However, this approach still provides a large incentive for an adversary since a 50% chance remains that they select the correct data supercluster. Also, they could select both and provide contingency plans in the 50% chance that they selected the incorrect one.

A possibly improved approach would be to introduce the synthetic observations into Laniakea at tracked but seemingly random locations. Using this approach, we may introduce a much larger number of synthetic observations into Laniakea that remain indistinguishable from the real observations. Therefore, it would be increasingly difficult to know which observations are real and which observations are synthetic to exploit them. This stenographic[34] approach limits the incentive structure of an adversary.

However, the tradeoff with using this alternative approach is that certain attributes will need to be flagged and duplicated across the synthetic data fields as well. These are typically those key fields such as a person's name, social security number, and so forth. Once these are duplicated across the synthetic observations as well, it will be extremely arduous to classify real from synthetic observations. Additional research needs to be conducted to validate this as a feasible approach.

Summary

Due to adversarial evolutionary pressures, the DoD requires an improved approach to conduct the PPBE process which is technologically-scalable, mathematically-flexible, and factually-rigid. The proposed methodology in this paper provided a potential approach to determine the ontology of important senior leadership questions, develop a strategy to map these to pertinent ADS via M^3 and Laniakea, and leverage analytical techniques that meet their assumptions and solution requirements via QuANTUM. The paper also indicated how someone might generate synthetic data via Perseus which would enable communities to participate in the platform's future development while addressing classification concerns with ADS aggregation. Lastly, the paper addressed a potential approach to addressing the critical flaws with data security by using stenographic encryption.

References

Carpenter, W. C., & Barthelemy, J.-F. M. (1993). A comparison of polynomial approximations and artificial neural nets as response surfaces. *Structural Optimization, 5,* 166–174.

Committee on National Security Systems. (2003). *National information assurance (IA) glossary* (CNSS Instruction 4009). Ft. Meade, MD: Author.

DoD. (2007). *Unique identification (UID) standards for a net-centric Department of Defense* (DoD Directive 8320.03). Washington, DC: Author.

DoD. (2013). *Estimating and comparing the full costs of civilian and active duty military manpower and contract support* (DoDI 7041.04). Washington, DC: Author.

GAO. (2005). *A glossary of terms used in the federal budget process*. Washington, DC: Author.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, J. (2014, June 10). Generative adversarial networks. Retrieved March 11, 2018, from https://arxiv.org/abs/1406.2661

Himmel, C. D., & May, G. S. (1991). Advantages of plasma etch modeling using neural networks over statistical techniques. *IEEE Transactions on Semiconductor Manufacturing, 6,* 103–111.

Hussain, A. S., Xuanqiang, Y., & Johnson, R. D. (1991). Application of neural computing in pharmaceutical product development. *Pharmaceutical Research, 8,* 1248–1252.

[34] Stenography is the process of concealing a data set within another data set.

Joint Chiefs of Staff. (2010). *Joint Communications System* (JP 6-0). Washington, DC: Author.

Libicki, M. C. (n.d.). *Sub rosa cyber war*. Santa Monica, CA: RAND.

Mascaro, J. (1962). *The bhagavad gita.* Penguin.

Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: An introduction to computational models of social life*. Princeton, NJ: Princeton University Press.

Miller, R. L., Benjamin, D. K., & North, D. C. (2002). *The economics of public issues*. Boston, MA: Pearson Education.

Mills, C., Carnell, M., & Wheat, B. (2001). *Leaning into Six Sigma: The path to integration of Lean Enterprise and Six Sigma.* Boulder City, CO: Publishing Partners.

Mitchell, M. (1998). *An introduction to genetic algorithms*. Cambridge, MA: MIT Press.

Mitchell, T. M. (1997). *Machine learning*. New York City, NY: MIT Press & McGraw-Hill.

Myers, R. H., Montgomerey, D. C., & Anderson-Cook, C. M. (2009). *Response surface methodology: Process and product optimization using designed experiments*. Hoboken, NJ: John Wiley & Sons.

O'Connor, T., & Wong, H. Y. (2012, February 28). Emergent properties. Retrieved March 12, 2018, from https://plato.stanford.edu/archives/spr2012/entries/properties-emergent/

O'Connor, T., & Wong, H. Y. (2015, March 21). Emergent properties. Retrieved March 1, 2018, from https://plato.stanford.edu/archives/sum2015/entries/properties-emergent/

O'Neil, J., & Conzemius, A. (2006). *The power of SMART goals: Using goals to improve student learning.* Bloomington, IN: Solution Tree Press.

Russell, S., & Norvig, P. (2015). *Artificial intelligence: A modern approach*. Harlow, England: Pearson Education.

Smalenberger, D. M. (2018, forthcoming). On the imitation of real observations via statistical collisions for informational fracking (SCIF).

Under Secretary of Defense, Comptroller. (n.d.). *Department of Defense Financial Management Regulation* (DoD 7000.14-R). Washington, DC: Author.

U. S. Air Force. (2017). *Life cycle management center business enterprise systems reference guide.*

Valiant, L. (1984). A theory of the learnable. *Communications of the ACM, 27*(11), 1134–1142.

von Clausewitz, C. (1832). *Vom kriege.*

Xiangsui, W., & Qiao, L. (1999). *Unrestricted warfare*. Beijing, China: PLA Literature and Arts Publishing House.

Disclaimer

The views expressed in this document are those of the authors and do not reflect the official policy or position of the United States Air Force, the United States Department of Defense, or the United States Government. This work is academic only and should not be used to imply or infer actual mission capabilities or limitations.

Panel 20. Program Management in an Age of Complexity

Thursday, May 10, 2018	
11:15 a.m. – 12:45 p.m.	**Chair: Major General Kirk Vollmecke, USA,** Program Executive Officer for Intelligence, Electronic Warfare, & Sensors ***Acquisition and Development Programs Through the Lens of System Complexity*** Antonio Pugliese, Stevens Institute of Technology James Enos, Stevens Institute of Technology Roshanak Nilchiani, Stevens Institute of Technology ***Developing a Sense of Reality Within Complex Program Management Environments*** Raymond Jones, COL, USA (Ret.), Naval Postgraduate School ***Fielding Better Combat Helmets to Deploying Warfighters*** Robert Mortlock, COL, USA (Ret.), Naval Postgraduate School

Major General Kirk F. Vollmecke, USA—became the Program Executive Officer for Intelligence, Electronic Warfare, and Sensors at Aberdeen Proving Ground, MD, in April 2016. In this position he is responsible for the development, acquisition, fielding, and life cycle support of the Army's portfolio of intelligence, electronic warfare, cyber, biometrics, and target acquisition programs. These capabilities provide the Soldier with the ability to detect, recognize, and identify targets, as well as to collect, tag, and mine intelligence which can be integrated into the tactical network to support force protection, maneuver, persistent surveillance, and provide a more detailed understanding of the battlefield.

MG Vollmecke was commissioned as a Second Lieutenant in May 1984 through ROTC as a distinguished military graduate of the Centre College of Kentucky, where he earned a Bachelor of Arts degree in Economics and Management. He also graduated from the Naval Postgraduate School in 1992 where he earned a Master of Science degree in Management with a concentration in Acquisition and Procurement Management. He is a 1999 graduate of the U.S. Army Command and General Staff and a graduate of the U.S. Army War College in 2004. MG Vollmecke is Acquisition Level III certified in Program and Contract Management.

Prior to his current position, MG Vollmecke served as the Deputy Program Executive Officer for Intelligence, Electronic Warfare and Sensors at Aberdeen Proving Ground, MD. He has also served as the Deputy Commanding General for the Combined Security Transition Command–Afghanistan (CSTC-A) overseeing the security assistance program for the Afghan National Defense Security Forces in support of Operations Enduring Freedom and Freedom's Sentinel. His acquisition assignments include the Deputy for Acquisition and Systems Management, Office of the Assistant Secretary of the Army (Acquisition, Logistics, and Technology), Washington, DC, in which he provided program management oversight of Army acquisition programs. Prior to that assignment, MG Vollmecke was the Commanding General of the Mission and Installation Contracting Command (MICC), Fort Sam Houston, TX, which provided Army commands, installations, and activities contracting solutions and oversight across CONUS. Before that, he served as the Deputy to the Deputy Assistant Secretary of the Army for Procurement to the ASA(ALT). He also served on the Joint Staff as the J-8 Chief, Capabilities and Acquisition Division, and before his tour on the Joint Staff in 2007, he was the Commander, Defense Contract Management Agency Iraq/Afghanistan supporting Operation Iraqi Freedom.

Other acquisition assignments include Headquarters Department of the Army Systems Coordinator for the Future Combat Systems (Brigade Combat Team) program, Executive Officer to the Assistant Secretary of the Army(AL&T); Commander, DCMA Boeing Philadelphia; Program Analyst for the Deputy Chief of Staff of the Army for Programs, Program Analysis, and Evaluation (PA&E) Directorate; Assistant Product Manager M2/M3 for the Bradley Fighting Vehicle Systems project office; Contingency Contracting Officer assigned to the U.S. Army Forces Central Command–Saudi Arabia under Operation Desert Falcon; and Weapon System Contracting Officer assigned to the Army Materiel Command's Communications-Electronics Command(CECOM), which included a deployment to Honduras, Joint Task Force Bravo. Prior to joining the Army's Acquisition Corps in 1991, he served in a variety of mechanized and light infantry battalion staff and company command positions.

Acquisition and Development Programs Through the Lens of System Complexity

Antonio Pugliese—is a PhD candidate in systems engineering at Stevens Institute of Technology, in Hoboken, NJ. He received his BSc and MSc in Aerospace Engineering and a postgraduate master in systems engineering from the University of Naples Federico II in Italy. His doctoral research is on structural complexity metrics for cyber-physical systems.

James Enos—received a BA in Engineering Management from the United States Military Academy in 2000 and an MS in Engineering and Management from the Massachusetts Institute of Technology in 2010. He is currently pursuing a PhD in Systems Engineering from Stevens Institute of Technology in Hoboken, NJ. Since 2000, he has served as an Army officer in various leadership roles and as an Assistant Professor in the Department of Systems Engineering at the U.S. Military Academy. His research interests include system architecture, social network analysis, and systems engineering within the Department of Defense.

Roshanak Nilchiani—is an Associate Professor of Systems Engineering at the Stevens Institute of Technology. She received her BSc in Mechanical Engineering from Sharif University of Technology, MS in Engineering Mechanics from University of Nebraska-Lincoln, and a PhD in Aerospace Systems from Massachusetts Institute of Technology. Her research focuses on computational modeling of complexity and system ilities for space systems and other engineering systems, including the relationship between system complexity, uncertainty, emergence, and risk. The other track of her current research focuses on quantifying, measuring, and embedding resilience and sustainability in large-scale critical infrastructure systems.

Abstract

The approach of the Department of Defense (DoD) to acquisition programs is strongly based on systems engineering. DoD Directive 5000.01 calls for “the application of a systems engineering approach that optimizes total system performance and minimizes total ownership costs” (DoD, 2007). Even when systems engineering best practices are employed, the cost of large systems is always increasing, and a large part of this increase is due to system complexity (Arena et al., 2008).

Part of this system complexity comes from the functionalities of the system, and is thus justified when these functionalities are required. The remaining contribution is due to unnecessary intricacies in the design, to local optimization, and to oversight in the system-level design. This complexity can lead to rising cost and schedule delays, and should be addressed properly. To overcome these issues regarding cost and schedule overruns, researchers have advocated for the adoption of a complexity budget (Sinha, 2014), which can help identify the effects of unintended interfaces between system elements. While most literature seems to agree about the existence of this issue, the solutions to the measurement of complexity are various and based on different approaches.

The purpose of this research is to develop metrics that will allow the DoD to evaluate a complexity budget, particularly in the phases of architecture and design development. The metrics are developed using a set of axioms that can be applied to cyber-physical systems, and they assume that the architecture of the system is known. Knowledge of the system architecture allows for a graph representation of the system and uses graph-theoretic approaches to the evaluation of the topology of the system. Concepts such as graph density and graph energy can be used to build metrics that allow to rank architectures, thus helping identify possible sources of complexity. Additionally, this approach allows engineers to look external to the system to identify the complexity required to interoperate with legacy DoD systems and systems under development. This research effort is limited to a snapshot of the

state of the system, but can be extended to a dynamical approach with a system changing state or changing its structure.

Introduction

Complexity in engineered systems is a double-edged sword. Part of it is due to the functionalities of the system, and part of it to the unnecessary intricacies which deviate the final design from an elegant solution, the optimal one. The excess complexity in engineered systems can potentially contribute to increased partial or systemic risks and increased fragility of the system in face of various shocks and environmental changes.

The first attempts at heavier-than-air flight were carried out by small teams of people that we would today call innovators. The goal of those systems was to achieve leveled flight over a relatively short distance. As time passed, the requirements for airplanes increased in almost all the applications, from military to commercial flight. The need to carry cargo, payloads, or passengers over increasing distances, in shorter time, at a viable cost, safely, and reliably has led to an increase in the complexity of these systems over the last 100 years. As a result, today's airplane manufacturers employ tens of thousands of people, and have a hierarchy of suppliers with an even larger total workforce. In addition, the development time for a new program has also increased due to the overall increase in complexity.

Airplanes are only one of the many examples of engineered systems where an increase in complexity is connected to an increase in cost as well as increased fragility and risks of the system. The costs associated with larger complexity are justified only when they are dictated by system requirements. These design decisions can contribute to the functionality of the system (i.e., functional requirements), or increase system-level characteristics such as resilience, reliability, or safety (i.e., non-functional requirements). According to Carlson and Doyle (2002), robustness is the maintenance of desired characteristics despite the failure or partial performance of some components of the system, and is correlated with complexity. As long as there is a reason for a design decision, and there cannot be a simpler solution obtaining the same effect at the system level, the increase in complexity is justified. When the increase in complexity is unintended and contributes to system fragility, then the design solution is not optimal and should be avoided or modified. Unfortunately, to determine the optimality of a design solution, it is necessary to have a deep knowledge of the specific application field, and to have a large set of possible solutions for comparison.

The Department of Defense (DoD) faces challenges in managing complexity, integration, and management of the complex network of systems that it has developed over the past 30 years. In 1996, the Vice Chairman of the Joint Chiefs of Staff proposed warfighting capability would be more reliant on systems of systems (SoS) and network centric operations (Owens, 1996). As such, DoD systems are becoming more and more complex, interconnected, and reliant on other systems to provide capability to the user. This creates a complex environment in which systems connect to each other through a variety of means that may not be initially evident to systems engineers. When these systems operate on the battlefield, they often cross service boundaries, but their development within the service makes collaboration difficult in traditionally hierarchal military structures (Dahmann & Baldwin, 2008). Additionally, the Government Accountability Office (GAO; 2015) found that the DoD lacked methods and tools for conducting portfolio management at the enterprise level for capabilities, and noted that there were gaps in the DoD’s ability to identify, understand, and assess the capability portfolio.

This paper presents and builds on a complexity theory, network analysis, and systems engineering to propose a method to understand the complexity budget of a network of systems. It examines how the addition of a new system to a network of legacy systems affects the complexity of the network. As an example, the paper examines the addition of the F-35 Joint Strike Fighter (JSF) to the network of DoD systems and its effect on the complexity of the network. It examines the complexity of the network before the DoD fielded F 35A/B/Cs, during the transition to the JSF, and post deployment after the DoD replaced the legacy systems with the F-35 variants.

Literature Review

This section presents a review of the relevant literature to include a discussion on systems engineering, complexity theory, and network analysis. The portion on systems engineering focuses on the foundation of systems engineering and the application of the ilities to help engineers manage complexity and the non-functional attributes of engineered systems. Additionally, the literature review includes a discussion on complexity theory and the impact of increases in technology and reliance on other systems. Finally, the literature provides an overview of network analysis techniques that serve as a basis for the quantification of complexity and analysis of the network of DoD systems.

Systems Engineering

As a discipline, systems engineering faces increased complexity of systems as technology progress and systems are more interconnected. In 2006, a workshop consisting of thought leaders from a variety of disciplines met to discuss the issue of complex systems, and one area that received substantial attention was the modeling of complex systems with an emphasis on the dynamic, networked nature of systems (Rouse, 2007). International Council on Systems Engineering (INCOSE) defines Systems Engineering as “an interdisciplinary approach and means to enable the realization of successful systems. It focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, and then proceeding with design synthesis and system validation while considering the complete problem” (INCOSE, 2007).

Systems engineers differ from traditional engineers in that they consider the system in its entirety, lead the conceptual design of systems, and bridge the gaps between traditional engineering (Kossiakoff et al., 2011). As such, systems engineers have developed a variety of means—system architecture, system of systems analysis, and enterprise architecture—to deal with complexity. To manage complexity and the qualitative nature of systems engineering, systems engineers have developed the ilities as a construct for assessing nonfunctional attributes of a system. Systems engineers have begun to recognize the criticality of these non-traditional design criteria and have begun to include them in the design of systems (McManus et al., 2009). However, these properties and attributes of a system often manifest themselves after engineers have designed and put the system into operation (de Weck et al., 2012). Further study of the ilities examines how system-level ilities begin to emerge from the subsystem level, where systems engineers can design in these non-functional attributes (Lee & Collins, 2017).

Complexity Theory

Wade and Heydari (2014) categorized complexity definition into three major groups, according to the point of view of the observer. When the observer is external to the system and can only interact with it as a black box, then the type of complexity that can be measured is called behavioral complexity, since it looks at the overall behavior of the system. When the observer has access to the internal structure of the system, such as blueprints and source code for engineered systems or scientific knowledge for natural systems, then the structural complexity of the system is what is being measured. If the process of constructing the entity is under observation, then it is the constructive complexity being measured, which is the complexity of the building process. This definition relates complexity to the difficulty of determining the output of the system.

Sheard and Mostashari (2011) developed a framework for the categorization of complexity types. Engineered systems have two types of complexity: structural and dynamic. Dynamic complexity can be short term or long term. Short term complexity is related to the operation of the system. System behavior can be unpredictable due to non-linear relationships among the system components. The environment can also play a major role on system behavior. Long-term complexity is related with the evolution of the system, its growth, and its adaptation to its environment which plays an important role in shaping the new generations. Structural complexity is instead interested in a snapshot of the system architecture and can be divided into three components: size, connectivity, and topology.

Metrics of structural complexity have been proposed in literature. The most common type of metrics is based on the concepts of entropy (Akundi, 2016; Gell-Mann & Lloyd,1996), information content, or logical depth (Fischi, Nilchiani, & Wade, 2016). Another common type of structural complexity metrics considers the spectrum (the set of eigenvalues) of the graph representation of the system. These metrics are known as spectral metrics and are the ones adopted in this research. The fist spectral metric, proposed by Gutman in 1978 (Gutman, 2011), is known as Graph Energy and is represented by

$$E_A(G) = \sum_{i=1}^{n} |\lambda_i| \tag{1}$$

where λ_i are the eigenvalues of the adjacency matrix of the graph *G*. A variation of this metric, proposed by Gutman as well (Gutman & Zhou, 2006), is the Laplacian Graph Energy, represented as

$$E_L(G) = \sum_{i=1}^{n} \left|\mu_i - \frac{2m}{n}\right| \tag{2}$$

where μ_i are the eigenvalues of the Laplacian matrix, n the number of nodes and m the number of edges of the graph *G*. Cavers, Fallat, and Kirkland (2010) provided a generalization of these two metrics that can be applied to any matrix representing a graph, which is represented by

$$E_M(G) = \sum_{i=1}^{n} \left|\lambda_i(M) - \frac{tr(M)}{n}\right| \tag{3}$$

where $\lambda_i(M)$ are the eigenvalues of the matrix *M*, and $tr(M)$ its trace.

Graph energy has been embedded in a structural complexity metric provided by Sinha (2014), as a contribution of the topology of the graph. The formula

$$C(n,m,A) = \underbrace{\sum_{i=1}^{n} \alpha_i}_{C_1} + \underbrace{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_{ij} A_{ij}\right)}_{C_2} \underbrace{\gamma E(A)}_{C_3} \qquad (4)$$

where α_i represents the inner complexity of each node, and β_{ij} the complexity of the edges, is based on the idea that structural complexity has three contributions: components, connections, and topology (Sheard & Mostashari, 2011).

Another type of spectral structural metric has been proposed by Wu et al. (2010) considers the eigenvalues of the adjacency matrix as an exponential function, and adjusts the value through a logarithmic scale

$$N_A(G) = \ln\left(\frac{1}{n}\sum_{i=1}^{n} e^{\lambda_i}\right) \qquad (5)$$

The coefficient $1/n$ is a way of normalizing the graph according to the number of nodes, which allows one to compare graphs of different sizes. This approach has been used by Sinha as well with the coefficient $\gamma = 1/n$. These metrics have been used as a starting point for the development of 12 metrics that consider the system as a graph and are based on the eigenvalues of a certain matrix representing this graph.

Capability Development in the DoD

The DoD generates requirements through the Joint Capability Integration and Development (JCIDS) process, which they then pass to the acquisitions community to develop and procure warfighting systems. As a part of this process, DoD systems engineers analyze the current state of legacy systems and determine how the new capability will integrate with these systems. The DoD designed the system to ensure validated military capability requirements support resourcing decisions for programs. The 2003 Joint Defense Capability Study first presented the concept of JCIDS and proposed a transition from requirements-based acquisition to a capability-based approach (Joint Chiefs of Staff, 2004). The JCIDS process supports the Chairman’s and the Joint Requirements Oversight Committee’s (JROC) statutory responsibilities to identify, assess, validate, and prioritize joint military capability requirements (Joint Chiefs of Staff, 2012a). The JCIDS process requires sponsors to generate three main documents—Initial Capability Document (ICD), Capability Development Document (CDD), and the Capability Production Document (CPD)—that support different phases in the development and acquisition process by providing traceability from warfighter capability requirements to fielded systems (Joint Chiefs of Staff, 2012b).

As part of the JCIDS process, the Joint Staff requires several DoD Architecture Framework (DoDAF) viewpoints to support the development of warfighter capabilities. Architecture frameworks assist decision makers by serving as a communication tool by presenting a manageable amount of information from a set of data to assist stakeholders in managing complex systems (Richards et al., 2006). System architects use DoDAF, one of several common frameworks, to capture multiple perspectives of a warfighting capability’s system architecture. All architecture frameworks include specific taxonomies, artifacts, and terminologies for describing a system to ensure standardization across multiple individual architectures (Friedenthal, Moore, & Steiner, 2012). DoDAF includes eight different viewpoints that capture data relevant capability requirements, integration, military operations, and program management aspects of a system (DoD Chief Information Officer, 2010). The DoD designed DoDAF to meet the needs of a diverse set of stakeholders and decision makers by abstracting essential pieces of information and presenting them in manageable pieces depending on their perspective (DoD Chief Information Officer, 2010). The required DoDAF products provide valuable data at the individual system level; however, they do not provide much insight into the larger, aggregated network of systems.

One shortfall of the DoDAF architectures used in capability development is that they do not capture a DoD-wide perspective of the interactions between individual systems. Several efforts have attempted to aggregate independent DoDAF products along mission threads; however, they still limit their approach to a subset of the entire DoD capability network of systems. Ring et al. (2009) proposed the Activity-Based Methodology, which aggregates DoDAF architectures into an integrated architecture that captures the organization, system, and role aspects of DoD systems. Another effort proposed aggregating independent architectures through a system, capability, and mission perspective by utilizing independent DoDAF viewpoints (Enos, 2014).

F-35 Joint Strike Fighter

The F-35 JSF is a joint, multi-role fighter and attack aircraft that is entering service with the Air Force, Navy, and Marines to replace a variety of legacy systems. The F-35 is a fifth-generation fighter aircraft that incorporates stealth technology into the design of the aircraft and uses a common airframe across all three versions of the aircraft (Church, 2015). The F-35A is the conventional take-off and landing version of the JSF that incorporates an advanced sensor package and situational awareness capability to drastically improve the effectiveness of the aircraft (U.S. Air Force, 2014). The Air Force plans to replace both the F-16 and A-10 with the F 35A beginning in 2016 as it fields their version of the F-35 in air-superiority, suppression of enemy air defense, and close air support roles (Church, 2015). The Marine Corps began fielding the F-35B short takeoff and vertical landing (STOVL) version of the JSF in that provides the capability to take off and land on extremely short runways. The Marine Corps plans to use the F-35B to replace both the F/A-18 Hornet and the A/V-8B Harrier II with the JSF (JSF Program Office, 2017). The Navy's version of the JSF, the F-35C, includes increased wing area and structural enhancements to support carrier landings and take offs. The Navy plans to replace the F/A-18 with the JSF to serve as its primary air superiority and attack aircraft (JSF Program Office, 2017).

Methodology

This section presents the methodology that the authors adopted in the formulation of new spectral structural complexity metrics, and the data collection strategy for the characterization of the complex tactical aircraft system of systems.

Development of Complexity Metrics

The metrics presented in this paper are all spectral complexity metrics, meaning that they are based on the eigenvalues of a certain graph representation of the system. To represent the graphs, three different matrices are used: the adjacency matrix, the Laplacian matrix, and the normalized Laplacian matrix. The adjacency matrix is the most frequently used representation of an architecture within the systems engineering domain. Also known as Design Structure Matrix (DSM; Yassine & Braha, 2003), or N^2 matrix, it is used to represent the interfaces and their arrangement, and allows one to make considerations on architectural modularity and clustering of components. The Laplacian matrix includes additional information with respect to the adjacency one, specifically regarding the degree of each component. The normalized Laplacian matrix has an interesting spectrum that is related to other graph invariants more than the spectra of the other two matrices (Chung, 1997). These three matrices are considered in their weighted variations, where edges and vertices of the graph carry different weights. The metrics are based on two similar concepts, graph energy and natural connectivity, which as seen in the previous section are both functions of the eigenvalues of the matrix representation of the system. A corrective coefficient $\gamma = 1/n$ to compare graphs with different number of nodes is included in the definition of natural connectivity (Wu et al., 2010) and in Sinha's (2014) structural complexity metric.

The metrics are applied to two sets of random graphs, generated through Erdõs-Rényi (ER) and Barabási Albert (BA) algorithms. The values of each metric are plotted against graph density, which is defined as

$$d = \frac{2m}{n(n-1)} \tag{6}$$

for undirected graphs, and as

$$d = \frac{m}{n(n-1)} \tag{7}$$

for directed graphs, where n is the number of nodes and m is the number of edges in the graph *G*.

Another graph indicator used in this research is graph diameter, defined as the maximum shortest path between all pairs of nodes in the graph. In absence of accurate information regarding the internal structure of nodes, which is usually the case in system of systems applications, where one organization cannot access data belonging to external actors, the complexity of the nodes can be approximated with the degree of the node $\alpha_i = \deg v_i$, and $\beta_{ij} = \sqrt{\alpha_i \alpha_j}$.

Metrics such as graph energy and natural connectivity, which have been introduced in the previous section, can be represented through the following formula

$$C(S) = f\left(\gamma \sum_{i=1}^{n} g\left(\lambda_i(M) - \frac{tr(M)}{n}\right)\right) \tag{8}$$

where $f_1(x) = x,\ g_1(y) = |y|,\ f_2(x) = \ln x\ ,\ g_2(y) = e^y$ are the possible values for the functions f and g, the coefficient γ can be $\gamma_1 = 1, \gamma_2 = n^{-1}$, and the matrix representation of the graph can be either $M_1 = A, M_2 = L, M_3 = \mathcal{L}$, which have been defined in our previous publication (Nilchiani & Pugliese, 2016).

Table 1 shows the metrics that can be derived from this formula through combinations of these parameters. Two sets of functions, two values for the coefficient γ, and three matrices, give 12 possible metrics. Throughout this paper, the metrics are referred to using acronyms: graph energy (GE), Laplacian graph energy (LGE), normalized Laplacian graph energy (NLGE), natural connectivity (NC), Laplacian natural connectivity (LNC), normalized Laplacian natural connectivity (NLNC), and where $\gamma = 1/n$, the acronym has a trailing n, such as in (GEn). These metrics will be applied in the next section to sets of random graphs, and to the TACAIR system of systems.

Table 1. Twelve Examples of Spectral Structural Complexity Metrics

	Adjacency Matrix	Laplacian Matrix	Normalized Laplacian Matrix
$\gamma = 1$	$E_A(G) = \sum_{i=1}^{n} \lvert\lambda_i\rvert$	$E_L(G) = \sum_{i=1}^{n} \left\lvert\mu_i - \frac{2m}{n}\right\rvert$	$E_{\mathcal{L}}(G) = \sum_{i=1}^{n} \lvert\nu_i - 1\rvert$
	$N_A(G) = \ln\left(\sum_{i=1}^{n} e^{\lambda_i}\right)$	$N_L(G) = \ln\left(\sum_{i=1}^{n} e^{\mu_i - \frac{2m}{n}}\right)$	$N_{\mathcal{L}}(G) = \ln\left(\sum_{i=1}^{n} e^{\nu_i - 1}\right)$
$\gamma = \frac{1}{n}$	$E_{An}(G) = \frac{1}{n}\sum_{i=1}^{n} \lvert\lambda_i\rvert$	$E_{Ln}(G) = \frac{1}{n}\sum_{i=1}^{n} \left\lvert\mu_i - \frac{2m}{n}\right\rvert$	$E_{\mathcal{L}n}(G) = \frac{1}{n}\sum_{i=1}^{n} \lvert\nu_i - 1\rvert$
	$N_{An}(G) = \ln\left(\frac{1}{n}\sum_{i=1}^{n} e^{\lambda_i}\right)$	$N_{Ln}(G) = \ln\left(\frac{1}{n}\sum_{i=1}^{n} e^{\mu_i - \frac{2m}{n}}\right)$	$N_{\mathcal{L}n}(G) = \ln\left(\frac{1}{n}\sum_{i=1}^{n} e^{\nu_i - 1}\right)$

TACAIR System of Systems

This section presents an overview of the methodology to develop three individual networks of systems that capture the "as-is," "transitional," and "to-be" networks. A variety of publicly available sources provide the necessary data to develop the network of systems and identify connections between the systems (Church, 2015; JSF Program Office, 2017). The network captures interoperability connections between the systems that include information flows, shared resources, and physical connections (Enos & Nilchiani, 2017). Table 2 presents an excerpt from the entire adjacency matrix for the tactical aircraft network of systems. A complete matrix for each of the networks captures the data required to analyze the complexity of the network.

Table 2. Excerpt From Adjacency Matrix

	A-10C	AIM-120	AIM-9X	F-16C	F-22	F-35A	GPS III	Link-16	JDAM	KC-46
A-10C			X				X	X	X	X
AIM-120				X	X	X				
AIM-9X	X			X	X	X				
F-16C		X	X				X	X	X	X
F-22		X	X			X	X	X	X	X
F-35A		X	X		X		X	X	X	X
GPS III	X			X	X	X				
Link-16	X			X	X	X				
JDAM	X			X	X	X				
KC-46	X			X	X	X				

The "as-is" network captures the systems that comprise the DoD's tactical aircraft system and consists of aircraft, munitions, sensors, and communication systems prior to the fielding of the F-35. The "transitional" includes all the legacy aircraft as well as the JSF and its connections that represents the DoD network as the Air Force, Navy, and Marine Corps transition to the F-35 from their legacy aircraft. Finally, the "to-be" network depicts the DoD's network of tactical aircraft and systems after the three services retire the systems the F-35 is scheduled to replace.

Figure 1 presents the graphical depiction of the “as-is” network of DoD tactical aircraft systems and represents the past version of the network prior to the deployment of the F-35 JSF variants. This network captures various types of systems that operate together to provide the DoD with tactical aircraft capability to include the aircraft, munitions, sensors, satellites, weapons, and command, control, communications, computer, and intelligence (C4I) systems. In the graph, the colors represent the various services, the shapes of the nodes represent the type of system, and the size of the node represents its degree. This network represents the DoD’s tactical aircraft systems prior to the development of the JSF and provides the baseline for analyzing the complexity of the network.

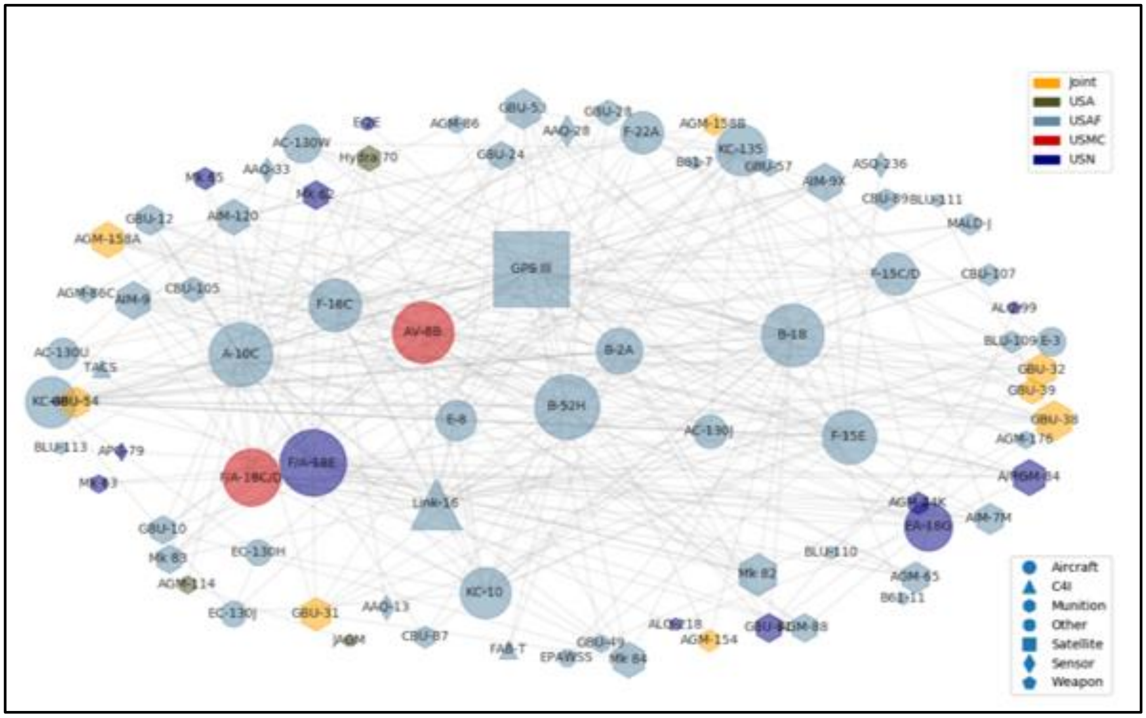


Figure 1. As-Is Network of DoD Tactical Aircraft Systems

Figure 2 presents the graphical depiction of the network and the connections that will be present during the transition from the legacy aircraft to the JSF variants. In this case, both the JSF and the aircraft the services plan to replace with the F-35 variants are included in the network along with any of their connections to other systems in the network. This version of the network provides a means to evaluate the complexity of the network during the transition period to the JSF which could impact resource expenditures, maintenance, supply operations, and tactical operations of the DoD.

Figure 3 presents the final version of the network and represents the “to-be” tactical aircraft network after the services retire the legacy systems that they are replacing with the F-35. In this case, the network does not include the retired A-10, F-16, F/A-18, and AV-8B systems from the Air Force, Navy, and Marine Corps. In addition to the four retired systems, the network removes systems that may not be retired but no longer connect to the network to include the Hydra Rockets, AN/APG-79 AESA Radar, and Mk 63 Sea Mine. This does not indicate that these systems could also be retired as they may be used by other systems; however, it does affect the complexity of the tactical aircraft network. This version of the network provides the means to calculate the complexity of the network after a complete transition to the JSF and can determine if the DoD increased or decreased the complexity of its tactical aircraft network.

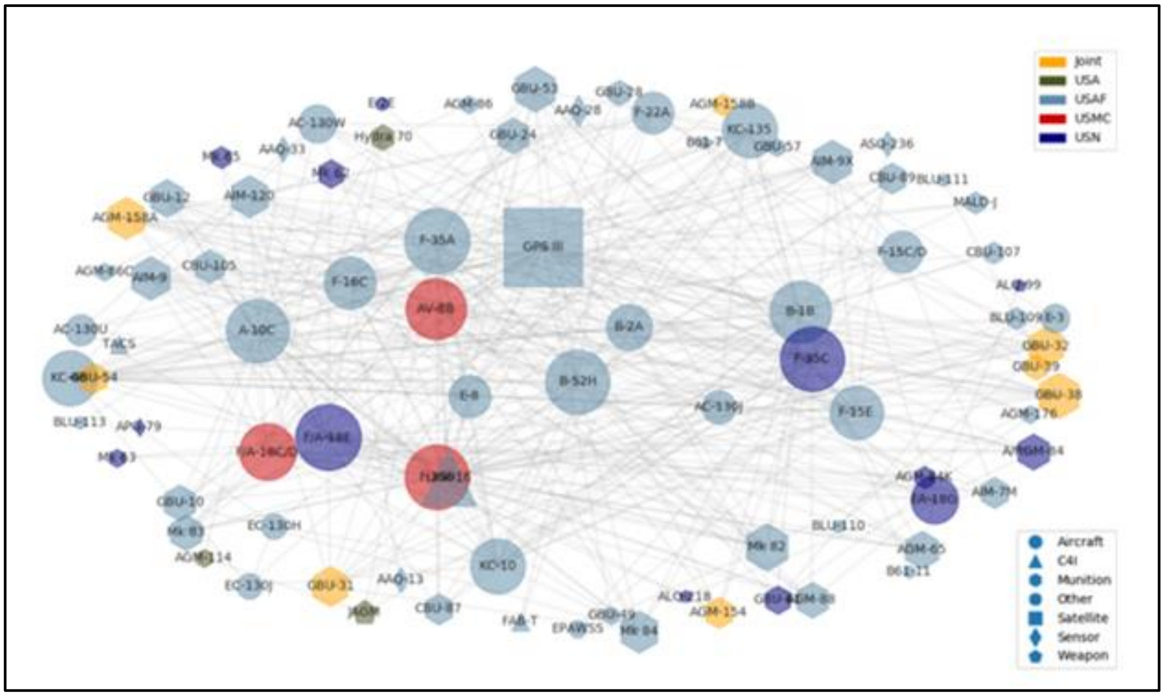


Figure 2. "Transitional" Network of DoD Tactical Aircraft Systems

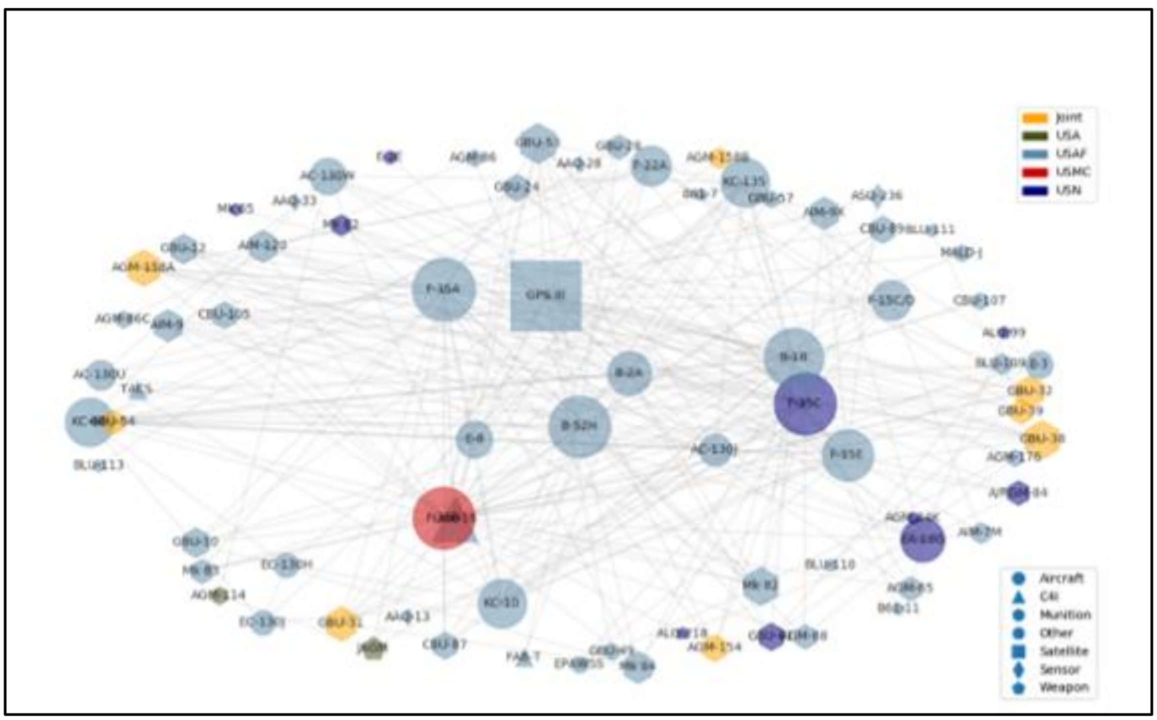


Figure 3. "To-Be" Network of Tactical Aircraft Systems

Analysis and Results

The metrics have been applied to two sets of random graphs, generated with ER and BA models respectively. The sets of graphs contain approximately 23,000 and 38,000 unique labeled graphs.

Figure 4 represents the values that the 12 spectral structural metrics assume when applied to the ER set of random graphs. Most of the metrics have a positive correlation with the number of nodes in the graph, meaning that the metric value is higher when the number of nodes is higher. This is the expected behavior for a complexity metric, and the two metrics that do not follow it, NLGEn and NLNCn, are not suitable as complexity metrics.

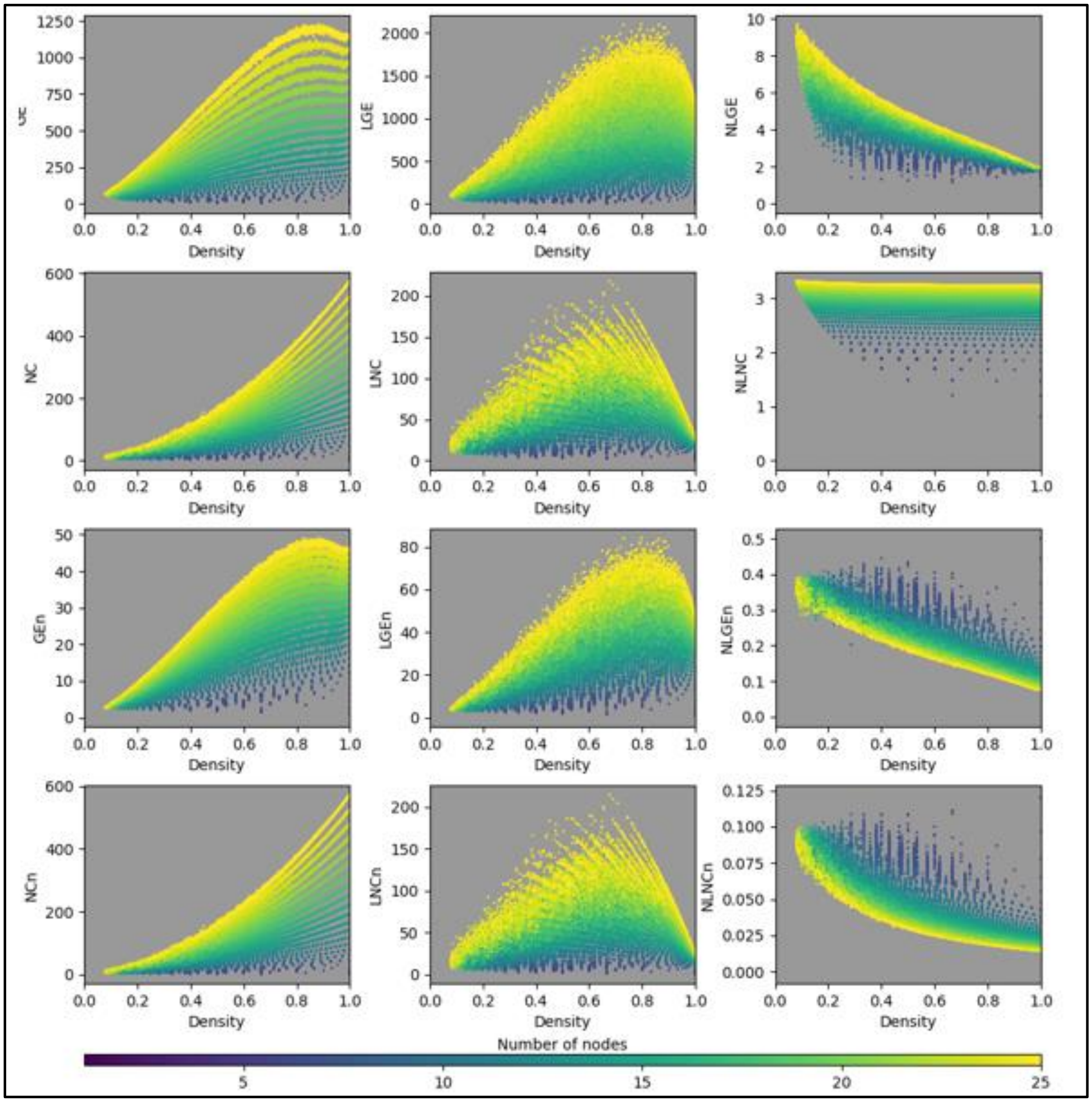


Figure 4. Metric vs. Density Plots With Color Scale According to Number of Nodes, for Each Metric, for Graphs Generated Using the Erdõs-Rényi Algorithm

From Figure 5 it is possible to see that for ER random graphs the diameter is high with low density graphs, and low when the density is high. This relationship is expected since the complete graph has diameter one and removing edges creates an increase of the shortest path between pairs of nodes. Although valid for ER graphs, the relationship between density and diameter is not general, since star graphs and path graphs with the same number of nodes have the same density, but the former have diameter 2 while the latter have diameter $n-1$. This means that for high n, the diameter of these two types of graphs is very different. This is one limitation of the ER algorithm, which will not generate star graphs, or graphs with highly skewed degree distributions, given its uniform probability of edge creation.

To overcome the limitations of the ER model, and to better mimic the topology of engineered systems with heterogeneous components, a set of graphs has been generated using the BA model. These graphs have a more skewed degree distribution, given by the preferential attachment strategy.

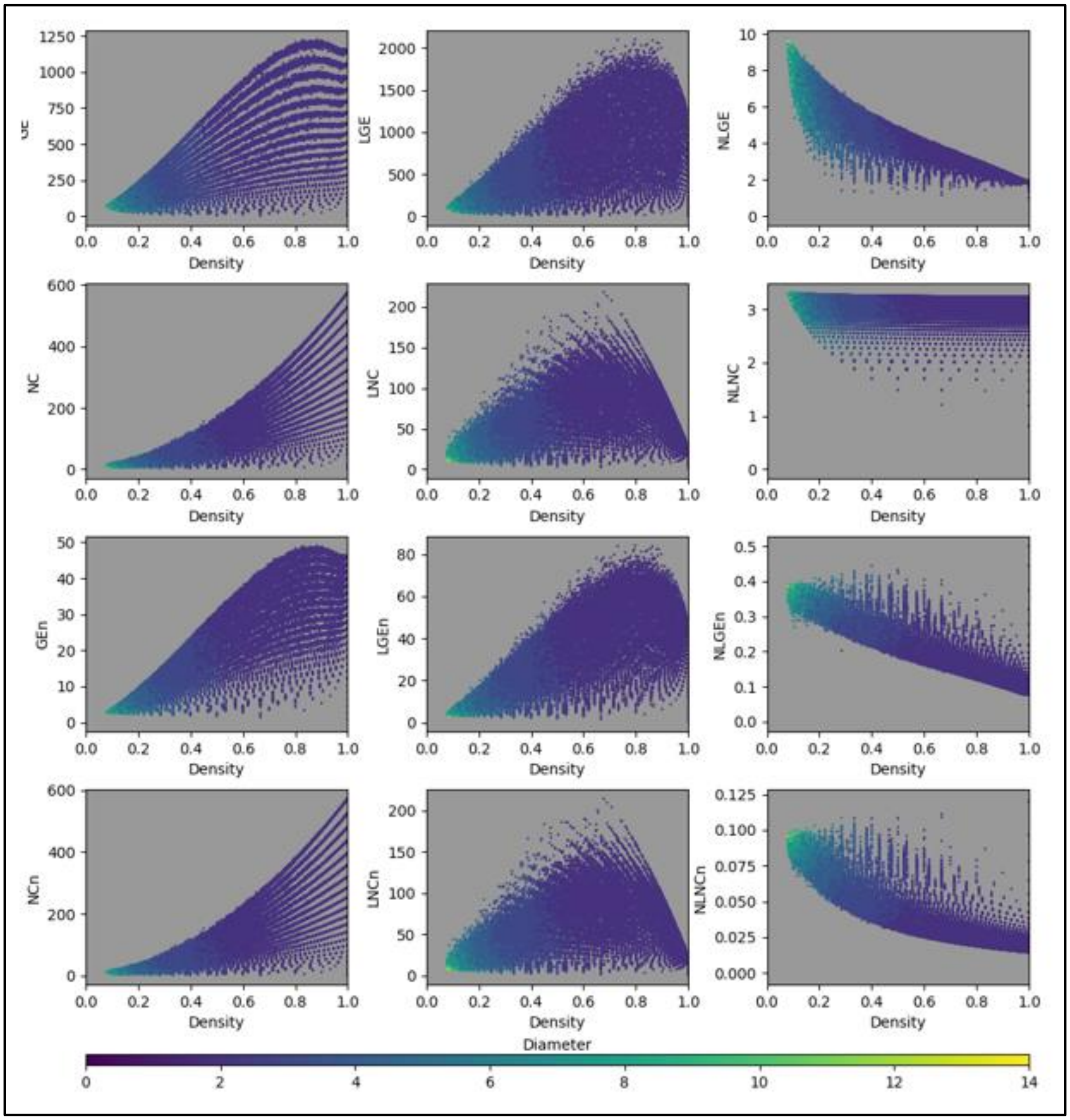


Figure 5. Metric vs. Density Plots With Color Scale According to Diameter, for Each Metric, for Graphs Generated Using the Erdõs-Rényi Algorithm

Figure 6 shows the metrics evaluated for the set of BA random graphs. Given the way the algorithm works, these graphs do not span the whole density range, but stop at $d = 0.57$. The main feature of these point clouds is a folding, a bifurcation, so that graphs with the same density will belong to two distinct sets with a high and low value of each metric respectively. This bifurcation gives meaning to the metrics, highlighting the fact that they are responsive to topological changes.

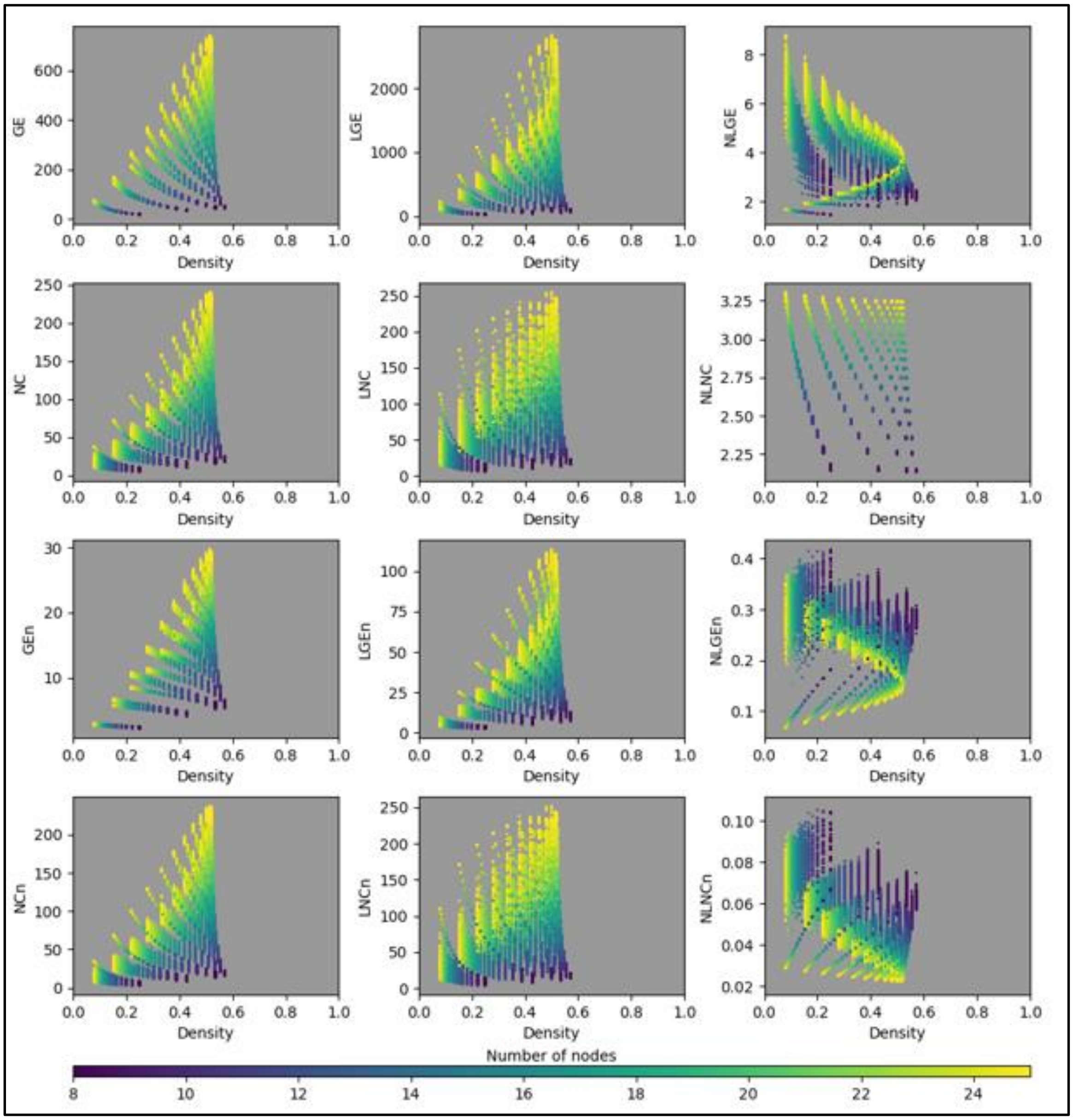


Figure 6. Metric vs. Density Plots With Color Scale According to Number of Nodes, for Each Metric, for Graphs Generated Using the Barabási-Albert Algorithm

Figure 7 shows that this bifurcation in BA random graphs is related to the diameter of the graphs. The diameter does not have the same trend as in ER graphs. Graphs with low density which have high diameter and low diameter exist. These two sets are represented by trees with high depth and stars, respectively. While a star topology is not common in engineered systems, since it is subject to bottlenecks and the complexity of the central node would tend to be too high, trees are common structures for engineered systems, where a certain level of decentralization is in order. Even in the presence of cycles, when the graph is not a tree anymore, a diameter value of 10 in a graph of 25 nodes is representative of engineered systems.

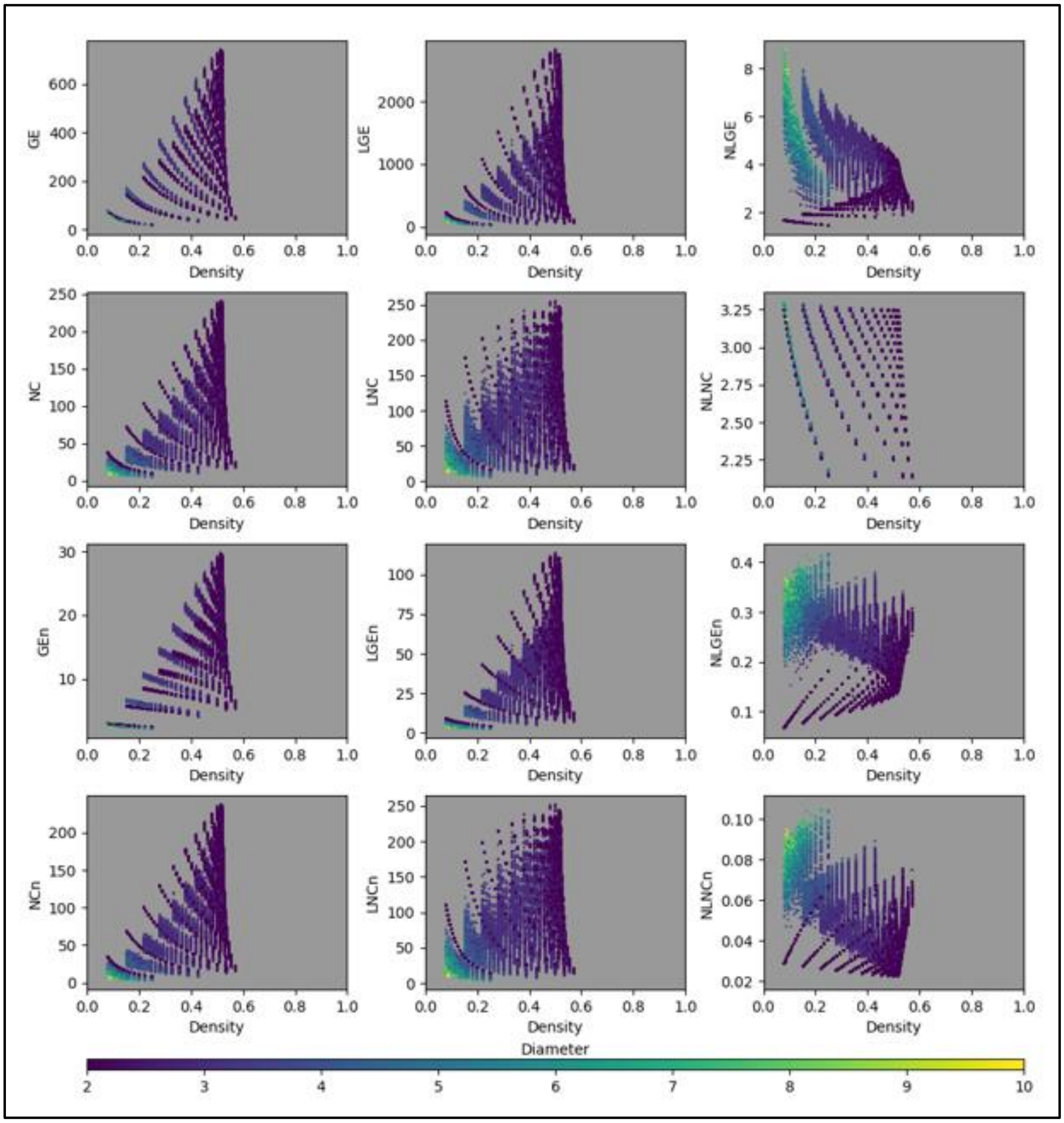


Figure 7. Metric vs. Density Plots With Color Scale According to Diameter, for Each Metric, for Graphs Generated Using the Barabási-Albert Algorithm

Analysis of the TACAIR System of Systems

The TACAIR system of system, in its three versions presented earlier, is undergoing radical changes. The introduction of the F-35 in the operational scenario and the subsequent retirement of legacy systems is causing modifications to the network topology. The number of nodes went from 82 to 85 and will go down to 77, and the number of interfaces went from 384 to 466 and will be 347 once the transition is complete. This leads to a density value going from 0.115 to 0.130, and to 0.118 in future. This density variation is not accompanied by a change in diameter which remains constant to 5, due to the centrality of the nodes that are being added and removed from the network. In this case, the metrics are beneficial to the network analysis, since they can tell more than the diameter about the topology of the network.

Figure 8 shows the metrics applied to the TACAIR system of systems. Other than NLGEn and NLNCn, which we have already ruled out as reliable complexity metrics, and NLGE, the other metrics agree that the introduction of the F-35 represents an increase in the complexity of the network. Most of the metrics, other than NC and NCn, also agree that the retirement of the legacy systems is beneficial for the network and will lead to a simplification of the overall network.

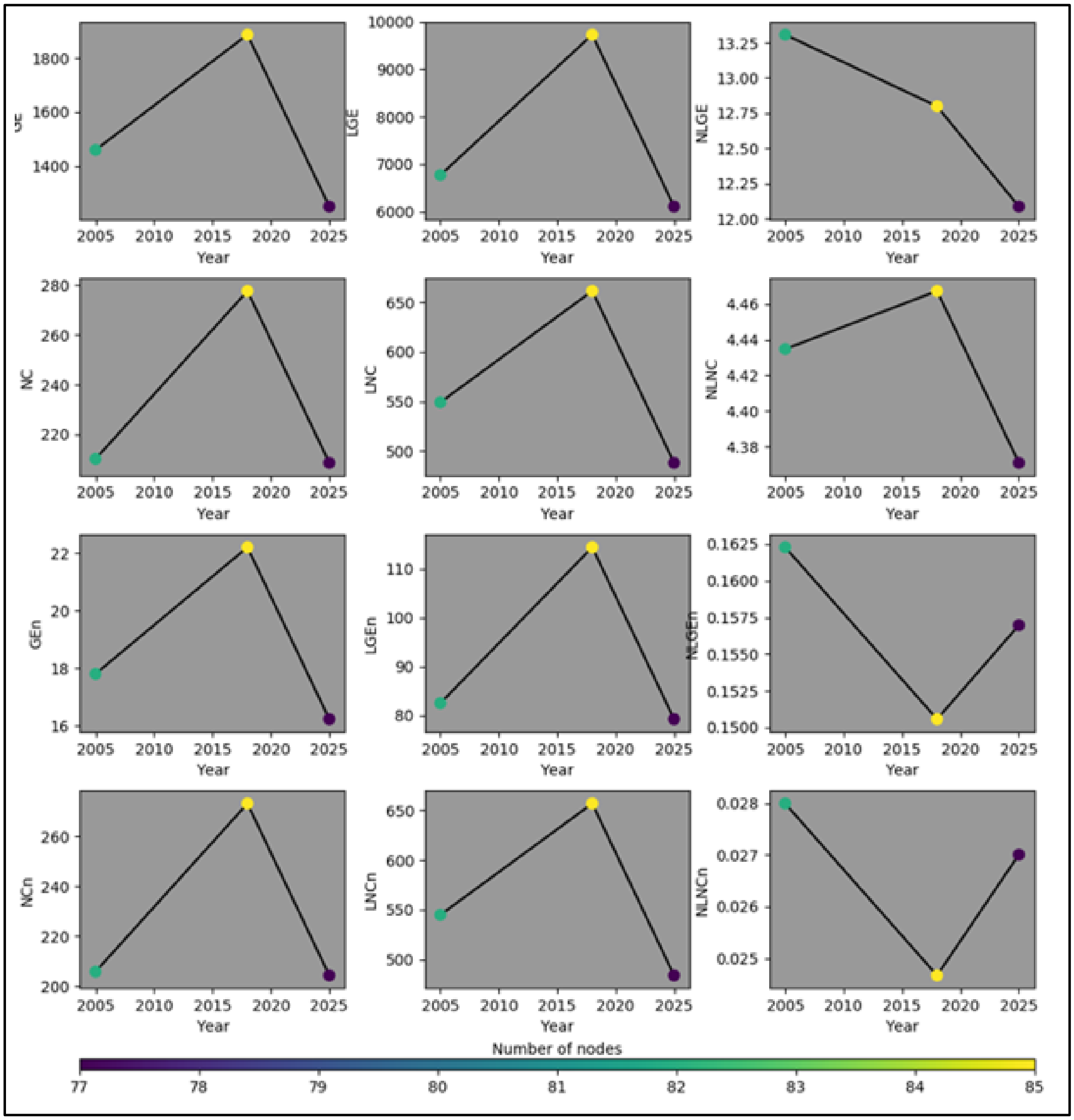


Figure 8. Evaluation of Spectral Structural Complexity Metrics With the Evolving Versions of the TACAIR System of Systems

Conclusion and Future Work

This paper presented an approach to the measurement of structural complexity that involves the measurement of the eigenvalues of a matrix representation of the system. Twelve spectral metrics have been created, based on features of existing metrics. The metrics have been applied to two sets of graphs, generated using the Erdõs-Rényi (ER) and Barabási Albert (BA) algorithms respectively. It is argued how the application of these algorithms to the generation of graphs representing engineered systems should be carried out together with considerations about the heterogeneity of the components of the system and the expected distribution of node degree. ER models having a close to uniform distribution of node degree are applicable to the representation of homogeneous graphs, such as networks of routers, in which all the components have the same tasks and functionalities. When specialization arises, and the components of a system are wildly heterogeneous, the degree distribution is highly skewed, and BA models are more appropriate.

The application to the TACAIR system of systems is an example of how the operational scenario can become complex thanks to the relationships between different types of systems, and how the introduction of new systems and the retirement of legacy ones can be beneficial to the management of the network, by streamlining the supplying of common resources and reducing the diversity of systems that achieve the same functionalities. Of course, this type of analysis can be improved when details about the architecture of each system are available, and the interfaces can be modeled with high fidelity regarding the timing and range of connections.

Limiting the approach to publicly available data allowed us to assume the point of view of an external actor who is interested in introducing a new system in an already existing environment. Examples of such systems can be the introduction of a new type of transportation system, such as the hyperloop concept, within the already existing network of air, sea, and land transportation systems, or the introduction of a new surgical tool to be used in conjunction with the existing set of operation room equipment.

In the future, if detailed data is available regarding one of the existing systems in the network, it would be possible to analyze the network and yield more insightful considerations about the retirement of such systems and the effect on the overall network.

References

Abbott, R. (2006). Emergence explained: Abstractions: Getting epiphenomena to do real work. *Complexity, 12,* 13–26.

Akundi, S. A. (2016). *Information entropy measures applied to hierarchial complex technical and soci-technical systems*. El Paso, TX: University of Texas at El Paso.

Arena, M. V., Younossi, O., Brancato, K., Blickstein, I., & Grammich, C. A. (2008). *Why has the cost of fixed-wing aircraft risen? A macroscopic examination of the trends in us military aircraft costs over the past several decades*. RAND.

Carlson, J. M., & Doyle, J. (2002). Complexity and robustness. *Proceedings of the National Academy of Sciences, 99,* 2538–2545.

Cavers, M., Fallat, S., & Kirkland, S. (2010). On the normalized Laplacian energy and general Randić index R-1 of graphs. *Linear Algebra and Its Applications, 433*, 172–190.

Chung, F. R. K. (1997). *Spectral graph theory*. Providence, RI: American Mathematical Society.

Church, A. M. (2015, May). Gallery of USAF weapons. *Air Force Magazine,* 80–103.

Dahmann, J., & Baldwin, K. (2008). Understanding the current state of US defense systems of systems and the implications for systems engineering. In *Systems Conference*. Montreal, Canada.

de Weck, O. L., Ross, A. M., & Rhodes, D. H. (2012). Investigating relationships and semantic sets amongst system lifecycle properties (ilities). In *Proceedings of the Third International Engineering Systems Symposium.* Delft, Netherlands.

DoD. (2007). *The defense acquisition system* (DoDD 5000.01). Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics.

DoD Chief Information Officer. (2010, August). *DoD architecture framework version 2.02.* Retrieved from http://www.dodcio.defense.gov/dodaf20.aspx

Enos, J. (2014). Synthesizing DoDAF architectures to develop the joint capability enterprise architecture. In *Systems Engineering.* Washington, DC.

Enos, J. R., & Nilchiani, R. R. (2017). Using social network analysis to quantify interoperability in a large system of systems. *In Proceedings of the 12th Annual System of Systems Engineering Conference.*

Fischi, J., Nilchiani, R., & Wade, J. (2016). System and architecture evaluation framework using cross-domain dynamic complexity measures. *In Proceedings of the 2016 IEEE Systems Conference (SysCon).*

Friedenthal, S., Moore, A. & Steiner, R. (2012). *A practical guide to SysML: The systems modeling language*. New York, NY: Morgan Kaufmann OMG Press.

Gell-Mann, M., & Lloyd, S. (1996). Information measures, effective complexity, and total information. *Complexity, 2*, 44–52.

GAO. (2015). *Opportunities exist to improve the Department of Defense's portfolio management.* Washington, DC: Author.

Gutman, I. (2011). *Hyperenergetic and hypoenergetic graphs. Selected Topics on Applications of Graph Spectra, Math. Inst., Belgrade*, 113–135.

Gutman, I., & Zhou, B. (2006). Laplacian energy of a graph. *Linear Algebra and Its Applications, 41*4, 29–37.

INCOSE. (2007). *Systems engineering handbook v3.1*. International Council on Systems Engineering.

Joint Chiefs of Staff. (2004). *Joint defense capability study: Improving DoD strategic planning, resourcing and execution to satisfy joint capabilities*. Washington, DC: DoD.

Joint Chiefs of Staff. (2012a). *Joint capabilities integration and development system* (CJCSI 3170.01H). Washington, DC: DoD.

Joint Chiefs of Staff. (2012b). *Manual for the operation of the joint capabilities integration and development system.* Washington, DC: DoD.

JSF Program Office. (2017). F-35 Lightning II background. Washington, DC: DoD. Retrieved from http://www.jsf.mil/f35/f35_background.htm

Kossiakoff, A., Sweet, W., Seymour, S., & Biemer, S. (2011). *Systems engineering principles and practice.* Hoboken, NJ: Wiley.

Lee, J. Y., & Collins, G. J., (2017). On using ilities of non-functional properties for subsystems and components. *Systems*.

McManus, H., Richards, M., Ross, A. & Hastings, D. (2009). *A framework for incorporating "ilities" in tradespace studies*. American Institute of Aeronautics and Astronautics.

Nilchiani, R. R., & Pugliese, A. (2016). *A complex systems perspective of risk mitigation and modeling in development and acquisition programs.*

Owens, W. A. (1996). *The emerging U.S. system-of-systems*. National Defense University Strategic Forum.

Richards, M., Shah, N., Hasting, D. & Rhodes, D. (2006). Managing complexity with the Department of Defense architecture framework: Development of a dynamic system architecture model. In *Conference on Systems Engineering Research*. Los Angeles, CA.

Ring, S., Nicholson, D., Thilenius, J. & Harris, S. (2004). *An activity-based methodology for development and analysis of integrated DoD architectures—"The art of architecture."* Bedford, MA: MITRE.

Rouse, W. (2007). Complex engineered, organizational and natural systems. *Systems Engineering,* 260.

Sheard, S. A., & Mostashari, A. (2011). Complexity types: From science to systems engineering. In *Proceedings of the 21st Annual of the International Council on Systems Engineering (INCOSE) International Symposium*.

Sinha, K. (2014). *Structural complexity and its implications for design of cyber-physical systems.*

U.S. Air Force. (2014, April 11). F-35A Lightning II. Washington, DC: DoD. Retrieved from http://www.af.mil/About-Us/Fact-Sheets/Display/Article/478441/f-35a-lightning-ii-conventional-takeoff-and-landing-variant/

Wade, J., & Heydari, B. (2014). Complexity: Definition and reduction techniques. In *Proceedings of the Poster Workshop at the 2014 Complex Systems Design & Management International Conference.*

Willcox, K., Allaire, D., Deyst, J., He, C., & Sondecker, G. (2011). *Stochastic process decision methods for complex-cyber-physical systems.*

Wu, J., Barahona, M., Yue-Jin, T., & Hong-Zhong, D. (2010). Natural connectivity of complex networks. *Chinese Physics Letters, 27*, 078902.

Yassine, A. & Braha, D. (2003). Complex concurrent engineering and the design structure matrix method. *Concurrent Engineering, 11,* 165–176.

Developing a Sense of Reality Within Complex Program Management Environments

Raymond D. Jones, COL, USA (Ret.)—retired as a Colonel from the U.S. Army in 2012 and is a Lecturer with the Graduate School for Business and Public Policy at the Naval Postgraduate School. His last assignment in the Army was as the Deputy Program Executive Officer for the Joint Tactical Radio System (JTRS). Additionally, he served as the Military Deputy for the Director of Acquisition Resources and Analysis in the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]), managed three Major Defense programs for the DoD in addition to his many operational and research and development assignments. He graduated from the U.S. Naval Test Pilot School in 1995 and is a 1983 graduate of the United States Military Academy. He has a Bachelor of Science degree in Aerospace Engineering, a Master of Science degree in Aeronautical Engineering from the Naval Postgraduate School, a Master's in Business Administration from Regis University, a Master's Degree in National Resource Strategy from the Industrial College of the Armed Forces, and is currently a PhD candidate with the Graduate School of Information Sciences at the Naval Postgraduate School in Monterey, CA.

Abstract

How leaders make decisions in complex and chaotic environments could have a significant impact on organizational performance. This study of leaders from across the Department of Defense (DoD) provides the foundation by which a more informed understanding of how program managers' sense of situational reality ultimately leads to timely and relevant decisions. This study specifically focuses on the emergence of four aggregate categories—sensemaking, trust, tacit knowledge, and explicit knowledge—that seem to shape the leader's reality and subsequent decision-making process in highly complex environments. I refer to the integrated nature of these categories as nousmaking, or making reality of the situation and choices based on one's sense of the reality. Ultimately, these factors determine the velocity and quality of the decisions leading to overall organizational effectiveness. Understanding the underlying nature by which leaders gain a sense of reality within the decision-making environment will help shape future organizational structures and processes as well as leader development.

Introduction

Despite the many Department of Defense (DoD) successes, many of the DoD's programs and operations are still vulnerable to underperformance and excessive cost growth during times of increasingly constrained budgets. Since 1975, there have been an annual array of studies, beginning with the Packard Commission, that have had virtually no impact on the ever-increasing trend of cost growth and substandard program performance. Successfully addressing these challenges can yield fiscal dividends that the Department could use to meet priorities such as readiness and modernization needs.

The DoD continues to struggle to overcome the many problems brought about by more than a decade of war and the need to accelerate the procurement of capability, while fighting on several fronts around the world. Often, the necessity of speed of delivery, resulting in underperforming programs, has spuriously suggested that program management is the root cause of program underperformance. The consequence of this assumption has been legislative language that tends to address program leaders' motivations and incentives, rather than the root causes of program managers making decisions that often have little impact on program performance.

Although the DoD has increased its procurement budget over the years, it consistently pays more and takes longer than planned to develop systems that do not perform as anticipated. The DoD spends over $100 billion a year in contracting for goods and services. Over the last few years, the DoD has made several broad-based changes to its acquisition and contracting processes to improve DoD–contractor relationships and rules and has given attention to acquisition reform initiatives with little real improvement. The most glaring example of this failure is the termination of the DoD's Joint Tactical Radio System, which cost more than $17 billion, with little return for the investment.

It is time to examine the root causes of DoD program challenges from a more scientific perspective, rather than from the traditional organizational theory and policy view. The policy changes that have attempted to create efficiencies by using commercial best practices, portfolio management, and additional oversight have failed to produce their intended results. A deeper understanding of how leaders make decisions and the mitigating impacts of those decisions is necessary to truly change the acquisition framework in a way that will result in an improved return on investment for defense materiel development programs. *The problem this research seeks to understand is the underlying nature of why a program manager's decision making does not consistently manifest in improved program performance*. This study is being conducted in two phases. Phase I of this study is a qualitative research effort based upon grounded theory. The results of this study will provide the basis for a quantitative study in which measurable factors such as organizational structure and policy will be examined with regard to the leader's ability to link a sense of situational understanding with the structural realities of the business environment. Commensurate with this problem are the questions that help guide this research, presented in the following section.

Phase I

This research is focused on the first two questions:

a. What is the underlying nature of how decision makers gain a sense of reality by which their decisions are subsequently informed within the unique construct of their functional framework?

b. How do program managers of Major Defense Acquisition Programs (MDAPs) make sense of complex and chaotic program environments, and does this differ from other professions that operate in complex environments?

This study is predicated on the basic assumption that there is an inherent process by which an individual makes decisions. This process involves a deliberate problem-solving methodology and a less-well-defined cognitive and interactive process that influences the ability of the decision maker to gain a sense of reality. While the overall research effort will be a mixed methods approach, the initial study is a qualitative descriptive approach based upon grounded theory. This report begins to address the first question in Phase I and will support subsequent research through which the initial theory will emerge in support of Phase II.

Phase I is grounded in the naturalistic tradition and using a longitudinal qualitative, ethnographic approach to better understand the dynamics and processes of individuals making decisions in a group environment under volatile, uncertain, complex, and ambiguous (VUCA) conditions. By understanding the constructs from with which a program manager derives a sense of reality and understanding of the nature of the world perhaps we can gain insight into how to better inform that reality, leading to more effective judgements and decisions.

We will focus on programs that are at Technology Readiness Level (TRL) six or higher, as defined by the DoD TRL Guide (Assistant Secretary of Defense for Research and Engineering, 2011). This study will examine many complex decision-making environments and compare the fundamental nature of these environments with each other and their relative effectiveness. By exploring a wide variety of complex and chaotic leadership environments, and ultimately cross-coding them, perhaps we can gain a more informed view of how individuals respond to adversity. Ultimately, this insight can lead to better organizational understanding and the changes that will have a greater chance of success.

Phase I leverages previous studies on decision makers in complex and chaotic environments such as Operational Detachment A (ODA) team leaders in highly volatile and ambiguous situations. While the scenarios are different than procurement environments, there are common themes which will help us to better understand why some decisions result in success and some in failure when they are eventually shaped by the functional construct within which they are made. The purpose of this analysis is to reveal a deeper understanding of the very nature of how individuals establish a sense of reality within the context of a complex ambiguous decision-making environment.

Driskell and Salas (1991) presented two conclusions that are extremely relevant to this research. First, "under stress, group members will defer more to the opinions, ideas, and actions of the group leader" (pp. 473–478). This implies that in a stressful time, support staffs will begin to defer more to the leader instead of being the unbiased and objective voice for the leader that informs the leader of the cost of operations, in terms of the manpower, resources, time, and risk involved. Driskell and Salas's conclusions could help us to understand how fundamental confidence is shaped based upon the leader support structure. Driskell and Salas (1991) also explained that "at the same time, the leader will be more likely to reject input from group members" (p. 473). This implies that if the staff were to remain impartial and act as a voice of reason for the leader, the leader who is under stress would disregard the guidance and counsel of the staff and make a decision based on either inadequate information (ignored or discounted information) or intuition (Riabacke, 2006). Hence, the dynamics of external influences become a factor in how leaders perceive and respond to their environment, possibly influencing their sense of reality within the construct of their situation.

Understanding how the complexity of the situation influences the decision maker at a base level will lead us to a richer understanding of the decision process as we begin to contextualize it within a functional context such as program management. The culture of an organization will influence how the organization makes decisions (Riabacke, 2006). Organizations can predict outcomes by examining the epistemic motivation of the staff. With a higher pro-social motivation, the staff or team will be more likely to search, encode, and retrieve information that is more conducive and consistent with group goals (De Dreu, Nijstad, & van Knippenberg, 2008). Furthermore, the research of Kruglanski and Webster in 1996 shows that the staff is likely to "seize and freeze" when it comes to a quick solution, rather than an accurate one, and that once the staff reaches closure, they are usually unmovable (De Dreu et al., 2008).

What Is Decision Making?

Research on decision making has focused more on the organizational and environmental influence of the leader and less on the inherent contextual interaction by which leaders make decisions. For example, models such as the Cynefin Framework (Snowden & Boone, 2007) are used to better understand the decision-making process in environments that range from simple to chaotic. This model describes chaotic environments as those in which relationships between cause and effect are impossible to determine because they are constantly changing and never develop a manageable pattern. Hence, Snowden and Boone (2007) suggest that the leader is simply reacting with the intent of eventually creating the conditions by which a pattern can emerge, migrating the environment into one of complexity rather than chaos.

Complexity tends to be viewed as something with many parts that interact with each other in many ways ("Complexity," n.d.). More specifically, complex decision environments tend to involve many interacting and non-linear elements, and can be retrospective when viewed from a historical perspective, resulting in agents that tend to constrain themselves over time (Snowden & Boone, 2007). While these definitions are important to understanding the environment within which the leader makes decisions, the research has not provided an understanding the cognitive processes by which the decision maker formulates a sense of perspective and understanding of the situational reality and subsequently translates this reality into effective decisions.

Kathleen Eisenhardt (1989) suggests in her study, *Making Fast Strategic Decisions in High Velocity Environments, t*hat performance (or effective decisions) is a function of speed which results from a number of key mediating processes, including accelerated cognitive processing. The ability to make "speedy" decisions of sufficient quality is directly related to the effectiveness of the decision. The notion that effective decisions are related to confidence in the decisions is the basis upon which this study is focused. How decision makers create the reality within which they develop a sense of confidence and conviction in their choices is fundamental to understanding the relative relationship between effective and non-effective outcomes.

Nousmaking

Initial findings indicate that there are four basic categories that decision makers seem to consistently exhibit when confronted with chaotic and complex problems. These emerging categories were observed in our initial round of interviews with Special Operations (ODA) soldiers and will be the basis of subsequent interviews of program managers' decision making in complex and chaotic environments. These four categories include sensemaking, trust, tacit knowledge, and explicit knowledge. Because of the strong interaction of these four categories with regard to influencing the ability of the decision maker to interpret and come to a state of reality (Nous), I refer to this interaction as *Nousmaking*, a necessary process for "speedy" and quality decisions that lead to enhanced performance or effective decision outcomes. Within Eisenhardt's (1989) model of decision making, Nousmaking would encompass the key mediating process, in particular cognitive processing. The interaction of these four aggregate categories was shown to be present in all of the decision environments described by the ODA leaders.

A situation can consist of random, unordered events that cloud judgment and may impact the problem-solving ability of the decision maker. While the defense program environment may not have the same immediate impact to life, the random and inconsistent nature of events can be just as relevant to the decision maker and can lead to second- and third-order effects, which can then lead to major adverse programmatic impacts. The increasingly complex nature of today's technical and programmatic environment, coupled with the uncertainty of future security threats to the nation, provide for a complex and chaotic environment, similar to other fields at their base level that are trying to understand the stimulus under which the decision maker is formulating a sense of perspective or reality. Additionally, the value of the decisions made in context with the environmental inputs and preferred outcomes can be a seemingly random series of events influenced by the VUCA nature of the environment.

Being able to arrive at a true meaning of the environment and see the reality of a situation is referred to as *Nous*, which in classical philosophy refers to the ability to understand what is true or real ("Nous," 1973). Nous is often referred to as the equivalent of perception that works within the mind (Rorty, 1979). This paper illustrates that in order to achieve a level of perception necessary to translate into an effective decision, there is an inherent level of understanding and processing that must occur, which includes sensemaking, trust, tacit knowledge, and explicit knowledge.

Theoretical Framework

This research leverages the data, information, knowledge, wisdom (DIKW) framework as a loose model upon which to understand the evolution of insight within the decision-making process (see Figure 1).

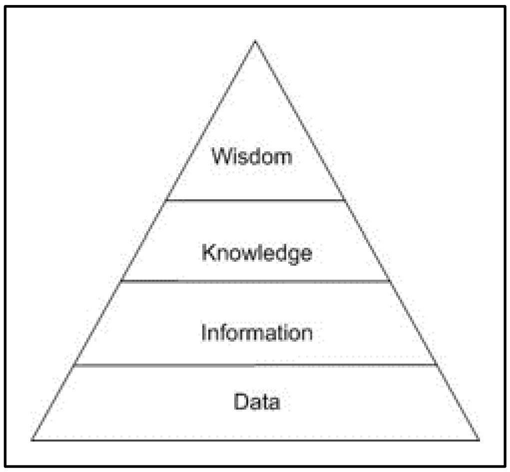


Figure 1. Data, Information, Knowledge, Wisdom (DIKW) Framework

The DIKW pyramid—also known as the DIKW hierarchy, wisdom hierarchy, knowledge hierarchy, information hierarchy, or data pyramid (Rowley, 2007)—refers loosely to a class of models for representing purported structural and/or functional relationships. This basic model proves useful in our research, in that it reflects the insight gained through a deliberate evolution from the "lifeless" unknown of pure data, to the novel insight of wisdom. As one is immersed in a situationally complex environment, making sense of it is predicated on the "data" one internalizes. Lacking any other context, this initial source of input is just as lifeless as the data described in the DIKW model. It is not until a higher level of context is applied to the data that the situation begins to come alive with regard to context. Additionally, this model makes no dispersions on the type of data or the functional environment in which it resides. This definitization begins to occur as the consumer of the data begins to shape it within the context of their environment and derives value and insight as the data transforms along the DIKW framework.

With this as a point of reference, Figure 2 represents a loosely constructed hierarchical model that represents the evolution of "knowing" in the decision-making process, which I refer to as the Decision Clarity Model (DCM).

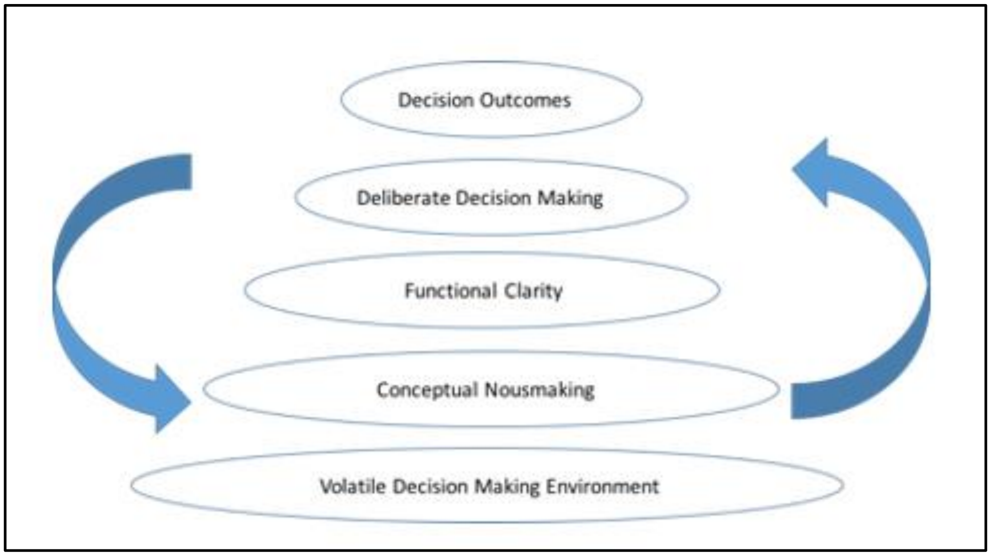


Figure 2. Decision Clarity Model

The DCM represents an evolution of knowing in that at the base level, there exists simply a random, context-free environment consisting of volatility, uncertainty, complexity, and ambiguity. The specific context at this point is not yet relevant other than to say that this environment can exist across many functional domains. It is not until the participant in this environment begins to perceive the environment and applies it to a particular functional construct that the decision-making context becomes relevant.

Conceptual Nousmaking is the point at which the participant begins to make sense of the complex environment through an internal struggle of what is real and relevant. This brings us to our first hypothesis: that there are four key attributes that influence how a decision maker understands and reacts to a particular complex environment.

> Hypothesis 1: There are four aggregate categories that shape and influence a program manager's understanding of a complex environment, consisting of sensemaking, trust, tacit knowledge, and explicit knowledge.

Table 1 summarizes the four categories and the respective attributes associated with each specific category.

Table 1. Conceptual Nousmaking Categories

Second Order Analytic Code	Aggregate Category
1. Retrospective	Sensemaking
2. Plausibility	
3. Social Identity	
4. Organizing	
5. Ability	Trust
6. Benevolence	
7. Integrity	
8. Experience	Tacit Knowledge
9. Know-how	
10. Codified	Explicit Knowledge
11. Logical	
12. Deduction	

Hypothesis 1 will be explored in subsequent interviews with program managers and other leaders that operate in complex environments. Continuous coding will be conducted from subsequent interviews of program managers. This will allow refinement and validation of the initial categories until we have reached a point of saturation. At this point, theory can be proposed upon which Phase II will be quantitatively assessed using the hypothetical deductive process.

The following definitions for the four categories show the inherent relationship with their associated attributes. The initial data collected during initial interviews correlates to the attributes that were derived from initial coding of interviews.

1. *Sensemaking* is the process by which people give meaning to experience and is characterized by the following properties (Weick, 1995):
    a. Identity—helps people identify who they are and shapes what they enact and how they interpret events (Currie, & Brown, 2003; Thurlow & Mills, 2009; Watson, 2009; Weick, Sutcliff, & Obstfeld, 2005).
    b. Retrospection—provides the conditions for sensemaking, such as attention and interruptions, which impact what people notice (Dunford & Jones, 2000).
    c. Organizing—is where individuals simultaneously shape and react to the environment they face. Thurlow and Mills (2009) suggest that individuals will project themselves into an environment and observe the consequences they learn about their identities and the accuracy of their understanding of the event.

d. Plausibility—is more relevant to sensemaking than accuracy since the world is filled with people who have multiple shifting identities (Weick, 1995). This reinforces the value of a larger study group and allows the researcher to explore the possibility of theoretical perspectives.

2. *Trust* is the willingness of an individual to be vulnerable to the actions of another based upon the expectations that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control the other party. While there are several terms associated with trust, three characteristics tend to appear frequently in studies associated with trust (Mayer, Davis, & Schoorman, 1995):
    a. Ability—is a group of skills, competencies, and characteristics that enable a party to have influence within some specific domain (Zand, 1972).
    b. Benevolence—is the extent to which a trustee is believed to want to do good to the trustor, independent of personal profit motive (Mayer et al., 1995). Additionally, Rosen and Jerdee (1977) considered the likelihood that the trustee would put the organization's goals ahead of his or her own goals.
    c. Integrity—is the trustor's perception that the trustee will adhere to a set of principles the trustor finds acceptable (McFall, 1987).
3. *Tacit knowledge* is knowledge that is difficult to transfer to another individual by means of writing or verbalization (Polanyi, 1958). Effectively, it is knowledge that one seems to have acquired and that cannot easily be transferred to another individual, even for extremely complex tasks or situations. Polanyi refers to tacit knowledge as "we can know more than we can tell" (Polanyi, 1966). Tacit knowledge can be characterized by the following:
    a. Know-how—involves learning and skill that was acquired through means other than writing them down. Knowing how, or embodied knowledge, is characteristic of an expert who acts and makes judgements without explicitly reflecting on the principles or rules involved (Schmidt & Hunter, 1993)
    b. Experience—is a key to tacit knowledge in that without some form of shared experience, it would be difficult for people to share each other's thinking processes (Lam, 2000), and thus it would be difficult to anticipate the actions of others, given a common framework and understanding.
4. *Explicit knowledge* can be readily articulated, codified, and accessed (Helie & Sun, 2010). Thus, explicit knowledge can be generated through logical deduction and acquired through both formal and informal means, such as practical experience within a relevant context.

Interfield Theory

Interfield theory is a cross disciplinary study that explores the common relationships between various fields (Darden & Maull, 1977). It is this theory upon which we are able to explore how the volatile environment and conceptual decision making in one discipline relates to another discipline. While the intent is to understand what influences the decision making of program managers in complex situations, this paper predominately explores the first two layers of the Decision Clarity Model, recognizing that the functional clarity level introduces a specific context to the Nousmaking within different disciplines. For example, the functional clarity a program manager experiences is based upon the defense acquisition framework, while the functional clarity for Special Operations soldiers is grounded in the combat framework. The DCM assumes that the first two layers of the process are neutral with regard to the situation. Within our definition of Nousmaking, the four categories of sensemaking, trust, tacit knowledge, and explicit knowledge support theoretical discussions within other disciplines, such as organizational theory, psychology, behavioral science, and so forth.

Interfield theory allows us to identify common patterns at the subconscious level of decision making that can subsequently lead us to a richer understanding of how decisions are made independent of policy and regulation. Introducing the functional clarity of the participant's unique operational framework, we will be able to separate the influence of the environmental framework from the innate process of rationalizing a situation. Once the Deliberate Decision Making and Decision Outcome layers are introduced into the scenario, it will become clear how nousmaking shapes the outcome of decisions and their relative impact. This leads to our second hypothesis:

> Hypothesis 2: Nousmaking is independent of the functional clarity of the operational environment in which choices are formulated.

Hypothesis 2 does not presume that the individual's personal experiences and bias are not relevant to formulating a sense of reality. We are simply suggesting that the aggregate categories that make up the Nousmaking process influence the individual's objective reality similarly, regardless of disciplines, and that the subsequent decision making and outcome are influenced and can be altered by the exigent factors of the functional environment in which the individual's reality has previously been established.

By establishing a demarcation between Nousmaking and decision making within a functional construct, a leader's ability to formulate a speedy high-quality decision is impacted by the ability to both establish a sense of reality as well as respond to the unique constructs of a particular functional setting, and one informs the other. This line of reasoning could lead us to a better understanding of why some leaders prevail and some do not, given the same functional constraints.

A large portion of a leader's ability to make a decision is his or her reliance on past experiences. Leaders are selected after a careful scrutiny of records and evaluations by a centralized panel of senior officers. Research shows that in an experienced-based choice, decisions are made from memories of past outcomes, concluding that memory biases may play a role in the overweighting of extreme outcomes and causing more risk seeking behavior, as demonstrated in the preceding paragraph (Ludvig, Madan, & Spetch, 2013). Ludvig et al. (2013) cite five studies that conclude that a bias exists in which "highly salient and emotional events are over weighted in memory tasks." Another conclusion reached is that extreme outcomes are more likely to be retrieved at the time of a decision and that this may be a heuristic used to simplify the situation at hand and to limit the number of outcomes considered (Ludvig et al., 2013). Ludvig et al.'s research would suggest that there are core processes at work at the base layer at which reality is created that may influence a leader's perspective and will necessarily influence or predetermine the decision strategy within the functional environmental constraint. Studying emergent patterns in transition from a predictable normal routine–centric environment to one of chaos and unpredictability may have significant relevance across various functional domains. The appreciation for the potential of chaos in decision making may have potential relevance in the understanding of both the nonlinearity of making decisions as well as the functional aspects of instability as a means for adapting to new situations in any VUCA environment. Understanding the chaotic and volatile decision-making environment of the battlefield may yield an increased clarity and potential for interpreting decision making in a variety of dynamic and nonlinear decision-making environments. According to Keil (1995), *nonlinearity* refers to behavior in which the relationships between variables in a system are dynamic and disproportionate, whose outcomes are subject to high levels of uncertainty and unpredictability.

Method

Interpretive Approach to Understanding Decision Making in Complex Environments

This study was predicated on the basic assumption that there is an inherent process by which an individual makes decisions and that this process involves a deliberate problem-solving methodology (Drucker, 1967) and a less-well-defined cognitive and interactive process, which influences the ability of the decision maker to arrive at a sense of clarity in ambiguous conditions. Similar to the Buddhist understanding of self and environment in which everything around us is a reflection of our inner lives and is perceived through the self and alters according to the inner state (SGI Quarterly, 1995), this study explored the notion that there are other intrinsic factors involved in decision making that influence the effectiveness of these decisions. Phase I research adopts a qualitative descriptive approach based upon grounded theory (Glaser & Strauss, 1967), in which the interpretations and experiences of the participants remained in the foreground, notwithstanding the fact that some of the interviewers tended to have similar backgrounds and experiences as those being interviewed. During the initial interviews, it was important to maintain a sense of separation from the interviewee in order to limit the bias toward preconceived understanding of the specific events being discussed. In keeping with the approach described by Gioia with regard to giving voice to the informants, it was important to recognize the researcher's expertise, and interpret this pattern in the data, thus providing the best opportunity for discovering new concepts or relationships between existing concepts (Gioia, Corley, & Hamilton, 2012).

While the overall research is focused on program manager decision making in complex and chaotic environments, I was initially interested in a variety of decision-making environments in order to begin to address both hypotheses presented in this paper. The first unit of measure was special operations forces in complex and chaotic conditions and how they developed a sense of reality within the context of these situations. This analysis supports the categorical definitization of the elements within which reality is shaped, or what we are calling Nousmaking.

This initial research study selected participants from a pool of available graduate students within the Defense Analysis Department of the Naval Postgraduate School. Candidates were solicited from the student body enrolled in the Defense Analysis program via email. Respondents were screened and selected based on a required set of criteria, resulting in the identification of 20 research participants. Each participant was interviewed for approximately 60–90 minutes during a semi-structured interview conducted in-person by one of nine identified researchers.

In order to participate in the research study, participants needed to satisfy a number of selection criteria. First, they had to be United States military officers who had served in a leadership position in Iraq or Afghanistan. Second, they had to have experienced complex decision-making situations while in a position of leadership. The final selection focused on Army Operational Detachment–Alpha (ODA) and Navy Sea, Air, Land (SEAL) team leaders. These team leaders tended to have significant exposure to chaotic combat environments, and were in positions to make critical decisions affecting both themselves and their teams. Team leaders from these units were in the rank of O3–O5, the equivalent of captains, majors, and lieutenant colonels in the Army or lieutenants, lieutenant commanders, and commanders in the Navy. This selection of personnel resulted in an exclusively male research pool, and excluded military officers from the Air Force and Marines, due to either inconsistent exposure to similar ground combat operations or lack of availability within the current student body.

Data

The intent of Phase I of this research was to interview a broad spectrum of leaders from varying complex environments and to build a baseline of common categories that exist between the various disciplines. While this initial study leveraged interviews from Special Operations leaders, follow-on interviews will look at a minimum of at least 30 program managers from a broad spectrum of programs that include both challenged programs as well as programs that are performing well against their predetermined baseline.

Data collection for this first round of interviews included three primary sources: (1) tapes and transcripts of the subject interviews; (2) briefings from subject matter experts regarding the operational concepts of ODA teams; and (3) interview debriefing with the entire research team as well as self-debrief and analysis of the data. As part of this research, we used conventional ethnographic analysis methods through the use of memos, notes, and subjective interpretations of the subject's experience depictions. The focus was mainly on the description of events through language by the interviewees to gain meaning to support the experimental interpretation. Analysis of the interviews in a group setting provided varying perspectives of the data, allowing me to explore alternative interpretations and category development. This provided the basis upon which a theoretical direction could be established. The interviews and subsequent interpretation provided a rich basis of data from which to begin to establish a theoretical understanding of the cognitive processes involved in decision making in VUCA environments.

Understanding and subsequent theory requires plausibility, direction, centrality, and adequacy (Charmaz, 2014). It was important to ensure that the descriptive data provided by the interviewees was plausible, lending itself to the development of emerging categories. Throughout this process, a method of constant comparison (Glaser & Strauss, 1967) was used from the many different subject interviews. The data from the interviews was coded, categorized, and evaluated until a systemic pattern began to emerge. The first order coding was the critical link to developing the emergent theory. Incident-by-incident coding was used to compare the relevant ideas identified in the various interviews (Charmaz, 2014). The initial coding helped to establish correlation between the incidents and was the basis from which a framework of understanding evolved.

In order to establish emerging themes, the informants' initial incident coding was put into context and compared against each other. As themes began to emerge, theoretical sampling was used to further elaborate and refine the initial categories. Initially, sensemaking seemed to have a significant effect on the participant's ability to shape the reality of the situation, but it became clear through further research and the memos that there might be a more complex set of variables helping to shape the participant's reality and subsequent decision-making process. Throughout the interviews, I kept asking myself why an individual with relatively few years of experience generally made the “right” decision under seemingly life-altering situations. This question will be critical to further examine as program managers begin to be interviewed and cross coded with the results from other leader’s experiences. Understanding the basis of this phenomenon could have significant impact in helping to shape the conditions for other complex decision-making environments.

While the respondents kept attributing their successes to their formal training, this simply did not reveal itself as the primary causal factor in the data. Through theoretical sampling from the various interviews, I was able to develop the properties of my categories until I reached a point of saturation, the point at which I was not able to develop new information from the data (Glaser & Strauss, 1967). Subsequent interviews of project managers will follow the same method, providing an even richer body of knowledge, which will add validity to the process of determining the overall aggregate categories of Nousmaking.

Data Analysis

Initial respondent coding began to reveal 12 second-order analytical codes that seemed to be interacting throughout the incidents under investigation. These included retrospection, plausibility, social identity, organizing, ability, benevolence, integrity, experience, know-how, codified knowledge, logical knowledge, and deductive knowledge. Through theoretical sampling and continuous probing of the data from all of the interviews, these 12 areas continued to emerge, leading me to four aggregate categories—sensemaking, trust, tacit knowledge, and explicit knowledge—as characteristics influencing the decision maker during the events being described. These categories seemed to have the closest alignment with the emergent themes and, upon further research, revealed themselves as the most plausible description of the process characteristics being described.

Second-order analytic codes (Table 1) are characteristics associated with the aggregate categories and were revealed during the specific events being described by the participant. As the second-order analytic codes began to emerge, it was useful to begin to search for categories that helped to explain my observations. The four categories—sensemaking, trust, tacit knowledge, and explicit knowledge—seemed to align with the emerging data.

While the timelines and circumstances for the various chaotic events varied across all of the subject interviews, the general nature was similar in that the respondents described chaotic and complex circumstances in which they had to make deliberate decisions based upon limited information. As I decomposed their situations and began to code their narratives, there seemed to be a finite set of characteristics emerging and interacting that helped to shape their actions. The decisions they made were both conscious and subconscious, in that often their deliberate actions without apparent deliberation seemed to be second nature. While virtually all of the respondents attributed this to “good training,” further analysis suggests the presence of more than just training.

The respondents consistently displayed the influence of all four aggregate categories during the time frame in which they were responding to immediate chaotic circumstance. While training manifested as explicit knowledge and allowed the respondents to perform certain actions with little thought, tacit knowledge and trust reinforced this knowledge with the sense that they simply “knew” what to do based upon their instincts. Charging an ambush, for example, in Scenario 2 was reflective of this innate knowledge: The decision was shaped by trust in self, perceived trust from superiors, and trust in the team that was reinforced by training and an evolved sense of the current situation, indicating the continuous interaction of sensemaking, trust, tacit knowledge, and explicit knowledge. Figure 3 shows a summary of the number of times I was able to identify the influence of the four aggregate categories during the specific chaotic decision-making window for each of the interviewees.

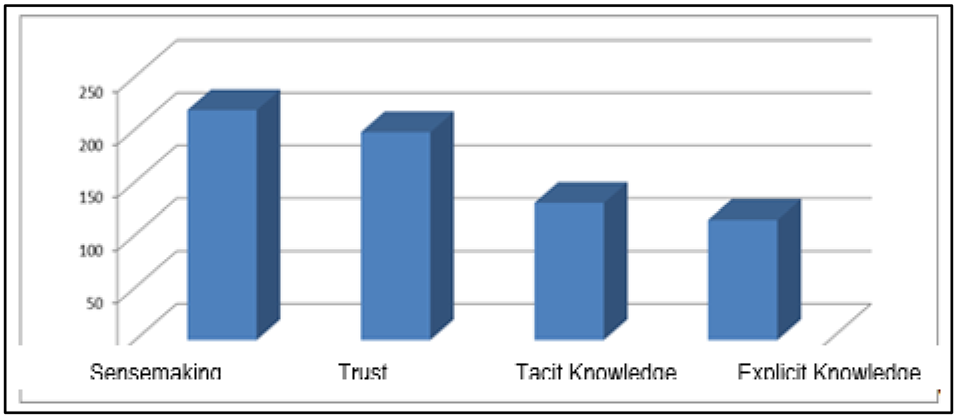


Figure 3. Summary of the Total Number of Observations

Sensemaking and trust seem to be the most prominent categories influencing the respondent during the specific decision-making events, with tacit and explicit knowledge manifesting significantly across all of the chaotic events in relatively equal value. Figure 4 further shows the distribution and number of observations of the second-order codes that emerged during the coding process that helped to define the aggregate categories. Examining the individual second-order codes in relation to each other, there appears to be a higher influence of trust and sensemaking when compared to the other attributes.

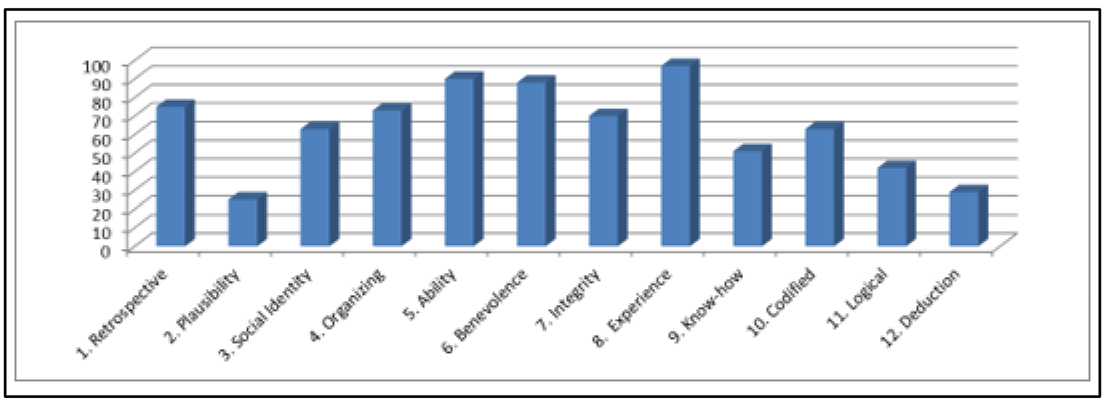


Figure 4. Second-Order Code Summary by Number of Observations for All Interviewees

Discussion

Sonenshein (2007) begins to address the notion that individuals make intuitive judgements in their construction of ethical decisions and suggests that responses to ethical issues are not always based on deliberate and extensive moral reasoning. One can extrapolate from his research that individuals also make intuitive judgements in ambiguous decision-making environments, which ultimately involve ethical and logical choices. While Sonenshein (2007) suggests that individuals are engaging in sensemaking (Weick, 1979) under conditions of equivocality, I further suggest that, consistent with Hypothesis 1, initial findings from these interviews reinforce the notion that decision makers are inherently influenced by the four aggregate categories of sensemaking, trust, tacit knowledge, and explicit knowledge.

What remains to be seen is whether or not these same categories manifest themselves across leaders from different functional environments. If individuals develop their sense of reality similarly, regardless of the functional environment in which they make decisions, this could lead us to questioning the emphasis of focusing on leadership issues as a root cause of defense program failure. One has to then turn to the actual functional constraints of the environment and assess the impact of the actual decision-making environment on program outcomes. If leaders develop their foundation and sense of clarity in similar ways, yet perform differently in different functional constructs, this might even suggest that leaders that are successful in one complex environment may be less successful in others such as the program environment.

The second round of interviews will consist primarily of current and former program managers. Their results will be compared to current data in an effort to establish a sense of validity to the theoretical construct. If data reveals itself as consistent with the emerging results of the Special Operations Forces interviews, this will reinforce the preliminary findings in support of Hypotheses 1 and 2 and set the conditions for Phase II, in which the functional construct of the individual’s environment is compared with the conceptual Nousmaking.

There does seem to be a relationship emerging between speed and quality of decision making. From the initial round of interviews, speed and quality seemed to manifest itself and have some relation to performance. It is still unclear what the relationship between speed, quality, and Nousmaking are with regard to the overall decision-making process. As more data is collected and the functional environmental constraints are applied to the process, this relationship will gain additional clarity. By understanding how an individual establishes a sense of reality and how the functional constructs of the individual’s environment interact with this sense of reality, we hope to better understand how individuals make effective decisions and how the outcomes of these decisions are impacted by speed and quality. This leads us to yet a third hypothesis:

> Hypothesis 3: Speed and quality of decision making have a direct relationship with the positive or negative impact of the decision and are influenced by the aggregate categories of Nousmaking.

The velocity and quality of the decision is influenced by the decision maker’s sense of reality and perceived outcome based upon the functional construct and clarity. Figure 5 represents the relationship between Nousmaking and decision effectiveness, as hypothesized above.

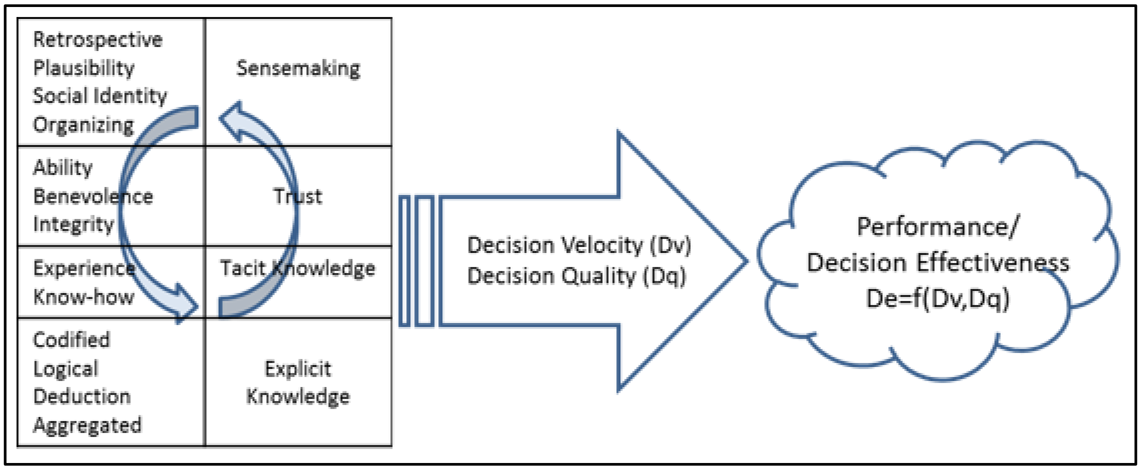


Figure 5. Interactive Relationship Between Four Aggregate Categories and Decision Effectiveness

If decisions are made rapidly but are not reinforced with a level of clarity that informs the quality of the decision, then the overall effectiveness of the decision could be compromised. Additionally, if the quality of the choices is sufficiently high, but the decision comes late, then the effectiveness is also hampered. For example, in the case of Sample Scenario 1, although the respondent could have reached a level of understanding regarding the choice to be made, his overall reality of the situation with regard to the urgency was not sufficiently realized. In this case, I suggest that the combined effect of the four categories, in which a reality of the environment due to the interaction of the four categories resulted in a high velocity and high quality decision, ultimately led to an effective decision and outcome.

Kathleen Eisenhardt (1989) revealed in her article, *Making Fast Strategic Decisions in High-Velocity Environments,* that fast decision makers use more, not less, information than slow decision makers. Additionally, the greater the number of alternatives that are considered simultaneously, the greater the speed of the strategic decision. Her research showed that executives immersed themselves in real-time information about their environment and their firm's operations. The result of this, according to Eisenhardt, was a deep personal knowledge of the enterprise that allowed for rapid decision making. Consequently, the greater the speed of the strategic decision process, the greater the performance in high-velocity environments (Eisenhardt, 1989).

The relationship between performance and speed is illustrated in Figure 6, in which Eisenhardt illustrates the interdependencies of the mediating processes necessary for speedy high performance decisions. The data presented in this paper takes Eisenhardt's reasoning a bit further by offering a definitive relationship between the tangible and intangible qualities of decision making in high velocity and chaotic environments and their relationship to effective decisions. Within the context of Eisenhardt's model, this would further explain the key mediating processes to reflect the relationship between the key mediating processes and the aggregate categories process described in this paper.

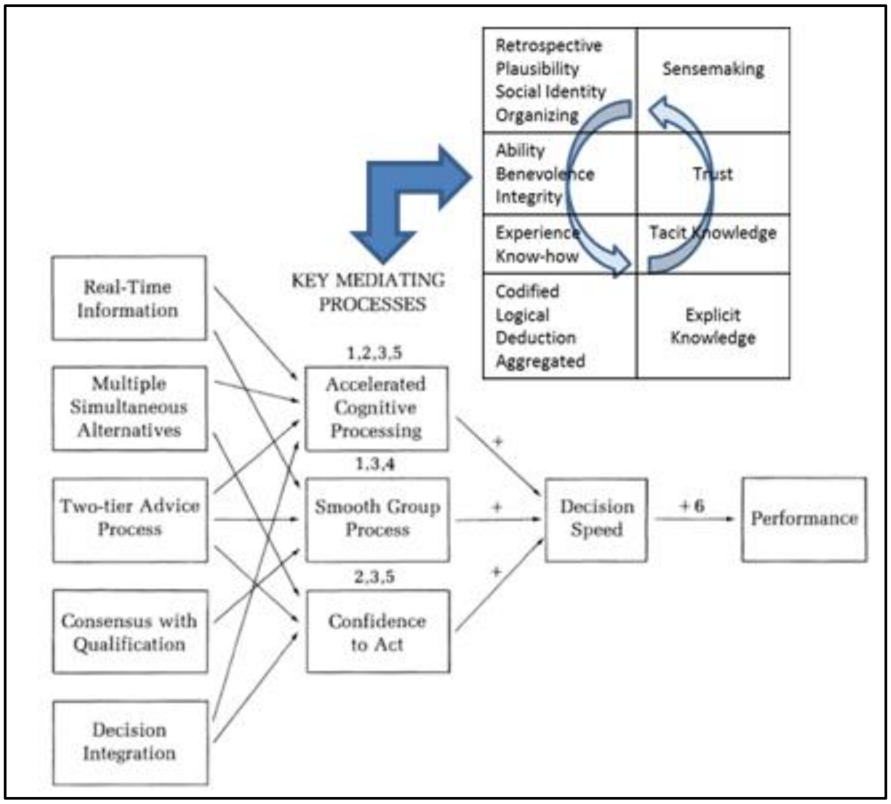


Figure 6. Relationship Between Eisenhardt (1989) Model of Strategic Decision Speed in High-Velocity Environments, With Aggregate Categories

In essence, the decision maker is determining what is real through the interaction of the four aggregate categories that emerged from the respondent data, and subsequently acting on this understanding. The degree to which the decision maker optimizes the aggregate categories and is able to make a timely and high quality decision determines the overall effectiveness of the decision. As we examine more interview data for project managers, we will be able to establish a theoretical basis from which to begin Phase II. Phase II will examine the causal relationship between Nousmaking and the functional environmental construct in which the individual makes decisions. The interdependent nature of Nousmaking and decision making within a functional environment will be revealed, allowing us to design subsequent experiments that examine the effects of varying either the Nousmaking or functional environment and the subsequent impact on the speed and quality of the decision-making process.

Conclusion

The theory presented in this paper represents a qualitative ethnographic study of a group that has the propensity to be required to make life-altering decisions in time-constrained chaotic and complex environments. While this data focused on SoF soldiers, the underlying factors of Nousmaking are presumed to be independent of the functional environmental construct of the participants. While one could dismiss their ability to operate in these environments successfully as a function of their significant training, the results of this study reflect a higher level of cognitive processing that leads to effective decisions and subsequent performance. Using the grounded theory approach to qualitative analysis, interviewees were asked to describe chaotic events during combat situations, in which they were required to make critical decisions. From their stories, I was able to conduct first- and second-order coding from which a theoretical construct emerged to help understand the nature of decision making in these environments.

Four aggregate categories emerged as being relevant to almost all of the scenarios described by the interviewees. These four categories included sensemaking, trust, tacit knowledge, and explicit knowledge. At some point during the decision-making windows, the interviewee exhibited signs that at least one of these categories was at play and facilitated the actions at the moment. Recognizing that decision making is a dynamic process, the interaction of these categories likely played a significant role in helping to shape the interviewee's reality of the environment, involving a perception of the current conditions; trust relationships with subordinates, peers, and superiors; and an innate self-confidence and confidence in their own skills. For the purposes of helping to describe this interactive relationship, I termed this process *Nousmaking*, or perhaps more simply, reality making that provides a sense of clarity in action and purpose.

Practical Impact and Future Research Opportunity

Chaotic and complex decision-making environments are not limited to combat scenarios. Disasters and emergency situations are examples of decision-making environments that have potentially similar characteristics as combat environments in that they reflect the unpredictability and nonlinearity of the situation relative to a more predictable steady state environment. The nonlinearity of these events in which human decision making is predicated by chaos may have certain similarities and patterns that can be studied with regard to their association with the individuals involved in the decision-making process. Complex and high risk business environments can also manifest themselves in a chaotic or unpredictable nature and could be subject to the same cognitive processes as combat. Eisenhardt's (1989) study of high tech companies began to explore the relationship between fast and slow decision making and their potential outcomes. If we better understood the internal influencers that lead to making effective decisions in ambiguous environments, perhaps future organizational and leadership theory and methods could be better tailored to the environment, leading to more predictable outcomes.

Future research should examine in much greater depth the theoretical nature of the Nousmaking process with the goal of mapping these interactions to their relative inputs and desired outputs. Although we will likely never accurately predict the nature of human decision making, better understanding of the integrated parts and their relationships to each other could provide greater insight into the ability to improving decision making across a full spectrum of complex environments.

References

Assistant Secretary of Defense for Research and Engineering. (2011, April). Technology Readiness Assessment (TRA) guidance. Retrieved from https://www.acq.osd.mil/chieftechnologist/publications/docs/TRA2011.pdf

Bourgeois, L. J., & Eisenhardt, K. (1988). Strategic decision processes in high velocity environments: Four cases in the microcomputer industry. *Management Sciences, 34*(7), 816–835.

Charmaz, K. (2014). *Constructing grounded theory*. Thousand Oaks, CA: SAGE.

Complexity. (n.d.). In *Merriam-Webster's online dictionary*. Retrieved from http://www.learnersdictionary.com/definition/complexity

Currie, G., & Brown A. (2003). A narratological approach to understanding processes of organizing in a UK hospital. *Human Relations, 56*(5), 563–586.

Darden, L., & Maull, N. (1977, Mar.). *Interfield theory. Philosophy of Science, 44*(1), 43–64.

De Dreu, C. K. W., Nijstad, B. A., & van Knippenberg, D. (2008). Motivated information processing in group judgment and decision making. *Personality and Social Psychology Review, 12*(1), 22–49.

Department of the Army. (2005*). Army planning and orders production: The military decision making process* (FM 5-0). Retrieved from https://militaryscience.usu.edu/files/uploads/Cadet_Toolbox/Army_Planning_and_Orders_Production_FM_5-0.pdf

Department of the Army. (2012). *Mission command* (ADP 6-0). Washington, DC: Army Publishing Directorate. Retrieved from https://armypubs.us.army.mil/doctrine/index.html

Driskell, J. E., & Salas, E. (1991). Group decision making under stress. *Journal of Applied Psychology, 76*(3), 473–478.

Drucker, P. (1967, January). The effective decision. *Harvard Business Review*.

Dunford, R., & Jones, D. (2000). Narrative in strategic change. *Human Relations, 53*(9), 1207–1226

Eisenhardt, K. (1989). Making fast strategic decisions in high-velocity environments. *Academy of Management Journal, 32*(3), 543–576.

Fredrickson, J., & Mitchell, T. (1984). Strategic decision processes: Comprehensiveness and performance in an industry with an unstable environment. *Academy of Management Journal, 27*(2), 399–423.

Gioia, D., Corley, K., & Hamilton, A. (2012). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research, Methods, 16*(1), 15–31.

Glaser B. G., & Strauss A. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago, IL.

GlobalSecurity.org. (n.d.). Special Forces Operational Detachment Alpha (SFOD-A). Retrieved from http://www.globalsecurity.org/military/agency/army/a-team.htm

Helie, S., & Sun, R. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychology Review, 117*(3), 994–1024.

Jean, G. (2006). Army training evolving to develop better combat leaders. *National Defense Magazine.* Retrieved from http://www.nationaldefensemagazine.org/articles/2006/11/30/2006december-army-training-evolving-to-develop-better-combat-leaders

Kiel, D. L. (1995, May 18). *Chaos theory and disaster response management: Lessons for managing periods of extreme instability*. Paper presented at California Research Bureau, California State Library Invitational Conference, Sacramento, CA.

Lam, A. (2000). Tacit knowledge, organizational learning and societal institutions: An integrated framework. *Organization Studies, 21*(3), 487–513.

Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2013). Extreme outcomes sway risky decisions from experience. *Journal of Behavioral Decision Making, 27*(2), 146–156. doi:10.1002/bdm.1792

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–734.

McFall, L. (1987). Integrity. Ethics, 98(1), 5–20.

Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of “unstructured” decision processes. *Administrative Science Quarterly, 21*, 246–275.

Nous. (1973). In *The shorter Oxford dictionary on historical principles*. Oxford, UK: Oxford University Press.

Polanyi, M. (1958) *Personal knowledge: Towards a post-critical philosophy*. Chicago, IL: University of Chicago Press.

Polanyi, M. (1966). *The tacit dimension*. Garden City, NY: Doubleday & Co.

Pratt, M. G. (2000). The good, the bad, and the ambivalent: Managing identification among Amway distributors. *Administrative Science Quarterly, 45*(3), 456–493.

Riabacke, A. (2006, November). Managerial decision making under risk and uncertainty. *International Journal of Computer Science, 32*, 453–459. Retrieved from http://www.iaeng.org/IJCS/issues_v32/issue_4/IJCS_32_4_12.pdf

Rorty, R. (1979). *Philosophy and mirror of nature*. Princeton, NJ: Princeton University Press.

Rosen B., & Jerdee, T. H. (1977). Influence of subordinate characteristics on trust and use of participative decision strategies in a management simulation. *Journal of Applied Psychology, 62*(5), 628–631.

Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information and Communication Science, 33*(2), 163–180.

Ryle, G. (1945). Knowing how and knowing that. In *Proceedings of the Aristotelian Society* (Vol. 46 [1945–1946]; pp. 1–16).

Schmidt, F. L., & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science, 2*, 8–9.

SGI Quarterly. (1995, April). *The oneness of self and environment*. Retrieved from http://www.sgi.org/about-us/buddhism-in-daily-life/oneness-of-self-and-environment.html

Snowden, D. J., & Boone, M. E. (2007, November). A leader’s framework for decision making. *Harvard Business Review*.

Sonenshein, S. (2007). The role of construction, intuition, and justification in responding to ethical issues at work: The sensemaking intuition model. *Academy of Management Review, 32*(4), 1022–1040.

Thurlow, A., & Mills, J. (2009). Change, talk and sensemaking. *Journal of Organizational Change Management, 22*(5): 459–579.

Vroom, V., & Yetton, P. (1973). *Leadership and decision making*. Pittsburgh, PA: University of Pittsburgh Press.

Weick, K. E. (1979). *The social psychology of organizing*. Reading, MA: Addison-Wesley.

Weick K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: SAGE.

Weick, K. E., Sutcliff, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking and organizing. *Organization Science, 16*(4), 409–421.

Zand, D. E. (1972). Trust and managerial problem solving. *Administrative Science Quarterly, 17*(2), 229–239.

Fielding Better Combat Helmets to Deploying Warfighters

Robert F. Mortlock, COL, USA (Ret.)—managed defense systems acquisition efforts for the last 15 of his 27 years in the U.S. Army, culminating in his assignment as the Project Manager for Soldier Protection and Individual Equipment in Program Executive Office for Soldier. He holds a PhD in chemical engineering from the University of California, Berkeley, an MBA from Webster University, an MS in national resource strategy from the Industrial College of the Armed Forces, and a BS in chemical engineering from Lehigh University. He is a recent graduate from the Post-Doctoral Bridge Program of the University of Florida's Hough Graduate School of Business. [rfmortlo@nps.edu]

Abstract

The development, testing, and fielding of combat helmets for United States (U.S.) Soldiers offers project management (PM) professionals an opportunity to analyze how programs begin, how they progress through development and testing, and finally, how new capability is fielded within the U.S. Defense Acquisition Institution. The case study centers on the U.S. Army's adoption of the Enhanced Combat Helmet (ECH) for Soldiers stationed in war zones around the world. The case study is applicable broadly to project managers, business managers, engineers, testers, and logicians involved in project management within the private sector, as well as specifically targeting acquisition professionals within the Government Defense Departments. The case study can develop critical thinking and analysis skills in the areas of project initiation, stakeholder management and decision-making with ambiguous and contradicting testing/field data. The ECH case is in two distinct parts: Part I allows PM professionals to analyze how to initiate a program with an increased chance of success of meeting desired objectives. Part II allows PM professionals to analyze how to determine a procurement and fielding recommendation without stakeholder consensus and with ambiguous data.

Executive Summary

This Enhanced Combat Helmet (ECH) case study encourages critical analysis of a U.S. Department of Defense (DoD) project at two key decision points: project start and production. The case centers on the development, testing, and procurement (also referred to as acquisition) of the ECH for U.S. Army Soldiers and U.S. Marines. Two things make this case study particularly interesting. First is that key project stakeholders are passionate about helmets because they save lives in combat, and all Soldiers and Marines consider themselves subject matter experts on helmets—resulting in wide applicability. Second is that the key decisions involved with the ECH effort involved ambiguous data within a complex acquisition environment—requiring decision-making under uncertainty. The ECH case study reinforces critical thinking in uncertain environments, documents lessons learned for sound project management for future application, and provides wide private sector exposure to the complexities of public sector acquisition and helmet manufacture in particular. The following are the learning objectives for this case study:

- Develop the ability to critically analyze a project at key decision points—**critical thinking**.
- Identify key stakeholders and outline their contribution to the pending decision—**stakeholder management**.
- Develop alternative recommended strategies or courses of action for the decision-maker—**decision-making with uncertainty or ambiguous data**.
- Compare alternative strategies and identify decision criteria used for the comparison—**decision-making with uncertainty or ambiguous data.**

- Identify second-order considerations or consequences of the recommended strategies—**strategic management/leadership.**

Part 1 of the case focuses on the decision to initiate the ECH program. Guidance from the warfighting community and senior leaders was clear: The top priorities are maximum protection and weight reduction. Specifically, the ECH had to address the rifle threat, be fielded as quickly as possible, and reduce the weight on Soldiers and Marines in combat. In this part of the case, program management professionals can compare the various courses of actions developed for the initiation of an ECH effort with the actual ECH program. Valuable insights can be gleaned as lessons learned. It may be possible to avoid strategy pitfalls, and project management teams may be better able to manage cost, schedule, and performance trade-offs—and ultimately deliver capability more successfully. Questions to consider include the following:

- Who are the key stakeholders in the ECH program initiation decision and how does he manage their expectations?
- Would the ECH program be considered a "technology push" or "capability pull" program, and what are the implications?
- How should the ECH requirements be set? Should increased protection or weight reduction be emphasized? What is the right balance between reductions of Soldier load (combat weight) versus greater Soldier protection?
- How does the Army set testing protocols for the ECH prior to development and manufacturing of a helmet based on a new technology?
- What are the advantages and disadvantages of various acquisition approaches for the development, testing, procurement and fielding of the ECH? What are the criteria used to compare the alternative approaches?

Part 1 of the ECH case study offers key fundamental defense acquisition and program management lessons, which include the following:

- All programs are held to the constraints of cost, schedule, and performance. However, programs that involve the application of a new technology inherently include high levels of integration, manufacturability, producibility, and quality risk. These programs should guard against being primarily schedule-driven. Time is required to optimize the requirements and testing protocols and to allow the widest possible participation in the program by interested and innovative manufacturers. In this case, an effort that originally planned to field helmets within a year was seeking a production decision almost fours year later. The industrial base suffered as the program settled on a sole-source contracting strategy without the benefits of competition to keep costs and schedule in check. A program that is knowledge-driven from a research and development effort that includes many competitors from the industrial base may have proven more beneficial and had a similar actual schedule timeline.
- PMs, decision-makers, and senior leaders should be realistic about the risks associated with development efforts that are primarily schedule-driven rather than knowledge-driven.

Part 2 of the ECH case study focuses on the decision to actually procure and field the helmets to Soldiers and Marines despite the objections of the testing and medical communities. The Army and the Marine Corps approved urgent requirements based on combat operations and the need for increased protection against enemy rifle threats. After passing testing and four years after program initiation, the ECH was ready for a full rate production decision. Despite passing testing against the requirements, senior leaders faced a difficult decision because not all key stakeholders interpreted the test results similarly, raising significant safety concerns. Specifically, the testing and medical communities believed that the ECH was not operationally effective or operationally suitable for fielding and that the risk of injury to Soldiers and Marines was unacceptable—Soldiers and Marines wearing the ECH could suffer life-threatening skull fractures. Questions to consider include the following:

- Who were the key stakeholders and how would he manage their expectations?
- How does the Army balance the importance of development test data versus field data from helmets that were battle damaged? Should developmental test results or field data carry more weight in decision-making? How can the same development test data be interpreted differently by stakeholders?
- Are the concerns of the testing and medical communities warranted?
- How should the Army address these concerns with Congress, the media, and the American public?
- What are the advantages, disadvantages, and second-order implications of various courses of actions for the path forward? What are the decision criteria?
- How do you quantify benefits such as saving a Soldier's life and compare these benefits with long-term, potential health problems like concussions or muscle-skeletal neck injuries from the weight of helmets?

Part 2 of the case study offers key fundamental defense acquisition and program management lessons, which include the following:

- Test data can be interpreted differently by key stakeholders—leading to ambiguity in the decision-making process. The PM is in a position to understand not only the business side of the project (cost and schedule), but also the engineering side of the project (technology, testing, and risks). With this knowledge, the PM needs to try to reduce the uncertainty associated with the test data and present an interpretation in an unbiased, rational manner.
- The extension of test data obtained in a controlled test environment to relevance in an operational setting needs to be viewed with caution about its applicability and viewed from the proper perspective—from the perspective of the ultimate customer, in this case the warfighter.
- The cost constraints of projects should not be minimized, which is particularly hard to do in schedule-driven projects with urgent requirements.
- The recommendation is easier for the decision-maker if all the stakeholders are engaged early and often in the process, if their concerns are addressed, and if they have some ownership and buy-in in the path forward. The PM is key to making this happen successfully through effective leadership and communication.

Enhanced Combat Helmet (ECH) Case Study

Coverage of returns of battle-damaged helmets (as shown in Figure 1) to Soldiers is a good news story. It underscores the importance of Soldier protective equipment for combat effectiveness and Soldier force protection. Soldiers are wearing the very best combat helmets that industry can produce. However, the efforts to modernize helmets, and all protective combat gear for that matter, face the same defense acquisition challenges that all programs within the DoD face: a complex, bureaucratic Defense Acquisition Institution. The accelerated pace of technology innovation, rapidly evolving threats, and declining defense budgets make program management within the DoD challenging but even more critical than ever. Defense acquisition operates in an uncertain, complex, and ambiguous environment but maintains a simple focus: develop, procure, and field advanced warfighting capability to Soldiers to enable technological superiority on the modern battlefield.

Note. The photo on the right is a photo taken by U.S. Army Program Executive Office (PEO) Soldier of a battle-damaged helmet returned to a Soldier in a ceremony at Fort Belvoir in 2016. In news coverage entitled "U.S. Army Soldier Reunited With Equipment That Saved His Life in Afghanistan," the reporter covers the Soldier's description of how his helmet saved his life. The photo on the left is another photo taken by PEO Soldier of a recovered helmet damaged by enemy fire in Afghanistan.

Figure 1. Why Are Stakeholders So Passionate About Helmets?
(Aubert, 2016)

Current Situation, Summer 2013

Monday Morning Project Management Office Staff Meeting

Chief Engineer, Project Office for Soldier Protection and Individual Equipment (PM SPIE): "Sir, we have an Enhanced Combat Helmet (ECH) update. We just learned that Director, Operational Test & Evaluation (DOT&E) sent Congress the ECH Beyond Low Rate Initial Production (BLRIP) Report, and recommended that the Army not buy or field the ECH. The report says the unit cost is too high and that Soldiers wearing the ECH would have an unacceptably high risk of dying from excessive backface transient deformation from threat bullets."

Project Manager, Soldier Protection and Individual Equipment (PM SPIE): "Hmm … that puts us right in the middle between the warfighters and the operational testers. Both Army Senior leaders and Congress rely on the independent assessment of DOT&E for good reasons. DOT&E has a lot of influence."

Chief Engineer: "Yes sir. Also, DOT&E received concurrence from the Army Surgeon General with their assessments and recommendations."

PM SPIE: "So, after a four-year joint development and testing effort with the USMC in which the ECH finally passed its requirements, now we have to get an Army decision on whether to buy and field the ECH against the recommendations of the testing and medical communities, who have legitimate safety concerns?"

Chief Engineer: "Yes sir. However, the warfighters and Army combat developers have been very involved in this effort, and they remain adamant that the ECH should be fielded to deploying Soldiers. The requirement remains over 35,000 helmets. The USMC is strongly in favor of buying and fielding the ECH as well."

PM SPIE: "What's the funding situation?"

Chief Engineer: "We have over $35 million in operations & maintenance (O&M) overseas contingency operations (OCO) funding reserved for the buy that must be obligated by the end of the fiscal year (FY)."

PM SPIE: "Okay. Well, you know the drill. DOT&E probably already has the ear of the Army Acquisition Executive (AAE). Because the ECH was a wartime directed requirement with high visibility, the AAE is the Milestone Decision Authority (MDA). Let's get together a solid briefing to review, and let's start scheduling the pre-briefs to the AAE staff. Also, we need to be prepared to provide the congressional committees an update with the Army's decision. There are many stakeholders involved with the ECH, and some will not be happy. So, we need to think about how this will play out with the media and senior leaders from all the stakeholders with a solid strategic communications plan."

Background

The protection of American Soldiers in combat remains a top priority for senior leaders in the U.S. Army, the DoD, and Congress. The DoD has committed considerable resources and funding over the years in research & development, resulting in advanced materials and manufacturing processes. These investments have paid off. The American Soldier going into battle today has technologically advanced, rigorously tested combat equipment. Soldiers know that their combat equipment works as intended. In the end, this increases the combat effectiveness of the Soldiers and their units. One can consider the force protection of Soldiers as a layered approach. The outer force protection layer for Soldiers is situational awareness. The middle force protection layer is concealment. The inner force protection layer is personal protective equipment, like helmets, eyewear, and ballistic vests with ceramic plate inserts. This case study centers on combat helmets, which provide Soldiers skull and brain protection against both ballistic threats (i.e., bullets) and blunt impact forces, and prevent mild traumatic brain injury and concussions.

Army Combat Helmet Evolution

Figure 2 graphically displays the evolution of Army combat helmets and shows the tradeoff between increased performance and cost over time (Mortlock, 2014). The combat helmets that Soldiers wear into battle show a constant improvement in performance over time. This improvement in performance has been the result of advances in material research and manufacturing techniques. Soldiers wore the M1 helmet, nicknamed the "steel pot," from the 1940s through the late 1970s. The M1 provided ballistic protection largely because steel is hard. The M1 helmet consisted of a pressed manganese steel shell with a webbing suspension that Soldiers fitted to their heads. However, the M1 helmet was heavy and uncomfortable, and it provided little blunt trauma protection.

Advances in material research provided the opportunity to increase ballistic protection at a reduced weight. The maturation of ballistic fabrics based on para-aramid polymer technology enabled the Army to replace the M1 with the Personnel Armor System for Ground Troops (PASGT) helmet in the mid-1980s. The PASGT helmets were in the 3–4 pound range (lighter than the M1) and provided increased ballistic protection. The shell of the helmet consisted of layers of ballistic aramid fabric, the most famous of which is DuPont's Kevlar®. Thus, the PASGT was nicknamed simply "Kevlar" or "K-pot." The ballistic aramid technology allowed helmets to provide not only fragmentation protection from explosions but also small caliber hand gun protection at a reasonable weight. Eventually, the Modular Integrated Communication Helmet (MICH) replaced the PASGT helmet on a limited basis. By the mid-2000s, the Advanced Combat Helmet (ACH) was the Army's primary helmet. The basis for both the MICH and ACH is para-aramid polymer technology. These helmets provided Soldiers important performance improvements like reduced weight and better blunt impact protection through an interior suspension system using foam pads versus webbing.

In the late 1990s and early 2000s, the U.S. Army Research Lab, the U.S. Army Research Development and Engineering Command, and commercial industry teamed to mature the next generation of ballistics materials, resulting in the development of high molecular weight polyethylene (HMWPE) ballistics fibers that could be weaved into fabrics with application to combat helmets. HMWPE are polymer materials with different performance characteristics than para-aramid polymer materials. Para-aramids are a thermoset polymer, which means that above certain temperatures the polymer breaks down, loses its properties and cannot be remolded back into its original state when cooled. On the other hand, HMWPE are thermoplastics, which means that above a certain temperature the polymer breaks down but it can be remolded into its original state when cooled. The application of HMWPE fiber material in helmets created the misperception that helmets made with HMWPE materials might easily lose their form under ballistic events and potentially jeopardize Soldiers' safety. Ultimately, the advantages of HMWPE helmets for reduced weight and greater ballistic capability outweighed this concern. The basis of future Army helmets, both the ECH and its eventual replacement, the Soldier Protection System (SPS) Integrated Head Protection System (IHPS), is HMWPE technology.

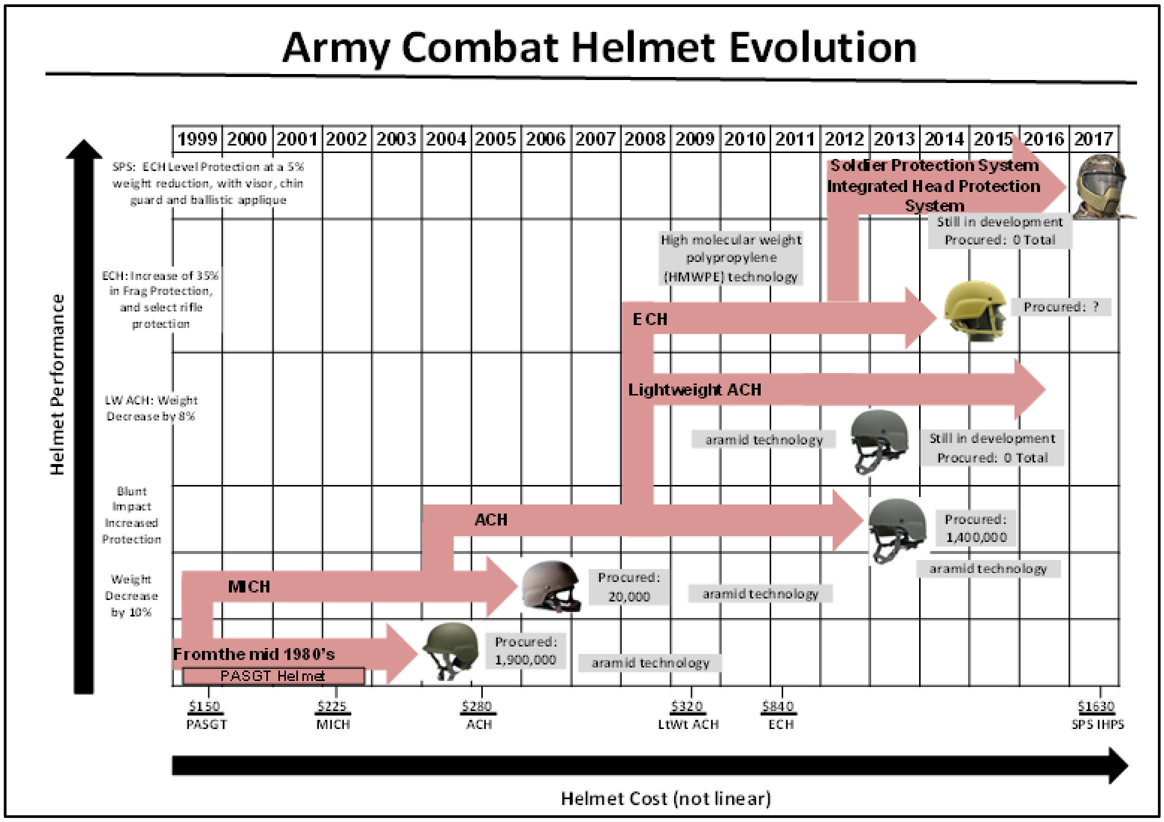


Figure 2. Evolution of Combat Helmets
(Mortlock, 2014)

Helmet Testing Basics[1] (Zheng, 2002)

The Army rigorously tests combat helmets worn by Soldiers against protocols to ensure they conform to stringent requirements to protect Soldiers against both blunt trauma and ballistic threats. Three ballistic properties particularly important for describing impacts to helmets are complete penetration (the bullet goes completely through the helmet), partial penetration (the bullet does not go completely through the helmet) and backface transient deformation (BTD—a measure for the amount the round's impact dents the helmet material; Zheng, 2002). The final performance of the helmet in testing and in combat depends both on the inherent properties of the materials used to develop the helmet and the processing techniques used to manufacture the helmet. Helmet requirements are performance-based. Each helmet manufacturer optimizes its design over time using a combination of materials (layers of polymer fibers woven into sheets with chemical binders) and different processes based on temperature, pressure, and time. The use of statistics is important in testing because testing simulates live combat, and the warfighter requires a high confidence that the helmets will perform as advertised. The testing must balance the need for statistical confidence with the costly and destructive nature of the testing.

---

[1] Refer to Appendix 2 for a helmet-testing tutorial.

Operational Field Data

As was presented previously, the Army collects battle-damaged helmets from Soldiers. Before returning them, the Army conducts forensic studies to better understand enemy threats and analyze the performance of the helmets to improve future designs. Figure 3 presents the data collected from combat operations in Iraq and Afghanistan from 77 helmets hit by small arms bullets (Mortlock, 2013). When the bullet completely penetrated the helmet, the Soldiers died nearly 75% of the time. When the bullets did not completely penetrate the helmet (partial penetration), the average permanent helmet deformation was about 9mm and the Soldiers all survived with relatively minor head/neck injuries and eventually returned to duty.

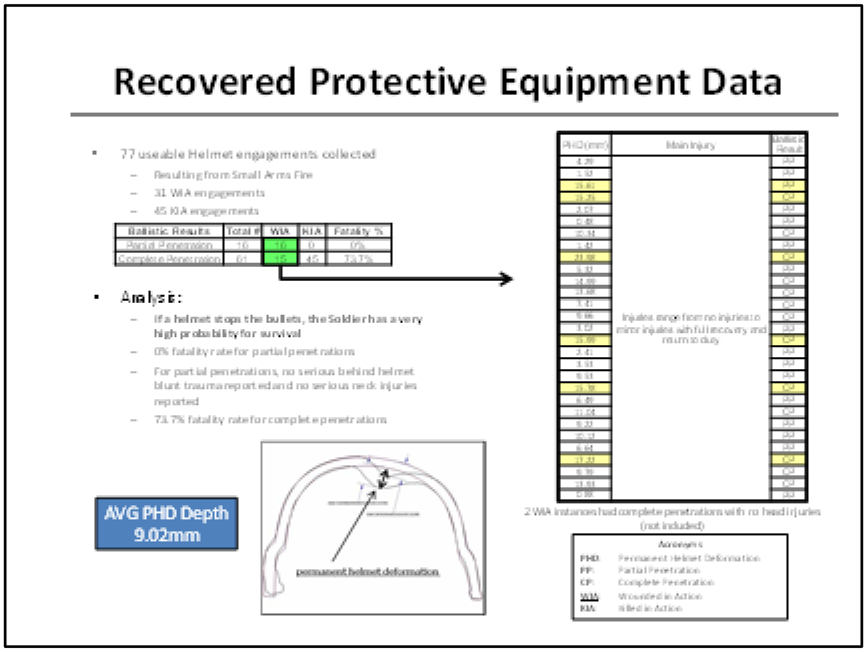


Figure 3. Recovered Protective Helmet Data
(Mortlock, 2013)

Part I: Project Initiation Decision, Early 2009 Timeframe

Colonel Bob Smith was recently assigned as the Project Manager (PM), Soldier Protection and Individual Equipment—the office responsible for developing, testing, procuring, and fielding helmets to Soldiers. Colonel Smith was a seasoned defense acquisition veteran with over 15 years of project management experience. During his preparation for this position, the guidance from the warfighting community and senior leaders was clear: the top priorities are maximum protection and weight reduction.

Colonel Smith was preparing for a key decision in the Pentagon regarding the start of a new helmet program, named the Enhanced Combat Helmet (ECH). Luckily, Colonel Smith's chief engineer for the program office was an armor expert, Dr. James Suchez. Dr. Suchez led the efforts to mature HMWPE technologies across the DoD and with commercial industry for the last decade. Dr. Suchez explained that the application of HMWPE to helmets allowed the Army to consider the following basic options for the new helmet requirements: (1) Maintain the protection levels of the current helmets with a reduced weight of up to 20% or (2) increase the protection levels but maintain (or increase) the weight of the helmet.

Colonel Smith knew that the Army senior leaders would rely on the advice and recommendations of the PM during the meeting. The final decision would be made by the Army Acquisition Executive (AAE). However, the AAE would likely turn to key stakeholders before making the final decision. The first stakeholder was the PM, and Bob was well prepared to discuss key considerations from a cost, schedule, performance, and technology perspective.

The second stakeholder was the warfighter representative, also called the “user” representative. The warfighter representative was a crusty old officer named Colonel Billy Johnson from Fort Benning, home of the U.S. Army Maneuver Center of Excellence. Colonel Johnson spent most of his time in the Army leading Soldiers in combat. Colonel Johnson took his job seriously as the approver of the requirements. He was passionate about the possibility for a new helmet because he believed that the current helmets were too heavy and uncomfortable. He also represented the warfighters currently downrange in combat, and was under extreme pressure to approve requirements for a new helmet to protect Soldiers not only against fragmentation and handgun rounds, but also against enemy rifle threats. Another key stakeholder was Colonel Harry Crisp, the representative from DOT&E. Any new helmet development program would fall under DOT&E oversight to approve the testing protocols used to ensure the requirements were met. DOT&E would also provide an independent assessment of the helmet’s operational effectiveness and suitability for Soldiers to Army senior leaders and document that assessment in reports to Congress. Colonel Harry Crisp had years of experience as a tester and evaluator of Army systems. The importance, influence, and visibility of DOT&E’s independent assessment were increased by the recent congressional and public concerns calling into question the adequacy of Soldiers’ protective equipment.

Colonel Smith knew that each of the stakeholders was passionate about a new helmet program. He realized that his role as the PM was not to advocate for a new program but to give advice about the underpinning technological possibilities; additionally, he needed to lay out the cost, schedule, and performance implications of various strategies for the development, testing, and procurement of the new helmet.

Two important determinants of program success are requirements definition and alignment of those requirements against capability gaps. Simply put, poorly defined requirements will set a project’s initial trajectory that will be difficult to fix later in the development cycle. Project initiation can be the result of a need from the warfighters generically called *capability pull*. Alternatively, the project might be the result of an innovative new technology without a specific identified warfighting application generically called *technology push*. The question of technology push or capability pull at program initiation often delays efforts and creates perception challenges among key stakeholders. The ECH effort was driven by the urgent need for a new helmet to address protection for Soldiers against rifle threats in combat, and enabled by the maturation of HMWPE technologies. The ECH requirements must balance acceptable minimum risk versus maximum safety for protective equipment, and weight reduction (Soldier load) versus protection (ballistic and blunt force). Colonel Smith knew that this balance would not be an easy compromise for any of the stakeholders.

During the meeting hosted by the AAE, Colonel Johnson was adamant that the ECH had to address the rifle threat, be fielded as quickly as possible, and reduce the weight on Soldiers in combat. Colonel Smith laid out the basic options that he had discussed with Dr. Suchez; the ECH would not be able to address the rifle threat and also reduce the helmet weight. Colonel Johnson was not happy, and doubted the validity of the technology assessment. He stated that, just a week prior, he received an industry brief from a company that claimed they could develop a helmet at reduced weights that also addressed increased threats. Dr. Suchez, also in attendance, spoke up and said that it was not unusual for industry to make claims that they could not back up, and that the application of a new technology into helmets is technically challenging from a manufacturing perspective. “It’s one thing to produce a prototype helmet in a controlled laboratory,” he said, “but completely different to produce many helmets from a manufacturing line that consistently perform against rigorous testing requirements.”

To address the schedule aspect of the program, Colonel Smith next laid out the options of pursuing a program of record (PoR) through the deliberate acquisition process versus pursuing a rapid acquisition process supported by a directed or urgent requirement. Establishing formal ECH PoR would involve a four-year time period. Year 1 would allow refinement, analysis, and approval of formal requirement documents and the development of testing protocols. Year 1 would also allow the Army to request development and procurement funding from Congress in the Army’s base budget for the program. Years 2 and 3 would involve development and testing of ECH prototypes resulting from competitively awarded contracts (probably cost-plus type contracts) to be awarded to industry companies. Year 4 would allow the Army to award procurement contracts to the successful companies for the manufacture and production of ECHs. Again, Colonel Johnson was not happy that it would take four years to get the ECH to Soldiers. The alternative to a PoR was to use the rapid acquisition process. In rapid acquisition, the Army could write a directed requirement (probably within a month) for the ECH, and the Army could award competitive contracts (probably fixed-price contracts for certain quantities with production options) to industry within six months. A rapid acquisition effort could be funded with overseas contingency operations (OCO) money, which was limited to procurement money and no development money. Another six months would be required to test the helmets. So, ECHs could be on Soldiers in just over a year. Colonel Johnson was much happier with the second strategy. But, Colonel Crisp was quick to point out that for the rapid acquisition options, the ECH requirements would not be underpinned by analysis, and the test protocols would have to rely on the protocols for current helmets because there would be no time to develop test protocols specifically for the ECH. Colonel Crisp noted this was particularly important for the ECH, which would rely on thermoplastic polymers. The ECH based on HMWPE might perform much differently than the current para-aramid based helmets. For example, ECHs might lose their rigidity after being shot once and offer much less protection from multiple shots. Also, the ECH may deform excessively, leading to head trauma and skull fractures. There were legitimate testing and safety concerns that would have to be addressed.

Colonel Smith tried to remain neutral. Both strategies had advantages and disadvantages. Decision-making involves defining and analyzing alternative approaches. It came down to the level of risk the Army was willing to accept. The ECH project initiation decision also encompassed setting future funding levels and procurement quantities, as well as addressing industrial base concerns, competition, and testing implications. From past experience, Colonel Smith understood that stakeholder management would be key to the success of the ECH program and that proper communication and collaboration increased the chances of program success.

The AAE was pleased with the frank dialogue between the key stakeholders and stated that enough information was presented for an informed decision on whether or not to initiate the ECH program. Before prioritizing resources for the ECH program, the ECH would need to be considered through the lens of the defense acquisition institutional framework presented in Figure 4 (refer to Appendix 1 for a description of the U.S. Defense Acquisition Institution). The PM has cost, schedule, and performance responsibilities, and manages the effort with the Defense Acquisition Management System. The PM's official chain of command is in the executive branch, but the PM also reports to Congress with program status updates and works through contracts with industry. The requirements generation system provides requirements and the resource allocation system provides funding. Depending on the program, the public and media perceptions may be important considerations.

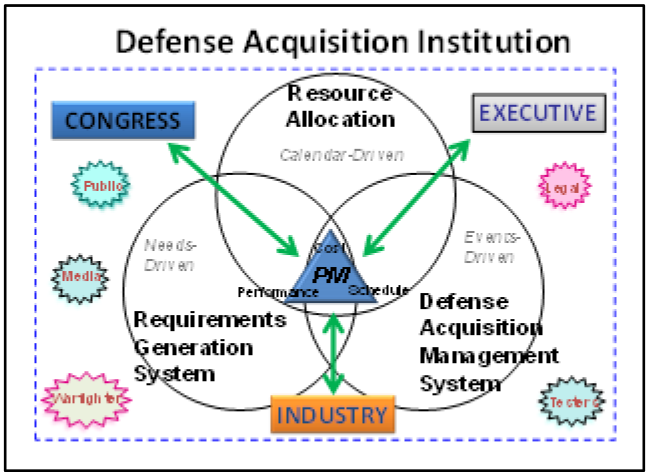


Figure 4. U.S. Defense Acquisition Institution

The ECH Program

The ECH program began in early 2009 ("Marine Corps Notice," 2009). Figure 5 depicts the ECH program timeline. The Army and the U.S. Marine Corps (USMC) approved urgent requirements based on combat operations and the need for increased protection against enemy rifle threats. The overseas contingency operations (OCO) account funded the ECH program. The acquisition procurement objectives were set based on the predicted numbers of deploying Soldiers. The Army and USMC set broad requirements to include a 35% increase in fragmentation protection, increased 9mm pistol protection, and rifle threat protection—all at the same weight of the ACH (Mortlock, 2013). The acquisition strategy was a single-step development in which competition was encouraged among industry manufacturers. The original request for proposal asked for each ECH vendor to deliver test data validating their claim that their ECH design met the combat helmet test protocols used at the time for the ACH and the new ECH requirements for rifle protection. Four vendors submitted proposals. However, only one vendor's design was acceptable. At the end of 2009, this vendor received a contract to produce ECHs to undergo government developmental testing with contract options for production deliveries after successful first article tests (FATs). In late 2010, after successful developmental testing, the Army and USMC approved the Milestone C to enter into low rate initial production (LRIP) with the selected vendor ("Enhanced Combat Helmet," 2009; "Marine Corps Notice," 2009). The LRIP decision permitted the production of a small number of helmets to undergo testing in order to validate that the contractor could successfully produce the helmets to performance requirements.

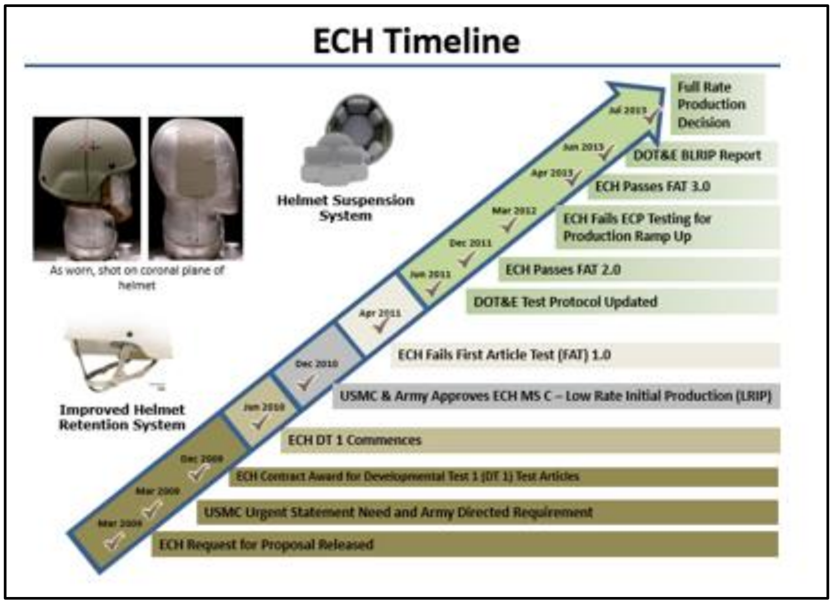


Figure 5. ECH Program Timeline

(Cox, 2013; "Enhanced Combat Helmet," 2009; "Marine Corps Notice," 2009; "New Helmets," 2010)

In late 2011, the ECH passed the second round of FAT. To meet an aggressive production schedule for the Army and USMC, the ECH vendor submitted an engineering change proposal for a second and third production line. It would take all of 2012 for the vendor to successfully pass the third round of FAT for all three production lines after working through issues between test sites, U.S. Army Test Center and National Institute Justice laboratories, as well as issues with the source of rifle rounds (Winchester vs. Hornady vs. Remington).

The ECH FAT results demonstrated that the ECH met its requirements and offered Soldiers the potential for greater protection compared to the ACH. Against a requirement for a 35% increase in fragmentation protection compared to the ACH, the ECH demonstrated an average increase of 53% (Mortlock, 2014). For the 9mm BTD requirements, the ECH demonstrated an average increase in performance of 10% over the ACH performance (Mortlock, 2014). Finally, against the chosen test rifle threat, the ECH demonstrated an over 153% increase in protection over the ACH for resistance to penetration (Mortlock, 2014). Of note is the fact that there was no BTD requirement against rifle threats for the ECH. The legacy ACH 9mm BTD requirements were too restrictive for rifle threats, and there was no basis to assign 9mm BTD requirements to rifle threats without injury data, which does not exist. To avoid jeopardizing the program due to unachievable or unrealistic requirements, rifle BTD testing occurred for government reference purposes only.

Part II: ECH Procurement and Fielding Decision, Summer 2013

After passing FAT and four years after program initiation, in summer 2013, the ECH was ready for a full rate production (FRP) decision, after which the ECH would be produced to the approved acquisition objective quantity. Each production lot of helmets would undergo lot acceptance testing (LAT) to verify continued compliance to specification requirements. After passing LAT, the Army could field helmets to deploying Soldiers.

The FRP decision would involve significant procurement money to buy and field the ECH. Despite FAT results in which the ECH demonstrated superior performance against the requirements over current helmets, Army leaders, specifically the AAE, faced a difficult decision. Not all key stakeholders interpreted the test results similarly, raising significant safety concerns for Soldiers. The DOT&E issued a congressionally mandated Beyond LRIP Report recommending that the ECH not be fielded to Soldiers. DOT&E believed that the cost per helmet (roughly 2.5 times the current helmet) did not justify the minimal performance increase (DOT&E, 2011, 2012, 2013). DOT&E was also concerned that the Army did not test the ECH against the most stressing or most prevalent enemy rifle threats. More importantly, DOT&E stated that Soldiers wearing the ECH in combat would face an unacceptable risk of head injuries due to excessive backface deformation caused by rifle rounds. The medical community, through the Army Surgeon General, supported the DOT&E recommendations. These concerns put the AAE in a difficult position. To further complicate matters, the AAE had just spoken to DOT&E, who emphatically stood behind their recommendation.

Again, the AAE convened the same Council of Colonels that had met four years earlier to discuss the decision to initiate the ECH program. Colonel Smith admitted the ECH program had not met the original timelines, but emphasized that the ECH had finally successfully passed testing and met its performance requirements. Colonel Smith also stressed that $35 million was at risk if the procurement decision passed the end of the fiscal year, which was nearing. Colonel Crisp noted that he understood the program history well and understood the challenges. However, in DOT&E’s opinion, the ECH was not operationally effective or operationally suitable for fielding to Soldiers. The risk of injury to Soldiers was unacceptable; in DOT&E and the Army Surgeon General’s opinion, Soldiers wearing the ECH could suffer life threatening skulls fractures from excessive BTD from threat rifle rounds. Additionally, Colonel Crisp noted that the ECH was not tested against the most stressing threats, bringing in question the validity of the requirements. Colonel Johnson was livid that there was even a question about the requirements. All stakeholders had agreed to the original requirements more than four years ago. Everyone had accepted the program risks. Now, three years later than planned, when the ECH finally passed testing, concerns were raised. Colonel Johnson stated that the warfighter community strongly recommended getting the ECH to Soldiers as quickly as possible.

Colonel Smith again tried to remain neutral to avoid the appearance that the PM was biased toward buying the ECH. However, he was compelled to provide the complete picture to the AAE for the most informed decision. His program office was also charged with the collection and analysis of battle-damaged helmets from Soldiers who had been shot in the head while wearing their helmets. Analysis of those helmets indicated that no Soldiers had died or suffered major injuries as a result of excessive backface deformation of the helmet. The average deformation observed was 35% of the 9mm BTD required of 25.4 mm (or coincidentally, exactly 1 inch). Colonel Crisp interrupted and stated that DOT&E placed no value on the results because they were not statistically robust, and were not done under strict testing conditions where the variables were controlled. Colonel Crisp also pointed out that the government’s own reference testing indicated that the BTD observed from the test rifle threat was 18% to 89% higher than the 9mm BTD requirement. Colonel Smith concurred with those numbers but indicated that he was not finished presenting the rest of the field data results, which indicated that nearly 74% of Soldiers died if the threat round completely penetrated the helmet. Again, Colonel Crisp dismissed that data, and again brought up the rifle threat round used in testing. Colonel Johnson asked a question about the operational safety margin built into the testing. Colonel Smith replied in the affirmative that the chosen rifle round was fired at the ECH at muzzle velocity and at 0% obliquity, operationally providing Soldiers a safety margin, because in combat, rounds are fired at considerable distance, slowing down in flight and striking at non-direct angles. Therefore, even though the chosen test round was not the most stressing rifle threat round, the ECH still provided considerable protection and 153% more protection from penetration than the current helmet against the rifle threat.

The AAE realized that the meeting of the Council of Colonels was probably at a point of agreeing to disagree. The AAE understood each of the positions clearly and thanked everyone for their candid and articulate input. In light of the data presented, should the Army buy and field the ECH?

References

Aubert, A. (2016, April 19). *U.S. Army soldier reunited with equipment that saved his life in Afghanistan* [Audiovisual media]. Aired on ABC7. Washington, DC.

Cox, M. (2013, July 30). Army, Marines to field better ballistic helmets. Retrieved from http://www.military.com

Director, Operational Test and Evaluation (DOT&E). (2011). Enhanced combat helmet. In *Director, Operational Test and Evaluation FY 2011 annual report* (pp. 123–124). Washington, DC. Retrieved from http://www.dote.osd.mil/pub/reports/fy2011/

Director, Operational Test and Evaluation (DOT&E). (2012). Enhanced combat helmet. In *Director, Operational Test and Evaluation FY 2012 annual report* (pp. 151–152). Washington, DC. Retrieved from http://www.dote.osd.mil/pub/reports/fy2012/

Director, Operational Test and Evaluation (DOT&E). (2013). Enhanced combat helmet. In *Director, Operational Test and Evaluation FY 2013 annual report* (pp. 171–172). Washington, DC. Retrieved from http://www.dote.osd.mil/pub/reports/fy2013/

Enhanced combat helmet goes back to research and development test. (2009, December 7). *Inside the Navy*. Retrieved from http://insidedefense.com

Marine Corps notice on combat helmet. (2009, February 27). *Inside Defense*. Retrieved from http://insidedefense.com

Mortlock, R. F. (2013, August 1). *ECH decision briefing* [PowerPoint presentation]. Briefing prepared for Army Senior Leadership. Washington, DC: Program Executive Office Soldier, Program Manager Soldier Protection and Individual Equipment.

Mortlock, R. F. (2014, February 13). *Body armor information paper* [Memorandum]. Fort Belvoir, VA: Program Executive Office Soldier, Program Manager Soldier Protection and Individual Equipment.

New helmets could be deployed in late FY11 if testing successful. (2010, August 13). *Inside the Navy*. Retrieved from http://insidedefense.com

Zheng, J. Q. (2002, October). *Ballistic testing basics* [PowerPoint presentation]. Fort Belvoir, VA: Program Executive Office Soldier, Program Manager Soldier Protection and Individual Equipment.

Appendix 1. U.S. Defense Acquisition Institution—Decision Framework

Within the DoD, the development, testing, procurement, and fielding of capability for the warfighter operates within a decision-making framework that is complex. Within the private sector, similar frameworks exist. The U.S. Defense Acquisition Institution has three fundamental support templates that provide requirements, funding, and management constraints. The executive branch, Congress, and industry work together to deliver capability with the program manager (PM) as the central person responsible for cost, schedule, and performance.

The government PM is at the center of defense acquisition, which aims to deliver warfighter capability. The PM is responsible for cost, schedule, and performance (commonly referred to as the “triple constraint”) of assigned projects—usually combat systems within the DoD. The executive branch of government provides the PM a formal chain of command in the DoD. The PM typically reports directly to a program executive officer, who reports to the Service Acquisition Executive (an assistant secretary for that service—either Army, Navy, or Air Force), who reports to the Defense Acquisition Executive (the Under Secretary of Defense for Acquisition, Technology and Logistics). Depending on the program’s visibility, importance and/or funding levels, the program decision authority is assigned to the appropriate level of the chain of command.

Programs within Defense Acquisition require resources (for funding) and contracts (for execution of work) with industry. Congress provides the resources for the defense programs through the annual enactment of the Defense Authorization and Appropriation Acts, which become law and statutory requirements. The PM, through warranted contracting officers governed by the Federal Acquisition Regulations, enters contracts with private companies within the defense industry. Other important stakeholders include actual warfighters, the American public, the media, and functional experts (like engineers, testers, logisticians, cost estimators, etc.), as well as fiscal and regulatory lawyers.

As a backdrop to this complicated organizational structure for defense PMs, there are three decision support templates: one for the generation of requirements, a second for the management of program milestones, and a third for the allocation of resources. Each of these decision support systems is fundamentally driven by different and often contradictory factors. The requirement generation system is driven primarily by a combination of capability needs and an adaptive, evolving threat. The resource allocation system is calendar-driven by Congress writing an appropriation bill—providing control of funding to the Congress and transparency to the American public and media for taxpayer money. The Defense Acquisition Management System is event-driven by milestones based on commercial industry best practices of knowledge points and off-ramps supported by the design, development, and testing of the systems as technology matures. Often integration and manufacturing challenges occur.

Appendix 2. Helmet Testing Basics (Zheng, 2002)

The Army rigorously tests combat helmets worn by Soldiers against protocols to ensure they conform to stringent requirements to protect Soldiers against both blunt trauma and ballistic threats. Typical battlefield ballistic threats include fragments from explosive devices and bullets from handguns and rifles. Within the DoD, System Threat Assessment Reports document relevant and existing helmet threats and these threats are validated by the National Ground Intelligence Center. With respect to fragmentation, the Army Research Laboratory proved that five fragment simulators represent 95% of the range of threat fragments Soldiers expect to face from exploding munitions. Fragment threats used in testing include the 2, 4, 16, and 64 grain right circular cylinders, as well as 17 grain fragment simulating projectile. Handgun threats include the 9 mm full metal jacket 124 grain, .357 Sig full metal jacket 125 grain, and the 44 Mag 240 grain (Mortlock, 2014). These threats are defined by the National Institute of Justice (NIJ). Rifle threats include eight different rounds to include 5.45 mm, 5.56 mm, and 7.62 mm rounds (both armor piercing and non-armor piercing varieties; Mortlock, 2014).

Helmet testing is a form of destructive testing because the helmets are non-recoverable after the testing. Generally, testing can focus on physical properties (like density or melting point), mechanical properties (like tensile strength or impact strength), and ballistic properties. Three ballistic properties particularly important for helmets are complete penetration (the bullets goes completely through the helmet), partial penetration (the bullet does not go completely through the helmet), and backface transient deformation (a measure for the amount the round's impact indents the helmet material).

Depending on the materials selected and the manufacturing process, each helmet will demonstrate a ballistic testing curve, represented in Figure 6. The frequency of complete penetration can be plotted against the striking velocity of the round. A striking velocity of V_0 is the highest velocity at which no rounds completely penetrate the helmet shell. A striking velocity of V_{100} is the velocity at which all rounds completely penetrate the helmet shell. The V_{50} striking velocity represents the velocity at which 50% of the rounds completely penetrate and 50% partially penetrate the helmet (Zheng, 2002). Figure 6 labels the zones of variation and non-variation. The variation zone represents a performance area for the helmet in which the helmet may provide the different levels of protection but demonstrate the same V_0 and V_{100} characteristics.

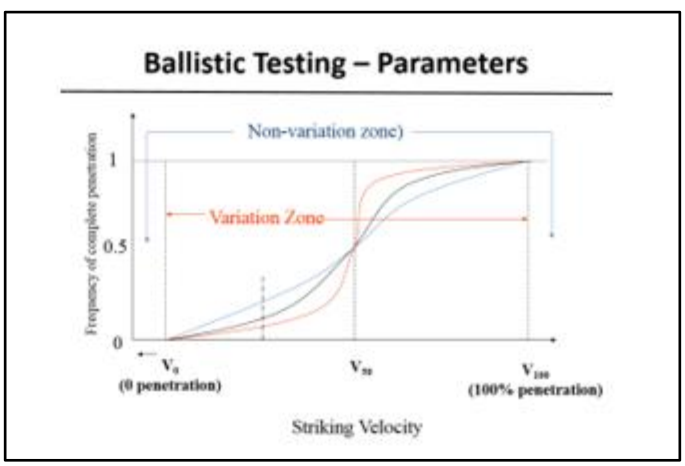


Figure 6. Ballistic Helmet Testing—Penetration
(Zheng, 2002)

V_0 is the “protection parameter” because it identifies the warfighter’s guaranteed protection level. It is an important parameter in production quality and control; however, it does not completely measure material performance and depends greatly on the production process. Generally, helmet manufacturers want to make the actual V_0 demonstrated by a helmet higher than the V_0 required to ensure a helmet passes testing (see Figure 7). V_{50} is the “material parameter” because it does not represent a guaranteed level of protection but is important in the optimization of the helmet design. There is a unique V_{50} for each helmet design. Generally, the design goal is to make V_{50} as high as possible and as close to V_{100} as possible.

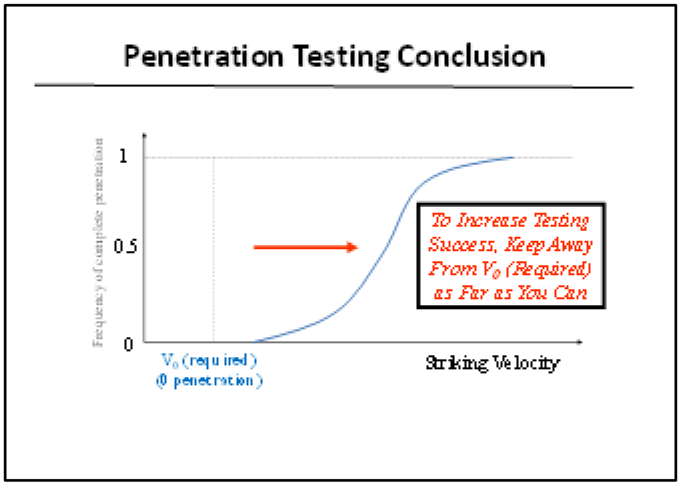


Figure 7. Penetration Testing Conclusion
(Zheng, 2002)

During ballistic testing, if a bullet only partially penetrates the helmet, testers measure the backface deformation using calipers or laser techniques. The lower the deformation exhibited by a helmet in testing, the lower the potential for injuries to the wearer’s head. Figure 8 is a pictorial representation of a sample backface deformation measurement. After a series of tests, testers plot the observed backface deformations for a helmet. This results in a distribution of values around an average value (Figure 9). The lower the average measured backface deformation compared to the required value, the more protection the helmets offer and the greater the testing success rate for the design and manufacturer (Figure 9).

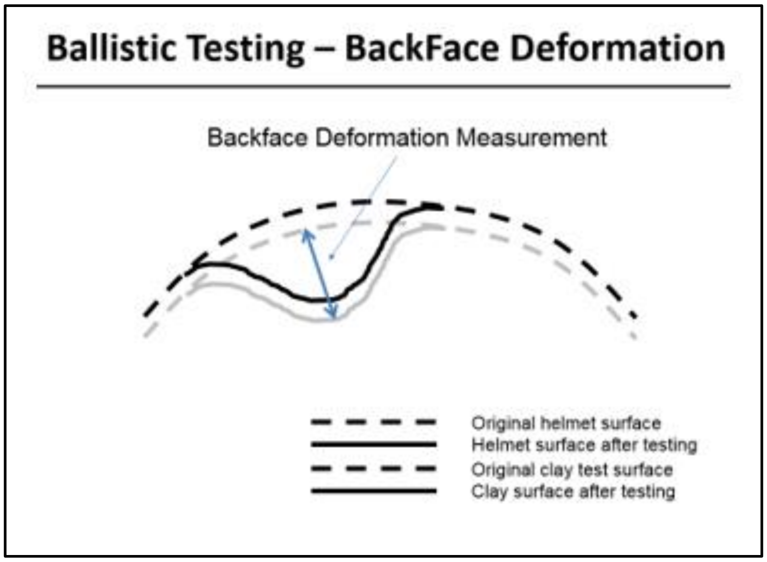


Figure 8. Ballistic Testing—Back Face Deformation
(Zheng, 2002)

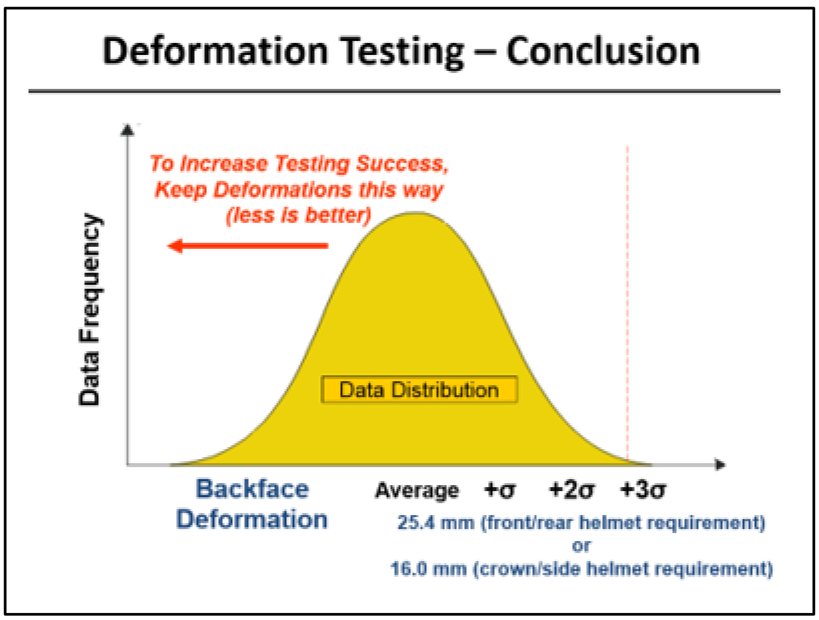


Figure 9. Deformation Testing—Conclusion
(Zheng, 2002)

There is an additional important point to understand about helmet testing with respect to battlefield operational relevance. In testing, the Army performs V_0 resistance to penetration and backface deformation testing with the threat rounds fired at the helmet at speeds representing threat weapon muzzle velocity and at angles of 0% obliquity. This represents a worst-case condition that is representative of extremely close combat scenarios. Under these conditions, the round strikes the helmet with the maximum force and the highest chance for penetration, but during combat, the enemy fires at various distances from their targets. Over these distances, bullets slow down and strike their intended targets at various angles. Therefore, in combat, bullets strike Soldier helmets at speeds significantly lower than muzzle velocity speeds and from non-perpendicular angles.

Panel 21. Considerations in Accelerating Technology Adoption in Defense Acquisitions

Thursday, May 10, 2018	
1:45 p.m. – 3:15 p.m.	**Chair: Rear Admiral Michael Haycock, USCG,** Assistant Commandant for Acquisition and Chief Acquisition Officer ***Applying a DEVelopment OPerationS (DevOps) Reference Architecture to Accelerate Delivery of Emerging Technologies in Data Analytics, Deep Learning, and Artificial Intelligence to the Afloat U.S. Navy*** CAPT Kurt Rothenhaus, USN, PEO C4I CDR Kristine De Soto, USN, PEO C4I Emily Nguyen, PEO C4I Jeffrey Millard, PEO C4I ***Using GAO's Technology Readiness Guide*** Shelby Oakley, Government Accountability Office Cheryl Andrew, Government Accountability Office ***Conducting Viability Assessments for Acquisition Planning*** Virginia L. Wydler, The MITRE Corporation Erin M. Schultz, The MITRE Corporation

Rear Admiral Michael Haycock, USCG—currently serves as the U.S. Coast Guard's Assistant Commandant for Acquisition and Chief Acquisition Officer (CAO). As CAO he directs efforts across all Coast Guard acquisition programs, acquisition support, personnel, finance, and research and development activities to execute the service's acquisition portfolio. Prior to reporting as CAO, he served as Program Executive Officer and Director of Acquisition Programs, where his duties included management oversight of all Coast Guard acquisition programs and projects for the modernization and recapitalization of surface, air, command and control, and logistics assets.

Before that, RADM Haycock was the Coast Guard's Assistant Commandant for Engineering and Logistics, and Chief Engineer. In that role, he oversaw all naval, civil, aeronautical, and industrial engineering and logistics for the Coast Guard's 23,000 facilities, 230 ships, 1,800 boats, and 200 aircraft. He also served as Commander of the Surface Forces Logistics Center, where he managed Depot Level Maintenance, Engineering and Supply Chain Support, and Technical Information Services necessary to maintain the Coast Guard's 1.5 million operational mission hours each year.

As CAO, Haycock leads an organization where he previously held positions as the Surface Domain Deputy Project Manager and the National Security Cutter Project Manager. RADM Haycock received his Level III DHS Program Manager Certification in January 2008. RADM, a permanent Cutterman with over 11 years of sea service, has served afloat as Damage Control Assistant aboard USCGC *Polar Star* (WAGB-10), Assistant Engineer Officer aboard USCGC *Boutwell* (WHEC-719), Engineer Officer aboard cutters *Mellon* (WHEC-717), and *Polar Sea* (WAGB-11), and Executive Officer and later Commanding Officer of USCGC *Sherman* (WHEC 720).

RADM Haycock's other tours ashore include Maintenance and Logistics Command Pacific as the Icebreaker Support Section Assistant, Naval Engineering Support Unit Seattle as the Project

Manager for the Polar Console Renewal Project, and Commanding Officer of the Naval Engineering Support Unit Honolulu.

He is a native of Delavan, IL, and graduated from the United States Coast Guard Academy in 1985 with a Bachelor of Science degree in Electrical and Electronics Engineering. He later earned a Master of Science degree in Naval Architecture and Marine Engineering and a Master of Science degree in Industrial and Operations Engineering from the University of Michigan in 1992.

RADM Haycock's military awards include the Legion of Merit, Meritorious Service Medal with Operational Device, Coast Guard Commendation Medal with Operational Device, Coast Guard Achievement Medal, Antarctica Service Medals, Arctic Service Medal, and several other unit and service awards.

Applying a DEVelopment OPerationS (DevOps) Reference Architecture to Accelerate Delivery of Emerging Technologies in Data Analytics, Deep Learning, and Artificial Intelligence to the Afloat U.S. Navy

CAPT Kurt Rothenhaus, USN—serves as the Program Manager of the PMW 160 Tactical Networks Program Office at PEO C4I in San Diego, CA. CAPT Rothenhaus holds a PhD in software engineering from the Naval Postgraduate School and has served as an Engineering Duty Officer since 2003. Previous assignments include service as Commanding Officer of SPAWAR System Center Pacific, Combat Systems/C5I Officer on USS *Harry. S. Truman* (CVN 75), and Deputy Program Manager for PMW/A 170 (Navy Communications and GPS Program Office). [kurt.rothenhaus@navy.mil]

CDR Kristine "Kris" De Soto, USN—currently serves as the Assistant Program Manager (APM; CANES Development) in PMW160 CANES Development. CDR De Soto was commissioned in 1999, following graduation from The Pennsylvania State University with a BS degree in computer science. CDR De Soto earned an MS in systems engineering with an emphasis in net-centric systems from the Naval Postgraduate School in 2009. Previous assignments include Combat Systems Officer on USS *John C. Stennis* (CVN 74), Protected Communications Manager Navy Multiband Terminal APM in PMW/A 170 Communications Program Office, and Littoral Combat Ship Integration Platform Manager in PMW 760. [kristine.m.desoto@navy.mil]

Emily Nguyen—serves as the Assistant Program Manager for PMW 160 Agile Core Services. Nguyen received her Bachelor of Science in computer science engineering from the University of Colorado and subsequently earned a Master of Business Administration at San Diego State University. Nguyen started as a software developer for Science and Technology projects sponsored by the Office of Naval Research and later transitioned one of the efforts to Marine Corps Systems Command which resulted in her role as the Project Manager at the Rapid Response Integration project providing agile application development of command and control technologies. [emily.nguyen@navy.mil]

Jeffrey Millard—serves as program and technical support for the PMW-160 CANES program and Agile Core Services subsystem. He holds a BA in chemistry from DePauw University and a PhD in physical chemistry from Indiana University. Prior to providing support to PMW-160, Dr. Millard developed mathematical/statistic models of the currency derivatives markets, shock physics material response models for satellite systems, electronics reliability models for missile navigation systems, and optical analyses laser-hardened aircraft canopies. He has designed electronic trading exchanges for non-standard financial assets, and coordinated nuclear weapons effects testing for Air Force and Navy programs. [jeffrey.millard.ctr@navy.mil]

Abstract

Twenty years ago, the Navy began expanding the use of commercial industry information technology (IT) to employ Internet Protocol (IP)–based client server and web-based technologies to improve software effectiveness and affordability on ships and submarines. Coupled with wideband satellite capabilities, these systems increased the Navy's ability to plan, communicate, command and control, and execute increasingly complex missions. With a sound foundation in commercial IT installed in the Fleet, the Navy is looking today to improve warfighting by leveraging emerging technologies in Data Analytics, Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL). These technologies have the potential to change how the Navy fights and will drive changes to the Fleet's Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) architecture and processes. This paper proposes a reference architecture, new processes, and tools to meet the dynamic nature of these emerging

technologies, to include employing the commercial DEVelopment and OPerationS (DevOps) construct. The reference architecture and processes have the potential not only to accelerate the modernization of the afloat Navy networking WAN/LAN infrastructure, but also to deliver important warfighting capabilities to support Command and Control, Intelligence, and Logistics software applications.

Introduction

In 2017, the Space and Naval Warfare System Center Pacific (SSC Pacific) presented senior Navy leadership a technical vision for the future of the afloat Navy in which AI technologies, Cyber, and increased cohesion in mission planning played a central role in warfighting success. In the video scenario presented, a Carrier Strike Group (CSG) in 2035 was deployed and tasked to conduct a humanitarian relief operation in a contested battlespace. Examining many of the concepts highlighted in Brynjolfsson & McAfee (2016), *Second Machine Age*, and the concept of improved human-machine teaming, the video focused on the interaction of the CSG staff as they wrestled with the development of courses of action (COAs) to balance rules of engagement, asset limitations, and force protection. During the planning session on the flagship USS *John F. Kennedy* (CVN 79), both CSG and ship leadership engaged with the ship's computing stack, nicknamed "Kennedy" via a natural language processing interface. Kennedy was able to not only understand the crew's instructions, but also to access tactical and logistic information resident aboard the ship and, using reach-back, interface with ashore command and control nodes. The information Kennedy processed spanned the tactical, operational, and strategic levels, providing the system requisite context across all domains.

In one vignette, SSC Pacific highlighted the potential use case for human-machine learning in a scene featuring a conversation between the staff Operations Officer and the Battle Group Commander regarding COAs. Having monitored the conversation and without prompting, Kennedy interjected with an independent COA not considered by the staff. In the past, human-machine teaming of this level was the stuff of science fiction. Today, however, advances across multiple technologies are making these capabilities a potential reality—albeit in incremental steps.

The Consolidated Afloat Network Enterprise Services (CANES) provides the computing infrastructure and the foundation of the Navy's Information Warfare Platform afloat (NEJ, 2018). To deliver a more mission-effective, cyber-secure and affordable afloat Information Warfare Platform, CANES is modernizing its software application hosting and application integration processes and tools to implement a cloud-enabled DevOps framework. Agile Core Services (ACS) is a critical sub-system of CANES that has two broad categories of capabilities—core services and data analytics. Both support data sharing and analytics for CANES' hosted applications. ACS also provides a platform for the rapid insertion and management of software and facilitates applications' access to commercial analytics tools optimized to support the maritime battlespace. As shown in Figure 1, ACS sits on the stack as part of the CANES system between software application logic and the computing system.

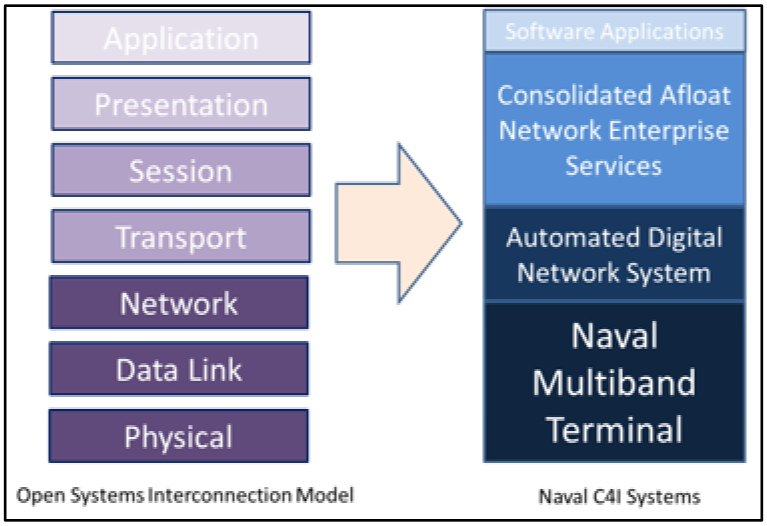


Figure 1. Open System Interconnection Model Mapping to Present Naval C4I Systems

An examination of the emergence of AI and ML technologies indicates that ACS also provides a framework to deliver these emerging technologies at the maritime computing edge afloat. AI technologies applied to command and control, intelligence, and logistics missions have the potential to provide new insights and speed of command. In this paper, we examine how the Navy can expand CANES to support AI and ML. In the first section of the paper, we examine ACS and its present capabilities. After that, we discuss how the current DevOps processes can support AI and ML. Next, we explore some of the commercial AI and ML technologies that exhibit initial potential to support Navy missions afloat. Finally, we cover potential Fleet use of the technologies and discuss implementation at sea.

Agile Core Services (ACS) Capabilities

Agile Core Services (ACS) provides a core set of software services that collectively create a shared framework for applications to build, deploy, and operate mission threads. ACS provides enterprise computing with integrated solutions of commercial off-the-shelf products creating an underlying service infrastructure to support the modernization of applications. Two primary services are provided to promote connection, collaboration, and communication between applications. These are Data Analytics and Common Services. Figure 2 provides an overall system resource flow description.

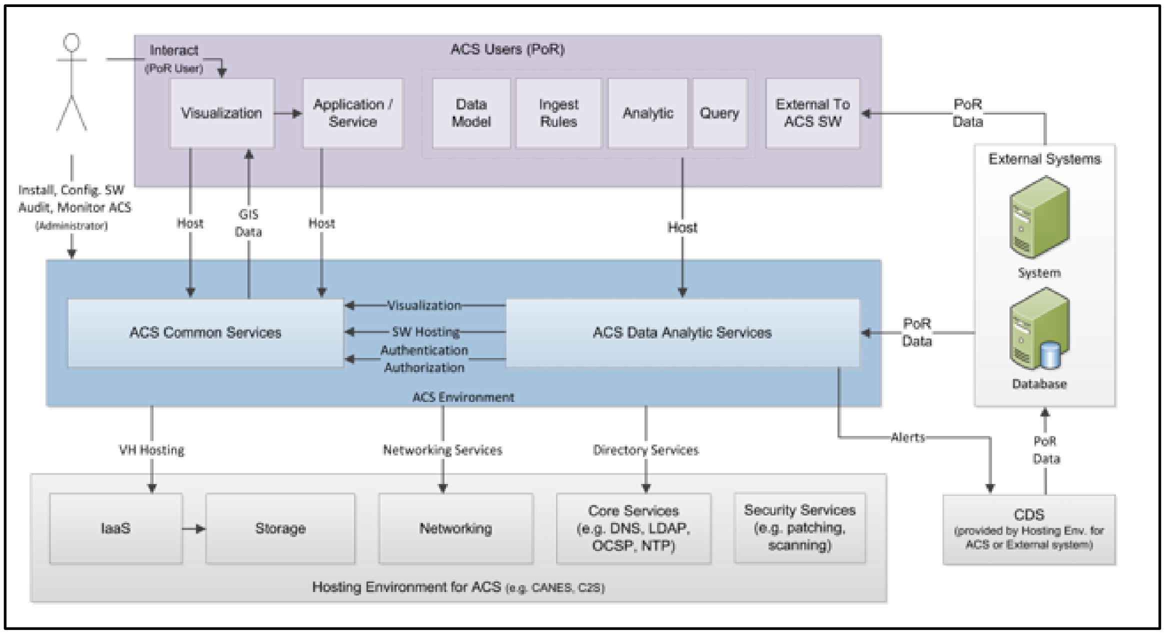


Figure 2. Agile Core Services Functional Diagram

ACS Data Analytics Services

The abundance of data within the Navy has driven the demand for ACS to provide a common data analytic architecture. By leveraging new industry technologies, ACS is able to provide a data analytic architecture that can process far greater volumes and variety of data at unprecedented velocities. The Data Processing and Analytics Framework utilizes clustering for the streaming, batch, and data processing to enable parallel processing for improved performance, robustness, and scalability.

Today, ACS leverages Apache's Spark and Storm, two common open source distributed compute engines. Both perform analytics and distributed compute tasks, but with distinctive implementations and focuses. Apache Spark provides fast cluster computing as a general-purpose distributed computing platform. It does provide limited stream processing; however, Apache Storm is specialized in reliably processing unbounded streams of data. Apache Storm provides real-time analytics, online ML, continuous computation, distributed Remote Procedure Call (RPC), and Extract, Transform, Load (ETL). A Storm topology consumes streams of data and processes those streams in arbitrarily complex ways, repartitioning the streams between each stage of the computation however needed.

The two engines described above are only a subset of the ACS data analytic framework. The following is a synopsis of all of the capabilities:

- Streaming Processing Framework for non-interactive manipulation and analysis of data moving at high velocity
- Batch Processing Framework for efficient, non-interactive analysis of big data stored at rest
- Shared Semantics Framework enabling a common vocabulary across heterogeneous data sources
- Application Programmer Interfaces (APIs) for data ingest, normalization, enrichment, and fusion

- Pre-Defined Query Processor to support interactive queries across heterogeneous data (structured, unstructured, and semantic)
- Alert Processor to generate alerts by comparing data against alert criteria

Identifying trends, finding patterns and relationships, and drawing conclusions are all reasons why data analytics has become vitally important. Given the variety of data available in the maritime environment, including operational data, content data (e.g., documentation, videos, and imagery), authoritative data (e.g., sensor data), and system-generated data (e.g., system logs), the data analytic solution has to be flexible enough to handle structured and unstructured data. With the diversity in data there are great opportunities to develop new decision aids that can help assist the Fleet fight and win.

Until now, many applications have leveraged data to do a specific task. As the Navy moves into the future, data will drive decisions by providing additional solutions that may have been overlooked due to the previous inability to analyze and provide relationships between overwhelming quantities of data. This is where a data analytic engine can facilitate advanced analytics and can accelerate the delivery of emerging technologies in DL and AI. Analytics include ML, data mining, and statistical analysis. When applied in real time and presented in an operational context, analytics can enhance the warfighters' ability to complete complex missions afloat. This unlocks many possibilities for the advancement in applications developed for the warfighter.

ACS Common Services

To support the agility required in present-day software development and to keep pace with the demand for new updates and patches, ACS Common Services provides a suite of services to aid in the modernization of applications. Each service is decomposed below.

Geospatial Data Access Layer and Geographic Information System (GIS) Services

A prevailing set of data that is common to many Navy applications is geospatial data. Geographic data is large in size and requires ample storage in the realm of terabytes. The Geospatial Data Access Layer and GIS Services provides a full suite of geographic data persistence, analytic, query, and mapping capabilities based around standard Open Geospatial Consortium (OGC) interfaces. By providing a common service for applications to retrieve geospatial data, this reduces the burden of hosting multiple map servers, each of which require maintenance while consuming terabytes of storage space reducing cost, man-hours, and storage space.

Mediation

The mediation service provides middleware platforms for hosting and integrating modular Java software components and sharing data via machine to machine messaging using topics and queues. Together these components support implementation of common enterprise integration patterns.

Visualization

Learning new applications can be incredibly challenging for Fleet Sailors due to the current portfolio of diverse user facing software. The visualization service provides a common user interface framework to help support applications with similar mission threads providing consistency to the look and feel of the application. A cohesive user interface presenting different types of data processed in various ways in a common structure enables warfighters to focus on decision making with the presented data instead of struggling with the complexities of using an application. Combining natural language with user interfaces will be the next step in user application interaction.

Platform as a Service (PaaS)

PaaS enables application services to change quickly, innovate easily, and remain competitive by supporting a services/microservices architecture reducing the complexity of building and maintaining the infrastructure typically associated with developing an application. A growing trend is the demand for cloud computing with microservices that can be scaled and deployed separately, enabling shorter release cycles. A PaaS service enables easy deployment of microservices by standardizing how services are executed, maintained, and orchestrated with a standard framework using containers to isolate elements of application deployment which reduces integration risks often seen in monolithic software architectures. As a result, service providers can more easily implement continuous delivery of updates to their services incorporating key practices of DevOps discussed in more detail in the next section.

Development Operations (DevOps) Processes, Capability Delivery

One of CANES' objectives is to lower the barrier of entry for applications that are hosted on or connect to CANES. Serving more than 100 applications from across the Department of Defense (DoD) and the Navy, CANES can drive affordability and increased speed to capability by providing application developers a common development environment, in some cases a cloud-based capability, along with the governance and processes to rapidly progress to the testing, information assurance, and fielding pipeline. To increase their ability to release capabilities, features, and patches out faster than ever before, industry established a software development culture and practice to bridge the gap between development and operations, which is known as DevOps.

Riding on the CANES platform, ACS provides an operational environment where it enables many of the tenets of the DevOps movement. As discussed previously, ACS provides a PaaS, in which automated development, deployment, provisioning, security, and other application lifecycle management tools are supported. In DevOps, PaaS enables applications to have a representative operational environment during development so that each step of taking a capability from development to operations can be automated. Automation accelerates the development cycle, increasing deployment frequency, while maintaining stability.

Beyond providing the technology and tools for application to leverage for DevOps, CANES/ACS has strategically aligned the people, processes, and tools to create the DevOps objective end state, which is depicted in Figure 3. It highlights several organizational and cultural changes that are described in detail in the next section to how CANES/ACS is developed and deployed operationally.

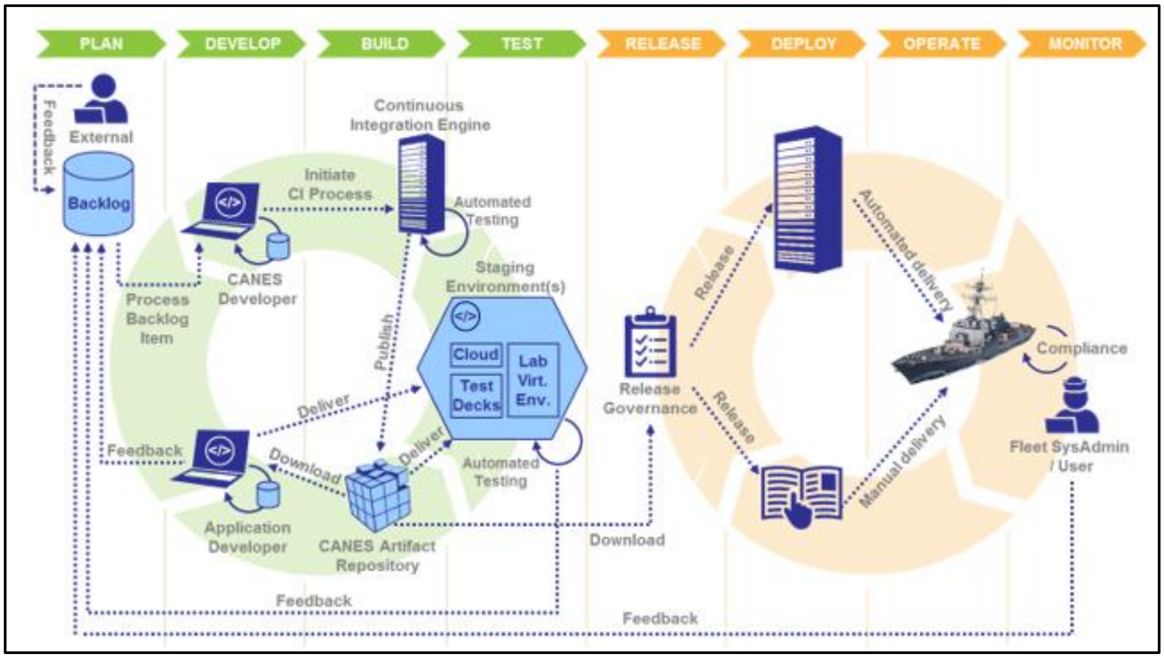


Figure 3. Afloat DEVelopment and OPerationS Framework

Planning

Each cycle starts with planning following the Agile Scrum methodologies. Work is described as product backlog items where it is prioritized based upon criticality and need. An important change that is depicted in the Figure 3 diagram is the feedback loop that comes from operational users and/or automated monitors afloat in the CANES/ACS platform that can provide suggestions and/or issues back to developers. As DevOps matures, this cycle between receiving feedback to deployment of changes to the system becomes shorter and shorter, transitioning from a lifecycle of years to possibly days.

Development/Build

Continuous Integration is a foundational part of DevOps that occurs during development in which developers regularly merge their code changes into a central repository (i.e., software version control), and automated builds and tests are executed. Both CANES and application stakeholders that deploy applications on CANES will drive affordability and improved interoperability by adopting this development practice to minimize integration efforts throughout the development and deployment process and to be able to continually deliver new capability as they become readily available.

Test

Since CANES provides a fully integrated computing infrastructure comprised of hardware and software, the staging environment supports various test environments. Applications at the software level can be tested against the latest version of CANES in a commercial cloud environment, and applications that are tested at a hardware level can test in a lab environment with representative hardware. The key benefit of the staging environment is to provide integration early on and throughout the development of CANES and integrated applications. The combination of a staging environment and automated testing is an indispensable part of quality assurance. Adopting this new test strategy will free up time that used to be spent on manual tests that were too slow to keep up with the rapid development that typically occurs and help teams focus on quality enhancements that until now have been addressed at the tail end of development.

Release/Deploy

Release will be a governance process to be defined based on existing processes and new options now available as automation is introduced. In addition, technologies provided by ACS such as PaaS help simplifies the deployment and management of modern web applications.

Operations/Monitoring

Operations is the last stage in DevOps where CANES is deployed and is operating on ship. As monitoring tools continually advance providing platforms that aggregate data and perform cross analysis, commercial tools are improving the user experience with easy-to-use UIs and system alerts to show correlations between events providing performance degradation or user experience issues quickly back to the developers is the crucial data in shortening the cycle between identified issues and deployment of fixes. When examining the Operations side of DevOps, it's important to highlight that unlike our commercial counterparts, our afloat networks are mobile, global, and engineered to be shot at both kinetically and are prime nation-state cyber targets. Furthermore, our systems are maintained by Sailors with at times limited access to shore support. These attributes drive an even more critical need to get the right system data from the platform and to use every interchange with the afloat platform right.

Commercial Technologies

To envision the CANES/ACS DevOps and tactical Data Analytics platform for the near future, it is instructive to look at some of the current and emerging trends in this space:

- Advanced DevOps/microservices capabilities
- Self-healing and self-protecting systems
- Serverless platforms and function as a service (FaaS)
- ML, especially deep machine learning
- Augmented analytics
- Game theory and ML/DL
- Advanced analytics processors

Advanced DevOps/Microservices Capabilities

As discussed earlier, ACS provides an operational environment that is integral to DevOps. In particular, the PaaS capability provides containers, where applications can build in an environment that encapsulates the necessary software dependencies and enables robust, fault-tolerant remote installations. While we have yet to fully exploit all of the benefits, the capability exists today (Farcic, 2017) to perform zero-downtime software upgrades, immediate rollback to prior versions if problems occur, and phased rollouts ("blue/green deployments"). Of particular interest to data analytics is the ability to perform "A/B testing." In this case, two (or more) different versions of a data analytic could run simultaneously and, based on the results, the more successful version would be retained. Blue/green and A/B deployments allow new analytics to be pushed to a ship and tested, evolving the analytics on a tactical platform without disrupting current missions.

Self-Healing and Self-Protecting Systems

The PaaS + Microservices architecture supports self-healing software, a design pattern for high availability systems that is actually fairly common in distributed data analytics platforms such as Hadoop. These self-healing systems are as follows (Bonér et al., 2014):

- Reactive—maintaining rapid and consistent response times
- Resilient to failure—through replication, containerization, isolation, and delegation
- Elastic—rapidly scaling under varying workloads
- Asynchronous—implementing loose coupling between components, so failures are isolated

By using ML algorithms for anomaly detection and classification, it will soon be possible to build systems that are self-protecting at the architecture, application, and OS levels (Yuan, 2016). This adaptive, intelligent defense, coupled with traditional security protections, will be very important in improving our defenses against the growing frequency, complexity, and sophistication of cyberattacks. In fact, the trend is toward self-managing, self-healing, self-configuring, and self-protecting software systems enabled by ML.

Serverless Computing/Function as a Service

Serverless computing allows highly efficient sharing and utilization of compute resources (Roberts, 2016). The term serverless is generally acknowledged as a misnomer—there are definitely still servers in the system—but the details of the infrastructure, servers, operating systems, and runtime environments have been abstracted away from the application developer, so they need not be concerned with the implementation or management of the underlying system. The serverless platform provider manages all the details of the environment and the dynamic allocation of compute resources. All the application provider does is deliver the code. This code is an event-driven, functional program (as opposed to object-oriented or procedural code), hence the alternative term *function as a service* (FaaS). Amazon Web Services Lambda, Azure Functions, and Google Cloud Functions are all examples of cloud-based FaaS. Apache OpenWhisk, originally developed by IBM, is an open source serverless platform that can be deployed on-premises.

In FaaS, an event (such as incoming data) triggers execution of the function, which can scale horizontally with varying load. The code is automatically containerized, deployed, scaled, and executed. The serverless architecture has built-in high availability and fault tolerance, and allows for very efficient utilization of compute resources. This resource efficiency is very appealing in the fixed-footprint compute environment of a tactical platform. Because FaaS compute is event driven and transient, there is a reduced attack surface and it is more resilient.

Not all applications are suited for FaaS. The best applications are for short computations, triggered by very short bursts of activity. Long-running processes, or ones with consistent loads, do not benefit much from FaaS. Nevertheless, FaaS is excellent in periods of rapid activity and is highly optimized for performance in such situations.

Deep Machine Learning

It is now commonplace to encounter ML applications in our day-to-day lives. Digital assistants such as Alexa, Siri, and Cortana are generally able to respond to questions and voice commands with reasonable accuracy. ML speech-to-text, natural language processing, and text-to-speech algorithms are all employed by these digital assistants. Photo library software on our personal computers can do facial recognition. Using clustering algorithms, recommender systems from Netflix or Amazon suggest what to watch or buy. DL is a subset of ML that brings us closer to AI. Whereas ML can perform natural language processing (NLP), finding common words, n-grams, and phrases, DL is more capable of natural language understanding (NLU). NLU not only identifies common words and phrases, it can analyze sentences and groups of sentences to discern topics and context and thus it comes closer to understanding conversation. Furthermore, speech-to-text ML can couple with audio ML analysis (e.g., speech-to-emotion) or image recognition (body language), providing further understanding beyond word recognition. Even tougher NLU problems such as identifying sarcasm and irony, which are difficult with purely speech-to-text analysis, are amenable to these hybrid analyses. It is feasible to use these technologies to construct a model of the social hierarchy of a group (think of a ship's Flag Command Center or a boardroom in a business). The implications of these technologies to assist decision-makers during periods of stress, for example, to determine if group dynamics are negatively impacting the decision-making process are tremendous. Similar to how better cockpit resource management made aviation safer, perhaps in the above scenario, these technologies could help decision-makers in real time make better strategic decision by better understanding the human teaming in the room.

Augmented Analytics

In the SSC Pacific vision video, when "Kennedy" suggested a COA not considered by the staff, this represented Augmented Analytics (AA; Su, 2017). AA uses ML, NLP, and other tools to perform data source selection, preprocessing (cleaning), analysis, insight generation, and presentation. Indeed, behind the scenes, Kennedy was presumably continuously selecting and analyzing data to generate actionable insights, and then communicated those insights to the ship's staff without the benefit of a human data scientist in the decision loop. No one commanded Kennedy to conduct an analysis or build a DL model, but rather it apparently was simply churning out insights on its own, waiting for an applicable situation to arise.

Today, AA maturity as a near-term capability is unlikely. If performed as an exhaustive survey of a vast parameter space, AA is very resource intensive, roughly akin to categorizing everything in a haystack, finding multiple "non-hay" objects in the haystack, recognizing one of these non-hay objects is a needle, and generating the insight that this sharp object can be used to puncture a balloon, and that having the insight that puncturing a balloon would be a useful tactical objective. Even so, AA has near-term potential in streamlining the tedious parts of analytics: data collection, cleaning, labeling, classification, and preliminary analysis, setting the stage for people to make focused inquiries and interpreting and generating insights from the results.

Game Theory and ML

The discussion of AA brings us to the intersection of Game Theory (GT) and ML/DL (Perez, 2016). If we (oversimplify) GT as modeling cooperation and competition (decision-making) strategies based on deterministic game rules, we can see the role of DL as discovering rules based on imperfect knowledge of the outcome of games. What does that mean? To better understand this, consider the use of adversarial neural networks (ANNs), one of many possible approaches in DL. With ANNs, we pit one neural network against another in a game. The classic example in image recognition is that we have one neural network (NN) trying to distinguish between real and fake images. At the same time, a second NN is trying to generate convincing fake images. As time goes on, the first NN is trained to distinguish real versus fake images, while the second NN learns how to make better counterfeits. In the end, both NNs "learn" what makes an image "real" and what makes it "fake." The result of the game is that both NNs learn from imperfect data to become better players.

These technologies have potential to apply to tactical missions as they mature to a broader range of human machine interaction. For example, developing a DL algorithm about how to avoid detection of a ship's electromagnetic signature could be modelled using ANNs under various environmental/weather conditions. These are the type of "games" Kennedy could play against itself to improve its algorithms. In Figure 4, we present a simplified NN that ingests data from various domains into the hidden layer and ultimately presents a probability of success of a specific mission course of action (COA).

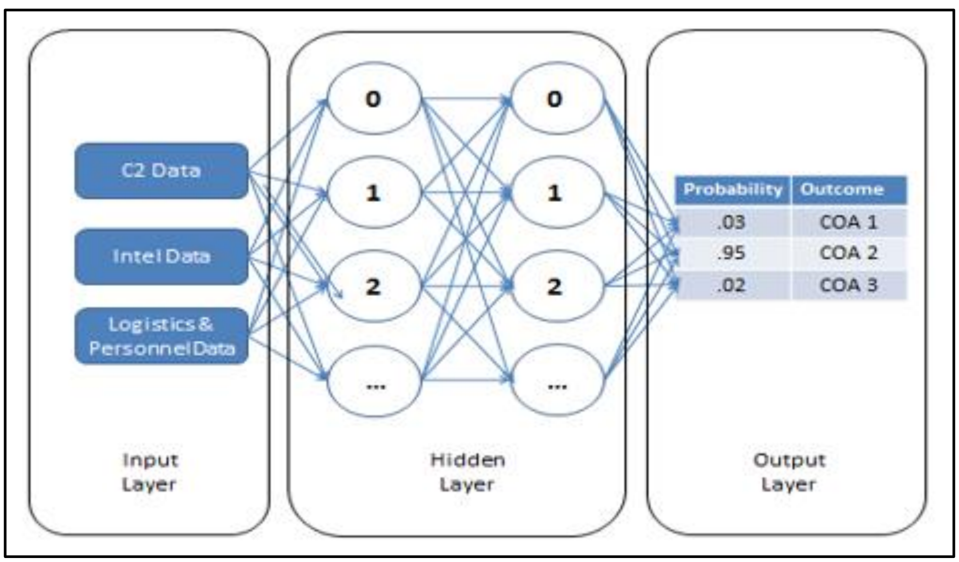


Figure 4. Simplified Example of Neural Network in a Maritime Tactical Network

Advanced Analytics Processors

Finally, all of these data analytics make high demands on the compute resources on ship. Part of the problem is that generic (e.g., x86) processors are optimized for general purpose computing and not data analytics. Graphical Processing Units (GPUs) and Field-Programmable Gate Arrays (FPGAs) are better suited for analytics processing, but Application-Specific Integrated Circuits (ASICs), purpose-built for NN (ML/DL) computations, such as Google Tensor Processing Unit, have the potential to greatly accelerate DL calculations. These have a much smaller footprint and lower power consumption (relative to NN compute capacity), so they are enablers of more data analytics capacity in a constrained footprint, like an afloat platform. If we expect our systems (Kennedy) to be performing AA and ANN behind the scenes, we will need to rely on purpose-built hardware to achieve the necessary throughput and allow the ship to remain tactically relevant even if disconnected from the shore infrastructure.

Overall, the near-term Data Analytics reference architecture will exploit DevOps capabilities to rapidly deploy new capabilities to the tactical edge. Looking toward the future, advances in technology will enable the platform to be self-healing and self-protecting, with highly efficient use of computing resources, including purpose-built processors. Potentially, we will leverage ML and DL for machine-human interaction and rely on AA to prepare and stage data for our analytics. Those analytics will be able to train themselves using game theoretic and DL technologies, and be ready to facilitate insights that will give our warfighters a critical advantage. As we engineer the CANES and ACS stack afloat to operate afloat in what is a very harsh maritime battlespace with intermittent and limited connectivity, we will need to examine how we integrate these emerging technologies afloat and allow them to be survivable and maintained by our Sailors afloat.

Fleet Employment and Management of AI, DL, and ML Technologies

In the 2018 National Defense Strategy, rapid technological change and challenges from our adversaries in every operating domain are identified as causes of our complex security environment. The strategic approach identifies several steps to build a more lethal force, including modernizing key capabilities in the C4ISR mission. The secretary of Defense states that investments will be prioritized for resilient and federated networks to assist in gaining information. AI, DL, and ML can support the objectives in the defense strategy as these technologies have the potential to give Navy warfighters the advantage in combat operations if they are deployed with the reliability and security that is needed for fleet missions.

In the book *The Master Algorithm*, Pedro Domingos (2018) discusses the increasing growth of ML technologies across a wide spectrum of activities. Domingos highlights the initial use in the 1980s of early ML in the financial sector, and the migration toward other commercial uses in the subsequent decades. Additionally, Domingos discusses the rapid rise and early employment of ML tools following the attacks of 9/11 by the DoD and other agencies, and how ML is not a single set of technologies or algorithms, but rather a multi-discipline body of knowledge or “tribes.” These “five tribes” that Domingos explores each approach the desired outcomes of ML from varied theoretical and applied approaches, suggesting that the state of practice today requires different technical approaches or in many cases a blending of technologies, based on the questions being asked. Hence, as we look toward ML solutions for the Navy, we will likely need to field a collection of capabilities that are optimized to our unique connectivity and maritime concept of operations.

The CSG organization provides an excellent use case of where these technologies can be applied to support Fleet operations. In the CSG organizational structure, the Battle

Group Commander's staff executes a series of functions required for the mission within each coded directorates. On the CSG flagship, the CVN or Nuclear Powered Aircraft Carrier, staff functions are supported by departments that exist within the shipboard organization. The Operations Department executes and supports current tasking and future missions while maintaining schedule. The Combat Systems Department is responsible for maintaining weapons and communications systems and ensuring the enterprise network is available. The Engineering and Reactor Departments are charged with ensuring safe and sustained propulsion, as well as maintaining all hull, machinery, and electrical equipment. The Air Department enables flight operations to occur hazard free and within environmental constraints of any geographical area. The Supply Department supports logistical needs in parts, materials, and consumable equipment. The Navigation Department is responsible for ensuring safe passage during open ocean transit, as well as in constrained and heavily trafficked waterways. The arrangement of departments onboard a CVN is similar to what may be found on smaller units, such as cruisers and destroyers—with the roles of each department differing slightly based on the specific mission of the ship. Weaving the activities of these departments, along with the Aircraft Carrier's mission to deliver air power via an embarked airwing and other ships in the strike group, provides a compelling use case for the potential capabilities ML can provide.

In examining each department's specific areas of responsibility, it is evident that there are myriad opportunities for AI, DL, and ML to contribute to the naval mission. Examining the use cases, we see applications across Command and Control, Intelligence, Cyber, Logistics, and Personnel, to use the data in these domains to improve decision making. For example, one subset within the Operations Department is Operations Information (OI). The Operations Specialists in OI Division comprise the watchteam that maintains a current operational picture (COP) that provides the mission commander situational awareness to make decisions for the entire CSG. Key assets in the mission are identified and tracked with tools that present the commander with a comprehensive view of the battlespace. Contacts on the display are identified as friendly, neutral, or foe. Since the environment is constantly changing, this mission area is one that has the potential to benefit from AI technologies and a DevOps framework to rapidly integrate and process new sensors, data feeds, and contingences.

The Navy has laid much of the groundwork associated with implementing AI technologies by employing data models and software patterns that leverage the commercial technologies that are many of the underpinning connectivity methods the newer AI technologies employ such as eXtensible markup language (XML) and other web-based connectivity tools (Rothenhaus IEEE, 304). To manage a growing demand for faster decision cycles and improved battlespace awareness, supporting systems such as Global Command and Control System Maritime (GCCS-M), as well as intelligence systems, the Navy can ensure our warfighters are not missing key elements of the operational picture due to over-complexity and human error by implementing intelligent systems able to process data automatically and present that correlation of data to Sailors and decision-makers. In the "Kennedy" vignette discussed earlier, tools are dynamic enough to adapt to emerging conditions, giving the warfighter the advantage of not having to focus on updating changing information, and instead allowing warfighters to focus precious attention on strategies to support mission success.

Further Research

Although AI is not a new field of study, new methods and approaches enabled by increased computing, storage, and data management techniques are driving renewed interest in the commercial sector and are fertile areas of future research in the Navy enterprise. The adoption of those technologies for the Navy presents additional areas for multi-disciplinary research in the areas of computer science, system of systems engineering, and acquisition with likely many more potential areas of study. Areas of study include questions about how to characterize the data the Fleet manages from the perspective of volume and variety. Many of the technologies highlighted in this paper rely on significant quantities of data to gain the insights they deliver. Although ships have large databases, they may not have the types of enterprise-level data that would permit meaningful analysis. Additionally, our ships operate in a unique connectivity construct ranging from full wideband connectivity to being completely disconnected. Future research in the areas of data strategies to support sharing of processing between on-board and off-board processing with a focus of graceful scaling from very elastic processing environments ashore to more limited processing on shipboard data centers could help future engineers design systems to support the Fleet’s unique requirements.

Conclusion

Navy warfare increasingly depends on advanced software capabilities, deployed across numerous platforms and systems. As the amount of information grows with the increase in quality and quantity of sensors and new autonomous platforms, the Navy is looking to leverage commercial information technologies to improve warfighting outcomes and enhance mission execution. Leveraging AI technologies, Navy Program Offices are examining methods to provide a ready platform to rapidly and affordably integrate and test emerging AI capabilities. As we integrate these types of technologies into an already complex system of systems afloat, it is critical to manage the complexity to ensure our Sailors can employ and support them. In the areas of command and control, intelligence, and logistics, AI and ML have the potential to deliver to the nation that integrates them first a tactical advantage. For the Navy, CANES and ACS are the target platform on which to integrate those technologies to deliver important warfighting capabilities across Command and Control, Intelligence, and Logistics software applications.

References

A Summary of the 2018 National Defense Strategy of the United States of America. (2018). Retrieved from https://www.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

Bonér, J., Farley, D., Kuhn, R., & Thompson, M. (2014). The reactive manifesto. Retrieved from https://www.reactivemanifesto.org/

Brynjolfsson, E., & McAfee, A. (2016). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* New York, NY: W.W. Norton.

Buduma, N., & Locasio, N. (2017). *Fundamentals of deep learning: Designing next-generation machine intelligence algorithms*. Sebastapol, CA: O'Reilly Books.

Domingos, P. (2018). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. New York, NY: Basic Books.

Farcic, V. (2017). *The DevOps 2.0 toolkit: Automating the continuous deployment pipeline with containerized microservices*. Retrieved from http://www.amazon.com

Perez, C. (2016). Game theory reveals the future of deep learning. Retrieved from https://medium.com/intuitionmachine/game-theory-maps-the-future-of-deep-learning-21e193b0e33a

Roberts, M. (2016). Serverless architectures. Retrieved from https://martinfowler.com/articles/serverless.html

Rothenhaus, K., Bonwit, B., Galdorisi G., & Stang, A. (2018). Distributed lethality, command and control software engineering, and naval laboratories. *Naval Engineers Journal* (5).

Rothenhaus K., Michael, B., & Shing, M. (2008). Architectural patterns and auto-fusion process for automated multi-sensor fusion in SOA system of systems. *IEEE Systems Journal.*

Su, B. (2017). Augmented analytics demystified: What it means and why it is the future of data analytics. Retrieved from https://medium.com/analytics-for-humans/augmented-analytics-demystified-326e227ef68f

Yuan, E. (2016). *Architecture-based self-protecting software systems* (Doctoral dissertation, George Mason University). Retrieved from https://seal.ics.uci.edu/dissertations/eric.pdf

Disclaimer & Distribution Statement

The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Distribution Statement A: Approved for public release, distribution is unlimited (1 APRIL 2018).

Conducting Viability Assessments for Acquisition Planning

Virginia L. Wydler—is an Acquisition Scientist at The MITRE Corporation, supporting Defense sponsors with acquisition strategies and source selections. She has over 25 years of experience in acquisition, both government and commercial. She is a former Navy Contracting Officer. She holds an MS in National Security Strategy, Industrial College of the Armed Forces; an MS in Acquisition and Contracting, Naval Postgraduate School; and a BS in business administration, University of Maryland. She has published in *Defense Acquisition Journal* and *National Contract Management Association Journal*. She is a Certified Professional Contracts Manager (CPCM) and Fellow with NCMA. She is DAWIA Level III Contracting. [vwydler@mitre.org]

Erin M. Schultz—is Technical Director at the Center for Acquisition and Management Sciences at The MITRE Corporation, providing subject matter expertise in the areas of contracting, source selection, proposals, and acquisition strategies for a variety of customers. A former Contracting Officer, Schultz also served as the Vice President of Proposal Development for a Fortune 500 IT corporation. She has over 25 years of contracting experience with both federal government and industry, specializing in information. Schultz is an instructor for acquisition and contracting courses for NCMA. She has taught acquisition and source selection courses for The MITRE Institute. [eschultz@mitre.org]

Abstract

Defense Department program managers and industry can streamline the acquisition process by conducting an early assessment of the "viability" of technical solutions during the market research phase of acquisition planning, or as the first step of the source selection. This research documents a viability assessment process as a best practice, provides an example of a successful application, and suggests metrics for measuring success.

Introduction

Fielding government systems can take decades under current acquisition processes (GAO, 2015). As a result, government program managers find it difficult to rapidly adopt new technologies that offer improved efficiency and effectiveness. One approach to improve the acquisition process would be to conduct an early assessment of the "viability" of solutions during market research, or as a first step of the source selection. Conducting an early assessment of industry capability before final source selection can lead to more informed acquisition strategies to rapidly adopt new technology. This paper documents a viability assessment method as a best practice, offers ideas on how to conduct an early assessment, and suggests metrics for measuring success.

Background

Program offices that deploy major systems find it difficult to evaluate and quickly adopt new technology due to the long time involved in acquiring systems under the Federal Acquisition Regulation (FAR). Commercial practice, by contrast, allows industry to test new technologies well before launching them into the market to determine if they are viable for use in products. Unfortunately, the federal procurement process subjects every proposed solution submitted under a federal government solicitation to the same rigorous evaluation, whether the solution is viable or not. Thus, the government must devote time and resources to performing detailed assessments of non-viable bids. Companies offering non-viable products may spend opportunity time and proposal costs to participate in a source selection they cannot win. There is also a risk that if they do win, only afterwards does the contractor (and the government) discover that the solution doesn't work. Therefore, the federal government is trying to find methods or approaches to incorporate technology innovation quickly and soundly from companies that are technically capable of meeting requirements (GAO, 2016a).

How do program managers and acquisition professionals become early adopters of technology and quickly assess the viability of an offer before entering the rigorous proposal evaluation process? Using a viability assessment approach gives the vendors an opportunity to submit their products and solutions before the final request for proposal (RFP) is released. They can receive a no-harm-no-foul evaluation from the federal government, which may offer the necessary "quick look" and instant feedback to the prospective vendor community.

Conducting a viability assessment early in the acquisition process would allow the federal government to shape the pool of companies that would likely submit acceptable proposals. The viability assessment would also help the government revise requirements, based on knowledge of industry capability, for an achievable outcome. Finally, vendors would know before they submit a proposal whether they should continue to pursue a contract that they have little chance of winning.

Purpose

This research paper offers an approach for acquisition professionals to make more defensible decisions and lower risks for better outcomes. It describes how government program offices can incorporate viability assessments into the competitive acquisition process. Specifically, it presents an approach to using viability assessments as part of an acquisition strategy to minimize the likelihood of unqualified vendors, and stimulate better competition for viable solutions. It emphasizes the importance of communications between industry and government during the acquisition process. It also offers standards for measuring success.

Acquisition Planning

The purpose of FAR Part 7, Acquisition Planning, is to ensure that the government meets its needs in the most effective, economical, and timely manner. During the acquisition planning phase, program offices develop key system features or required functionality and convey them as a priority to industry through announcements issued prior to the formal source selection phase. The Requests for Information (RFI) announcements provide valuable information to industry on government-planned requirements. The RFI responses help government personnel, both acquirers and users, understand potential solutions available from vendors. The acquisition team can then leverage insights to determine how best to structure and shape the appropriate acquisition strategy. The RFI responses support development of contract strategies and incentives before releasing the final RFP. However, RFI responses generally offer paper solutions, marketing materials, and little detail about potential risks in performance.

In accordance with Federal Acquisition Regulation (FAR) 7.105(a)(7), the acquisition team must conduct risk analysis to support the acquisition planning process. This risk assessment is critical in developing source selection criteria and evaluation factors (Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics [OUSD(AT&L)], 2016). As part of the early phases of the acquisition planning process, the acquisition team needs to involve vendors in risk mitigation activities. Conducting viability assessments can create a better dialog between industry and government for risk reduction and risk mitigation strategies that may not be revealed until after award. Shaping the pool of prospective offerors before formal solicitation so that it includes only "viable" vendors would reduce risk and improve the quality of the proposals submitted.

Market Research

In accordance with FAR Part 10, Market Research, acquisition offices conduct market research to identify contractors with the capability to provide the required service or product before soliciting bids or proposals for a contract. The market research also offers an opportunity for the government to refine requirements. Some market research communication techniques recommended in FAR Part 10 include the following:

- Releasing a Request for Information (RFI) to solicit industry input on requirements and potential acquisition strategies and gauge current industry interest and market capabilities
- Contacting knowledgeable individuals in government and industry regarding market capabilities
- Gathering market information on specific products and potential suppliers
- Gathering market pricing and technical information from commercial or government sources
- Gathering market information on industry practices, supply and demand trends, and other relevant factors.

These techniques can point to the ability of acquisition offices to understand the viability of products and risks associated with meeting the government requirement before releasing the final RFP to industry. FAR Part 10 encourages the use of viability assessments by “gathering market information” in a structured approach, based on the recommended techniques shown above.

Communicating With Industry

The federal government’s ability to achieve a successful contract award depends upon credible communications with offerors. Yet program offices often feel unsure of how to conduct such communication. To demystify **misunderstood processes and procedures**, the Office of Federal Procurement Policy (OFPP) released a series of Myth-Busting memos to foster productive communication between the federal government and industry from the beginning phase (including market research) of acquisition planning through the final phase of the source selection process:

- *Myth-Busting: Addressing Misconceptions to Improve Communication with Industry in the Acquisition Process* (February 2, 2011)
- *Myth-Busting 2: Addressing Misconceptions and Further Improving Communication During the Acquisition Process* (May 7, 2012)
- *Myth-Busting 3: Further Improving Industry Communication with Effective Debriefings* (January 5, 2017)

These documents emphasize a strong dialog between industry and government, especially during the early phases of the acquisition process. They also encourage government to be more open with industry regarding their requirements during events, such as industry days and bidders’ conferences, to gain insight into market conditions and technical capabilities. Viability assessments can be considered another effective tool in communicating with industry regarding government requirements and the need for quality solutions.

Acquisition Strategy

> Rapid down selection, which narrows the field of potential bidders, has dramatically accelerated government acquisition efforts and can be applied to cyber contracting. It must be done early in the acquisition process, well before a formal RFP and preferably long before a draft RFP. Otherwise, companies already may have invested considerable resources in pursuing an opportunity and will be reluctant to forgo their activities. (Gilligan, 2016a, p. 41)

Including a viability assessment as part of the acquisition strategy gives program offices a better understanding of the marketplace and how their requirements will match vendor products. This strategy also allows industry to become aware of the competitive nature of the acquisition by obtaining early feedback on how competitive their solution will be viewed. As a result, industry is in a better position to choose which solicitations to pursue, since they normally have more contract opportunities to bid on than they have resources to support a proposal.

An acquisition strategy that utilizes viability assessments can reflect a stronger analysis of alternatives in fulfilling customer requirements. It accommodates the FAR Part 7, Acquisition Planning, and Part 10, Market Research, by encouraging two-way communication between industry and government. It offers a valid down-selection method under FAR 15.202 for source selection.

RFI Versus Multi-Step RFP

Viability assessments can be conducted at two points in the acquisition process. Both have merit, depending on industry capability, market conditions, and technology maturity.

The RFI

Releasing a Request for Information (RFI) to solicit industry input on requirements and potential acquisition strategies and gauge current industry interest and market capabilities is one of the recommendations for market research under FAR Part 10. Rather than require paper responses that offer no feedback to industry, an RFI can include an assessment of capability based on a set of criteria or scenario. Feedback can be provided to industry on an advisory basis. Any results would not preclude industry from bidding on the formal RFP. However, this early look allows vendors to decide if they pursue the business considering their product solutions.

The Multi-Step RFP

FAR 14.501 addresses two-step sealed bidding as a combination of competitive procedures designed to obtain the benefits of sealed bidding when adequate specifications are not available. FAR 15.202 Advisory Multi-Step process allows for viability assessments as part of the solicitation process. The language in FAR 15.202 specifically defines the advisory nature of the assessment and allowing industry to make an informed decision to pursue a competition:

- The agency may publish a pre-solicitation notice that provides a general description of the scope or purpose of the acquisition and invites potential offerors to submit information that allows the overnment to advise the offerors about their potential to be viable competitors.
- The pre-solicitation notice should identify the information that must be submitted and the criteria that will be used in making the initial evaluation.

Information sought may be limited to a statement of qualifications and other appropriate information (e.g., proposed technical concept, past performance, and limited pricing information). At a minimum, the notice shall contain sufficient information to permit a potential offeror to make an informed decision about whether to participate in the acquisition.

- This process should not be used for multi-step acquisitions where it would result in offerors being required to submit identical information in response to the notice and in response to the initial step of the acquisition.
- The agency shall evaluate all responses in accordance with the criteria stated in the notice, and shall advise each respondent in writing either that it will be invited to participate in the resultant acquisition or, based on the information submitted, that it is unlikely to be a viable competitor.
- The agency shall advise respondents considered not to be viable competitors of the general basis for that opinion. The agency shall inform all respondents that, notwithstanding the advice provided by the Government in response to their submissions, they may participate in the resultant acquisition.

Building viability assessments into the acquisition process can strengthen the government's knowledge of the industry and technical capabilities before award. This would reduce the risk of design problems and allow adjustments to government specifications before award.

Source Selection

Conducting viability assessments during source selection can provide invaluable insight into how the marketplace will respond to federal government requirements. Since viability assessments are "advisory" (the contracting officer never directly discourages companies from bidding), they do not restrict competition or the ability of companies to submit offers for award consideration.

Competitive Approaches

Incorporating viability assessment results into a competition can be done either through feedback from the RFI results, including any challenge events, or the first step of the multi-step process under FAR Part 15.202. Federal procurements use several evaluation methods to select a winning proposal. Viability assessment information can be incorporated into these methods.

- **Traditional evaluation process.** Under the traditional process, the contracting office releases an RFP to industry; industry provides written proposals and receives written evaluations; discussions, usually in writing, resolve discrepancies or address weaknesses; and the government makes an award based on best value. All communication is in the form of paper proposal submissions. Even though many contracting offices release a draft RFP to obtain comments and answer questions to "fix" the RFP anomalies, many times this process does not reflect a robust dialog or a preview of proposed solutions.
    - A viability assessment conducted before the traditional competitive process can allow a healthy exchange between industry and government before the formal RFP, allowing adjustments to requirements before award.

- **Sample Scenarios with proposal.** Some acquisitions include sample Task Orders with the RFP. Companies respond to scenarios after the formal selection process starts. This process leaves little room to adjust scenarios or requirements based on learning.
    - Viability assessments can replace scenarios to allow that learning, feedback, and adjustments to requirements before source selection.
- **Demonstrations after award by winning vendor.** Some acquisitions require the successful vendor to prove that its solution performs as required through a demonstration or prototype after award has been made. If the first awardee is not successful, the award goes to the next offeror, who has a chance to demonstrate that its product will meet the requirement. This creates added churn and expense for industry and government.
    - Viability Assessments can capture and refine risk prior to award.

Viability assessments provide evaluation of solutions and feedback to industry and allow adjustments to government requirements. Viability assessments can capture and refine risk prior to source selection and award, reducing the need for several rounds of negotiations.

Protests

The Government Accountability Office (GAO) Report on Protests for fiscal year (FY) 2016, identified the following reasons for sustaining protests, in descending order of importance:

1. unreasonable technical evaluation
2. unreasonable past performance evaluation
3. unreasonable price/cost evaluation
4. flawed selection decision (GAO, 2016b)

The GAO report stated that in many cases agencies failed to meaningfully consider the merits of the evaluations or proposed prices. In response to this situation, the National Defense Authorization Act 2017 requires government agencies to submit a report on protests based on quantity and quality of information that vendors received both *before and after award* that affected the decision to file a protest. This congressional requirement underscores the attention being paid to communications with industry. Acquisition officials can improve the quality of information in proposals by responding to offeror concerns early in acquisition planning.

Companies can protest requirements included in the formal RFP and the final award decision. However, industry does not protest an RFI or feedback on their responses to that RFI, since it is advisory in nature. Viability assessments allow companies to understand why their solutions are unsatisfactory early in the acquisition process, which in turn can avoid submission of proposals that receive a negative evaluation and the potential of protest by unsuccessful offerors.

A viability assessment conducted during the RFI process may generate a sounder RFP package. A viability assessment conducted as part of the first step of an RFP down-select, the government provides an unsuccessful offeror all the documentation related to the evaluation, including evaluation results generated during a viability assessment. In the latter case, the vendor may decide not to protest the award because it understands that the company's proposed solution has little chance to win the award.

Designing a Viability Assessment

This section describes how program offices can design and execute a viability assessment during the acquisition planning phase. The key to designing the viability assessment is to incorporate the method into the existing acquisition planning process. This would include defining the requirements in terms of outcomes or functional characteristics, defining the format of responses, developing a scoring method, advertising the event through a request for information (RFI) to industry, evaluating submissions, and providing feedback to industry.

Defining the Requirements of a Viability Assessment

Describe the requirement in terms of the distinctive characteristics, quality attributes, or property features that are especially important for the system or service. The information requested should focus on critical characteristics of the requirement, not provide a full detailed specification. As an example, software is typically described in terms of features; the Institute of Electrical and Electronics Engineers (IEEE) defines the term *feature* in IEEE 829 as "a distinguishing characteristic of a software item (e.g., performance, portability, or functionality)" (IEEE, 2008).

Viability assessments may be used for information technology (IT) requirements, especially those that lend themselves to commercial product attributes and characteristics. Chapter 10, Vendor Lock, of *Open Systems Architecture (OSA) Contract Guidebook for Program Managers* provides a strategy to evaluate the viability of software before the RFP phase (OUSD[AT&L], 2013, p. 11).

Define the evaluation criteria and scoring mechanism. The team should decide whether scoring criteria for each characteristic should be numeric point scoring (for example, on a scale from 1 to 10), or adjectival (using qualitative descriptors ranging from "outstanding" to "unacceptable").

Format for Responses

Program offices can define the responses for viability assessments based on several types of information from industry. These responses can follow a design or characteristic, as those used in a market research request, or a sophisticated challenge or scenario. Formats can take the form of a white paper stating relevant product capability or technical approach, or a more complex response, such as a demonstration of a proposed solution (The MITRE Corporation, n.d.).

The more extensive the responses, the more expensive the process for both sides. The RFI or RFP first step should instruct vendors to submit only a technical statement of features and characteristics of the product, and vendors should be expected to perform a small demonstration to validate the claims made in the statement.

Evaluation Criteria

Criteria can provide objective reference points for the government's analysis. Some notional criteria based on characteristics of the system or service can include the following:

- Maturity level of commercial off-the-shelf (COTS)-based products
- Web-enabled solutions
- Product that can be configured or minimally customized to support unique requirements
- Product that can integrate into an enterprise infrastructure

- System that can interface with other systems internal and external to being acquired
- System that can interface with a data warehouse
- Innovative, cost-effective solution and implementation methodology to achieve desired objectives and results within the context of the solicitation

Scoring Responses

After the contracting office receives responses, the evaluation team scores the responses to each question and documents notes to support the scoring. Program offices can analyze the responses and then score them for purposes of feedback to industry. The generally accepted methods used in source selections to evaluate technical requirements can easily be adapted to viability assessments. These methods are as follows:

- Pass/fail. The government reviews the merits of submissions based on various criteria that lend themselves to a yes or no result. This may work well with known IT system characteristics, but would present a challenge for emerging, unproven technology.
- Color scoring. The government evaluates submissions based on grading criteria that depict levels of acceptability based on a color-coding scheme. This scoring can reflect strengths, weaknesses, and deficiencies in the product.
- Adjectival scoring. The government evaluates submissions by assigning adjectives that describe the level of quality. This scoring can also reflect strengths, weaknesses, and deficiencies in the product.
- Risk. The government can incorporate consideration of risk levels into the assessment if a proposed solution could cause problems with performance or overall system sustainment. Risk is normally described as low, moderate, or high.
- Not Addressed. If a vendor's response does not address a certain aspect of the requirement, the government provides no feedback on that portion of the requirement that is not address.

The Department of Defense (DoD) *Source Selection Procedures* contain evaluation criteria that suggest consideration of overall proposal risk in conjunction with strengths, weaknesses, significant weaknesses, uncertainties, and deficiencies in determining technical ratings for a source selection (OUSD[AT&L], 2016). Section 3.1, Evaluation Activities, Table 3, describes these criteria. Table 1 adapts that DoD table to apply specifically to viability assessments.

Table 1. Combined Technical/Risk Rating Method Rating

Color	Adjective	Rating Description
Blue	Outstanding	Viability assessment submission indicates an exceptional approach and understanding of the requirements and contains multiple strengths. Risk of unsuccessful performance is low.
Purple	Good	Viability assessment submission indicates a thorough approach and understanding of the requirements and contains at least one strength. Risk of unsuccessful performance is low to moderate.
Green	Acceptable	Viability assessment submission meets requirements and indicates an adequate approach and understanding of the requirements. Risk of unsuccessful performance is no worse than moderate.
Yellow	Marginal	Viability assessment submission has not demonstrated an adequate approach and understanding of the requirements. Risk of unsuccessful performance is high.
Red	Unacceptable	Viability assessment submission does not meet requirements, and thus, contains one or more deficiencies. Risk of unsuccessful performance is unacceptable. Product would not receive a successful award during full and open competition.

The evaluation team then provides a summary of the results, outlines discriminating differences, and compiles the scores for the different vendors, from highest to lowest, to identify the most viable solutions. The team can use a consensus process to mitigate outlier scoring or differing opinions. After reviewing the scores and rationale, the contracting officer and program manager determine the "most viable" companies.

Providing Feedback

The contracting officer notifies all respondents of their evaluation. Those companies that did not submit a viable solution are informed that they do not stand a good chance of success "with that proposed solution" for the pending solicitation. The contracting officer can offer to provide feedback on how the company scored in the evaluation. However, at no time does the contracting officer directly discourage that company from submitting an offer on the subsequent RFP. This process gives offerors whose solution was deemed non-viable the opportunity to decide whether to submit a bid once the formal solicitation process begins.

Measuring Value

Program offices can use the metrics shown in Table 2 to measure the value of applying viability assessments during the acquisition planning phase. Continuous review of known acquisition and contracting metrics will ensure that programs apply this method effectively and that it yields benefits.

Table 2. Viability Assessment Metrics

Metric	Used to Measure
Time to Award	Reduced lead time to reflect efficiency
Number of High-Quality Proposals	Reduced number of proposals that were not responsive to requirements, because the viability assessment made vendors realize they could not meet the requirements
Cost Control	Reduced cost over life of the system based on viable solutions
Number of Protests	Industry satisfaction with the process as being competitively fair
Timely Performance or Delivery	Contract performance resulting in shorter development and delivery timelines
Government/Contractor Relationship	Partnership between government and contractor to solve technical problems in program delivery; fewer technical changes or engineering change proposals Positive impact on acquisition process; less rework

Conclusions

Viability assessments can help the federal government instill confidence in the acquisition process to maximize competition among viable offerors. Industry can better understand requirements before submitting a formal offer, while gaining additional insight from their assessment results to improve future proposals. Contracting officers and program managers can apply this approach to refine their requirements before the formal RFP is released, thereby improving the likelihood that the proposed solutions would meet program needs.

Acquisition offices can use viability assessments when vendor solutions will likely vary widely and many companies can be expected to submit proposals. Such variance in solutions could occur when market conditions reflect many new entrants or when legacy systems require unique upgrades or emerging, unproven technology.

The following criteria would indicate use of viability assessments as an appropriate approach:

- Use of this method improves government knowledge of the marketplace and an ability to rapidly adopt new technology.
- An early engagement between government and industry leads to better quality proposals in the final evaluation process.
- This method complements Chapter 10, Vendor Lock, of *Open Systems Architecture Contract Guidebook for Program Managers* (OUSD[AT&L], 2013) by providing another method to evaluate viability of software before RFP phase.

References

Federal Acquisition Regulation (FAR), 48 C.F.R., Part 7, Acquisition Planning (2018).

Federal Acquisition Regulation (FAR), 48 C.F.R., Part 10, Market Research (2018).

GAO. (2015, February). *Acquisition reform: DOD should streamline its decision-making process for weapon systems to reduce inefficiencies* (GAO-150192). Washington, DC: Author.

GAO. (2016a, August). *GAO technology readiness assessment guide: Best practices for evaluating the readiness of technology for use in acquisition programs and projects* (GAO-16-410G). Washington, DC: Author.

GAO. (2016b, December). *Re: GAO bid protest annual report to Congress for Fiscal Year 2016* (GAO-B-158766). Washington, DC: Author.

Gilligan, J. (2016a, November). The cyber implications of acquisition speed: Part VII. *Signal Magazine.*

Gilligan, J. (2016b, November). Shrinking pool of prospective bidders increases acquisition agility. *AFCEA Signal Magazine.*

IEEE. (2008). *IEEE standard for software and system test documentation* (829-2008). New York, NY: Author.

The MITRE Corporation. (n.d.). *Challenge-based acquisition handbook* (2nd ed.; PRS 13-3525). McLean, VA: Author.

Office of Federal Procurement Policy. (2011, February 2). *Myth-busting: Addressing misconceptions to improve communication with industry during the acquisition process*. Washington, DC: Author.

Office of Federal Procurement Policy. (2012, May 7). *Myth busting 2: Addressing misconceptions and further improving communication during the acquisition process*. Washington, DC: Author.

Office of Federal Procurement Policy. (2017, January 5). *Myth-busting 3: Further improving industry communication with effective debriefings.* Washington, DC: Author.

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2013). *Open systems architecture (OSA) contract guidebook for program managers* (Version 1.1, Appendix 10). Washington, DC: Author.

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2016, April 1). *DOD source selection procedures*. Washington, DC: Author.

Appendix: Viability Assessment in Practice

DHS ICE Investigative Case Management System

The Department of Homeland Security (DHS) Immigration and Customs Enforcement (ICE) conducted a competitive acquisition to modernize an investigative case management information technology (IT) system, using viability assessments as part of their RFI and RFP process.

DHS ICE TECS Requirement

DHS ICE needed to modernize its mission-critical major IT development program, ICE Treasure Enforcement Communications System (TECS). This program provides the next-generation system for ICE Homeland Security Investigations (HSI) special agents.

Approach and Methodology

The RFI Process: DHS ICE released an RFI to industry through the Federal Business Opportunity (FEDBIZOPPS) website to solicit their input on potential contract types and performance-based requirements. DHS ICE issued a subsequent RFI to industry through Government Wide Acquisition Contracts (GWACs) and Multi Agency Contracts (MACs) to solicit capabilities and evaluate responses, using a specific set of questions and requirements for the desired capabilities. Because of this market research, DHS ICE decided to conduct a full and open competition for the TECS modernization.

The Multi-Step RFP Process: DHS ICE released RFP HSCETC-14R-0002 on May 2, 2014 and provided the ICE TECS system modernization objectives to modernize and simplify the case management technological infrastructure and improve data integration and information sharing. The DHS ICE conducted a public Industry Day on May 8, 2014, and provided the following acquisition strategy:

- HSCETC-14-R-00002 will be competed IAW FAR Part 15 as a full and open competition using a multi-step advisory down-select process.
- Offerors must participate in Step 1 (OCD) to be eligible to participate in Step 2 (Full Proposals).
- The full set of requirements will be provided to those offerors who participated in the Operational Capabilities Demonstration (OCD) and choose to proceed to Step 2 of the competition.

System Criteria: Within the RFP Section C, ICE defined the system by the following criteria:

- Mature COTS-based, web-enabled solution that can achieve delivering a production ready solution for formal integration testing for initial operation of the system
- System that can be configured or minimally customized to support unique requirements
- System that can integrate into the DHS and ICE enterprise infrastructure
- System that can interface with other specified systems that are internal and external
- System that can interface with the ICE Data Warehouse.

DHS ICE then required the offerors to provide, as part of their proposals, a Performance Work Statement (PWS) that would be evaluated and incorporated into the winning award.

Evaluation Process

DHS ICE RFP, Section M, included the following evaluation criteria:

M.3.1-STEP 1 (OPERATIONAL CAPABILITIES DEMONSTRATION) EVALUTION CRITERIA

Each Offeror's Operational Capabilities Demonstration (OCD) will be evaluated using the following factor to determine their current solution's viability:

FACTOR 1: Maturity of the Offeror's current system to meet requirements of OCD:

The Offeror's ICM system must demonstrate a high level of existing capability to meet the requirements of the OCD. Step 1 is meant to provide an advisory down-select to Offerors with the potential for success at an acceptable level of risk.

The system proposed by the Offeror during the OCD will be evaluated as to whether they have demonstrated the ability to meet the requirements of the preplanned scenario provided by the government, and whether they are likely to be a viable competitor in accordance with the criteria set forth below. These criteria will be used to rate each Offeror's OCD.

The degree to which the proposed investigative case management system demonstrates existing ("out of the box") capability to substantially meet the system capabilities of the preplanned scenario of the OCD with a potential for success in Step 2 at an acceptable level of risk. The Contracting Officer will notify each Offeror, in writing, by e-mail of the results of their OCD evaluation. Offerors who are notified that they are unlikely to be a viable competitor are encouraged to evaluate their likelihood of receiving an award and decision to continue to Step 2.

Any Offeror who provides an OCD in Step 1 may participate in Step 2.

M.3.2-STEP 2 EVALUTION CRITERIA

Proposals will be evaluated using the following three (3) factors to make a best value determination:

FACTOR 1: Technical: Sub-factor 1: Technical Approach; Sub-factor 2: Management Approach

FACTOR 2: Past Performance

FACTOR 3: Business & Price

DHS ICE RFP, Section L, included instructions on the OCD and the evaluation process:

L.3.1 OCD Instructions. HSCETC-14-R-00002 will be competed as a multi-step advisory down-select process. The requirements in the RFP posting are a sub-set of the entire requirements package which contains Law Enforcement Sensitive (LES) requirements. The requirements provided in the RFP are sufficient to allow Offerors to prepare for the Operational Capabilities Demonstration (OCD). The full set of requirements (including LES information) will be provided to those Offerors who participated in the OCD and choose to proceed to Step 2 of the competition. Offerors must participate in Step 1 to be eligible to receive the LES information and participate in Step 2.

L.3.2. STEP 1 OCD. To continue to Step 2 and receive the full list of requirements, each Offeror must perform an OCD to test and validate their system's maturity and capability in meeting the requirements of the ICE Investigative Case Management (ICM) System. These demonstrations will require each Offeror to execute a predefined scenario of critical capability and then allow the government operators an opportunity to execute additional tasks that encompass the same system capabilities as the predefined scenario. This scenario will require the Offeror to demonstrate system capabilities.

The result of Step 1 is an advisory down-select. After all OCDs have been conducted, the Offerors will be notified in writing as to whether they appear to be a viable competitor for Step 2. OCDs will be evaluated in accordance with Section M.2.2-STEP 1. Although all Offerors that participated in the OCD are eligible to participate in Step 2, Offerors who appear to be nonviable based on the evaluation of their OCD are provided an opportunity now to make the business decision as to whether it is in their best interest to continue to Step 2.

L.3.3 STEP 2 WRITTEN PROPOSALS. Law-Enforcement Sensitive Information: Those Offerors who choose to proceed to Step 2 shall submit an e-mail to ICE with their intent to participate in Step 2; this e-mail must be received by 2:00 pm EDT on the third business day of receipt of the viable/non-viable down-select letter. All Offerors will have 30 days from receipt of the LES information to provide Step 2 written proposals:

Results

DHS awarded the ICE ICM Solution contract on September 26, 2015, which was 60 days after receipt of Step 2 proposals. DHS made the contract award with no protests of the award decision.

Disclaimer and Distribution Statement

The views, opinions, and/or findings contained in this report are those of The MITRE Corporation and should not be construed as an official government position, policy, or decision, unless designated by other documentation.

Approved for Public Release; Distribution Unlimited. 17-4802-4

Panel 22. Enhancing Small Business Participation in Defense Markets

Thursday, May 10, 2018	
1:45 p.m. – 3:15 p.m.	**Chair: Emily Harman,** Director, Office of Small Business Programs, Department of the Navy ***The Impact of 8(a) Small Business Graduation*** David J. Berteau, Professional Services Council Amanda Swanson, Professional Services Council ***New Entrants and Small Business Graduation in the Market for Federal Contracts*** Samantha Cohen, The Center for Strategic and International Studies Andrew Hunter, The Center for Strategic and International Studies Gregory Sanders, The Center for Strategic and International Studies ***Examining Small Business Set Asides: Evidence and Implications for Small and Mid-Sized Suppliers in Federal Procurement*** Trevor L. Brown, The Ohio State University Amanda M. Girth, The Ohio State University

Emily Harman—is the Director, Office of Small Business Programs (OSBP), for the Department of the Navy (DoN), serving as chief advisor to the Secretary on all small business matters. She is responsible for small business acquisition policy and strategic initiatives.

Harman joined the Secretary of the Navy staff as a member of the Senior Executive Service in August 2015 and has over 30 years of federal service. Prior to receiving this appointment, she served as Associate Director of the Naval Aviation Systems Command's (NAVAIR's) OSBP from November 2005 to August 2015.

Harman's previous experience includes serving as a Division Director in the Major Weapons System for Air-Antisubmarine Warfare, Assault, Special Mission Programs Contracts Department, and as the Multi-Mission Helicopters Program Office's (PMA-299) Contracting Officer. Harman has NAVAIR experience as a Services Contracting Officer, as well as Contracting Officer for the AV-8B Weapon Systems Program Office (PMA-257).

Prior to joining NAVAIR in 1997, Harman served as a Contracting Officer for the Naval Supply Systems Command's (NAVSUP's) Fleet and Industrial Supply Center (FISC), Norfolk Detachment Washington. Harman served as a Supply Corps Officer in the Navy from 1985–1992 and retired from the Naval Reserves. She served onboard the USS *Emory S. Land* (AS-39) and earned the Supply Corps Surface Warfare pin. Her other duty stations include Supreme Allied Command Atlantic, Commander in Chief U.S. Atlantic Fleet, United States Naval Academy, and FISC Norfolk Detachment Washington.

Harman is a member of the DoD Acquisition Professional Community and is Level III certified in Contracting. A Certified Professional Contracts Manager through the National Contract Management Association, she holds a Bachelor of Science degree in Physical Science from the United States Naval Academy, and a master's degree in Management/Acquisition and Contract Management from the Florida Institute of Technology. Harman is a member of Leadership Southern Maryland's Class of 2010.

Harman is a graduate of NAVSUP's Corporate Management Development Program, NAVAIR's Senior Executive Leadership Development Program, and the Federal Executive Institute. Harman has a number of personal and command decorations including the DoN's Meritorious Civilian Service Medal, DoN's FY2010 Acquisition Excellence Award, and the 2015 Public Servant Award from the St. Mary's County Chamber of Commerce.

The Impact of 8(a) Small Business Graduation

David J. Berteau—is Professional Services Council (PSC) President and CEO, with 400 member companies of all sizes providing federal contract services. Berteau was ASD for Logistics and Materiel Readiness and served 14 years in the Defense Department, under six defense secretaries. Earlier, Berteau served at the Center for Strategic and International Studies (CSIS), Syracuse University's National Security Studies Program, and SAIC. He is a Fellow of the National Academy of Public Administration and taught graduate courses for 14 years at the Maxwell School, Georgetown, and the LBJ School. [berteau@pscouncil.org]

Amanda Swanson—serves as PSC's Research and Analysis Associate. She conducts federal spending analysis on the behalf of both member companies and PSC leadership to support the association's advocacy efforts. Swanson also works closely with PSC's senior staff to explore emerging issues in defense policy and acquisition. Previously, she worked as a research consultant with the United States Coast Guard. Swanson completed her Master of Public Policy at George Washington University in 2018 and her Bachelor of Arts in Linguistics at the University of Kansas in 2014. She is a native of Erie, CO. [swanson@pscouncil.org]

Abstract

The 8(a) small business set-aside program is designed to support small business participation in the federal market and to foster small business growth. However, 8(a) graduates do not fare well over time, with more than 60% no longer receiving federal prime contract obligations less than 10 years after graduation from set-aside eligibility. Those who were still federal prime contractors gained very little additional government business, with the average contract obligation up only 3.3% to $6.25 million, from $6.05 million (showing a decline when adjusting for inflation). Additionally, 8(a) graduates still depended on set-asides for more than half of their federal prime contract dollars. Overall, they are not rewarded for graduation.

Introduction

For decades, it has been the policy of the U.S. government to support and promote the growth of small businesses in the American economy. Part of that support comes from awarding government contracts to businesses that fall below certain size standards and are therefore eligible for special consideration. Constraining eligibility for contracts to certain businesses are sometimes called "set-aside" programs. One group of such set-aside programs is the 8(a) Business Development program under the U.S. Small Business Administration (SBA). Under 8(a), the federal government limits competition for certain contracts to businesses that participate in the 8(a) Business Development program.[1]

Qualified businesses that are certified to compete for 8(a) set-aside contracts do not retain that eligibility forever. Participating companies need to complete annual reviews to maintain good standing in the program, and a firm's certification will last for a maximum of nine years. At the end of that period, companies may still compete for and win government contracts, but they will no longer be eligible for 8(a) set-aside contract awards, though in certain circumstances they may later regain eligibility.

[1] For more information about 8(a) requirements and eligibility, see https://www.sba.gov/federal-contracting/contracting-assistance-programs/8a-business-development-program.

Through the SBA, the federal government tracks and reports on 8(a) set-aside contracts. What happens, however, to companies whose certification eligibility has ended? How successful are they in pursuing government contracts? Do they even stay in business? It appears that the SBA cannot answer these questions, because once a firm graduates from 8(a) eligibility, it's not tracked and reported on.

The Professional Services Council, under its Foundation, undertook an examination of these questions (illustrated in Figure 1). This paper describes the research question, the methodology, and the initial research results. It draws some preliminary conclusions and outlines further research needed.

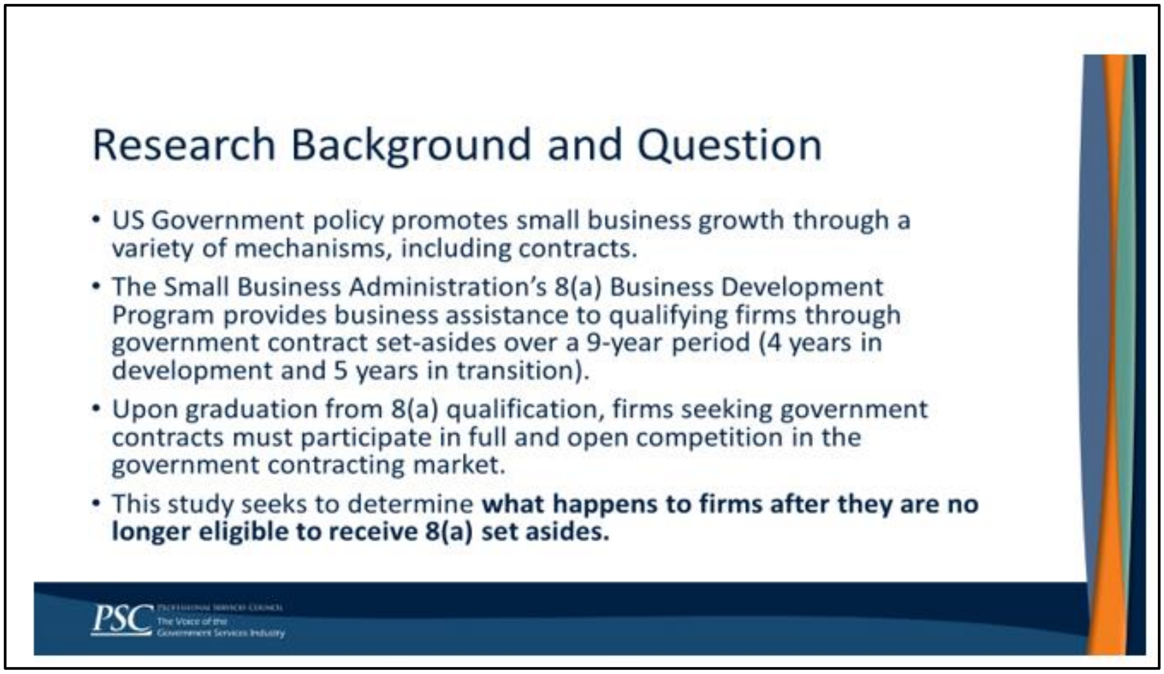


Figure 1. Research Background and Question

The 8(a) Program allows disadvantaged businesses[2] to compete for set-aside contracts, amongst other benefits. Firms may remain in the 8(a) program for up to nine years and may graduate from the program early in the case that they exceed size limits before the nine-year period of eligibility has expired.

The expiration of 8(a) eligibility for a company is commonly referred to as "graduation." From that point, previously certified firms may continue to pursue full and open competition for government contracts. This means they will need to be ready to compete with significantly larger and better resourced competitors. This study seeks to determine what happens to firms after they are no longer eligible to receive 8(a) set-asides.

Research specific to the 8(a) Business Development Program and its outcomes is relatively limited. However, there is a body of research that more broadly explores the role of

[2] "Disadvantaged" is defined by Title 13, Part 124 of the Code of Federal Regulations (CFR) and focuses on socially and economically disadvantaged individuals.

set-asides in the government contracting industry both in the United States and internationally.

A 2013 study by Jun Nakabayashi sought to estimate additional cost to government of procurement via set-asides for public construction projects in Japan. The study estimated that roughly 40% of small and medium contracting firms would fall out of the market if set-asides were taken away. The study concluded, therefore, that procurement cost would likely increase as competition decreased (Nakabayashi, 2013). That conclusion may require further analysis and verification.

Further, a 2011 comparison of small business contracting in the United States and in Europe found an increasing interest in fostering small business procurements in the European Union and some of its member countries (Kidalov, 2011). Small business set-asides and other means of supporting the competitiveness of these firms may be an important component of a healthy federal contracting market. However, that success can be jeopardized if, after graduation, those firms have difficulty remaining competitive or even staying in business.

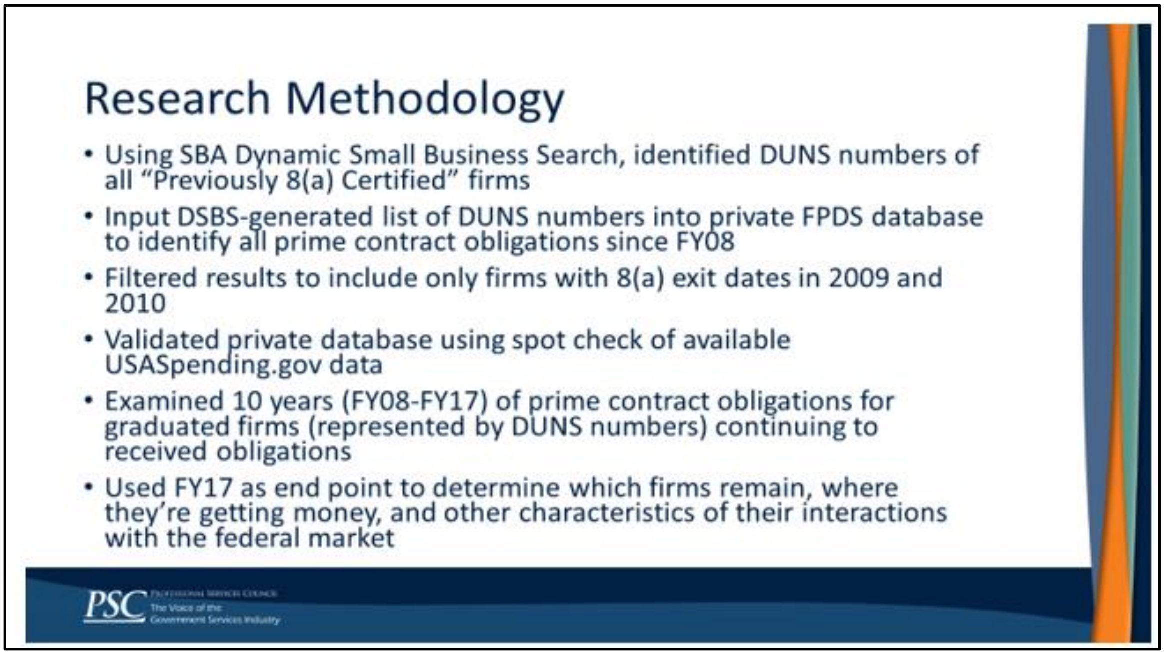


Figure 2. Research Methodology

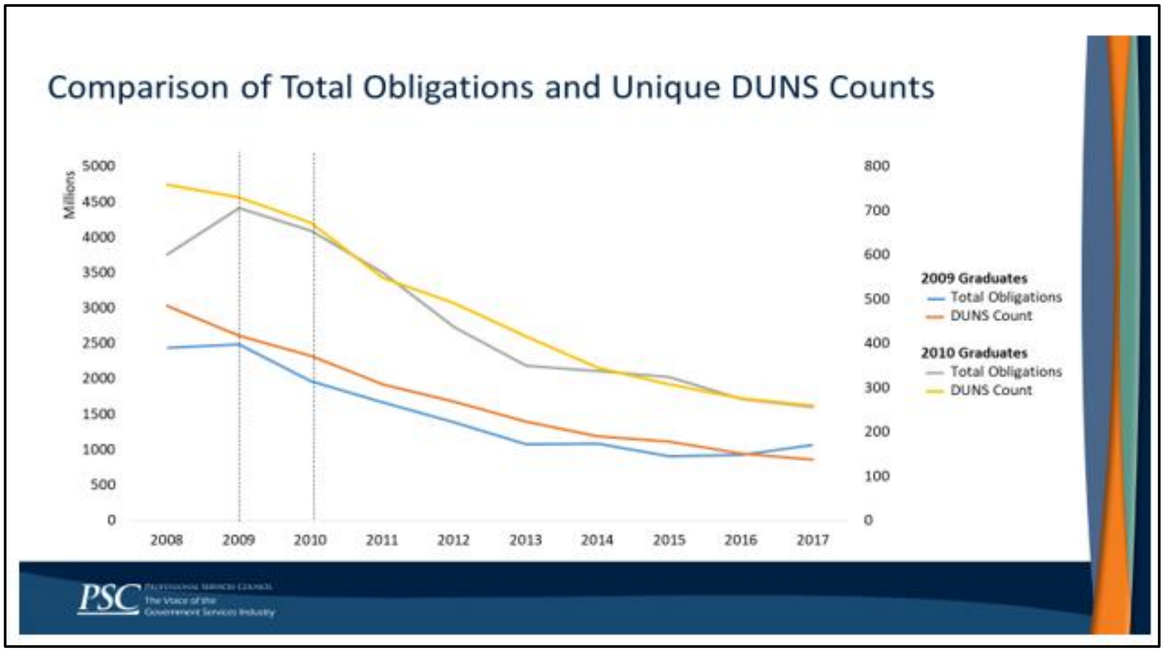
For purposes of this study, we narrowed our focus to those firms that graduated from 8(a) eligibility in 2009 and 2010. These two groups of firms will henceforth be referred to as "graduating classes," designated by their year of graduation.

We first identified firms of interest using the Dynamic Small Business Search (DSBS), furnished by the Small Business Administration. The search consisted only of previously 8(a) certified firms. The search did not include any firms with active 8(a) certifications. Due to the large number of firms that met this criterion, the DUNS numbers were gathered in groups of roughly five geographic states at a time. The resulting output from the DSBS included DUNS numbers and corresponding locations and 8(a) exit dates.

After compiling a complete list of previously 8(a) certified DUNS numbers, all relevant contract data since FY08 were queried within a proprietary database of federal obligations data, originally extracted from the Federal Procurement Data System (FPDS). To ensure the fidelity of the data, we spot-checked a sample of the selected DUNS numbers against obligations data available at USASpending.gov.[3]

We examined 10 fiscal years' worth of data for DUNS numbers continuing to receive obligations (Figure 3). We then used FY17 as an end point to examine who was still in the market, who their customers are, and other characteristics of their interactions with the federal market.

There are limitations when using obligations data provided by FPDS, with occasional errors and other inconsistencies within the data. Our work over the years with FPDS data shows, however, that such errors tend to be consistent over time and have little influence on long-term trends. Therefore, our analysis focuses on overarching trends within the data.



Note. This figure shows total contract obligations in current dollars, to the 2009 and 2010 graduates.

Figure 3. Comparison of Total Obligations and Unique DUNS Counts

One obvious metric for measuring the success of the 8(a) program is the number of firms remaining in the federal contracting market several years after their graduation from set-aside eligibility. There are, however, significant limitations to this approach. First, it is possible for a single firm to have more than one DUNS number, dependent on the services they provide and the customers they supply. It is also possible that a simple count of unique DUNS numbers will overestimate the number of firms participating in the market. Second, differences in spelling and format of vendor name entries in FPDS make it difficult to prevent overestimation using vendor names. Finally, there is nothing within FPDS data to indicate

[3] Because the data provided by USASpending are updated regularly, the obligations totals provided there and those provided by the proprietary database did not match exactly. However, the differences were minimal and deemed insignificant to the analysis.

why a DUNS number is no longer receiving prime contract obligations in later years. If a DUNS number is no longer receiving obligations after graduation, the firm may have failed and subsequently gone out of business, or it may have merged, been acquired, became only a subcontractor, or simply left the federal market. Further research will be required.

However, a count of unique DUNS numbers may provide insight into broader outcomes of the 8(a) program. The 2009 graduating class included 418 unique DUNS numbers. Members of the 2009 graduating class received nearly $2.5 billion in prime contract obligations in FY09. By FY17, only 138 unique DUNS numbers from the class of 2009 remained, representing a 67% decrease. In the same year, total obligations for this group had declined to $1.07 billion. Interestingly, this group experienced growth in total obligations over FY16, suggesting a potential rebound.

The 2010 graduating class included 673 unique DUNS numbers. These firms received roughly $4.1 billion in prime contract obligations in FY10. By FY17, only 291 of those unique DUNS numbers received prime contract obligations. At 57%, the attrition rate was slightly lower than that of 2009 graduates. Total prime contract obligations for the group declined to $1.61 billion in FY17 and did not show similar signs of rebounding.

There is a similar pattern in the decline in total prime contract obligations by class following graduation from 8(a) eligibility. For both graduating classes, the initial decline is steep, followed by a flattening out. Overall, only 39% of company-based DUNS numbers were still receiving prime contracts from the federal government in FY17. Contract obligations overall declined nearly as much, with FY17 totals being 41% of the funding in the year of graduation. In addition, the average size of contract obligations per DUNS number rose very little, from $6.05 million in graduation year to $6.25 million in FY17. Adjusted for inflation, the value of such obligations actually declined.

Total Obligations: Department of Defense

	2009	2011	2013	2015	2017
2009 Grads					
Army	737,864,519	356,425,288	214,620,537	105,488,964	122,804,981
Navy	302,180,864	205,579,187	141,215,994	113,880,067	105,556,071
Air Force	231,395,296	150,942,102	84,576,524	97,717,288	150,525,213
DoD Total	*1,433,090,507*	*831,879,516*	*505,882,192*	*401,129,564*	*516,657,411*
2010 Grads					
Army	1,289,416,635	878,477,779	454,728,895	505,338,917	415,838,489
Navy	815,069,314	588,838,673	362,027,167	269,982,826	255,715,927
Air Force	429,784,321	235,350,617	106,877,148	116,101,649	110,308,171
DoD Total	*2,680,988,601*	*1,868,888,722*	*990,488,557*	*946,492,795*	*822,010,449*

Note: Obligation amounts identified by "contracting agency"

PSC Professional Services Council The Voice of the Government Services Industry

Figure 4. Total Obligations: Department of Defense

In examining the general decline of 8(a) graduates, it is important to note that this decline does not occur uniformly across graduating classes or contracting agencies. Figure 4 shows trends within the Department of Defense (DoD). There are differences in how 8(a) graduates have fared in working with each military department. The seven-year decline was relatively uniform for 2010 graduates, between 10% and 11% per year for each military department. For 2009 graduates, the largest decline occurred in contracts with the Department of the Army, with a compound annual growth rate of -14.12% over the seven-year period from FY11 to FY17. Alternatively, the seven-year decline was less than 1% for contracts with the Department of the Air Force. Overall, 2009 graduates experienced a 6.58% annual decline in obligations from the DoD between FY11 and FY17.

Additionally, 2009 graduates experienced a significant rebound in prime contract obligations from FY15 to FY17, with the largest growth occurring in contracts with the Department of the Air Force at 15.49%. Contracts with the DoD grew 8.8% over the same period for 2009 graduates. The 2010 graduates did not experience the same rebound between FY15 and FY17, however it is possible that these firms may experience similar growth in FY18.

Total Obligations Amongst Other Agencies

	2009	2011	2013	2015	2017
2009 Grads					
Dept. of State	26,102,914	34,629,639	18,427,001	20,085,687	3,225,553
Veterans Affairs	23,702,739	15,130,194	19,458,063	7,763,811	3,373,248
Dept. of Energy	34,166,782	15,989,055	2,538,206	1,666,730	5,638,490
Other	967,001,895	777,594,461	533,311,172	482,901,276	542,147,095
2010 Grads					
Dept. of State	78,324,034	72,138,760	79,716,915	128,726,033	76,662,238
Veterans Affairs	89,524,637	54,443,534	96,202,468	88,899,068	28,881,349
Dept. of Energy	52,605,523	35,268,168	8,`173,652	19,792,959	36,706,706
Other	1,517,879,364	1,470,772,671	1,016,156,189	843,952,314	643,326,531

PSC Professional Services Council The Voice of the Government Services Industry

Figure 5. Total Obligations Amongst Other Agencies

Figure 5 shows declines contract obligations for 8(a) graduations from the Department of State, the Department of Veterans Affairs, and the Department of Energy. There is no uniform trend across these agencies.

For the Department of Energy, 2009 graduates experienced a seven-year decline of roughly 14%; they experienced 50% compound annual growth in the three-year period from FY15 to FY17. Prime contract obligations to 2010 graduates working with the Department of Energy were essentially flat from FY11 to FY17, however these firms experienced 23% in compound annual growth from FY15 to FY17.

The Department of State showed a significant decline in prime contract obligations to 2009 graduates over both the three-year and seven-year periods. However, though they experienced a decline from FY15 to FY17, contract obligations to 2010 graduates have remained flat at the Department of State over the seven-year period from FY11 to FY17.

Neither graduating class experienced growth in prime contract obligations with the Department of Veterans Affairs. The largest decline occurred for 2010 graduates between FY15 and FY17 at 31%. The 2009 graduates experienced a 24% decline over the same period.

The number of 2009 and 2010 8(a) graduates receiving federal prime contract obligations receiving federal prime contract obligations declined more than 60% by 2017, and the total value of their contracts by nearly the same amount. However, to understand the viability of 8(a) graduates, it is also important to consider the sources of their contract dollars. Figure 6 shows those sources for 2009 graduates.

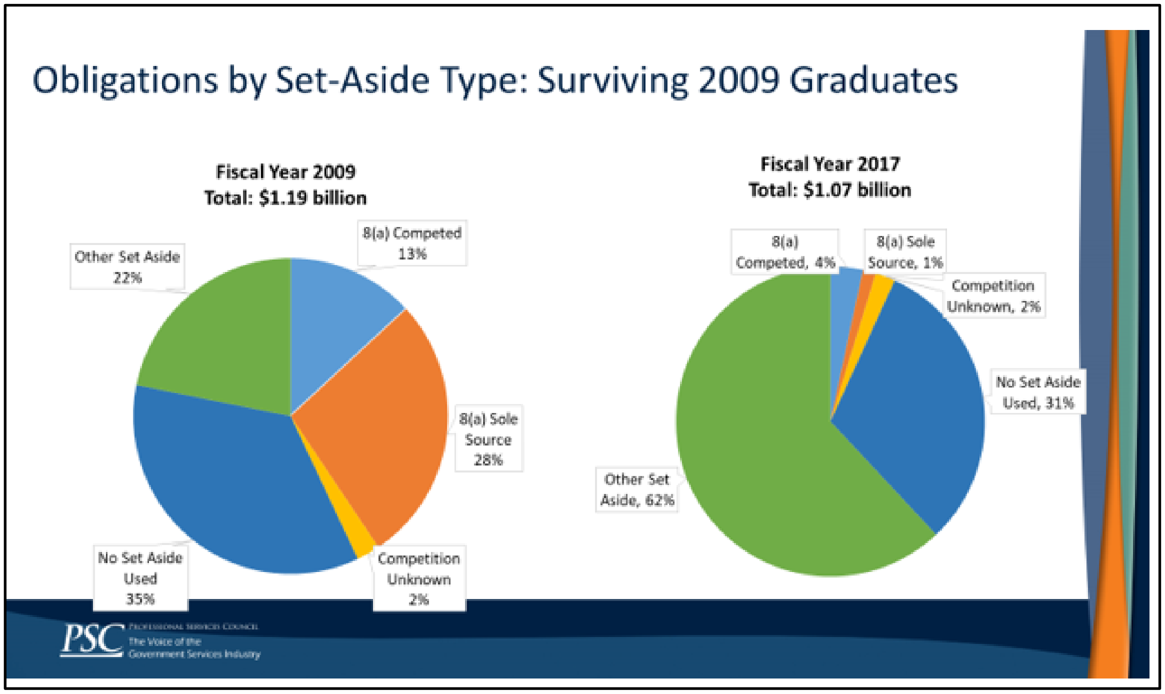


Figure 6. Obligations by Set-Aside Type: Surviving 2009 Graduates

The 2009 graduates received nearly 50% of their prime contract obligations from 8(a) program set-asides in their graduation year. It is also worth nothing, though, that roughly one-third of prime contract obligations for these firms came from full and open competition in FY09.

By FY17, however, the percentage of obligations for 2009 graduates from full and open competition remained the same as in FY09, roughly one-third. The proportion of obligations from other set-asides, on the other hand, more than tripled compared to FY09, from 19% to 62%. In other words, many of these businesses depended on other set-aside programs. A majority of the funds obligated through other set-asides came from general small business set-asides.

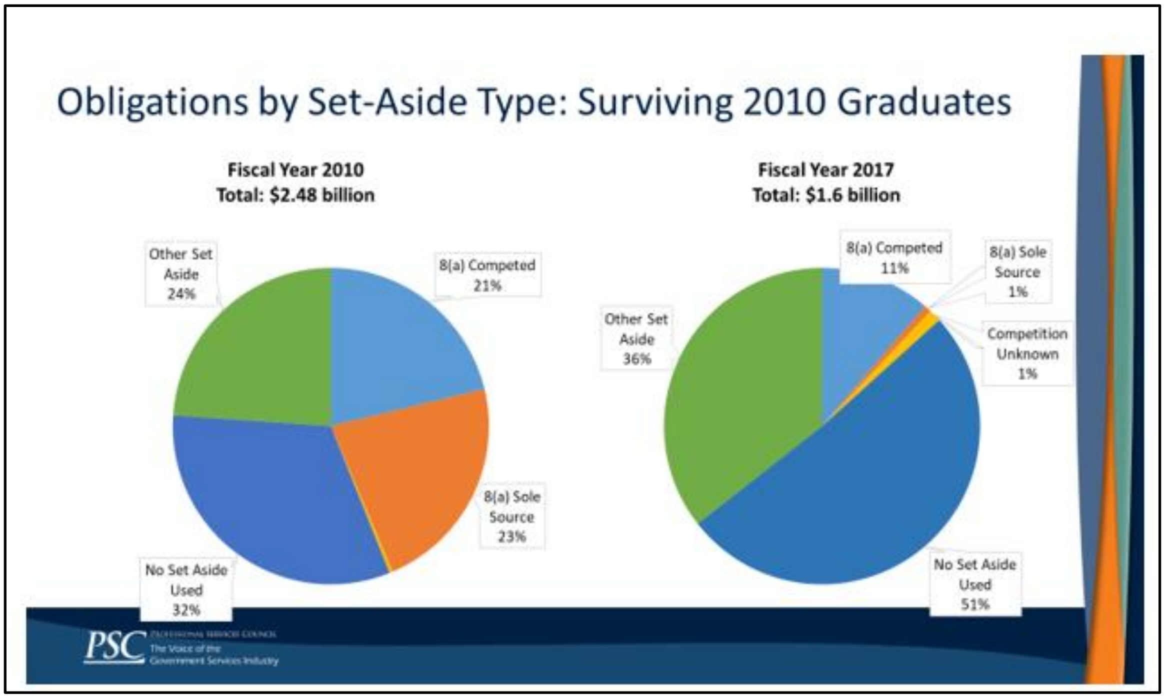


Figure 7. Obligations by Set-Aside Type: Surviving 2010 Graduates

In FY10, 44% of prime contract obligations for 2010 graduates came through 8(a) set-asides. This proportion is nearly identical to 2009 graduates in their graduation year. The 2010 graduates also received a near identical proportion of obligations through full and open competition in FY10. They received a slightly larger proportion through other set-asides.

Unlike 2009 graduates, however, by FY17 there was significant growth in the proportion of prime contract obligations awarded through full and open competition, from 32% in FY10 to 51% in FY17. Contract obligations from other set-asides rose 50%, a far smaller increase than for 2009 graduates.

Conclusion

The 8(a) small business set-aside programs are designed to support small businesses and foster their growth. However, 8(a) graduates do not fare well over time, with more than 60% no longer receiving federal prime contract obligations. Those who were still federal prime contractors grew very little, with the average contract obligation up only 3.3% to $6.25 million from $6.05 million (if we adjusted for inflation, the average would show a decline). Finally, 8(a) graduates still depended on set-asides for more than half of their federal prime contract dollars. Overall, they are not rewarded for graduation.

This suggests that, in general, even 8(a) graduates who are still receiving federal prime contracts are not growing. Further research is needed, particularly to identify which firms have grown and to determine the nature of that growth. Until then, it seems that 8(a) set-asides do not position graduates for growth.

References

Kidalov, M. (2011). Small business contracting in the United States and Europe: A comparative assessment. *Public Contract Law Journal, 40*(2), 443–509.

Nakabayashi, J. (2013). Small business set-asides in procurement actions: An empirical analysis. *Journal of Public Economics, 100*, 28–44.

Acknowledgements

The authors would like to thank Kevin Frankovich of Goodstone Research, LLC for his data contributions to this project.

New Entrants and Small Business Graduation in the Market for Federal Contracts

Samantha Cohen—is a Research Associate with the Defense-Industrial Initiatives Group at CSIS. Her work focuses on managing and analyzing data to identify relationships among policies, defense spending, and the related impacts on the United Sates and national security. Her recent research focuses on designing and managing international joint development programs, new entrants' survival rates and business graduation in the market for government contracts, and defense acquisition trends. Cohen holds a BS in economics from American University in Washington, DC, and an MS in economics from Katholieke Universiteit (KU) in Leuven, Belgium.

Andrew Hunter—is a Senior Fellow in the International Security Program and Director of the Defense-Industrial Initiatives Group at CSIS. From 2011 to 2014, he served as a senior executive in the DoD, serving first as Chief of Staff to Under Secretaries of Defense (AT&L) Ashton B. Carter and Frank Kendall, before directing the Joint Rapid Acquisition Cell. From 2005 to 2011, Hunter served as a professional staff member of the House Armed Services Committee. Hunter holds an MA degree in applied economics from Johns Hopkins University and a BA in social studies from Harvard University.

Greg Sanders—is a Fellow in the International Security Program and Deputy Director of the Defense-Industrial Initiatives Group at CSIS, where he manages a research team that analyzes data on U.S. government contract spending and other budget and acquisition issues. In support of these goals, he employs SQL Server, as well as the statistical programming language R. Sanders holds an MA in international studies from the University of Denver and a BA in government and politics, as well as a BS in computer science, from the University of Maryland.

Contributor

Marielle Roth—is a research intern with the Defense-Industrial Initiatives Group at CSIS. Prior to joining CSIS, she interned with the Department of Homeland Security (DHS) in the Office of Policy, focusing on DHS's PPBE system and opportunities for "jointness" within the department; START (National Consortium for the Study of Terrorism and Responses to Terrorism), researching patterns of domestic radicalization; and AT&T: Government Solutions, developing security applications of AI for use in the military. Roth holds an MA in security studies from Georgetown University and a BA in mathematics from Goucher College.

Abstract

This paper garners information crucial to understanding business growth for new entrants and small businesses who contract with the federal government. This information is then used to evaluate entrances, exits, and status changes among commercial and federal vendors with the purpose of comparing challenges faced by small businesses with those of larger ones. Measuring market trends over time and in multiple sectors shows how the challenges facing small businesses, such as market barriers to entry and imperfect competition, keep them from growing. The final results compare the survival rates between small and medium or large new entrants contracting with the federal government and analyzes the graduation rates for those small new entrants who grew in size during the observation period.

Introduction

Promoting small businesses has been a key issue for economic policy makers since the Great Depression occurred almost 90 years ago. This focus is not surprising given that small businesses have been referred to as the backbone of democracy, as their success unequivocally fosters an equal distribution of wealth (Bean, 1996). Furthermore, an entrepreneur's ability to create new companies and enter new markets is a signal of a healthy economy as the abundance and prosperity of small businesses and new entrants are clear indicators of market sustainability, improving both public and private interests. In recent years, small business policy makers have focused on emerging obstacles, especially for those businesses newly entering the heavily regulated market for federal contracts. For instance, the DoD's desire to access non-traditional vendors galvanized the inception of the DoD's Defense Innovation Unit Experimental (DIUx). Another example is Title 15 of U.S. Code § 657a, otherwise known as the HUBZone Program that provides set asides for qualified businesses that might not otherwise be able to effectively compete for federal contracts.

This paper studies entrances, exits, and status changes of six samples of newly-entered federal vendors and specifically DoD vendors. Each sample observes a set of new entrants in each year from 2001 to 2006 and how they fared over the following 10-year period. For example, the first sample looks at new entrants in 2001 and measures their success through 2011, while the last sample looks at how new firms in 2006 fared through 2016. The study team additionally investigates how these outcomes change between small and medium or large businesses.

There is a wide body of literature studying the ability for new entrants, and specifically small businesses, to survive in different industrial sectors. Scholars studying this issue have identified various industry-level, macroeconomic-level, and firm-level characteristics that affect new entrants' and small firms' ability to survive. In the context of public procurement, there is only a small amount of literature focusing on the relationship between small businesses and federal contracting. To break new ground in this critical but understudied domain, the study team observed a large longitudinal sample of firms that offers complete information on firm entries, firm exits, and other available firm-level characteristics.

The study team garnered firm-level information from the years 2000–2017. The study team posed four research questions to guide the exploration of the data:

1. What are the survival rates for new entrants in the market for federal contracts?
2. How do these survival rates compare with the survival rates for new entrants in the defense industrial base specifically?
3. How do these survival rates change between small and medium or large businesses?
4. What firm-level characteristics differentiate small from medium or large businesses?

This paper seeks to answer these questions by first reviewing the existing literature that studies new entrants' ability to survive and specifically how small businesses fare in this context. Second, by outlining the characteristics that have been found to shape a new entrant's ability to survive based on the literature. Third, by describing and analyzing the data that the study team gleaned from the Federal Procurement Data System (FPDS) and the System for Award Management (SAM). Finally, by offering a discussion of the results and drawing conclusions from the findings.

U.S. Government Policies and the Existing Literature

Small Businesses and New Entrants

Federal policies take a range of approaches to promote entrepreneurship, competitive markets, and small businesses. In addition to alleviating anti-trust threats and providing technical assistance, small-business policy aims to utilize public acquisition dollars as a tool for enhancing demand for small businesses in the market for federal contracts. Under the current policy, federal government agencies are subjected to an overall goal of spending 23% of their prime-contract dollars with small businesses (Moore, Grammich, & Mele, 2014). Generally, both public and academic spheres have acknowledged that the market for government contracts has high barriers to entry and can be a turbulent environment for small businesses, even once penetrated. The government responds to this market failure with various small-business set asides aimed to improve the relationship between small-business vendors and the government.

Given these theoretically favorable opportunities for small businesses in the market for federal contracts, it is worthwhile to study the chance of survival for new entrants in these markets. On the one hand, policy makers should be aware of the success rates for small businesses in the market for federal contracts to better adjust or implement policy when needed. On the other hand, small businesses who might utilize the policy advantages provided to them should be aware of the likelihood of success in certain markets before entering them.

While policy makers are concerned with making the market for government contracts accessible to small businesses, a 2008 survey found that when small businesses were asked to rank 75 problems in order, where 1 equals most concerning and 75 equals least concerning, small business participants listed being awarded a federal contract as 71st out of 75, on average (Kovacic, 1992). Whether small businesses view their participation in the market for federal contracts of high significance, well-rounded participation in this market is important so that the market for federal contracts remains healthy despite its monopsonistic and monopolistic nature. Without legal pressure, the market for federal contracts can become easily concentrated for a variety of reasons. First, many products and services bought by the federal government function at a large scope, making it difficult for small businesses to serve as a prime contractor for certain items (Kovacic, 1992). Second, barriers to entry in the market for federal contracts exist. For instance, navigating the highly regulated nature of federal contracting requires large structural and personnel investments by any businesses looking to sign a federal contract for the first time (Kovacic, 1992).

Given the historical priority placed by policy makers on both the amount of small business participation in, and the general health of, the market for federal contracts, the variables associated with successful small business contracting deserve empirical examination. Furthermore, the investigation of new entrants in the market for federal contracts goes hand in hand with the small business issue because previous research has shown that size impacts a new firm’s ability to succeed (Agarwal & Audretsch, 2001). Identifying which variables are associated with successful market participation will help small businesses and new entrants to target practices that enhance their ability to enter the market for federal contracts and further improve the health of the economy.

Variables Associated With New Entrants' Success

The current body of literature that studies the ability for new entrants to survive has identified three buckets in which the characteristics affecting survival rates of new entrants exist: firm-level characteristics, industry-level characteristics, and macroeconomic-level characteristics. This section discusses the existing literature's findings on these characteristics in support of the methods this paper uses to study small business new entrants in the market for federal contracts.

Firm-Level Characteristics

Size

The theories on how size affects new entrants' survival have evolved over time. Scholarship studying new firm survival initially accepted Gibrat's law, which states that firm survival and subsequent growth is independent of firm size (Agarwal & Audretsch, 2001). This law was challenged, however, by subsequent scholars studying small businesses and firm survival. For instance, Evans (1987); Hall (1987); Dunne, Roberts, and Samuelson (1988); Audretsch and Mahmood (1995); and Grammich et al. (2011) found that small firms have a higher likelihood of exiting the market compared to larger firms. Moreover, Geroski (1995) argues that the preponderance of support for the evidence that small firms are more likely to exit the market has become a stylized fact. The literature thereafter follows this view and as a result, includes variables measuring firm size when analyzing survival rates and growth for new entrants.

Size critically affects a business's ability to survive because small businesses are disadvantaged by their inability to operate at the minimum efficient scale level of output from the beginning (see discussion from Agarwal and Audretsch, 2001). Small firms experience a cost disadvantage compared to their larger, incumbent competitors and are therefore more likely to fail. In the context of public procurement, Flynn, McKevitt, and Davis (2015) find that within the definition of small businesses, there are further subsets of size that differentiate micro-businesses from small businesses in general and that these two groups tend to experience different survival and growth rates when participating in public tendering.

The literature review findings on the effect that size has on new entrants' survival rates provides the foundation for the study team's comparison of survival rates for small and medium or large firms. Analyzing this comparison in the context of federal contracting, with the dataset gleaned by the study team, is novel and looks to inform the public procurement community on the success of their efforts to improve the environment for new entrants and small businesses in federal contracting.

Firm Age

Firm age is an important variable in this analysis for two reasons. First, the association of firm age and survival of new entrants has been deemed as another stylized fact by Geroski (1995) and can also be found in analyses by Evans (1987) and Audretsch (1991). Geroski (1995) lays out a rationale for this phenomenon:

> Since the process of information acquisition is costly and time consuming, many new entrants are likely to under-invest in information gathering. Further, to the extent that market opportunities change post-entry, the types of actions which entrants need to undertake in order to survive and prosper are also likely to change. The implication is that the growth and survival prospects of new firms will depend on their ability to learn about their environment, and link changes in their strategy choices to the changing configuration of that environment.

In other words, new entrants can only know so much at the time of their entry into the market. It is necessary for these firms to spend time in the market to garner information that can be used and analyzed to improve business development. Some information can only be gleaned over time, making firm age an important variable to consider when analyzing new entrants' survival rates. While financially-robust new entrants might be better positioned to obtain information at an earlier stage, they will likely increase efficiency and capacity with age and experience.

Firm Ownership and Demographics

The study team incorporates the firm-level characteristic of firm ownership in the model estimating likelihood of survival because there are various small business policies issued by the U.S. government that create set asides depending on the nature of firm ownership. There are multiple categories of contracting assistance programs available to certain disadvantaged groups and locations. For example, the U.S. government created the 8(a) Business Development Program to aid small, disadvantaged businesses to participate in the market for federal contracts. The U.S. government has these policies in place due to the disadvantage that small, minority-owned businesses face in competing for federal contracts (Small Business Administration, n.d.).

Firm Nationality

The U.S. government has policies in place that regulate the content that federal contractors can procure from foreign sources. For instance, the Buy American Act (41 U.S. Code §§ 8301–8305) requires federal contracting agencies to prefer domestic materials and services for public use or public works in the United States (41 U.S. Code Chapter 83). Additionally, Buy American Laws such as the Berry and Kissel Amendments mandate nearly exclusive use of U.S. content in certain products. The Berry Amendment requires that the DoD specifically purchases certain items such as textiles, food, shoes, and hand/measuring tools exclusively from domestic suppliers, and the Kissel Amendment extends this to the Department of Homeland Security (Congressional Research Service, 2017). Due to these regulations, the study team explores the relationship between firm nationality and survival in the market for government contracts because these policies indicate that location could impact these relationships.

Industry-Level Characteristics

The characteristics that shape each industry create environments that have differing effects on the ability for new entrants to enter and survive. For this reason, policy makers and scholars who study new entrants account for the differing environments across industries. For instance, the SBA's definition of a small business varies depending on industry sector. Furthermore, scholars who have studied survival rates for new entrants tend to acknowledge these differences by implementing variables measuring industry-level characteristics that have been shown to affect a new entrant small business's likelihood of survival (Audretsch, 1991; Audretsch & Mahmood, 1995; Reijonen, Tammi, & Saastamoinen, 2016). Certain industry-level characteristics that the literature has focused on are degree of competition, innovation rate, industry growth rate, and capital intensity in an industry. While controlling for these industry-level characteristics is beyond the extent of this paper, future iterations will control for these factors by measuring the primary industry that a firm contracts in through NAICS identification.

Degree of Competition

As one of the pillars supporting a healthy market, the degree of competition impacts the conditions facing new entrants and their ability to survive in a market. Competitive markets provide more opportunity for growth, which enables firms to more easily reach the minimum efficient scale. Audretsch and Mahmood (1995) posit that risk is higher for new entrants when the cost-price margin of an industry is high and that this result increases in non-competitive markets that are highly concentrated because the larger incumbents have more control over price and supply. When measuring survival rates for new entrant small businesses, it is important to account for the degree of competition of the industry in which the new entrant participates.

Innovation Rate

The innovation rate in the industry entered by the firm is an important variable cited in the current literature as having an impact on small business net entrants' survival rates. Technological or informational conditions that dictate the amount of innovation necessary to succeed in an industry influence the ability for new entrants to survive in a market. This idea has been explored by Winter (1984) and Audretsch and Mahmood (1995). Winter (1984) finds that industries differ, with some operating as a "technological regime" and others as an "entrepreneurial regime." Industries characterized as a "technological regime" are more favorable to established incumbent firms who already have the capital and knowledge base to effectively innovate and survive. Conversely, "entrepreneurial regimes" foster innovative success by new entrants and small businesses, giving new entrants an innovative advantage over their incumbent competitors. Audretsch and Mahmood (1995) empirically test how hazard rates for new entrants depend on innovation rates. They estimate that new entrants face a higher risk of failure in highly innovative environments, although their results are not statistically significant.

Industry Growth Rate

Industry growth rates have been shown to affect survival rates because growth rates have been shown to increase price-cost margins (Bradburd & Caves, 1982). Like the industry characteristic degree of competition, industry growth rates influence the price-cost margins that in turn impact the operations of companies in that industry. Heightened price-cost margins create environments where participating firms can survive when operating at a suboptimal level of scale, thus influencing the ability for new firms to survive (Audretsch & Mahmood, 1995).

Capital Intensity

Theoretically, high capital intensity makes it harder for new entrants and especially small businesses to survive and grow in an industry. This is because it is more difficult to acquire the necessary resources needed to operate in a capital-intense environment before operating at the minimum efficient scale. Moreover, incumbent firms in capital-intense industries likely operate with economies of scale, giving them an advantage over newly-established competitors. On the one hand, Audretsch (1991) found that the likelihood of survival for small, newly-established firms is lower in capital-intensive industries that are dominated by scale economies. On the other hand, industries exhibiting high investments in human capital, with higher wages, are a reflection of the tendency to invest heavily in labor-related costs such as training and firm-specific skills, and tend to have a higher likelihood of survival (Audretsch & Mahmood, 1995).

Macroeconomic-Level Characteristics

The third and final set of characteristics that may influence a firm's likelihood of survival pertains to macroeconomic variables. The state of the economy influences business success across all levels of business size and thus must be controlled for when estimating the survival rates of new entrants. The point in time of the business cycle, the unemployment rate, and inflation rates all influence factors such as investment, GDP, employment, and demand. Previous work on this topic has acknowledged these relationships by including variables describing various macroeconomic characteristics, such as the unemployment and real interest rates, to control for these effects and estimate the impact of new entrant size on likelihood of survival more accurately.

Data and Specification

The study team collected the data for this report from the Federal Procurement Data System (FPDS) and the System for Award Management (SAM). The study team gleaned data on a yearly basis measuring a wide variety of variables on new entrants in the market for federal contracts from these two sources and merged them together by firm. The result is a longitudinal data set that provides information on firms entering and exiting the market on an annual basis over the period from 2000 to 2017.

The study team subsets the collected data to six analytical samples of new entrants in the market for both government-wide and DoD-specific contracts. Each sample includes all new entrants starting in year t where t = 2001–2006. Each sample is tracked over the decade following t. To define new entrants, the study team uses the registration date in SAM to indicate when a firm entered the market for federal contracts. To define exits, the study team uses the last signed date within the 10-year study period from FPDS.

Given the information on entries and exits, the study team calculates the i-year survival rates for each of the six samples of new entrants where i can equal three, five, or 10. The survival rate is equal to the number of firms that survived in the ith year divided by the total number of firms that entered in the baseline year. These calculations are made for all new entrants, small-business new entrants, and medium- or large-business new entrants. The same calculations are made for those new entrants contracting specifically with DoD. Furthermore, the study team calculates the graduation rates of small businesses specifically for each of the six samples. The study team considers small business graduation to occur through either organic firm growth or acquisition by a larger company. In this analysis, a small business is considered to graduate if, during the 10-year observation period, it changes from small-business status to medium- or large-business status for the majority of contract obligations that it has with the federal government after its first contract as a medium or large firm had been signed. The graduation rates are then calculated by dividing the number of graduated firms over the 10-year observation period by the total number of small firms that entered the market in the baseline year.

With respect to the firm-level variables, the study team uses firm size, firm location, and firm ownership status. Firm size is defined by the variable "Contracting Officer's Determination of Business Size" from the FPDS database, which denotes whether the contracting officer concluded that the firm satisfies the small business size criterion for the contract's assigned NAICS code. The firm location variable is binary, determining whether the firm is domestically or internationally located. This is defined by the "country" variable that is given from the SAM database. Firm ownership status is described by four variables: woman owned, veteran owned, minority owned, and foreign owned. Using the FPDS section on contractor data, woman owned is defined by the "Woman Owned Business" variable and veteran owned is defined by "Veteran Owned Business." Minority owned is defined as inclusion in any of the following categories: "Minority Owned Business," "American Indian Owned Business," "Alaskan Native Owned Corporation of Firm," "Native Hawaiian Owned Organization or Firm," "Tribally Owned Firm," "Asian-Pacific American Owned Business," "Black American Owned Business," "Native American Owned Business," "Subcontinent Asian (Asian-India) American Owned Business," "Hispanic American Owned Business," and "Other Minority Owned Business." Finally, foreign owned is defined by the field "Foreign Owned and Located." There were some cases where observations were dropped due to missing values in the firm-level characteristic fields; however, these numbers were never at high enough magnitude to raise alarm.

Results: New Entrants in the Market for Federal Contracts

The number of new entrants that entered the market for federal contracts and the market for DoD contracts specifically each year from 2001 to 2016 is reported in Figures 1 and 2. The overall trend for the entire time period is similar for new entrants in the market for all government contracts and for new entrants contracting with the DoD specifically, where the number of new entrants entering the market to contract with the government decreases each year on average. The trends from 2005 to 2016 are more constant, while there are much larger numbers of new entrants in the first four years. Interestingly, the number of new entrants for all federal agencies increases dramatically in 2004, yet this does not happen for new entrants contracting with the DoD. The high rate of decrease of new entrants contracting with the federal government in the beginning of the period of study is not sufficiently explained in this report, lending itself to future research. It could be due to factors such as a reporting phenomenon, policy implementation, or the state of the economy. The consistency of this phenomenon, with declines happening in the DoD sample even as contract spending is surging combined with significant overhauls of FPDS during the same period,[1] leads the study team to suspect that a change in the reporting of DUNS numbers most likely explains a significant portion of this phenomenon. Until the precipitous decline is better understood, the paper will focus on what trends within each sample rather than trying to explain the differences between years.

[1] FPDS-NG launched at the start of 2005, though a variety of data changes preceded the change. http://www.govexec.com/technology/2004/12/new-procurement-data-system-to-debut-at-end-of-month/18247/

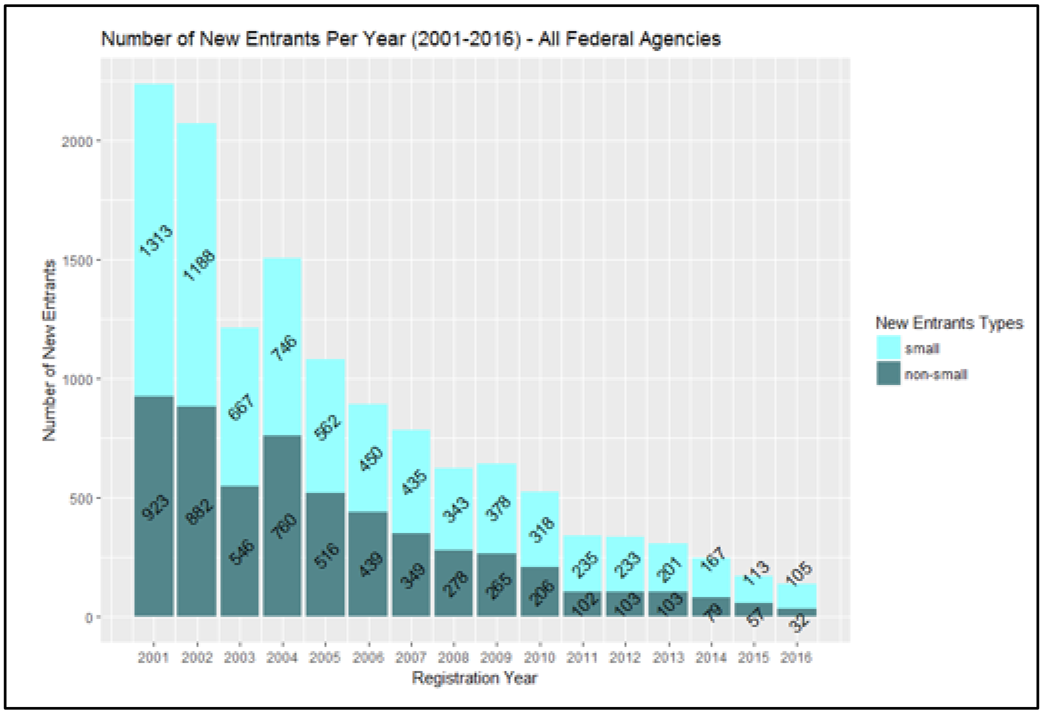


Figure 1. Number of New Entrants per Year (2001–2016)—All Federal Agencies

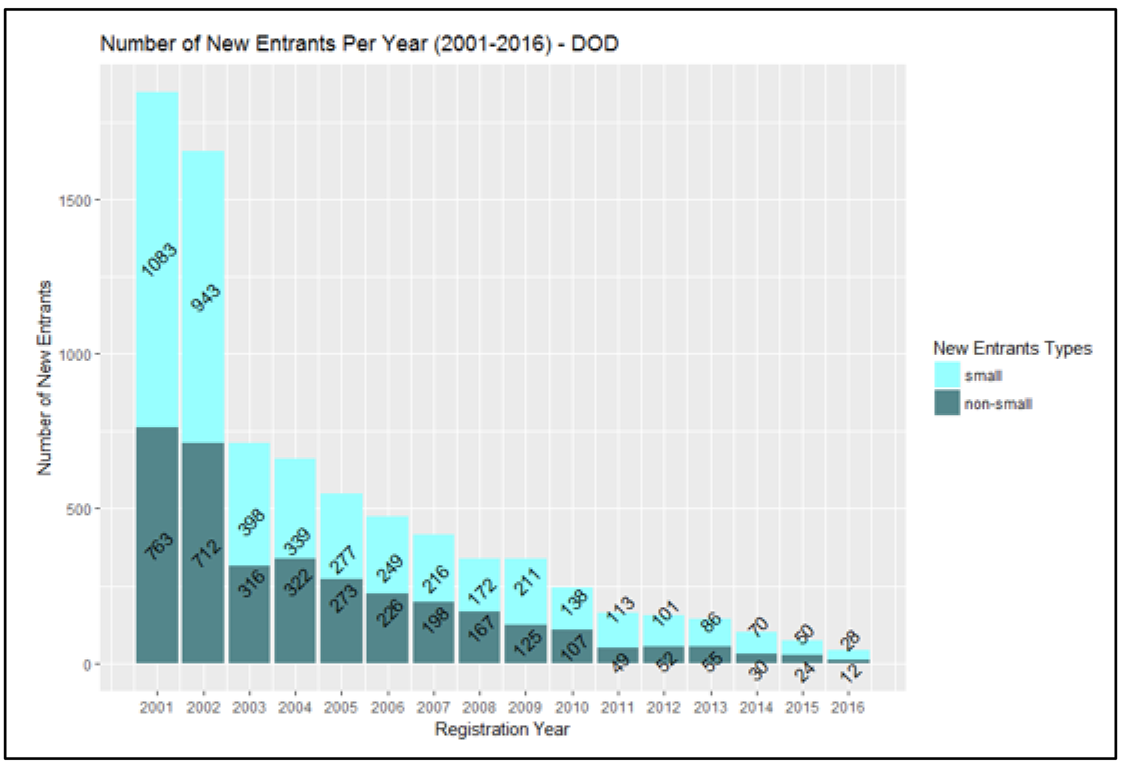


Figure 2. Number of New Entrants per Year (2001–2016)—DoD

2001 Sample of New Entrants

The differences in firm-level characteristics between small and medium or large vendors is displayed in Table 1. Small firms were around 14 years younger than their medium or large competitors, on average. Nearly all small and medium or large firms were domestically located, although the difference in means is significantly different from zero. Small businesses tended to have higher rates of minority, veteran, and woman owners, on average.

The sample of businesses that started contracting with the federal government in 2001 had relatively high three-year survival rates, where the survival rate for all new entrants after three years was 78%. The small businesses in this sample had lower survival rates than the medium or large firms, with three-year survival rates of 76% for small businesses and 81% for medium or large firms. After five years, the survival rates decrease to 62%, 61%, and 64% for all firms, small firms, and medium or large firms, respectively. The 10-year survival rates decrease by a much greater margin, and small new entrants have a higher survival rate than medium or large new entrants, at 20% and 19%, respectively. The graduation rate for small businesses contracting with all federal agencies is approximately 4%.

The new entrants that contract with the DoD have similar results in that the difference in survival rates between firms at three years and five years is less severe than it is between firms at five years and 10 years. Small new entrants contracting with the DoD also have lower three- and five-year survival rates than medium or large new entrants. After 10 years, however, small new entrants have a higher survival rate than their medium or large counterparts, although the difference is hardly over one percentage point. The graduation rate for small businesses contracting with DoD is 2.4%.

Table 1. 2001 Mean Firm-Level Characteristics Between Small and Medium or Large Firms

	Small	Medium or large	T-Stat Ha: diff≠0
Firm Age	19.12	33.46	-10.51***
Foreign Owned	0	0	-1.73
Woman Owned	0.1	0.012	9.74***
Veteran Owned	0.05	0.01	7.05***
Minority Owned	0.06	0.01	7.82***
Domestic Location	0.99	0.95	5.86***
*Significance Level: *p<.1 **p<.05 ***p<.01*			

Table 2. 2001 New Entrants' Survival Rates

All Federal Agencies				DoD			
Observations: 2237	**All New Entrants**	**Small New Entrants**	**Medium or Large New Entrants**	**Observations: 1846**	**All New Entrants**	**Small New Entrants**	**Medium or Large New Entrants**
3-Year	78.01%	76.09%	80.72%	**3-Year**	71.56%	69.81%	74.05%
5-Year	62.27%	61.31%	63.60%	**5-Year**	55.47%	55.31%	55.70%
10-Year	19.58%	20.49%	19.16%	**10-Year**	13.98%	15.14%	12.32%
Graduation Rate		3.53%		**Graduation Rate**		2.38%	

Source: FPDS and SAM

2002 Sample of New Entrants

The mean differences in firm-level characteristics for the 2002 sample of new entrants are displayed in Table 3 and follow very similar patterns to the 2001 sample. Additionally, the 2002 sample of new entrants (displayed in Table 4) exhibit similar results to the 2001 sample. The three-year survival rate is relatively high, at around 75%. Different from the 2001 sample, small and medium or large new entrants in 2002 show nearly equal three-year survival rates, varying within 1 percentage point of one another. Both small and medium or large new entrant's five-year survival rate reduces by about 15 percentage points compared with the three-year rates, but the difference between the two groups remains at less than 1%. The 10-year survival rate is much smaller for both small and medium or large new entrants, and small new entrants have a higher survival rate than medium or large new entrants after 10 years. The graduation rate for small businesses contracting with the all federal agencies in 2002 is 3%.

Small new entrants contracting with the DoD have slightly higher three-, five-, and 10-year survival rates than medium or large new entrants in 2002. Approximately 70% of all new entrants survive after three years, with 54% surviving five years, and about 15% of the 2002 new entrants still alive after 10 years. These rates are slightly lower than the survival rates for the 2002 sample of new entrants contracting with all federal agencies. The graduation rate for small new vendors over the decade after they entered in 2002 is 2%.

Table 3. 2002 Mean Firm-Level Characteristics Between Small and Medium or Large Firms

	Small	Medium or Large	T-Stat Ha: diff≠0
Firm Age	19.69	32.25	-9.65***
Foreign Owned	0	0	N/A
Woman Owned	0.11	0.01	9.79***
Veteran Owned	0.05	0	6.97***
Minority Owned	0.07	0.01	7.94***
Domestic Location	0.99	0.92	7.69***
*Significance Level: *p<.1 **p<.05 ***p<.01*			

Table 4. 2002 New Entrants' Survival Rates

All Federal Agencies				DoD			
Observations: 2070	All New Entrants	Small New Entrants	Medium or Large New Entrants	Observations: 1655	All New Entrants	Small New Entrants	Medium or Large New Entrants
3-Year	75.41%	75.17%	75.74%	**3-Year**	70.21%	70.84%	69.38%
5-Year	60.19%	60.69%	59.52%	**5-Year**	54.08%	55.14%	52.67%
10-Year	20.82%	22.22%	18.93%	**10-Year**	16.44%	17.39%	15.17%
Graduation Rate		3.24%		**Graduation Rate**		2.11%	

Source: FPDS and SAM

2003 Sample of New Entrants

On average, medium or large new entrants were approximately 13 years older than their small firm competitors in 2003. Additionally, small firms had higher levels of woman, veteran, and minority owners (see Table 5). The 2003 sample of new entrants contracting with all federal agencies have lower survival rates than the previous two samples examined for all year categories. The survival rates of the 2003 sample do not vary by a high magnitude between small new entrants and medium or large new entrants in the three- or five-year categories. However, the 10-year survival rate for small new entrants is just over 4 percentage points higher than that of medium or large new entrants, at 13% for small new entrants and 9% for medium or large new entrants. The graduation rate for the 2003 sample of small new entrants is 1.1%.

The 2003 sample of new entrants who contract specifically with the DoD shows similar changes between the three-, five-, and 10-year rates to those for all federal agencies, although the rates are lower across the board. Again, small new entrants have higher survival rates than their medium or large competitors, but the differences between the two groups is never larger than two percentage points. The graduation rate for those small businesses who started contracting with the DoD in 2003 is 0.42%.

Table 5. 2003 Mean Firm-Level Characteristics Between Small and Medium or Large Firms

	Small	Medium or Large	T-Stat Ha: diff≠0
Firm Age	17.07	30.77	-7.25***
Foreign Owned	0	0.01	-2.01*
Woman Owned	0.13	0.01	8.45***
Veteran Owned	0.05	0	5.83***
Minority Owned	0.11	0.01	7.64***
Domestic Location	0.99	0.83	9.37***
*Significance Level:*p<.1 **p<.05 ***p<.01*			

Table 6. 2003 New Entrants' Survival Rates

All Federal Agencies				DoD			
Observations: 1214	All New Entrants	Small New Entrants	Medium or Large New Entrants	Observations: 714	All New Entrants	Small New Entrants	Medium or Large New Entrants
3-Year	62.52%	62.67%	62.27%	**3-Year**	56.16%	56.53%	55.70%
5-Year	46.79%	47.83%	45.60%	**5-Year**	41.60%	42.71%	40.19%
10-Year	12.03%	13.94%	9.71%	**10-Year**	8.26%	8.79%	7.59%
Graduation Rate		1.07%		**Graduation Rate**		0.42%	

Source: FPDS and SAM

2004 Sample of New Entrants

The mean differences of firm-level characteristics between small and medium or large firms can be viewed in Table 7 and follow the same patterns as the previous three samples. The survival rates for the 2004 sample of new entrants (see Table 8) continue the trend of decreasing consistently from the three-year rate to the 10-year rate. However, a new trend exhibited by the 2004 sample shows the difference in survival rates between small and medium or large new entrants is much greater than the three previous samples. For instance, the three-year survival rate for small new entrants is 6 percentage points higher than medium or large firms, while the five-year rate for small new entrants is about 11 percentage points higher, and the 10-year survival rate for small new entrants is approximately 6 percentage points higher. The graduation rate for the 2004 sample is 1.1%.

The 2004 survival rates of new entrants who contract specifically with the DoD follow similar patterns but are across the board lower than those of new entrants contracting with all federal agencies. Small new entrants contracting with the DoD have higher survival rates than their medium or large competitors across all year categories. Additionally, the lowest survival rate out of all samples examined thus far occurs for medium or large new entrants contracting with the DoD after 10 years, at approximately 4%. On average, 0.91% of newly entered vendors contracting with the DoD in 2004 survive after 10 years.

Table 7. 2004 Mean Firm-Level Characteristics Between Small and Medium or Large Firms

	Small	Medium or Large	T-Stat Ha: diff≠0
Firm Age	16.65	29.35	-7.71***
Foreign Owned	0	0.01	-1.88*
Woman Owned	0.1	0.01	7.85***
Veteran Owned	0.06	0.01	6.22***
Minority Owned	0.09	0.02	6.49***
Domestic Location	0.99	0.84	10.63***
*Significance Level: +p<.15 *p<.1 **p<.05 ***p<.01*			

Table 8. 2004 New Entrants' Survival Rates

All Federal Agencies				DoD			
Observations: 1507	**All New Entrants**	**Small New Entrants**	**Medium or Large New Entrants**	**Observations: 661**	**All New Entrants**	**Small New Entrants**	**Medium or Large New Entrants**
3-Year	58.79%	61.93%	55.66%	**3-Year**	55.07%	60.56%	49.85%
5-Year	41.61%	47.05%	36.32%	**5-Year**	37.07%	43.79%	30.68%
10-Year	10.02%	13.27%	6.84%	**10-Year**	7.11%	10.25%	4.13%
Graduation Rate		1.13%		**Graduation Rate**		0.91%	

Source: FPDS and SAM

2005 Sample of New Entrants

Table 9 displays the mean differences in firm-level characteristics between small and non-small firms who became federal vendors in 2005. These results follow similar patterns to the previous samples; however, the mean difference in age between small and medium or large firms is at its highest in absolute value. The 2005 survival rates, shown in Table 10, follow the same trends as the 2004 sample, where small new entrants have higher survival rates for all three years examined. Again, the magnitude of difference is relatively high—small entrants have 7, 11, and 5 percentage point higher survival rates for three-, five-, and 10-years, respectively. The 2005 sample of new entrants has the lowest three-year survival rates out of all the samples examined thus far, at 54%, 58%, and 50% for all new entrants, small new entrants, and medium or large new entrants, respectively. On average, 0.93% of newly entered small firms contracting with all federal agencies in 2005 graduated from small-firm status.

The new entrants who entered the market in 2005 and contract with the DoD exhibit very similar results to those new entrants who entered the market in 2005 and contract with all federal agencies. Just over half of the new entrants survive after three years, around 40% survive after five years, and less than 10% survive after 10 years. Small new entrants tend to survive at higher rates than medium or large new entrants, much like the other samples examined. For those new businesses working with the DoD in 2005, 0.73% of them graduated from small-business status in the 10-year observation period.

Table 9. 2005 Mean Firm-Level Characteristics Between Small and Medium or Large Firms

	Small	Medium or Large	T-Stat Ha: diff≠0
Firm Age	15.61	31.53	-8.34***
Foreign Owned	0	0.02	-3.02***
Woman Owned	0.09	0.01	5.90***
Veteran Owned	0.07	0	5.92***
Minority Owned	0.13	0.03	6.64***
Domestic Location	0.98	0.72	12.92***
*Significance Level: +p<.15 *p<.1 **p<.05 ***p<.01*			

Table 10. 2005 New Entrants' Survival Rates

All Federal Agencies				DoD			
Observations: 1078	All New Entrants	Small New Entrants	Medium or Large New Entrants	Observations: 550	All New Entrants	Small New Entrants	Medium or Large New Entrants
3-Year	54.17%	57.65%	50.39%	**3-Year**	54.36%	58.48%	50.18%
5-Year	40.82%	46.26%	34.88%	**5-Year**	38.55%	45.85%	31.14%
10-Year	8.16%	10.68%	5.43%	**10-Year**	7.45%	9.39%	5.49%
Graduation Rate		0.93%		**Graduation Rate**		0.73%	

Source: FPDS and SAM

2006 Sample of New Entrants

The final sample of new entrants studied are those that entered the market in 2006. The mean differences in firm-level characteristics between small and non-small vendors that began federal contracting in 2006 follow the same patterns as the other five samples (see Table 11). Moreover, Table 12 shows that the survival rates of this sample for those contracting with all federal agencies are very similar to the sample studying new entrants that entered the market in 2005. About 60% of small new entrants survive after three years, while slightly over 50% of medium or large new entrants survive after three years. The difference in the five-year survival rate between small and medium or large entrants in 2006 is the largest across all years sampled, at 15 percentage points. Small new entrants have higher survival rates for all three-year categories. On average, 0.56% of the small firms that started contracting with the federal government in 2005 graduated.

The new entrants that entered the market in 2006 and contract with the DoD follow similar patterns in their survival rates to those that contract with all federal agencies. The 10-year survival rate for medium or large new entrants is the lowest of all samples studied, at around 4%. Consistent with every year sampled, small new entrants contracting with the DoD have higher survival rates than their medium or large competitors. The five-year survival rate for small new entrants is much higher (almost 17%) than the five-year survival rate for medium or large new entrants. For those new businesses working with the DoD in 2006, 0.42% of them graduated from small-business status in the 10-year observation period.

Table 11. 2006 Mean Firm-Level Characteristics Between Small and Medium or Large Firms

	Small	Medium or Large	T-Stat Ha: diff≠0
Firm Age	14.11	29.99	-7.33***
Foreign Owned	0	0.05	-4.25***
Woman Owned	0.12	0	7.59***
Veteran Owned	0.07	0.01	4.90***
Minority Owned	0.15	0.02	6.91***
Domestic Location	0.98	0.71	12.04***
*Significance Level: +p<.15 *p<.1 **p<.05 ***p<.01*			

Table 12. 2006 New Entrants' Survival Rates

All Federal Agencies				DoD			
Observations: 889	All New Entrants	Small New Entrants	Medium or Large New Entrants	Observations: 475	All New Entrants	Small New Entrants	Medium or Large New Entrants
3-Year	56.47%	61.11%	51.71%	**3-Year**	53.05%	59.84%	45.58%
5-Year	42.52%	50.00%	34.85%	**5-Year**	38.95%	46.99%	30.09%
10-Year	6.52%	8.67%	4.33%	**10-Year**	5.89%	7.63%	3.98%
Graduation Rate		0.56%		**Graduation Rate**		0.42%	

Source: FPDS and SAM

Discussion

The above results show a severe decline in the number of new entrants entering the market each year from 2001 to 2016. This result is surprising and merits further attention. The study team hypothesizes two factors that could be influencing this result. The first factor that could be significantly decreasing the number of new entrants contracting with the federal government each year is the reporting practices of Dun and Bradstreet or SAM. Second, the large reduction in new entrants working with the government over the study period could be purely due to an outside factor, such as policy or economic conditions. With this in mind, an analysis of the above survival rate results shows that when contracting with the federal government, new entrant small businesses tend to have higher survival rates than their medium or large competitors over three-, five-, and 10-year periods. A similar pattern persists for those new entrants contracting specifically with the DoD. As previously discussed, these results may be an outcome of the U.S. government's small business policies.

The implications of this result are multifaceted. On the one hand, these results suggest that small business policy successfully aids newly entered small businesses because they tend to survive at higher rates than newly entered medium or large firms. On the other hand, this could imply that small businesses face a perverse incentive regarding their business model. Growing firms produce the most jobs and provide more competition because they have reached minimum efficient scale for a wide range of products and services, fulfilling one of the goals of the small business promotion system. However, if the likelihood of survival in the market for federal contracts decreases as a firm grows, newly entered firms contracting with the federal government might not pursue a business model for profit maximization through growth because they would lose their small-business set aside privileges, inhibiting their ability to contract with the government.

Policy makers should pay attention to these perverse incentives when working with small businesses. These results imply that the small business policy that aims to aid small businesses in contracting with the government could be successful; however, the benefits of these policies may be exclusively limited to companies that stay small. Consequently, highly consolidated sectors where the government is reliant on a small number of large businesses, especially a risk for the DoD, might be cut off from a potential source for new competitors, as graduation from small business status is a major obstacle for most firms, who cannot compete with competitors like the big 5 for government contracts without the support of policy.

Compared to the existing body of literature studying the success between small business and medium or large business new entrants, these results are surprising. As previously discussed, the existing literature found that small businesses tend to have lower rates of success as new entrants than their medium or large competitors in different industrial sectors. The key difference here is the focus on federal contracts, and although the study team at this time cannot conclusively state if this change in small business new entrant success is due to contracting with the federal government, it will be a focus of future research.

Furthermore, the results should be taken into consideration with the following limitations in mind. First, these results paint a purely descriptive picture of the success rates for small and medium or large businesses contracting with the federal government. In other words, the calculation of the survival rates fails to control for other factors that could contribute to the success or failure of new entrants contracting with the federal government. Therefore, the reported results could be biased and an outcome of other factors not considered. The study team intends to address this limitation in future work by modeling the hazard rates of new entrants over time. Second, and as previously discussed, the study team is suspicious of potential reporting errors that might be a contributing factor to the large drop-off in new entrants in the beginning of the study period. The study team plans on working towards investigating the data further by cross-referencing with internal data sources and speaking with external experts.

Next Steps

The study team will continue this investigation through a variety of paths. First, the study team is working towards strengthening the statistical capacity of the calculations by examining survival between small and medium or large new entrants through a proportional hazard model. This model will allow the study team to control for the various firm-level factors that could contribute to a new entrant's success, regardless of whether that firm is small or not, as shown by previous research on this issue. Furthermore, this will allow for the control of the industry-level characteristics that, as determined by the existing literature, influence small and medium or large firms' ability to survive, depending on what industry they are operating in. By expanding the analysis in this way, the study team will be better positioned to draw conclusions regarding which factors contribute to the differences between small and medium or large new entrants' survival rates as vendors with the federal government.

Second, the study team has identified further areas of exploration that could productively contribute to the analysis of contracts for small business new entrants and their medium or large competitors when working with the federal government. Primarily, calculating the percent of dollar obligations that go to surviving firms, exiting firms, and graduating firms will increase the reliability of the study's results. These calculations will be made for existing firms in each sample period so that the proportion of work contracted to new entrants by dollar amount is known. Additionally, the study team will explore the survival rates of all existing firms and compare these to the rates calculated for new entrants to further increase the reliability of the study.

Finally, the study team will investigate the 2001–2002 phenomenon that was only lightly touched on in this paper, and will attempt to explain why there are significantly more new entrants contracting with the federal government than compared to the other years examined. Two areas of exploration could confirm whether these high counts are accurate relative to the rest of the observed period; data reporting practices and policy implementations will be examined as possible contributing factors to the phenomenon. The results of this research will be reported in a subsequent technical report that CSIS will publish later this year.

References

15 U.S. Code § 657a, HUBZone program. Retrieved from https://www.law.cornell.edu/uscode/text/15/657a

41 U.S. Code Chapter 83, BUY AMERICAN. Retrieved from https://www.law.cornell.edu/uscode/text/41/subtitle-IV/chapter-83

Agarwal, R., & Audretsch, D. B. (2001). Does entry size matter? The impact of the life cycle and technology on firm survival. *Journal of Industrial Economics, 49*(1), 21–43. Retrieved from http://www.jstor.org/stable/3569744

Audretsch, D. B. (1991). New-firm survival and the technological regime. *The Review of Economics and Statistics, 73*(3), 441–450.

Audretsch, D. B., & Mahmood, T. (1995). New firm survival: New results using a hazard function. *The Review of Economics and Statistics, 77*(1), 97–103. Retrieved from http://www.jstor.org/stable/2109995

Bean, J. J. (1996). *Beyond the broker state: Federal policies toward small business, 1936–1961*. University of North Carolina Press.

Bradburd, R. M., & Caves, R. E. (1982). A closer look at the effect of market growth on industries' profits. *The Review of Economics and Statistics, 64*(4), 635–645.

Congressional Research Service. (2017). *Defense primer: The Berry and Kissell Amendments.* Retrieved from https://fas.org/sgp/crs/natsec/IF10609.pdf

Dunne, T., Roberts, M. J., & Samuelson, L. (1988). *Patterns of firm entry and exit in U.S. manufacturing industries.* Wiley on behalf of RAND (Vol. 19).

Evans, D. S. (1987). The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *Journal of Industrial Economics, 35*(4), 567–581.

Flynn, A., McKevitt, D., & Davis, P. (2015). The impact of size on small and medium-sized enterprise public sector tendering. *International Small Business Journal, 33*(4), 443–461. https://doi.org/10.1177/0266242613503178

Geroski, P. A. (1995). What do we know about entry? *International Journal of Industrial Organization, 13*, 421–440. Retrieved from https://ac-els-cdn-com./016771879500498X/1-s2.0-016771879500498X-main.pdf?_tid=d18e6fec-bb2b-11e7-b27d-00000aacb362&acdnat=1509118433_6e9358e50c2341b4a5ffd9889f58f149

Grammich, C. A., Edison, T., Moore, N. Y., & Keating, E. G. (2011). *Small business and defense acquisitions: A review of policies and current practices*. Santa Monica, CA: RAND.

Hall, B. H. (1987). The relationship between firm size and firm growth in the U.S. manufacturing sector. *Journal of Industrial Economics, 35*(4), 583–606.

Kovacic, W. E. (1992, February). Regulatory controls as barriers to entry in government procurement. *Policy Sciences, 25*(1), 29–42. Retrieved from http://www.jstor.org/stable/4532242

Moore, N. Y., Grammich, C. A., & Mele, J. D. (2014*). Small business and strategic sourcing lessons from past research and current data.* Santa Monica, CA: RAND.

Reijonen, H., Tammi, T., & Saastamoinen, J. (2016). SMEs and public sector procurement: Does entrepreneurial orientation make a difference? *International Small Business Journal, 34*(4), 468–486. https://doi.org/10.1177/0266242614556661

Small Business Administration. (n.d.). 8(a) Business Development Program. Retrieved from https://www.sba.gov/federal-contracting/contracting-assistance-programs/8a-business-development-program

Winter, S. G. (1984). Schumpeterian competition in alternative technological regimes. *Journal of Economic Behavior and Organization, 5*, 287–320.

Disclaimer

The Center for Strategic and International Studies (CSIS) does not take specific policy positions; accordingly, all views expressed in this presentation should be understood to be solely those of the author(s).

Examining Small Business Set Asides: Evidence and Implications for Small and Mid-Sized Suppliers in Federal Procurement

Trevor L. Brown—is a Professor and Dean of the John Glenn College of Public Affairs at The Ohio State University. Brown's research on public procurement and contract management has been published in a variety of peer-reviewed academic outlets (e.g., Cambridge University Press, *Journal of Policy Analysis and Management*). He has also produced numerous consulting and project reports for the U.S. Navy, the Pew Center on the States, and the IBM Center for the Business of Government. Brown served for 20 years in a variety of contract management capacities for a technical assistance contract with the U.S. Agency for International Development. [brown.2296@osu.edu]

Amanda M. Girth—is an Associate Professor at the John Glenn College of Public Affairs at The Ohio State University. Her research on performance and accountability in government contracting is published in the *Journal of Public Administration Research and Theory*, *Journal of Supply Chain Management,* and *Public Administration Review*. She is the Editor-in-Chief of the *Journal of Strategic Contracting and Negotiation*, a peer-reviewed, international publication. Previously, Girth was a manager for a global consulting firm, where she supported information technology initiatives at the U.S. Department of State and U.S. Agency for International Development. She also served in Michigan state government. [girth.1@osu.edu]

Abstract

Our study of the federal small business set aside program assesses the impact of small business set asides on supplier competitiveness, program participation, and firm growth. Federal procurement policy distinguishes suppliers as either small or not small. Small businesses benefit from set asides and other programs offered by the U.S. Small Business Administration (SBA), whereas large companies have internal capacity, scale, and extensive past performance history to compete for procurements. Mid-sized suppliers are too big to qualify for set asides, yet do not have parity with large firms.

We focus our research on suppliers utilizing the small business set aside program and are particularly interested in firms that are "advanced small" or recently graduated suppliers (i.e., grew beyond the size standard prescribed by the SBA). We examine the federal small business set aside program and assess the impact of small business set asides on supplier competitiveness, program participation, and firm growth. We analyze 977 suppliers that participate or had participated in small business set aside procurements. We find the majority of suppliers stay small, and approximately 5% of small businesses grow to mid-sized.

Introduction

Federal procurement policy distinguishes suppliers as either small or not small. Small businesses benefit from set asides and other programs offered by the U.S. Small Business Administration (SBA), whereas large companies have internal capacity, scale, and extensive past performance history to compete for procurements. Mid-sized suppliers are too big to qualify for set asides, yet do not have parity with large firms. Anecdotal evidence of this disparity exits (perhaps best underscored by the work of trade associations such as the Association for Corporate Growth, Mid-Tier Advocacy, GTSC-Lion's Den, and the development of the bi-partisan Congressional Caucus for Middle Market Growth). However, there is a dearth of empirical evidence on both the structural barriers that exist for middle market firms and the effects of their competitive disadvantage. Before we can understand challenges for mid-sized suppliers, first we need to examine the marketplace where small businesses can thrive: small business set aside procurements. Our study begins to clarify industry narratives by analyzing the contours of the competitive federal procurement market for small and mid-sized suppliers.

Inequities in the public procurement market are not an insignificant concern. The scale and scope of federal procurement is vast, with over 5,000 different types of products procured (Brown, 2013) and over $438 billion in contracts obligated in 2015 (accounting for approximately 2.5% of gross domestic product [GDP]). The National Center for the Middle Market reports "middle market" firms account for one-third of private sector GDP and one-third of U.S. jobs. However, it is unclear whether mid-sized firms are correspondingly represented in the federal procurement market. A study by the Center for Strategic & International Studies (CSIS) suggests the answer may be no (Ellman, Morrow, & Sanders, 2011). The CSIS found that mid-sized market share of federal professional services contracts is shrinking. Mid-sized contractors claimed 40% of the total value of federal professional services contracts in 1995, but only 30% in 2009. During the same time period, large contractors increased their market share from 41% to 48%, and small business market share increased from 19% to 22%. Understanding the barriers to competition, purported disparities, and structural policy effects that impede middle market firms' ability to compete for federal contracts will in turn help us to understand their ability to capture market share, grow business, and deliver value to federal agencies.

This study examines the federal small business set aside program and assesses the impact of small business set asides on supplier competitiveness, program participation, and firm growth. Our study is based on a random sample of 977 firms with a small business set aside contract action in 2005. We include firms with contracts for products and services, which vary in complexity from simple product procurements to more complex services contracts (e.g., information technology systems). We follow these firms for a decade in order to better understand their contracts and the operating environment for small and mid-sized suppliers.

Our paper proceeds as follows. First, we provide context for the study by describing the federal policy environment for small and mid-sized suppliers. Next, we address our data and methodological approach. Then, we share our analysis and discuss policy implications and opportunities for future research. We suspect that firms that successfully transition out of the small business marketplace have unique ways of overcoming the "benefit cliff" they encounter as they grow out of a sheltered small business market, and this research lays the foundation for further study of these dynamics. We consider whether current policies governing procurement hamper mid-sized firm competitiveness in the federal procurement market and dampen U.S. economic growth.

Set Aside Policies in Federal Procurement

The first substantive guidance directed to federal agencies to contract with small businesses originated in the U.S. Senate in 1940 with the Special Committee to Study and Survey Problems of Small Business Enterprises, and in the U.S. House of Representatives in 1941 with the Select Committee on Small Business. The committees were created to protect the interests of small business owners, recognizing the need for a thriving small business community for innovation, economic growth, and national security. The Small Business Act of 1953 explicitly stated government prime contracts and subcontracts should be awarded to small businesses, and later the Small Business Act of 1958 created the SBA, an independent agency within the executive branch. Permanent committees were later established by both chambers: the Congressional Committee on Small Business and the

Senate Committee on Small Business and Entrepreneurship (U.S. Small Business Committee, n.d.; U.S. Senate Committee on Small Business & Entrepreneurship, n.d.; DoD Office of Small Business Programs, n.d.).

The policy goals for small businesses in federal procurement are multifaceted.[1] The initiating legislation created a competitive marketplace for small businesses to participate in federal procurement and win government awards; small business set aside procurements meet this policy objective. Government procurements are also required to allocate a percentage of all awards to small businesses. Firms bidding for these set asides must adhere to strict regulations to qualify as "small business concerns." Although there are many exceptions and stipulations delineated in the Federal Acquisition Regulation (FAR) that determine how contract officers must classify the size of firms, the two primary criteria are the 12-month average number of employees and three-year average receipts. The Code of Federal Regulations (CFR) requires the SBA to calculate these size standards for each line of business specified in the North American Industry Classification System (NAICS). For example, according to the current size standards, an iron and steel forging company (NAICS 332111) may be considered small if it has an average of 750 employees or fewer. A management consulting firm (NAICS 541611) may be classified as small if it has a three-year average of no greater than $15 million in revenue. In response to concerns that the SBA size standards failed to adapt to the changing economy, Congress passed the Jobs Act in 2010, requiring the SBA to review all size standards and make necessary adjustments to reflect market conditions at least once every five years (SBA, 2017).

The SBA also establishes procurement goals for federal agencies. Government-wide, 23% of the contract value of prime contracts is set aside for small businesses awards. There are goals within that subset, such as 5% of prime and subcontracts are to be awarded to woman-owned small businesses, and 5% of prime and subcontracts are to be awarded to small disadvantaged businesses, among others. Agencies also biennially negotiate their targets with the SBA in order to meet government-wide goals. In fiscal year (FY) 2017, goals ranged from 10% at the Department of Energy to 73% at the SBA. Additionally, federal agencies have annual goals for subcontracts. For example, the DoD's prime contract goal is 22% in FY2017, but the subcontracting goal is 34%.

One of the other key policy objectives of the Small Business Act is to promote small business in order to foster economic growth. Yet as suppliers grow towards their NAICS thresholds, they encounter a "benefit cliff" that disincentivizes growth, counter to this goal.

---

[1] The SBA serves the interests of small business beyond those discussed here relating to federal procurement. For example, the SBA has developed several financing and loan tools and set aside procurement policies to support small business growth. One of the most popular programs is the 7(a) Loan Guarantee which allows small businesses that are otherwise incapable of obtaining private sector financing access to funding up to $5 million. These funds may be used for a wide range of applications including the purchase or repair of capital, expansion or building of structures, and refinancing existing debt (Murray, 2013). A similar program, the 504 Certified Development Company loan, offers long-term fixed-rate financing specifically for the purchase of fixed assets for expansion or modernization (SBA, n.d.). Other innovative solutions such as the Small Business Investment Company (SBIC) and Surety Bond Guarantee programs offer growth-phase firms access to investment capital and bonding that they would otherwise be too small to acquire.

While the SBA might support economic growth among the smallest of small businesses (see Footnote 1), federal procurement policy is arguably less effective in supporting economic growth.

In the absence of robust research on firm behavior in sheltered markets, we turn to another policy domain for insight on benefit cliffs and unintended consequences. Comparisons can be drawn between the small business set asides growth disincentive and the benefit cliff observed in social welfare programs. Consider Temporary Assistance for Needy Families (TANF), the federal program implemented by states that provides financial welfare support. TANF includes work requirements and income thresholds to qualify and maintain benefits. Recipients benefit if they remain below the income threshold, as the loss in benefits is greater than the gain from increased wages (Bargain & Doorley, 2011; Randolf, 2014). In some cases, recipients maintain lower wages to keep ancillary benefits, such as child care, or forgo raises or promotions in order to stay under wage limits (Rutgers Center for Women and Work, 2016). Evidence shows that families also respond to benefit changes, altering employment decisions (Bargain & Doorley, 2011; Hoynes, 1996). It is clear that a similar dynamic exists for suppliers in the small business set aside program. Set aside procurements create incentives for suppliers not to grow beyond the NAICS size standard thresholds in their lines of business.

Public management research sheds light on a different dynamic at work in set aside procurements. One of the underlying tenets of contracting for goods and services is to harness the competitive force of the market. As with most government policies, unintended adverse effects can result in the pursuit of overcoming market failures (Vining & Weimer, 2005). In this case, set aside programs restrict competition, contribute to weakly competitive procurements, and thereby limit the range of cost, quality, and delivery options for goods and services procured under said programs (Brown, 2007; Girth et al., 2012). When markets are constrained, purchasers have fewer choices to balance different, and sometimes competing, purchasing goals (Brown, Potoski, & Van Slyke, 2013; Johnston & Girth, 2012). Taken together, these conflicting objectives between efficiency, equity, and effectiveness illustrate the intricacies of multifaceted policy interventions such as the small business set aside program.

Methods and Data

Our purpose is to understand the contours of the small business set aside marketplace, and we do this through descriptively analyzing federal contracts. We examine small business behavior by gathering data on 977[2] suppliers that had a small business set aside contract in 2005 (i.e., they have at least one contract action associated with a small business set aside contract). Data was drawn from the Federal Procurement Data System–Next Generation (FPDS). The FPDS is a repository of all non-classified prime contract activity with any action exceeding $3,000 in value for federal agencies. The FAR requires contract officers to enter contract information into the FPDS-NG and update as required. Contract actions include the individual records created when a contract is initiated and subsequently modified. We stratify the sample such that 60% of contracts are DoD contracts

[2] We initially extracted a sample of 1,025 suppliers. In some cases, contract actions reported in FPDS-NG were missing key data elements, such as product or service code, principal NAICS code, or contracting agency; this reduced our sample to 977 suppliers.

to mirror federal spending. The sample includes firms with contracts for products and services varying in complexity, from simple product procurements to more complex services contracts. The unit of analysis is firm-year, and contracts data from the FPDS-NG is aggregated to account for contract activity for each fiscal year.

Data on firm attributes was procured from Dun & Bradstreet. This data is reported annually and appended to FPDS data to create our dataset. We also conducted content analysis and cross-referenced data with the System for Award Management (SAM), Dynamic Small Business Search (DSBS) database, and open source material.

Analysis

While following 977 suppliers over a decade, we observed patterns in their engagement with federal agencies. As shown in Table 1, about one-third of suppliers consistently maintained contracts with federal clients in every year studied. The remaining firms had irregular activity (e.g., contract actions 2005–2006, no contract actions 2007–2009, and contract actions 2010–2014). In some cases, suppliers were only inactive for one year. In other cases, we observed suppliers with two or more continuous years of inactivity, which indicates the firm either discontinued serving federal customers or did not survive.

Table 1. Contracting Patterns

Description of supplier activity	*Number of suppliers*
Contract activity for all years	303
One year with no contract activity	71
Two or more continuous years with no contract activity	312
Irregular contract activity	291
Total	977

Our data shows that suppliers that are eligible for set asides utilize that advantage. Yet it also suggests that small businesses are successfully obtaining contracts on the open market. Although we have 9,770 observations (977 firms and 10 years of contracts data), we only have 5,995 observations in the data set with contract activity.[3] Table 2 contains descriptive statistics of variables in our dataset.

[3] As Table 1 illustrates, we only have 303 suppliers with contract activity in all 10 years. This means 3,030 observations for those suppliers as the unit of analysis is firm-year. The remaining observations represent varying scenarios from Table 1 (one supplier might only have contracts data for 2005, another might have contracts data for every year except 2014). This is why our total observations of suppliers with contracts data reduces to 5,995.

Table 2. Descriptive Attributes

Variable	Source	25th percentile	50th percentile	75th percentile	Mean	Std. Dev	Range
Percent set asides	FPDS	0.01	0.43	1	0.47	0.41	0-1
Agency diversity	FPDS	1	1	2	2.31	3.24	1-44
NAICS diversity	FPDS	1	1	2	1.70	1.15	1-10
Manufacturing	DB	0	0	0	0.21	0.40	0-1
Transportation	DB	0	0	0	0.05	0.22	0-1
Wholesale trade	DB	0	0	0	0.18	0.38	0-1
Retail trade	DB	0	0	0	0.05	0.22	0-1
Finance	DB	0	0	0	0.02	0.13	0-1
Construction	DB	0	0	0	0.10	0.30	0-1
Public administration	DB	0	0	0	0.00	0.02	0-1
Services	DB	0	0	1	0.39	0.49	0-1
Woman-owned	DB	0	0	0	0.23	0.43	0-1
Minority-owned	DB	0	0	0	0.16	0.37	0-1

- Percent set asides: We define contracts as small business set aside contracts to include set aside categories targeting small businesses, such as woman-owned small business, 8a, service disabled veteran owned small business, and the like. The variable measures the percent of contract actions specified as small business set asides compared to total contract actions in a given year. We see that an average of 43% of a supplier's portfolio are related to small business set aside awards. The distribution of this variable is like a hockey stick—at the 75th percentile, set asides are 100% of supplier portfolios.
- Agency diversity: We calculate the number of agencies (at the corporate level) with which a supplier has contract actions recorded for a given year. This shows that most small business suppliers serve one or two distinct federal agencies.
- NAICS diversity: We track the number of NAICS codes affiliated with contract actions for each year. This value represents the number of different NAICS associated with all contract actions for a supplier in a given year. NAICS are aggregated to the first two digits (e.g., naics53, naics54). Most firms operate in one or two NAICS categories.
- Lines of business/Standard Industrial Classification (SIC): Primary SIC is developed by the federal government and is reported to Dun & Bradstreet. The first two digits of a four-digit code rolls up to one of 11 categories. Seven of the 11 categories are in this sample: *services, manufacturing, transportation and public utilities, retail trade, wholesale trade, finance, insurance and real estate,* and *construction*. About 40% of the firms in this sample identify their primary SIC as services.
- Ownership: Approximately 23% of the observations are attributed to woman-owned suppliers. Approximately 16% of the observations are attributed to minority-owned suppliers.

In addition to capturing descriptive attributes of the contracts and suppliers in our dataset, we cross-checked suppliers in SAM, DSBS, and other open source materials. While FPDS data specifies whether a supplier obtained a set aside contract over this time period, supplier data on firm size is not as reliable. Dun & Bradstreet has a small business indicator among firm attributes, but this is also not a consistent indicator of the firm's status, as it compares reported annual revenue or employee totals to primary NAICS. This process does not capture the other NAICS (beyond primary) that the firm can utilize to win set aside contracts, and revenue and employee data is often incomplete or in error for small businesses in Dun & Bradstreet's data. As a result, we gathered information from additional resources to determine whether firms remained a small business or grew into the middle market and reported the results in Table 3.

Table 3. Supplier Growth

Description of supplier activity	*Number of suppliers*
Survived and remain a small business	574
Survived and are no longer a small business	46
Did not survive	183
Unable to substantiate	174
Total	977

Approximately 59% of suppliers with a small business contract action in 2005 remain small and eligible for set asides in 2017. Fewer than 5% of suppliers grew to mid-sized. Nineteen percent of suppliers that had a set aside contract action in 2005 no longer exist. We are unable to substantiate the status of approximately 18% of suppliers that had a contract action in 2005. These are suppliers that (a) no longer register in SAM, (b) have no recent contract activity in the FPDS, and (c) have no publicly available data to support approximating firm size.

Among those 46 firms that grew to mid-sized, we find that half grew through acquisition. We also see small businesses acquiring other small businesses and remaining small in some NAICS. Of the 43 supplier acquisitions we find through our content analysis, 20 are still small businesses.

Discussion

Our interest in this research lies in the design and implementation of the small business set aside program. Our analysis shows the tension between programmatic goals that established the SBA and set aside programs: economic growth versus equitable access. In this section, we explore the implications of these tensions in practice. We discuss implications of our descriptive analysis and present future research opportunities in light of our exploratory findings. We raise data limitations encountered with this study, which has bearing on future work, and explore approaches to research design in light of data constraints.

The vast majority of suppliers in our sample shelter in the small business set aside market. They fail to grow beyond the sales or employee thresholds in the product or service areas for which they have self-certified. In most cases, firms in our sample elect to stay small. These suppliers recognize the value of the constrained federal market established for small businesses. Their clients also value their small business status, allowing for more desirable procurements as they help to achieve the agency's small business goals. In other cases, we suspect small businesses are unable to harness the resources, whether financial or managerial, to grow. In yet other instances, prior study finds that firms respond to undesirable consequences resulting from growth. Despite evidence to the contrary, firms fear increased size can make firms more vulnerable to surviving crisis (Davidsson, Achtenhagen, & Naldi, 2006).

When suppliers make strategic decisions to stay small in order to retain small business status for federal procurements, they artificially constrain growth in order to stay under the revenue or employee thresholds for specified NAICS. If the intent of federal policies to support small business is to encourage economic growth and innovation, then firm behavior does not necessarily align with these goals. The behavior, albeit rational on the part of small business, is an unintended consequence of creating markets and subsidizing subsets of industry. Yet this phenomenon reinforces questions about the behaviors of suppliers that are not able to thrive as they grow into the middle market, and either intentionally constrain to meet small business thresholds in subsequent years or fail to win contracts when competing outside of the set aside market.

Several policy implications begin to form in light of this exploratory research, including policy priorities to support lasting transition to the middle market. Policy options might include modified size standards aimed to benefit mid-sized firms (particularly those at the lower threshold of the middle market). One concern with simply raising current size standards, or increasing the number of years in the rolling average to determine qualification, is that action fails to address underlying issues for emerging small businesses and is a temporary remedy for only those firms on the edge of mid-sized.

There are also other ways the federal government could support mid-sized suppliers that are neither large nor small. Agencies increasingly rely on federal schedules and multi-award vehicles. Creating a unique vehicle for mid-sized suppliers is one alternative advanced by trade associations supporting mid-tier suppliers. Another policy option is to create subcontracting benefits for prime contractors that utilize mid-sized suppliers. That is, rewarding proposals with mid-tier suppliers on the subcontracting team, or compelling large firms to diversify their teams by including mid-sized suppliers in addition to meeting existing small business requirements.

While there may be little drive among policymakers or administrators to create additional regulation, there appears to be a desire to address some of the structural challenges faced by mid-tier suppliers. The Chairman of the House Small Business Committee, Steve Chabot, recently noted,

> After a small business has proven its success by growing out of its small size standard, it exists in a murky limbo—it is too large to benefit from small business set-asides, yet is too small to compete with billion dollar firms. (U.S. Small Business Committee, 2017, p. 1)

Even with political support, we conceive of a number of roadblocks to these alternatives. First, contract officers are already stretched thin, and policies to support the middle market would likely increase regulatory burden on acquisitions staff. Second, large firms benefit from status quo policies and would likely mobilize and challenge any regulatory changes that strengthen the middle market at the expense of the large suppliers. Third, it is unclear where the line should be drawn between emerging small business at the margins of the middle market and larger mid-tier suppliers. Who should make that determination, how, and where is the appropriate regulatory home for middle market suppliers?

More broadly, the underlying issue for policymakers to consider is simply the way in which procurement policy and regulation recognize federal suppliers as either "small business" or "not small business." By ignoring the middle, the government may be losing an opportunity to secure value by not actively seeking contracts with mid-sized firms. It is well understood that small businesses are higher risk suppliers. Unlike large firms, their internal processes are immature, and resource shocks can have profound effects on a small enterprise and its clients. Large firms largely mitigate that risk, but can be costly. Mid-sized firms can be well suited to provide value at lower risk than small firms and lower cost than large. In any case, federal agencies are likely missing opportunities to secure value by treating all firms that are not small as large and not incentivizing contracting with mid-sized suppliers.

Future Research

The purpose of an exploratory study is to describe the current state and then propose further research to advance knowledge. With this in mind, we propose a number of possible avenues for future study. To begin, small and mid-tier businesses would benefit from understanding the success factors for the unusual firms in our study that started as small businesses and then successfully grew beyond their size standard and into the middle market. A qualitative study designed to solicit interviews from principals of the 46 firms we identified would shed light on success strategies.

Further, the initial intent of our research was to determine growth factors for suppliers successfully transitioning to the middle market, yet we have been unable to answer this research question due to inadequate data specification in the FPDS and from Dun & Bradstreet. Extensive primary data collection is needed to address this shortcoming. One approach to this is to leverage the qualitative findings from the aforementioned principal interviews to develop a survey instrument for wider distribution to small and mid-sized suppliers.

Next, we designed our study to gather contracts data across a wide range of contracts and did not constrain our sampling procedure to a specific NAICS in order to replicate the procurement environment. Replicating and then extending this study by analyzing a sampling of contracts within a single NAICS, to include the full range of suppliers (small and not small), would provide a different perspective on the supplier dynamics in the federal procurement market. Although this process would limit generalizability, this would help to shed light on the competitive landscape for all suppliers within a controlled category.

Finally, our analysis of contract patterns showed that one-third of businesses consistently contract with federal agencies. While some firms went out of business during the time period studied, others no longer participate in the federal marketplace. Future study on the reasons for the lack of continued participation could help us understand barriers to federal contracting, particularly among firms with a diverse client base.

Conclusion

In practice, acquisition officials are asked to deliver contracts that meet best value, low cost, or other performance objectives, and to meet broader political objectives that can affect (constrain) eligible suppliers. Public sector contracts are not simply a tool to increase efficiency; they can also serve to promote other public values. Procurement policies that target specialized groups, such as small businesses, minority-owned, or women-owned firms, are designed to promote equity and representativeness. In short, public procurement is a way in which governments can promote social policy goals, such as increased opportunity for underrepresented groups (McCrudden, 2004). It is this tension amongst competing values that motivates our interest in this study of small business set asides.

Our research represents a first step in capturing the structural dynamics involved in the design, implementation, and evaluation of competitive practices in federal agencies aimed at promoting small business participation and growth. For small and mid-sized suppliers, the analysis shows the market dynamics do not favor growth. Further analysis is needed to determine success factors for outlying suppliers that succeed in the middle market.

The results also have implications for policymakers. The Small Business Act has succeeded in carving out a competitive space for small businesses seeking federal procurements. The SBA's policies support growth among the smallest of firms, but appear to fall short in one of the Small Business Act's other goals, which is to encourage economic growth. There is yet another dynamic that should concern policymakers and administrations, which is if further analysis confirms that middle market firms are, on balance, unable to compete in the federal procurement market, then agencies are likely missing critical opportunities to secure value.

References

Bargain, O., & Doorley, K. (2011, May 26). Caught in the trap? Welfare's disincentive and the labor supply of single men. *Journal of Public Economics*, 1096–1110.

Beck, T., & Demirguc-Kunt, A. (2006). Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking & Finance, 30*(11), 2931–2943.

Brown, T. L. (2007, September 19). Testimony on the Small Business Administration's contracting programs, Small Business Committee, U.S. House of Representatives, Washington, DC.

Brown, T. L. (2013). *A guide for agency leaders on federal acquisition: Major challenges facing government*. Washington, DC: IBM Center for the Business of Government.

Brown, T. L., Potoski, M., & Van Slyke, D. M. (2013). *Complex contracting: Government purchasing in the wake of the U.S. Coast Guard's Deepwater Program*. New York, NY: Cambridge University Press.

Davidsson, P., Achtenhagen, L., & Naldi, L. (2006). What do we know about small firm growth? In *The life cycle of entrepreneurial ventures* (pp. 361–398). Boston, MA: Springer.

DoD Office of Small Business Programs. (n.d.). Legislation. Retrieved April 1, 2017, from http://business.defense.gov/legislation/

Ellman, J., Morrow, D., & Sanders, D. (2011). *Structure and dynamics of the U.S. federal professional services industrial base.* Center for Strategic & International Studies.

Girth, A. M., Hefetz, A., Johnston, J. M., & Warner, M. E. (2012). Outsourcing public service delivery: Management responses in noncompetitive markets. *Public Administration Review, 72*(6), 887–900.

Hoynes, H. W. (1996, March). Welfare transfers in two-parent families: Labor supply and welfare participation under AFDC-UP. *Econometrica*, 295–332.

Johnston, J. M., & Girth, A. M. (2012). Government contracts and “managing the market:” Exploring the costs of strategic management responses to weak vendor competition. *Administration & Society, 44*(1), 3–29.

McCrudden, C. (2004). Using public procurement to achieve social outcomes. Natural *Resources Forum, 28*(4), 257–267.

Murray, K. (2013, October 30). SBA's 7(a) loan program explained. Retrieved from https://www.sba.gov/blogs/sbas-7a-loan-program-explained

Randolf, E. (2014). *Modeling potential income and welfare assistance benefits in Illinois: Single parent with two children households and two parents with two children household scenarios in Cook County, City of Chicago, Lake County and St. Clair County*. Chicago, IL: Illinois Policy Institute.

Rutgers Center for Women and Work. (2016). *New Jersey's benefits “cliff effect” and economic self-sufficiency fact sheet.* New Brunswick NJ: Author.

U.S. Senate Committee on Small Business & Entrepreneurship. (n.d.). History. Retrieved April 1, 2017, from https://www.sbc.senate.gov/public/index.cfm?p=History

U.S. Small Business Administration (SBA). (n.d.). Office of financial assistance: Resources. Retrieved from https://www.sba.gov/offices/headquarters/ofa/resources/4049

U.S. Small Business Administration (SBA). (2017). *A report on the first five-year comprehensive review of small business size standards under the Small Business Jobs Act of 2010.* Washington, DC: Author.

U.S. Small Business Committee. (n.d.). Committee history. Retrieved April 1, 2017, from http://smallbusiness.house.gov/about/

U.S. Small Business Committee. (2017, November 14). Mid-sized businesses: Too big to be small and too small to be big [Press release]. Retrieved March 20, 2018, from https://smallbusiness.house.gov/news/documentsingle.aspx?DocumentID=400479

Vining, A., & Weimer, D. (2005). Economic perspectives on public organizations. In D. Ferrin, L. Lynn, & M. Pollitt (Eds.), *Oxford handbook of public management* (pp. 209–233). Oxford, England: Oxford University Press.

Acknowledgments

This material is based upon work supported by the Naval Postgraduate School Acquisition Research Program under Grant No. N00244-16-1-0049. The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. government.

Professors Brown and Girth are Fellows of the National Center for the Middle Market at the Fisher College of Business, The Ohio State University, and acknowledge the center’s support for this research.

Disclaimer

Draft. Do not cite.

Panel 23. Designing Cybersecurity Into Acquisition Programs

Thursday, April 27, 2017	
1:45 p.m. – 3:15 p.m.	**Chair: Rear Admiral Edward Anderson, USN,** Deputy Commander, Fleet Readiness Directorate, Space and Naval Warfare Systems Command ***Applying Cause-Effect Mapping to Assess Cybersecurity Vulnerabilities in Model-Centric Acquisition Program Environments*** Jack Reid, Massachusetts Institute of Technology Donna H. Rhodes, Massachusetts Institute of Technology ***Automated Methods for Cyber Test and Evaluation*** Valdis Berzins, Naval Postgraduate School ***Cybersecurity: Converting Shock Into Action*** Paul Shaw, Defense Acquisition University Robert Tremaine, Defense Acquisition University

Rear Admiral (Sel) Edward Anderson, USN—was born in Glendora, CA, and is a 1990 graduate of United States Naval Academy with a Bachelor of Science in Systems Engineering. He holds a Master of Business Administration with Honors from Indiana University and is a graduate of the Aspen Institute seminar on Global Leadership.

His sea tours include division officer, USS *Newport News* (SSN 750); strategic weapons officer, USS *Ohio* (SSBN 726) (Blue); and executive officer, USS *Jefferson City* (SSN 759). Four years after departing his executive officer tour, Anderson assumed command of *Jefferson City* in July 2008.

Ashore, Anderson has served as space control officer in Cheyenne Mountain, U.S. Space Command; submarine operations officer at Commander, Allied Naval Forces South in Naples, Italy; submarine operations and anti-submarine warfare training officer for Commander, Strike Force Training Pacific; deputy director for Navy programs and congressional liaison for Undersea Warfare and Strategic Programs, Navy's Office of Legislative Affairs; assistant program manager, Common Submarine Radio Room; and major program manager for Undersea Integration (PMW 770).

Anderson was selected for flag officer in March 2017 and was then selected to serve as deputy commander, fleet readiness, Space and Naval Warfare Systems Command in May 2017.

Anderson's decorations include the Legion of Merit, Defense Meritorious Service Medal, Meritorious Service Medal (three awards), Joint Commendation Medal, Navy and Marine Corps Commendation Medal (two awards), Joint Achievement Medal, Navy and Marine Corps Achievement Medal (four Awards), and various campaign, unit, and service awards.

Applying Cause-Effect Mapping to Assess Cybersecurity Vulnerabilities in Model-Centric Acquisition Program Environments

Jack Reid—is a graduate student with the Systems Engineering Advancement Research Initiative (SEAri) at the Massachusetts Institute of Technology. Reid is earning master's degrees in both Aeronautics & Astronautics and Technology & Policy. His research interests concern the design and management of complex sociotechnical systems, particularly with regard to the anticipation of emergent and cascading behavior. He received a BS in Mechanical Engineering and a BA in Philosophy from Texas A&M University and has experience with RAND Corporation and Sandia National Laboratories. [jackreid@mit.edu]

Donna H. Rhodes—is a principal research scientist at the Massachusetts Institute of Technology, and director of the Systems Engineering Advancement Research Initiative (SEAri). Dr. Rhodes conducts research on innovative approaches and methods for architecting complex systems and enterprises, designing for uncertain futures, and human-model interaction. Previously, she held senior management positions at IBM, Lockheed Martin, and Lucent. Dr. Rhodes is a Past President and Fellow of the International Council on Systems Engineering (INCOSE), and INCOSE Founders Award recipient. She received her PhD in Systems Science from T. J. Watson School of Engineering at Binghamton University. [rhodes@mit.edu]

Abstract

Digital engineering approaches are increasingly used in acquisition of systems, changing the current paradigm from documentation-centric to model-centric. Not only are these systems highly vulnerable to cyber threats, so too are their enabling environment and digital assets. While good practices have emerged to support the shift to model-centric program acquisition, such programs experience perturbations over their life cycles that introduce new vulnerabilities that may lead to cascading failures. Cybersecurity vulnerabilities are of particular concern given digital transformation and increasing threat actors, making vulnerability assessment essential throughout acquisition program life cycles. This paper discusses ongoing research that seeks to provide program managers with the means to identify cybersecurity vulnerabilities within model-centric programs (along with other model-related vulnerabilities) and determine where interventions can most effectively be taken. The research builds on recent work in developing a reference model for model-centric program vulnerability assessment that uses the Cause-Effect Mapping (CEM) analytic technique. This research investigates cybersecurity specifically, using CEM and other dynamic analysis approaches, including a prototype for proactive assessment of cybersecurity and evaluation of potential interventions.

Introduction

Digital transformation changes how systems are acquired and developed through the use of model-centric engineering practices and toolsets. While offering great benefit, new challenges arise from both technological and socio-cultural dimensions. This drives the need to examine and address vulnerabilities not only for products and systems, but also for the model-centric environments necessary for their acquisition and development. Recent research has investigated the use of Cause-Effect Mapping (CEM) as a mechanism for better enabling program managers and system engineers to anticipate and respond to programmatic vulnerabilities as related to model-centric environments. A Reference CEM for model-centric enterprises resulting from the work shows promise for considering the cascading vulnerabilities and potential intervention options. In ongoing research, additional investigation aims to refine the Reference CEM and analytic approach for cybersecurity-focused program vulnerability assessment.

Motivation

Modern society has many needs and problems that can only be addressed through large-scale socio-technical engineering programs (e.g., defense systems, multi-modal transportation systems, energy delivery system of systems, health-care management systems). The use of model-centric approaches, modeling and simulation, and "digital twins" is increasingly used in acquisition of such systems, changing the current paradigm from documentation-centric to model-centric. Not only are these systems highly vulnerable to cyber threats, so too are their enabling environment and digital assets. While good practices have emerged to support the shift to model-centric program acquisition, such programs experience perturbations over their life cycles that introduce new vulnerabilities that may lead to cascading failures. For instance, perturbations may be caused by policy change (leading to IP disagreements), economic factors (leading to training cuts), or disruptive technology (leading to outdated infrastructure). Early detection and intervention of vulnerabilities can mitigate disruptions and failures. The research seeks to contribute to the vulnerability assessment state of practice for acquisition programs, both public and private, that increasingly depend on digital assets and model-centric environments.

Background

The following subsections describe model-centric engineering, cyber-security vulnerability assessment, and cause-effect mapping. A companion paper (Reid & Rhodes, 2018) provides additional background information.

Model-Centric Engineering

Acquisition program management is grounded in management science and a sound set of practices evolved over decades; however, new challenges arise as acquisition becomes increasingly model-centric. Baldwin and Lucero (2016) state, "The DoD sees value in adopting digital engineering design and model-centric practices, enabling a shift from the linear, document centric acquisition and engineering process toward a dynamic digital, model-centric ecosystem."

Model-Centric Engineering (MCE) has been defined as "an overarching digital engineering approach that integrates different model types with simulations, surrogates, systems and components at different levels of abstraction and fidelity across disciplines throughout the lifecycle" (Blackburn et al., 2017). MCE involves using integrated models across disciplines, subsystems, life-cycle stages, and analyst groups. It uses models as the "source of truth" to reduce document handoff and allow for more continuous evaluation. This reduces communication time and rework in response to requirement changes. Most discussions of MCE focus on engineering practices and methods to overcome implementation difficulties. In any system, however, engineering is only a piece of the problem. Numerous human factors, business, and organizational issues exist. Current program managers have significant experience with modern engineering processes and use this experience to identify and mitigate vulnerabilities. No such experience exists with MCE, however. This fact, coupled with the increased integration of models, means that emergent uncertainties (policy change, budget cuts, disruptive technologies, threats, changing demographics, etc.) and related programmatic decisions (e.g., staff cuts, reduced training hours) may lead to cascading vulnerabilities within MCE programs, potentially jeopardizing program success. New tools are needed to enable program managers to identify model-centric program vulnerabilities and determine where interventions can most effectively be taken.

Cybersecurity Vulnerability Assessment

MCE, with its focus on digitization, integration, and collaboration, has the potential to increase the cybersecurity vulnerability of an enterprise. A *vulnerability* is the means by which the hazard might disrupt the system, thus it is through the vulnerability that the system is susceptible to the hazard. Vulnerabilities are best expressed as the causal series of events connecting a hazard to system failure. This is a generalization of common, field-specific usages of the term. MITRE's Common Vulnerabilities and Exposures (CVE) database defines a vulnerability as "a weakness in the computational logic (e.g., code) found in software and some hardware components (e.g., firmware) that, when exploited, results in a negative impact to confidentiality, integrity, OR availability" (The MITRE Corporation, 2015). In this definition, the same components can be seen: some structural means or "weakness" that can result in system disruption or "negative impact" if a hazard is present or the vulnerability is "exploited." For example, the infamous Spectre security vulnerability is described by CVE as "systems with microprocessors utilizing speculative execution and branch prediction may allow unauthorized disclosure of information to an attacker with local user access via a side-channel analysis" (The MITRE Corporation, 2017). This is a neat summary of the hazard (an attacker), the means (side-channel analysis using speculative execution and branch prediction), and the disruption (unauthorized disclosure of information).

Risk and vulnerability assessment methods have not failed to adapt to novel cybersecurity concerns. The aforementioned CVE database has been public since 1999. Quality assurance testing (essentially the verification and validation of software) has been around since the beginning of commercial software. Software penetration testing (where security experts intentionally seek to break a software product) has been the industry norm for more than a decade (Arkin, Stender, & McGraw, 2005). Black-box mutational fuzzing and concolic execution are being used to automatically test for certain types of software vulnerabilities (Schwarz, 2018). Formal verification tools, initially limited to pure software domains such as cryptography (Meadows, 1994), has been rapidly advancing and finding applications in hardware (Kern & Greenstreet, 1999) and business processes (Morimoto, 2008), as well as fields that straddle the software-hardware-environment boundaries (Kamali et al., 2016). The methods listed here just scratch the surface of approaches security researchers and engineers are taking to identify and resolve such technical cybersecurity vulnerabilities.

Beyond these specific testing methods, assessment frameworks have progressed as well. System-Theoretic Process Analysis (STPA) has adjusted, adapted, and been applied to handle cybersecurity vulnerabilities associated with additive manufacturing (Pope & Yampolskiy, 2016), Internet of Things (Pope, 2017), Air Operations (Young, 2013), and Mission Operations (Young & Porada, 2017). More recently, there have also been efforts to combine compiler technology with STPA to automatically detect vulnerabilities in software-controlled systems (Pope, 2018).

While cybersecurity vulnerabilities in operational systems remain alarmingly common, from the trivial (Hanselman, 2012) to the critical (Gressin, 2017), there is some evidence that software is becoming more secure, at least in terms of defects per line of equivalent source code (Pope, 2017). In many cases, however, the acquisition or development process itself needs to be protected from outside threats and endogenous failures. Be it military information or technology-related trade secrets, there is real value in attempting to penetrate much earlier in the life cycle in order to either steal secrets (Hanna, Smythe, & Martin, 2018; Raymond, 2017) or to disrupt production (Statt, 2018).

Defense acquisition programs have already instituted a variety of means of ensuring the security of their work. Some of these means were originally instituted to address other forms of threats but have turned out to be effective in addressing cybersecurity as well. These methods include relying on the security clearance process, the use of Sensitive Compartmented Information Facilities (SCIFs), restrictions on the use of media storage devices, separate networks such as SIPRNet and NIPRNet that are isolated or semi-isolated from the internet, and general compartmentalization of critical information. Some (non-U.S.) defense agencies have gone so far as to revert to using typewriters where able in order to avoid security breaches and leaks (Irvine & Parfitt, 2013).

Unfortunately, many of these historically successful methods are in conflict with the more straightforward implementations of many components of an MCE environment. For example, the use of SCIFs has been quite successful in preventing unauthorized access to data. The typical use of a SCIF in the design process, where a small number of engineers work on a task isolated from the outside world, is not directly compatible with an MCE environment structured around model integration and collaboration across teams and locations. While this problem has been previously considered and ways to mitigate this conflict have been proposed (e.g., Reid & Rhodes, 2016), no silver bullet to resolving these tensions exists and it is likely that the increased use of MCE will result in both the exacerbation of current vulnerabilities and the creation of new ones. Furthermore, most means of assessing such vulnerabilities are aimed at assisting software and systems engineers to identify and remove cybersecurity vulnerabilities from the end system. New methods for enabling project and program managers to perform cybersecurity assessments of their enterprise and engineering environment are needed.

Cause-Effect Mapping

Cause-Effect Mapping is a vulnerability assessment tool that consists of a mapping of causal chains that connect an exogenous hazard to a system degradation or failure, termed a *terminal event*. Each chain represents a specific vulnerability, sometimes called a *vulnerability chain* in order to emphasize that vulnerabilities are not discrete events. Terminal events are broadly defined and include any form of value loss. Interventions are actions that eliminate or mitigate a vulnerability, and are represented on the map as points that break the causal chain. An example CEM (that lacks interventions) can be seen in Figure 1.

The hazards are external to the perspective of the defined user, and are thus sometimes called *external triggers*. An *intermediary event* is any unintended state change of a system's form or operations that could jeopardize value delivery of the program.

A CEM is not created for a system, but for a specific class of decision-maker. The hazards (referred to as "spontaneous events" in Figure 1) are exogenous from the point of view of the decision-maker that the CEM was made for. In this way, CEM avoids the "blaming someone else" problem by making all hazards exogenous. The decision-maker only has control over the intermediary events. While she may not be at fault for any of the vulnerabilities, it is still her responsibility to address them.

CEM is fundamentally a qualitative analysis method, though it can be readily adapted into a quantitative form by adding probabilities of transition to each intermediary. CEM provides immediate insight into which parts of the system warrant more detailed modeling using other methods.

The basic steps to create a new CEM are not application specific and are as follows:

1. The stakeholder herself lists potential hazards posed to the program.
2. She then traces the consequences of each of these hazards through the intermediary events to the final terminal events.
3. The process is then done in reverse: She looks at the terminal events, adds in any that are still missing, and works backwards on how they might come about.
4. She then examines the causal connections between each intermediary event to see if there are any additional connections not previously noticed.
5. Finally, she consults lessons learned databases, case studies, and other experts to generate additional hazards, intermediary events, causal connections, and interventions, as well as to verify existing ones.

Any of these steps can take place either formally, using automated tools to enumerate possible vulnerabilities, or informally, relying upon the stakeholder's own experience.

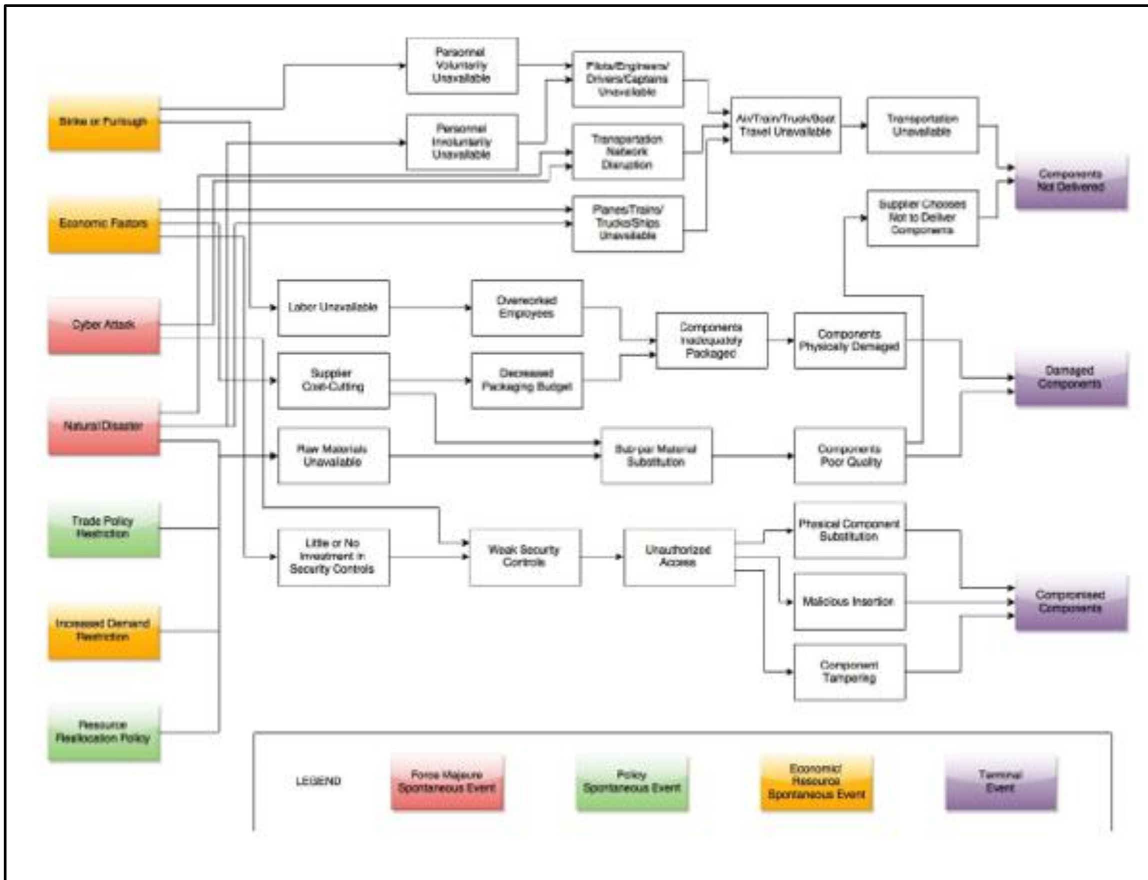


Figure 1. Example CEM of a supply chain
(Rovito & Rhodes, 2016)

CEM has previously been applied in a case study of a Maritime Security System of Systems (Mekdeci et al., 2012) and to a supply chain case (Rovito & Rhodes, 2016). More recently, an earlier phase of this research developed a Reference CEM for use by program managers to assess enterprise-level vulnerabilities in the MCE environment (Reid &

Rhodes, 2018). This work, which was based upon literature reviews, interviews with experts, and other sources, sought to provide program managers with an entry point for considering such vulnerabilities. Additionally, the steps to create a CEM for one's own program were outlined and some potential use cases discussed. These use cases are as follows:

(A) By a Program Manager: Assessing potential future vulnerabilities and plan possible interventions

(B) By a Program Manager: Determining specific vulnerabilities to address in response to the presence of a specific hazard

(C) By the Program Organization: Changing program processes to mitigate or eliminate vulnerabilities

(D) By Researchers: Organizing and classifying vulnerabilities into various categories or types

Most users of CEM tend to find it most useful in identifying high priority intervention points and new vulnerabilities. Other benefits of note include increased understanding of the causal path and the interrelationships between vulnerabilities. While the resultant reference CEM was quite detailed in some respects, such as both vulnerabilities and interventions involving model curation, it was less well developed in others, notably cybersecurity, as can be seen in Figure 2.

Use (A) is most relevant for novice program managers or program managers using MCE for the first time. A senior program manager or team of program managers creates a CEM for their organization's program process. This CEM can then be provided to the novice for study and reference. The program manager can then learn what can go wrong and how to intervene. In this case, the CEM could be tied to a Lesson's Learned database, such as NASA's Lessons Learned Information System (NASA Office of the Chief Engineer, 1994). This enables concrete examples and consequences to be linked to each vulnerability. One of the important factors here is that the CEM does not just present potential interventions, but it also places them in the appropriate part of the causal sequence. This enables the program manager to not only know how to intervene, but at what point.

In Use (D), CEM is used to organize and classify vulnerability chains. Two obvious classifiers are terminal events and hazards. Which is used to organize a CEM depends on whether the user wants to examine the causal chains forward or backwards. Beyond this, however, more complicated classifiers are possible. As can be seen in Figure 2, external triggers that result in similar vulnerability chains are grouped together. By "similar," we mean that these vulnerability chains either involve many of the same intermediary events or that they involve the same part of the program. For instance, most of the intermediary events involving model curation and trust are located close to one another in the center-top of the figure. Once these groupings have been identified, they can be considered together, such as the "Belt-tightening" grouping, and common means of intervention considered.

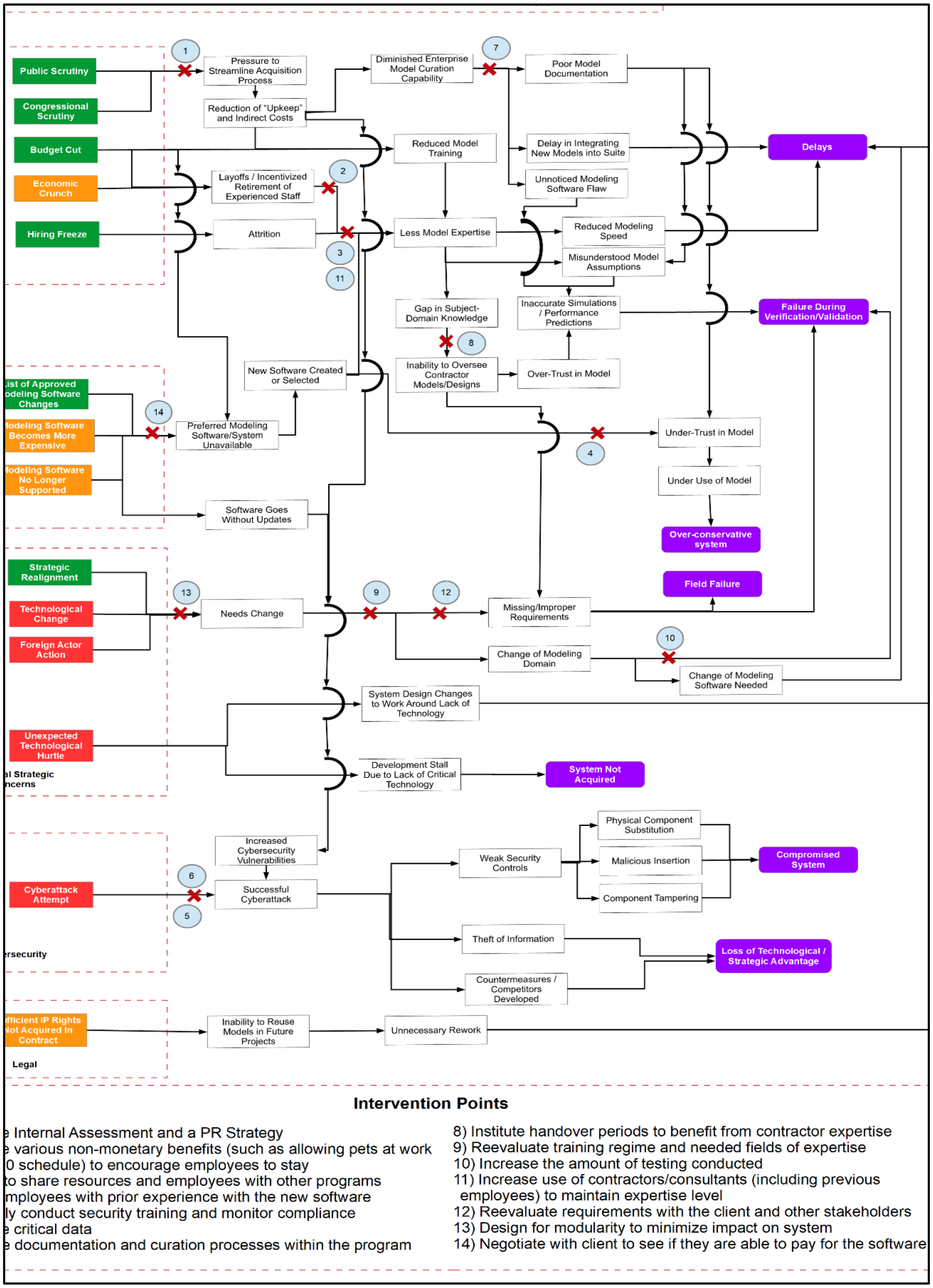


Figure 2. Preliminary Reference CEM for Model-Centric Vulnerabilities With Example Intervention Points

Strengthening Cybersecurity Aspects of a CEM for Model-Centric Programs

As was discussed in the previous section, the MCE Reference CEM shown in Figure 2 was generated using literature reviews and interviews with experts, among other sources. The cybersecurity portion of it was adapted, mostly unchanged, from previous work on supply chains (Rovito & Rhodes, 2016). Cybersecurity is a rising international concern and is of particular relevance with the increasing digitization associated with MCE environments. As a result, further development of that portion of the Reference CEM was desired.

To accomplish this, an ongoing series of interviews with systems engineers and program managers from a variety of fields, including defense, aerospace, manufacturing, and semiconductors, is being conducted. These interviews have sought to provide insight into the following questions, in the context of a model-centric enterprise:

1. To what extent are program managers aware of programmatic vulnerabilities?
2. How do program managers conceptualize these vulnerabilities?
3. How do program managers respond to these vulnerabilities?
4. What vulnerabilities are present in MCE programs?
5. What cybersecurity vulnerabilities does MCE pose?

The first four questions were the primary focus of that previous phase of research. In this phase of the research, the focus is on the fifth question as a means of expanding the cybersecurity component of the Reference CEM shown in Figure 2. When it came to the topic of cybersecurity vulnerabilities in general, the interviewees commonly raised the following issues:

- Cybersecurity needs to be thoroughly considered much earlier than it commonly is, preferably in the proposal generation stage.
- Program managers and systems engineers are sometimes intimidated by cybersecurity issues and thus seek to pass them onto specialists later in the acquisition process.
- MBSE and MCE toolset developers and proponents have not done a thorough enough job of considering programmatic cybersecurity vulnerabilities, though the tools are typically quite effective at designing for cybersecurity in end systems.
- Despite all of the above, according to the interviewees, traditional programmatic cybersecurity defensive practices tends to quite effective. This is due primarily to the conservative approaches most defense-related engineering groups use, as discussed in the Cybersecurity Vulnerability Assessment section. The increased use of MCE, particularly for multi-site collaboration, could change this.

The above points, many of which were commonly stated by the same expert, are clearly nuanced and complicated, with both points of success and failure. These points, along with more specific comments from the interviewees, resulted in an expanded cybersecurity CEM that can be seen in Figure 3. Note that in its full form, this would still be a part of the general Reference CEM shown in Figure 2. Here it is shown isolated for clarity.

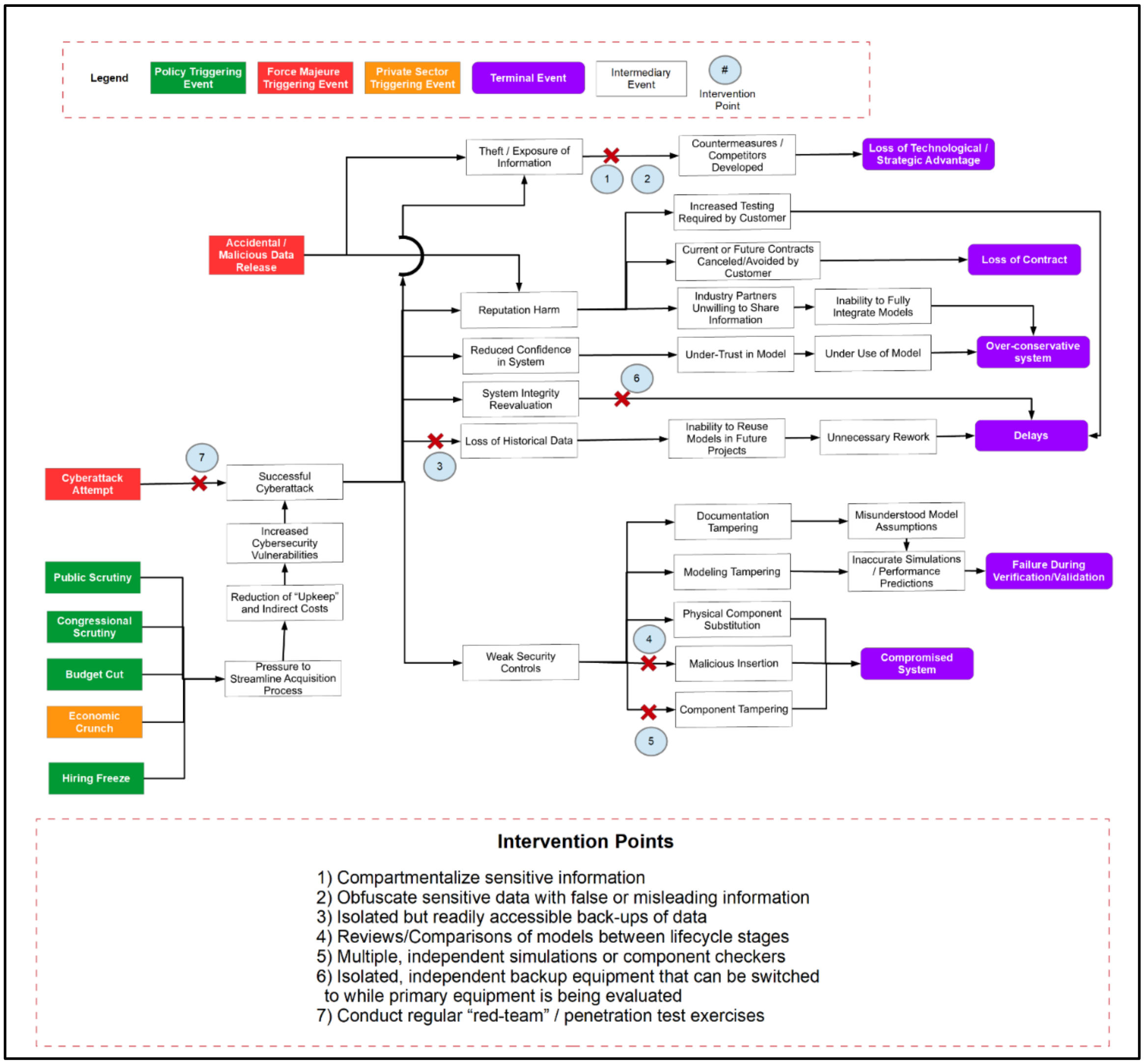


Figure 3. Reference Cybersecurity CEM (Preliminary)

Discussion

Some of the vulnerabilities and interventions shown in Figure 3 are not unique to MCE environments. Some of the vulnerabilities will simply be exacerbated by the increased use of MCE environments and processes. Some of the interventions will require new, creative means of implementing. For instance, Intervention Point #1 in Figure 3 is "Compartmentalize sensitive information." Clearly this is already done with the use of SCIFs and the Need-To-Know (NTK) information framework. However, such methods may not be feasible if the benefits of model integration and collaboration offered by MCE are desired. Instead, new methods must be developed. An example of one such possibility is the Federal Drug Administration's (FDA's) Sentinel Initiative, which involves querying a distributed system and receiving anonymized, aggregate data back (Office of Surveillance and Epidemiology, 2010). Such a system may allow modeling software to communicate across domains and locations, while still ensuring that even if one location is breached, only some information is exposed.

This Reference CEM does omit vulnerabilities and interventions that are entirely unchanged, however. For example, practices like the security clearance system and restricting the usual of digital storage media will remain effective interventions that are not significantly impacted by MCE environments.

One set of vulnerabilities that came up repeatedly in both the interviews and was observed in the class activity dataset were those that passed through the reputation harm intermediate event, as shown in Figure 4. Despite the frequency that the potential for this vulnerability was raised, few interventions were proposed for post-breach. This suggests that program managers and systems engineers could use more training in how to respond to breaches, particularly prominent ones, instead of just how to prevent them. While in the private sector there is evidence suggesting that the reputation harm incurred by a prominent breach does not significantly impact the firm (Lange & Burger, 2017), contractors to the government are known to suffer significant financial penalties due to breaches, even when such a breach is unrelated to their government duties (Braun, 2014; Overly, 2017). In a defense acquisition environment, there is thus significant incentive to having program managers (and the enterprise as a whole) well prepared to respond to major breaches.

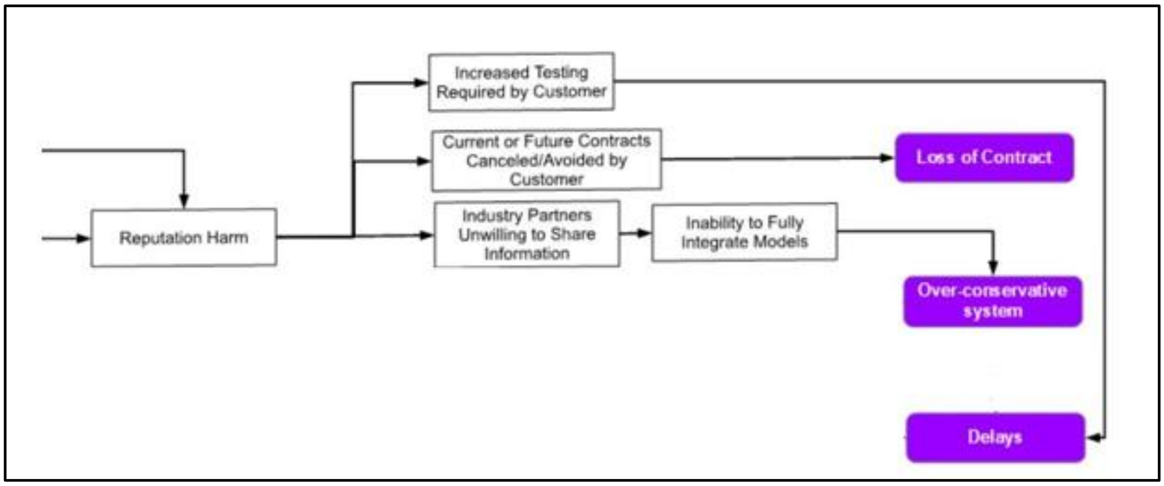


Figure 4. Reputation Harm Vulnerabilities, Section of Figure 3

CEM is intended to supplement, rather than replace, existing vulnerability assessment methods, particularly when it comes to cybersecurity. In this way, it can help fulfill the requirements set by NIST's Risk Management Framework (RMF; Ross et al., 2016) and the DoD's Defense Federal Acquisition Regulation Supplement (DFARS; Manufacturing Extension Partnership, 2017b). These regulations have shifted how government contractors handle cybersecurity. Previously, one-time assessments were completed and defensive practices instituted. Now the process is more dynamic. Contractors have to continuously assess threats and develop countermeasures as they arise, both with regards to the end-system and to the enterprise. CEM can potentially assist in this by serving as a reference that can be revisited as new threats arise.

Future Directions

As the research progresses, three directions of future research are being pursued. The first is to conduct a second round of interviews with other stakeholders in the acquisition process. The second is to evolve an interactive version of the Reference CEM. The third is to compare vulnerabilities present in MCE environments with those present in other, comparable fields.

Future Interviews

The interviews thus far have been with program managers and system engineers (the people who "live in" in the MCE environment). As this research proceeds, a future round of interviews with MBSE and MCE toolmakers and leaders of enterprise model-centric environments is planned. Several of the interviewees expressed an interest in increased enterprise-level ownership of MCE environments. Additionally, a few expressed concern about the degree of security in MCE toolsets. Thus it is worth talking to such individuals about their perspectives on vulnerabilities in MCE environments.

Interactive Tool

An interactive version of the CEM, which enables easy sorting and adding vulnerabilities, is desired. This would make the method more accessible, similar to how NIST's Cybersecurity Assessment Tool (Manufacturing Extension Partnership, 2017a) makes the RMF (Ross et al., 2016) more approachable to small manufacturers. Additionally, it could serve as a platform for future usability testing of CEM in MCE programs. In future research, an interactive demonstration prototype will be generated to synthesize the research outcomes and show how these can be used in practice.

Healthcare Industry Comparison

There is some indication that program managers may be well served by observing fields that are somewhat analogous to defense acquisition in order to derive helpful metaphors (Karas, Moore, & Parrott, 2008) or lessons learned (German & Rhodes, 2016).

The healthcare industry shows promise for such an analogy to cybersecurity in MCE environments. The healthcare industry deals with sensitive information, computer equipment, and high pressure environments. All of these are present at numerous stages of operation. Patient records have to be transferred from one system to another and be available to medical practitioners. Researchers need to be able to query systems in order to provide improved medical treatment but cannot violate individuals' privacy. They must do all this and more while under constant threat of cyberattack, as recent events have shown (Ryckaert, 2018; Woollaston, 2017; Zetter, 2016).

Engineers and researchers have made significant headway in making medical devices more interoperable with one another, particularly when it comes to sharing data securely (Goldman, 2014). Increasingly, model-based methods are being used to assess and design medical systems (Pajic et al., 2014). As was related in the Discussion section, the FDA's Sentinel Initiative seeks to enable active querying of medical data while preserving individual privacy.

All of these endeavors are strikingly similar to the challenges currently faced in defense acquisition. This suggests that there may be benefit in conducting a systematic comparison of the two fields. The healthcare industry, along with other fields, will be examined for potential metaphors and lessons learned that are applicable to understanding vulnerabilities in MCE environments.

Conclusions

Acquisition programs increasingly use model-centric approaches, generating and using digital assets throughout the life cycle. Recent advancements support new model-centric practices, yet uncertainties can lead to model-related vulnerabilities jeopardizing program success. Extending recent research (Reid & Rhodes, 2018) on vulnerability assessment of model-centric programs to cybersecurity, anticipated results are empirically-grounded cybersecurity vulnerabilities related to model-centric acquisition programs, and a prototype using a CEM reference model with dynamic analytic tools.

References

Arkin, B., Stender, S., & McGraw, G. (2005). Software penetration testing. *IEEE Security and Privacy, 3*(1), 84–87. https://doi.org/10.1109/MSP.2005.23

Baldwin, K. J., & Lucero, S. D. (2016). Defense system complexity: Engineering challenges and opportunities. *The ITEA Journal of Test and Evaluation, 37*(1), 10–16.

Blackburn, M., Verma, D., Dillon-Merrill, R., Blake, R., Bone, M., Chell, B., … Evangelista, E. (2017). *Transforming systems engineering through model-centric engineering. Hoboken, NJ: Systems Engineering Research Center*. Retrieved from http://www.sercuarc.org/wp-content/uploads/2014/05/A013_SERC-RT-168_Technical-Report-SERC-2017-TR-110.pdf

Braun, S. (2014, September 10). OPM plans to terminate contracts with USIS. *Federal News Radio.* Retrieved from https://federalnewsradio.com/management/2014/09/opm-plans-to-terminate-contracts-with-usis/

German, E. S., & Rhodes, D. H. (2016). Human-model interactivity: What can be learned from the experience of pilots with the glass cockpit? In *Conference on Systems Engineering Research*. Huntsville, AL.

Goldman, J. M. (2014, November). Solving the interoperability challenge. *IEEE Pulse.* Retrieved from https://pulse.embs.org/november-2014/solving-interoperability-challenge/

Gressin, S. (2017). The Equifax data breach: What to do. Retrieved March 27, 2018, from https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-do

Hanna, J., Smythe, C., & Martin, C. (2018, January 24). China's Sinovel convicted in U.S. of stealing trade secrets. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2018-01-24/chinese-firm-sinovel-convicted-in-u-s-of-trade-secret-theft

Hanselman, S. (2012). Everything's broken and nobody's upset. Retrieved from https://www.hanselman.com/blog/EverythingsBrokenAndNobodysUpset.aspx

Irvine, C., & Parfitt, T. (2013, July 11). Kremlin returns to typewriters to avoid computer leaks. *The Telegraph*. Retrieved from https://www.telegraph.co.uk/news/worldnews/europe/russia/10173645/Kremlin-returns-to-typewriters-to-avoid-computer-leaks.html

Kamali, M., Dennis, L. A., McAree, O., Fisher, M., & Veres, S. M. (2016). Formal verification of autonomous vehicle platooning. *Science of Computer Programming, 1,* 1–19. https://doi.org/10.1016/j.scico.2017.05.006

Karas, T. H., Moore, J. H., & Parrott, L. K. (2008). Metaphors for cyber security. *Sandia Report*. Albuquerque, NM: Sandia National Laboratories. Retrieved from http://evolutionofcomputing.org/Multicellular/Cyberfest Report.pdf

Kern, C., & Greenstreet, M. R. (1999). Formal verification in hardware design: *A survey. ACM Transactions on Design Automation of Electronic Systems, 4*(2), 123–193. https://doi.org/10.1145/307988.307989

Lange, R., & Burger, E. W. (2017). Long-term market implications of data breaches, not. *Journal of Information Privacy and Security, 13*(4). https://doi.org/10.1080/15536548.2017.1394070

Manufacturing Extension Partnership. (2017a). Cyber risk management. Retrieved March 29, 2018, from https://www.nist.gov/mep/cyber-risk-management

Manufacturing Extension Partnership. (2017b). DFARS cybersecurity requirements. Retrieved March 29, 2018, from https://www.nist.gov/mep/cybersecurity-resources-manufacturers/dfars800-171-compliance

Meadows, C. A. (1994). Formal verification of cryptographic protocols: A survey. *In International Conference on the Theory and Application of Cryptology* (pp. 133–150). Springer, Berlin, Heidelberg. https://doi.org/10.1007/BFb0000430

The MITRE Corporation. (2015). Terminology. Retrieved February 20, 2018, from https://cve.mitre.org/about/terminology.html

The MITRE Corporation. (2017). CVE-2017-5753. Retrieved February 20, 2018, from https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2017-5753

Morimoto, S. (2008). A survey of formal verification for business process modeling (pp. 514–522). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69387-1_58

NASA Office of the Chief Engineer. (1994). NASA Public Lessons Learned System. Retrieved July 13, 2017, from https://llis.nasa.gov/

Office of Surveillance and Epidemiology (Ed.). (2010). *The Sentinel Initiative*.

Overly, S. (2017, October). IRS temporarily suspends contract with Equifax. *Politico*. Retrieved from https://www.politico.com/story/2017/10/12/irs-equifax-contract-suspended-243732

Pajic, M., Mangharam, R., Sokolsky, O., Arney, D., & Goldman, J. (2014). Model-driven safety analysis of closed-loop medical systems. *IEEE Transactions on Industrial Informatics, 10*(1), 3–16. https://doi.org/10.1109/TII.2012.2226594

Pope, G. (2017). A hazard analysis technique for the internet of things (IoT) and mobile. In *STAMP Workshop*. Cambridge, MA.

Pope, G. (2018). Combining STPA with compiler technology to identify vulnerabilities and hazards in software-controlled systems. In *STAMP Workshop*. Cambridge, MA.

Pope, G., & Yampolskiy, M. (2016). A hazard analysis technique for additive manufacturing. In *Better Software East Conference*. Orlando, FL. Retrieved from https://arxiv.org/ftp/arxiv/papers/1706/1706.00497.pdf

Raymond, N. (2017, August 31). U.S. charges Chinese-Canadian citizen with trade secret theft. *Reuters2.* Retrieved from https://ca.reuters.com/article/topNews/idCAKCN1BB2K8-OCATP

Reid, J. B., & Rhodes, D. H. (2016). Digital system models : An investigation of the non-technical challenges and research needs. In *Conference on Systems Engineering Research.* Huntsville, AL.

Reid, J. B., & Rhodes, D. H. (2018). Assessing vulnerabilities in model-centric acquisition programs using cause-effect mapping. In *15th Annual Acquisition Research Symposium.* Monterey, CA: Naval Postgraduate School.

Ross, R., Dempsey, K., Pillitteri, V. Y., Jacobs, J., & Goren, N. (2016). Risk management. Retrieved March 29, 2018, from https://csrc.nist.gov/projects/risk-management/risk-management-framework-(RMF)-Overview

Rovito, S. M., & Rhodes, D. H. (2016). Enabling better supply chain decisions through a generic model utilizing cause-effect mapping. In *Proceedings of the 2016 Annual IEEE Sytems Conference*. IEEE.

Ryckaert, V. (2018). Hackers held patient data ransom, so Greenfield hospital system paid $50,000. *The Indianapolis Star*. Retrieved from https://www.indystar.com/story/news/crime/2018/01/17/hancock-health-paid-50-000-hackers-who-encrypted-patient-files/1040079001/

Schwarz, E. (2018). Automating vulnerability discovery in critical applications. Retrieved from https://www.sei.cmu.edu/research-capabilities/all-work/display.cfm?customel_datapageid_4050=6487

Statt, N. (2018, March). Boeing production plant hit with WannaCry ransomware attack. *The Verge*. Retrieved from https://www.theverge.com/2018/3/28/17174540/boeing-wannacry-ransomware-attack-production-plant-charleston-south-carolina

Woollaston, V. (2017, May). The NHS trusts and hospitals affected by the Wannacry cyberattack. *Wired*. Retrieved from http://www.wired.co.uk/article/nhs-trusts-affected-by-cyber-attack

Young, W. E. (2013). A system saftey approach to assuring air operations against cyber disruptions. In *STAMP Workshop*. Cambridge, MA.

Young, W. E., & Porada, R. (2017). System-theoretic process analysis for security (STPA-SEC): Cyber security and STPA. In *STAMP Workshop*. Cambridge, MA.

Zetter, K. (2016, March). Why hospitals are the perfect targets for ransomware. *Wired*. Retrieved from https://www.wired.com/2016/03/ransomware-why-hospitals-are-the-perfect-targets/

Acknowledgment

This material is based upon work by the Naval Postgraduate School Acquisition Research Programs under Grant No. N00244-17-1-0011.

Automated Methods for Cyber Test and Evaluation[1]

Valdis Berzins—is a professor of computer science at the Naval Postgraduate School. His research interests include software engineering, software architecture, reliability, computer-aided design, and software evolution. His work includes software testing, reuse, automatic software generation, architecture, requirements, prototyping, re-engineering, specification languages, and engineering databases. Berzins received BS, MS, EE, and PhD degrees from MIT and has been on the faculty at the University of Texas and the University of Minnesota. He has developed several specification languages, software tools for computer-aided software design, and fundamental theory of software merging. [berzins@nps.edu]

Abstract

Cyber security of mission-critical software is a relatively new concern that is difficult to measure and hence difficult to incorporate effectively in software development contracts. The DoD has typically relied on black-box approaches to software testing. However, cyber vulnerabilities, particularly those deliberately injected into systems, are often statistically invisible with respect to affordable levels of black-box testing, which implies that they cannot be effectively detected using conventional testing techniques. This motivates augmenting traditional testing approaches with additional types of test and analysis procedures. This paper explores application of automated testing and other automated analysis methods to reduce cyber risks. We analyze several types of undesirable software behaviors and identify automated methods that could detect them within practical limits on time and computational resources.

Overview: Cyber Testing Challenges

Failures Are Not Random

The quality objectives for cyber concerns are seemingly similar to those for software meant to operate in uncontested environments, but on closer examination, there is a fundamental difference with far-reaching consequences for testing and evaluation. In both cases, we wish to minimize the risk of improper software behavior, consistent with the policy set in ICD 503 (Office of the Director of National Intelligence, 2008). However, "risk" has very different meanings in the two contexts.

For uncontested environments, failures act like random processes, and an appropriate risk concept is a statistical combination of severity of consequences from each type of potential mishap weighted by their frequency of occurrence, consistent with the formulation in DoD (2012). The ideas of "safety" and "reliability" are based on this point of view, which equates risk to the expected loss, damage, or injury when averaged over time. The unstated assumption of most work in this camp is that the probability distribution for mishaps is stationary, which means that the frequency of occurrence is stable over long periods. This approach is reasonably consistent with the properties of failures that are due to unintentional events with unpredictable variations, such as equipment wearing out, human errors, electronic noise, background radiation, and so on.

[1] The views presented in this paper are those of the author and do not necessarily represent the views of DoD or its components.

In contested environments, failures are not random, and frequency of occurrence can vary dramatically depending on external circumstances, driven by the existence of an adversary whose deliberate behavior depends on variable conditions, such as the following:

- Are we at war?
- How much profit/military advantage/political value would a successful attack provide?
- Are sufficient resources available for successful attack?
- How much risk of prosecution or counterattack is there?

In this case, which matches cyber-attacks, game theory provides a better model of risk than statistics. The associated underlying assumption is that there is a capable adversary who will choose courses of action that maximize damage. Consequences of this paradigm shift include the following:

- Focus of risk management shifts from minimizing expected loss to minimizing worst-case loss.
- Scope of risk management expands from mitigations concerned solely with the software to those that address both the software and the adversary.
- Risk assessment becomes sensitive to surprises due to new adversary tactics.
- Unlikely conditions and rarely traversed paths through the code can no longer be ignored.

Causes and Effects Will Be Hidden

A consequence of Rice's theorem is that perfect cyber certification is impossible. This theorem is a well-known result in computability theory that says any non-constant property of program behavior is undecidable. Non-constant means the property is true for at least one program and false for at least one other program, which holds for all software security properties of interest. "Undecidable" means there is no systematic method (algorithm) that will always produce a correct decision and will always terminate in a finite amount of time. This theorem applies to both testing and static analysis of the source code.

Since any workable certification procedure must fit into a definite schedule, it must operate within some reasonably short bounded time, say less than a year. The theorem therefore implies that all practical certification procedures produce imperfect decisions: Any such procedure must produce some false positives and some false negatives, or fail to reach a conclusion for some inputs.

Capable adversaries expect defenders to search for vulnerabilities, know that detection will be imperfect, and design their attacks to make them difficult to find. Some ways to hide are as follows:

- Small footprint: Design malicious behavior so that it will be triggered in only one of a huge number of possible execution conditions. The triggering condition can be made statistically invisible very easily because the search space has exponential size[2] (see examples in Berzins et al., 2015).
- Fragmentation: Malicious behavior resulting from interaction between widely separated parts of the code, such as exception handlers and multiple threads with no logical connection. Such non-local interactions are difficult to detect in large systems.
- Delayed manifestation: Corrupt the code or data in ways that will not affect behavior until much later, possibly waiting for a statistically invisible external trigger.
- Timing: Information content of behavior is correct, but is delayed sufficiently to cause failures. Software that controls physical components is susceptible to this hazard.
- Parasitic effects: Breaking the model of computation so that logically correct source code can produce damaging behavior.

Consequences Are Physical

Much work on cyber security focuses on information—how to keep it confidential, free from corruption, authentic, and so forth. However, risk-based approaches are guided by severity of consequences, and relevant consequences are physical. Thus, system context must be considered as part of risk management, and if context is expanded far enough, all critical software is part of some cyber-physical system and either controls or influences its behavior. The following are examples of possible consequences of software faults:

- Dangerous physical events involving controlled equipment, such as collisions between moving vehicles or discharge of weapons
- Disrupting defensive capabilities of a military platform in a conflict, increasing vulnerability to kinetic attack
- Revealing the location of a military unit or identity of a covert operative, exposing them to attack
- Revealing military plans, enabling adversaries to target weak points and increase damage
- Economic and political analogs of the above

[2] 2^b cases to test, where *b* is the number of bits in all input and state variables combined.

Threats Can Morph

The most serious cyber risks are due to corruption of software at runtime. Runtime corruption of the executable code can result in unlimited damage at the discretion of the adversary. Not only is severity of consequence potentially the worst possible, but also, detection of this type of threat is especially challenging because the potentially damaging behavior is inserted after certification processes are completed. The destructive payload is not present either in the source or in the executable that is analyzed, so there are no adverse consequences to detect in the initial uncompromised configuration of the system under test.

Recovering From Mishaps

The above lines of reasoning imply that software defenses will never be perfect, especially for systems of practical size, for which exhaustive analysis is impractically expensive both in terms of money and in terms of available time (billions of years may not be enough).

Strategies involving defense in depth can be useful for extending the time that systems under cyber-attack can remain operational, but they cannot provide complete immunity to attack by a capable adversary. This well-known approach to cyber-defense needs to be augmented with self-healing capabilities—so that systems can recover from partially damaged states and continue operating despite partially successful cyber-attacks. Even better are adaptive immune responses that reconfigure a system damaged by attack so that it is no longer vulnerable to a replay of the previously successful attack. The objective should be to increase the time and cost of successful attacks to the point where they are not affordable by adversaries.

Some self-healing capabilities are practical at the current state of the art, and continued research to improve this approach is recommended.

You Don't Know What You Don't Know

Adversaries are constantly finding new ways to attack systems. This makes it difficult to write development contracts for secure software, because contracts inherently define fixed responsibilities for the contractors, but the set of actual threats is incompletely known and open-ended. Supporting a rapid repair capability requires agility both in the software and in the contracting approach.

This seemingly unsolvable problem can be addressed by suitable application of Open Systems Architecture and Technical Reference Frameworks. Rapid reconfiguration can be supported by an architecture that accommodates all needed configurations without changing the architecture. In this context, architecture can be considered to consist of the aspects of a dynamic system *that do not change*. Architecture for cyber-resilient systems should include standardized structures and requirements for supporting functions related to resilience, such as runtime monitoring and self-healing functions.

As an example, consider the high-risk threat of runtime code compromise. Services called out in associated parts of a Technical Reference Framework should include facilities for the following:

- Secure, authenticated distribution of software updates
- Runtime monitoring of executable code to detect unauthorized changes
- Restoring corrupted code to an authorized configuration
- Restoring the execution state to a valid configuration and resuming execution with the restored code configuration

These services and modules should all conform to a stable standardized interface specified in the architecture, so that best-in-class components providing these services can be shared across different systems and future technology improvements related to these critical issues can be readily incorporated by software module swapping, ideally without stopping operation of the system. The example illustrates a vision of how rapid reconfiguration capabilities could be specified in a fixed architecture fragment that could be called out in fixed and definite development contracts.

Insider Threats

Turn-Key Malware

Malicious insiders may build some types of malware into software before delivery. These people are part of the development team and have full access rights to the code. Examples of this type of malware include "Easter Eggs," which are extraneous bits of functionality that are typically triggered by some single special input value, often one that is extremely unlikely to be encountered as part of a normal workload. Although many known instances of Easter Eggs have done little harm, the ability to detect the pattern is a cyber-testing concern because Easter Eggs can also hide extremely damaging extra capabilities, such as enabling unauthorized access or unauthorized modifications to a system.

Testing Difficulties

Common testing practices in the DoD rely heavily on black-box testing, in which test cases are designed based on the requirements, without knowledge of the structure of content of the source code. The method is widely used because it is reasonably good for checking that the delivered software has the behavior specified in the requirements, which is a primary concern in acceptance testing. It may also be the only viable testing approach if the development contract does not include rights to access the source code for the developed software.

Unfortunately, black-box testing is not effective for checking the absence of undesired extra functionality, such as deliberately planted cyber vulnerabilities. Since there is often only a single test case that could demonstrate the existence of an embedded Easter Egg and the number of possible test cases is usually astronomical, the odds of detection by black-box testing are practically none (see Berzins et al., 2015, for quanititative details).

Solutions

Clear-box testing with respect to the statement coverage criterion can detect Easter Eggs in a practical manner that can be readily specified as a development requirement in a contract. The statement coverage criterion requires that every statement in the source code must have been executed by at least one test case. Relatively low-tech tools that count the number of times each statement in the code has been executed can check compliance with this requirement; the requirement is met if all of these numbers are greater than zero, which can also be checked by a simple piece of software. Many compilers include options for measuring statement coverage, for example, the gcov facility in the gcc tool set, which directly reports the percentage of source statements that have been covered by a test ("Monitoring Statement Coverage," n.d.).

Difficulties with this approach include finding test cases that can exercise rare paths and handling unreachable sections of the code. Although the general problem of finding test cases that trigger particular paths in the code does not have an effectively computable solution, experience with fuzz testing shows that constraint solver tools can handle a majority of the cases that arise in practice (Cadar et al., 2008). Additional tactics that can be useful for exercising rare paths are to seek module-level test cases that reach the statements in question, as opposed to system-level test cases. This provides more direct control over the execution state of the module and simplifies the constraints that need to be solved to generate the needed test cases, enabling a larger fraction of the cases to be solved automatically.

In some cases, unreachable code as well as extraneous code that does not affect the outputs of software services can be identified by using a form of dependency analysis known as software slicing (Berzins, 2014). These automated methods can help diagnose parts of the code that could not be exercised by test cases. The remaining cases are a small fraction of the code and may be few enough to be affordably examined by human analysts.

Outsider Threats

Runtime Code Modification

As noted above, runtime code modification is the most serious cyber risk in any system, because its severity of consequences includes the consequences of all other cyber risks. From a game-theoretic viewpoint, which focuses on worst-case risk, we expect an adversary to inject the most damaging exploit available to them if they chose a runtime code modification attack.

The signature of a runtime code modification vulnerability is not presence of inappropriate software behavior, but rather the existence of a possible path for executable code (or data that affects code behavior) to be eventually modified without authorization. The triggering condition as well as the inappropriate actions in such a path may involve interactions that bypass the official interfaces of the system to be certified, and may not be present in the high-level models programmers use to design and check their code. For example, the triggering event may involve corruption of system memory from a logically unrelated function or process. This implies that this type of malicious behavior may be completely invisible to black-box testing in the initial uncorrupted state of the software, not just statistically invisible.

Another consequence of this cyber risk is that the critical parts of the code cannot be localized within current development approaches, especially in the context of programming languages without garbage collection and provisions for memory protection. Any part of the software that manipulates pointers can be a potential avenue for attacks that modify code at runtime, and these parts are spread throughout most systems. This makes it very difficult to focus the most intensive testing and analysis efforts on just the "security critical" parts of the system, and greatly increases testing cost for high confidence systems.

Detection

Following the principle of defense in depth, we suggest a layered approach to detection, coupled with mitigations that combine preventive and remedial measures. This section focuses on detection. The following are possible measures:

- Software update service analysis
- Architecture conformance checking
- Memory allocation checking
- Memory reference checking
- Runtime monitoring of executable code

Many current systems include explicit interfaces for upgrading the software to new versions or distributing patches. This part of the system is cyber-critical, expected to be a focal point for attack, and should be subjected to the most intense degree of testing and analysis possible, at multiple levels.

- At the requirements and architecture level, check whether the service is required to authenticate authority to update the software and check integrity of the transmitted code, and whether the methods for doing so are the strongest available at the time. The latter review has to be repeated regularly to account for future development of improved methods for providing these capabilities.
- At the source code level, do static analysis of the implementations of these methods, up to and including constructing/checking mathematical proofs of the security properties of the protocols and algorithms used for transmitting and installing the software updates. Also check that the source code matches the algorithms and protocols that were proved, if the proofs are done based on some representation other than the actual source code for the service.
- At the executable code level, do penetration testing by highly competent red teams aimed at these services, and check that the executable code matches the source code and is free from extra functions. This last step is needed to guard against possible compromise of the compilers, linkers, and loaders used to build the software, which could be corrupted to add back doors into the executable code they produce, which could be specifically targeted at just the implementation of the software update service. Such back doors could bypass the authentication and integrity checks that exist in the source code and provide unauthorized access to the software update service when activated by adversaries.

Checking conformance to architecture includes checking that interfaces and executables do not contain any extra services or interaction paths, beyond those specified in the system interfaces. Such extra services could be avenues for execution of malware, and extra interactions could be paths for triggering malware or exfiltrating its results. Architecture conformance has two levels: checking actual source code interfaces for extra services, and checking source code of services that are specified in the architecture for extra interactions, such as reading or writing from files, network locations or global variables that are not included in the interface specification for the service. This latter check can be done via dependency analysis algorithms such as data flow analysis and software slicing (Berzins & Dailey, 2009).

Memory allocation checking consists of checks for references to pointers that have not yet been initialized or that point to memory areas that have already been deallocated ("dangling pointers"). This is a common problem in programming languages without automatic garbage collection, and there exist commercial tools for doing such checking, including Insure++ and Valgrind ("Parasoft Insure++," n.d.; "Valgrind's Tool Suite," n.d.).

Memory reference checking is a runtime check that all pointer references refer to a non-null object of the proper type and that all array references are within the range of declared array bounds. Compilers of some languages can do this, and for some such as Ada, the runtime checks are required except for contexts in which the compiler can prove they are unnecessary because violations are impossible. Requiring use of such facilities would make it more difficult for adversaries to create exploits that corrupt memory containing executable code and critical data. Proof systems such as Spark can check properties such as these mostly automatically ("Spark Pro," n.d.).

Runtime monitoring of executable code is an active check at runtime that the executable code matches the most recent version that was installed (Berzins, 2014). There should be a Technical Reference Framework that specifies a standard service for doing this, which can be easily incorporated in the architecture of any mission-critical system. This would enable development of standard software implementations of this service that could be used in any system conforming to the architecture. These implementations could have variants that work with different operating systems and different programming languages.

Mitigation

Detection of vulnerabilities and attacks is not sufficient for achieving reliable system operation—mitigation and recovery are needed as well. Some mitigations for code corruption attacks are as follows:

- Using pure code segments contained in read-only hardware. This preventive measure would make runtime code modification attacks impossible, at the expense of specialized hardware and prohibition of automatic installation of software updates. This approach is not a new idea—it was used in very early systems that had magnetic drums as secondary memory (predating magnetic disks). A modern version could use erasable programmable read-only memory (EPROMs), which could allow updates but only with human intervention via physically exposing the memory chips to UV light to erase them and enable them to be reprogrammed. This would make software updates less automatic, but could provide two-factor authentication for updates and guarantees of absence of change between such updates. This mitigation is relatively expensive and cumbersome, but it could provide very high levels of runtime code protection to critical applications that really need it.
- Restoration of code from ROM. A lighter-weight version of the above mitigation is to require runtime monitoring of the executable code, as described in the previous section, together with a facility for restoring the code from a backup copy in read-only memory (such as a locked CD ROM), restoring a safe execution state, and continuing the operation of the software. This solution can be implemented using existing hardware, but would require more time to restore operation after a failure. A stronger version of this idea would use a backup copy with a different code layout, which would reduce the chances that replay of the same attack would succeed again, thereby increasing time to next failure.
- Disabling reflective language capabilities. Some modern programming languages, including Java, provide capabilities for runtime inspection and modification of interfaces and implementations. Access to such capabilities makes an adversary’s job much easier. Development contracts should either require the use of a programming language without reflective capabilities, or require the developer to demonstrate that those capabilities have been removed from the system.
- Use of programming languages with garbage collection should reduce exposure to the threat of code and data corruption. The memory allocation and recycling facilities of languages and their supporting systems (compilers, runtime libraries, linkers, loaders, etc.) should be intensively checked for faults that could lead to memory corruption in applications constructed using those languages and systems.

Conclusions

There is no silver bullet when it comes to cyber security, and no such thing as a completely secure system. The best practical solutions involve a layered set of defenses and mitigations that increase the time and effort it will take an adversary to compromise the system and decrease the time to detect a compromise and restore dependable operation. An appropriate goal would be to make system compromise prohibitively expensive for most, if not all, potential attackers.

This paper defines a risk concept appropriate for gauging cyber threats, identifies cyber risks with greatest risk exposure, and suggests corresponding methods for detection and mitigation. In addition to methods that make the systems more difficult to compromise, we recommend further investigation of mitigations that address both the system and the adversary. This would include stronger methods for authenticating access to systems and networks, along with facilities for recording and linking people's identities to evidence of potential wrongdoing that could support deterrence in the forms of legal prosecution of individual wrongdoers and determined public action against state-sponsored attacks.

References

Berzins, V. (2014). Combining risk analysis and slicing for test reduction in open architecture. In *Proceedings of the 11th Annual Acquisition Research Symposium* (NPS-AM-14-C11P07R03-038; pp. 199–210.). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Berzins, V., & Dailey, P. (2009.) How to check if it is safe not to retest a component. In *Proceedings of the Sixth Annual Research Symposium—Acquisition Research: Defense Acquisition in Transition* (pp. 189–200). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Berzins, V., Van Benthem, P., Johnson, C., & Womble, B. (2015). Use of automated testing to facilitate affordable design of military systems. In *Proceedings of the 12th Annual Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Cadar, C., Dunbar, D., & Engler, D. (2008). KLEE: Unassisted and automatic generation of high-coverage tests for complex systems programs. In *Proceedings of the Eighth USENIX Conference on Operating Systems Design and Implementation* (pp. 209–224). Retrieved from http://static.usenix.org/legacy/events/osdi08/tech/full_papers/cadar/cadar_html/paper.html

DoD. (2012). *Standard practice system safety* (MIL-STD-882E). Retrieved from http://www.system-safety.org/Documents/MIL-STD-882E.pdf

Monitoring statement coverage with gcov. (n.d.). Retrieved from https://www.cs.odu.edu/~zeil/cs333/website-s12/Lectures/wbtesting/pages/gcov.html

Office of the Director of National Intelligence. (2008). *Information technology systems security risk management, certification, and accreditation* (Intelligence Community Directive Number 503). Retrieved from https://www.dni.gov/files/documents/ICD/ICD_503.pdf

Parasoft Insure++. (n.d.). Retrieved from http://www.parasoft.com/jsp/products/home.jsp?product=Insure

SPARK Pro. (2012). Retrieved from https://www.adacore.com/sparkpro

Valgrind's tool suite. (n.d.). Retrieved from http://valgrind.org/info/tools.html

Cybersecurity: Converting Shock Into Action (Part 1)

Paul Shaw, CAPT, USN (Ret.)—is a Professor of IT (cyber security emphasis) at Defense Acquisition University. Shaw retired from the Navy after 30 years of service (active and reserve). He is a former Naval Aviator and Aeronautical Duty Officer (AEDO). Shaw is a Doctorate of Science (DSc) candidate in Cybersecurity at Capitol Technology University (expected completion: 2018). He has also earned five master's degrees (systems engineering, IT, international relations, finance, and management) and a BS in ocean engineering (USNA).

Robert Tremaine—is the Associate Dean for Outreach and Mission Assistance at the Defense Acquisition University West Region. He has over 30 years of experience in air, missile, and space weapon systems acquisition. Col Tremaine holds a BS from the U.S. Air Force Academy and an MS from the Air Force Institute of Technology. He is Level III Defense Acquisition Workforce Improvement Act certified in both Program Management and Systems Planning, Research, Development, and Engineering. Col Tremaine is a graduate of the Canadian Force Command and Staff College in Toronto, Ontario, Canada; and the U.S. Army War College in Carlisle Barracks, PA. He also completed a military research fellowship in association with the Harvard Business School.

Introduction

BIGBADABOOM-2. That's the name of a recent cybersecurity breach affecting 5 million stolen credit card and debit card holders (O'Brien, 2018). Unfortunately, these breaches are becoming all too common. At an alarming rate, nation states and malign actors are better equipped to conduct cyberattacks than ever. The risk is growing. Some adversaries will be able to disrupt critical infrastructure against the United States in a crisis short of war (Coates, 2018). To make matters worse, cyber threat actors are more threatening and their abilities more sophisticated. While "abilities" are just as important to defend against cyberattacks, attitudes are just as vital when it comes to the selection of the required learning strategies given their connection to necessary cybersecurity behaviors. Unfortunately, the DoD's current approach for the acquisition community won't easily fulfill the stated and implied security and resilience imperatives anytime soon unless attitudes (a critical catalyst) start to change. The learning strategies required that embody it trace back to Bloom, Krathwohl, and Harrow—all research leaders in their respective fields. Their works speak to the importance of the affective domain (i.e., the way our attitudes affect our learning behaviors). This study explores the impact of the DoD's overall implied cybersecurity learning strategy and associated actions taken to date—all intended to safeguard the efficacy of the DoD's weapon systems and supporting infrastructure. Also included is a case study discussion to demonstrate the cybersecurity actions taken by one particular organization to better prepare themselves for their assigned cybersecurity duties despite the DoD's good intentions. The learning outcome of this case study could serve as a forerunner for other DoD acquisition organizations as they consider how to implement a robust, effective and sustainable cybersecurity program. The researchers firmly believe that the DoD will be hard pressed to achieve the desired gains in security and resilience without recognizing that the critical cybersecurity behaviors and concomitant attitudes at the individual, team, and organizational levels come first. And, that might come as a shock.

Background

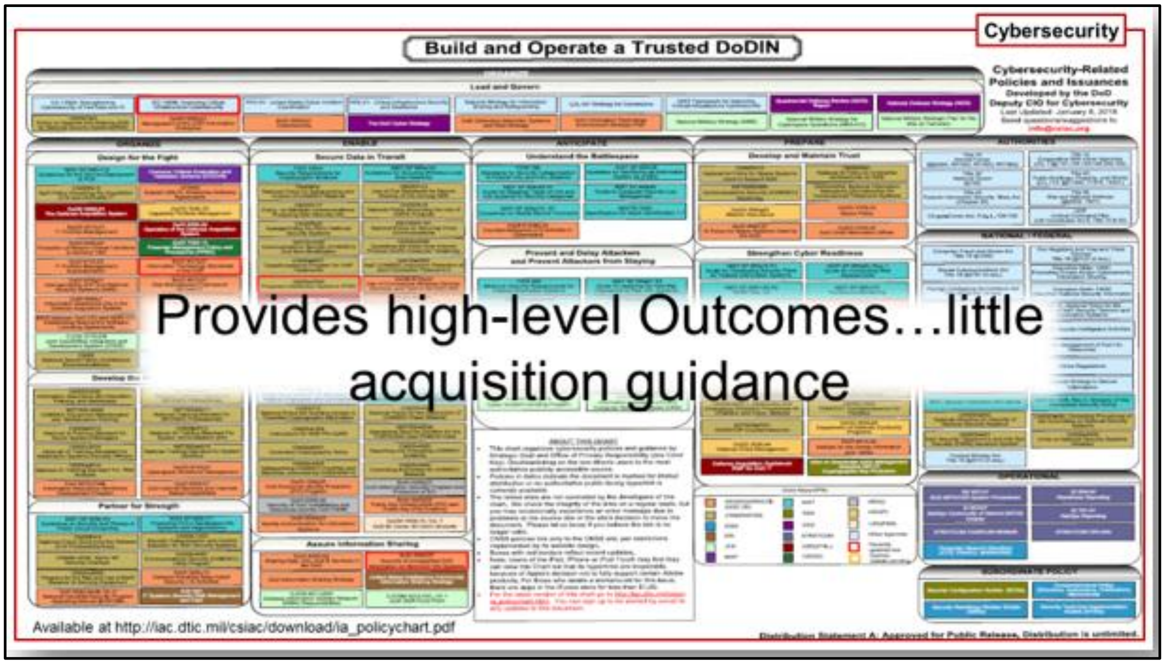
Like any emerging challenge as complex as cybersecurity, organizations will test the outer edges of their learning envelopes. To better guide this research pursuit, the authors used four specific questions to better isolate these and other learning implementation limitations currently found in the DoD's cybersecurity learning strategy. The answers were both informative and instructive:

1. **Have the DoD's actions (e.g., policy directives, tools, methods, etc.) met the stated and implied expectations for protection and resilience in the acquisition community?**

   Not really. Results of the independent assessments of the DoD's cybersecurity reports from the Defense Science Board and the DoD's testing community collectively signaled severe concerns about whether or not the DoD can accomplish its core missions and keep its critical assets intact. In 2015, a RAND Corporation report found that "cybersecurity risk management does not adequately capture the impact to operational missions and that cybersecurity is mainly added onto systems, not designed in (Snyder et al., 2015, p. ix)." Rand went on to say that the policies governing cybersecurity are better suited for simple, stable, and predictable environments leading to significant gaps in cybersecurity management. The consequences could include the following:

   - prescriptive solutions for military system cybersecurity that favor security controls over more sound system security engineering,
   - emphasis of processes and security controls for information technology systems over more tailored military systems solutions,
   - implementation of tactical security controls over more strategic mission assurance imperatives, and
   - overreliance on standardized and formalized security control compliance as a means to achieve cybersecurity (Snyder et al., 2015, p. viii).

To communicate cybersecurity imperatives, the DoD Chief Information Officer (CIO) regularly updates (about every three weeks) the policies affecting cybersecurity in a summary chart called "Build and Operate a Trusted DoDIN" (see Figure 1). Aside from the microscope nature of the details, the chart is largely outcome-based only (http://iac.dtic.mil/csiac/ia_policychart.html). Many of these high level outcomes exceed the security capabilities of the DoD's current systems capabilities recommended by the DoD DIACAP and now Risk Management Framework (RMF) process. Many of the other outcomes are either slightly mentioned in current acquisition documents or absent. The disconnect is readily apparent. How programs actually capture the DIACAP/RMF processes appear to be more compliant dominant and presumably driven by official approval of the system instead. These programs also tend to depend on a "cookie cutter" approach where they sometimes use a template overlay for security controls. Without thinking more critically about every likely eventuality along with leveraging the testing community's expertise to confirm operational objectives beforehand, these same programs face looming cost and schedule risks. The OCX program reinforces the repercussions when they do. Raytheon fully underestimated the cybersecurity requirements by discounting the impact of COTS and free and open source software. It represented one of several factors that contributed to a multi-year schedule delay and cost increase estimated to exceed $1 billion (Kendall, 2016).



Note. This chart is updated frequently by the DoD CIO.

Figure 1. Build and Operate a Trusted DoDIN

What are the metrics and have they been effective?

The answer is no.

- Most systems have rudimentary security requirements for implementing metrics. They generally include the following:
  - exercising logical access controls with certain frequency,
  - managing software inventories at certain intervals,
  - implementing information security management in accordance through prescribed methods, and
  - monitoring/detecting data exfiltration.

  While NIST 800-53 security controls recognize these type of metrics as a good start, programs fall short of implementing a dynamic evaluation approach that includes testable standards with the proviso that they need to evolve as a part of a system's inherent system security architecture.
- Most Program Office requirements fall short of testing at levels that mimic likely operational conditions and scenarios. Systems designers normally concentrate on the threats to and subsequent actions required in the context of information exchanges within their system where they believe they could be more easily exploited. Alternately (and more effectively), the testing community uses external stimuli they expect to see surface in an operational environment that could easily (and frequently) exploit security and resilience gaps. They don't treat systems as adiabatic in any way, shape or form. This effect is especially evident when high level requirements go beyond basic system behaviors (from inside the system to the system's exchange of information requirements).

- In an independent assessment conducted by the Office of Management and Budget (OMB) with *Federal Information Security Modernization Act of 2014: Annual Report to Congress* for FY 2016, the DoD's information security program received an uncomplimentary rating (p. 44). From a scale of 1 (lowest) to 5 (highest), the DoD earned grades on the lower end of the scale compared to all federal agencies that earned grades on the higher end, ranging from 4 to 5. The DoD's grades were consistently on the low end:
    - Identify—Level 2 Defined
    - Protect—Level 2 Defined
    - Detect—Level 1 Ad Hoc
    - Respond—Level 2 Defined
    - Recover—Level 2 Defined

Can these shortcomings be overcome? It requires a change of approach, culture, and workforce attitudes.

Is the DoD headed in the right direction?

The answer is partly.

- The DoD has reinforced "cybersecurity as a requirement for all DoD programs across the life cycle" (DoD CIO, 2014, p. 155).
- The DoD has recognized that all systems must manage "risk commensurate with the importance of supported missions and the value of potentially affected information or assets" (DoD CIO, 2014, p. 2). Moreover, the DoD's cyber strategy emphasizes the need to:
    - not defend every network against all threats;
    - identify, defend, & prioritize most important for mission;
    - be able to operate in degraded & disrupted environments; and
    - use technology & innovation to stay ahead of threat (p. 13).
- The DoD emphasizes the need for systems to be both secure and resilient. Security mechanisms afford a defense against a cyberattack or allow a system to maintain operations. Resilience can reset a system, even if the cyberattack is not detected or understood (Defense Science Board, 2016), or allow it to operate in a degraded mode. Critical cyber components could implement resilience for performance of critical functions, regardless of fault cause or nature (Defense Science Board, 2013).

  Systems like WIN-T changed their thinking to incorporate threat-based engineering and developed multiple threat models. They assumed comprise and adopted a continual testing process strategy and cybersecurity that became an inherent part of the engineering processes across their systems. Cybersecurity was no longer a separate solution.

More specificity is warranted. The DoD's high-level policy has many cybersecurity elements at the outcome level to guide programs, although noticeable gaps in acquisition guidance persist for the effective implementation of key objectives in a meaningful way (i.e., how to better respond to realistic conditions that the operational test community will impose). The DOT&E annual report dated January 2018 indicated that "despite improvements in network defenses, *almost every assessment and test demonstrated that DoD network defenses still contain exploitable problems* that provide cyber adversaries opportunities for access to DoD networks" (p. 318). If the DoD were to compel the acquisition workforce to go beyond a "compliance construct" for cybersecurity, more systems might just pass various Adversarial Assessments in Operational Test and fulfill Operational Commanders' mission assurance needs. This requires a change of approach, culture, and workforce attitudes.

What industry best practices should the DoD adopt and why?

Industry best practices have concentrated their efforts on resilience, trustworthiness, and continual testing. Intel, Google, Microsoft, Netflix, and others have boosted their security posture by going beyond traditional security activities that focus just on system protection. For example,

- Intel employs a Trusted Execution Technology to ensure their operating kernels are of a known trusted state.
- Google verifies that all servers in their data centers operate from a globally distributed trusted image.
- Microsoft's evolving security posture continually evaluates threat activities.
- Netflix conducts cybersecurity testing with the Simian Army in continuous mode and digitally stresses their content delivery infrastructure to influence responsive systems engineering actions.

All these companies have adopted a security posture of adaptability and innovative thinking in response to impending cyber threats. They don't think for a second they won't be compromised. Their active measures are also consistent with comments made by the Director of Operational Test and Evaluation FY 2016 Annual Report, where he said, "Cybersecurity tests will demonstrate active defense from attacks, measure the effectiveness of the cyber defenses, and assess the mission impacts resulting from cyber-attacks" (Behler, 2018, p. 447). These cybersecurity strategies align with the September 2016 Defense Science Board report on Cyber Defense Management, which suggested "examining the attack data to determine what is working well, what is not, where changes need to be made, and where investment is required to better defend against troublesome or emerging threats to move beyond a compliance approach towards a more dynamic performance evaluation" (p. 11). Will this type of thinking eventually become pervasive in the DoD? It requires a change of approach, culture, and workforce attitudes.

Assumptions

As with any research study, assumptions generally help characterize the research constraints as well as the prevailing environmental domain. For cybersecurity, it's no different. While strikingly provocative, the following assumptions reinforce today's cybersecurity operating envelope:

- Cybersecurity is a decaying function—static cybersecurity assures a declining security posture.
- NO SYSTEM is without malware—every system has an inherent vulnerability just waiting to be exploited.
- Organizations rely too much on technology for security and don't sufficiently consider the people and process components.
- The seemingly most secure system often fails to acknowledge that it can be affected by a higher-level threat (e.g., any system can be misconfigured).

- Cybersecurity policy stands at the outcome level; acquisition guidance and implementation below the outcome level is subjective (i.e., outcome level is typically characterized as “design for the fight”).
- Most programs undershoot “adequate security”—many operate under a false sense of security until they discover they did not sufficiently manage realistic and likely operational risks.
- The DoD may not be proactive enough to exploit its own systems to withstand advanced threats.

Research Tools Used

Several tools were ideally suited for this research pursuit. The first, a high-level logic model (see Figure 2), would set the flow, narrow the focus, and underpin the researcher’s end-in-mind. In the past two decades, its usefulness has also been recognized by others. Clarke (2004) used diagramming since they link categories with categories to form a substantive theory of action that shows “at a glance if outcomes are out of sync with inputs and activities.” They help researchers “make sense of relationships that may not have been previously explicit” (Buckley, Waring, 2013). Spaulding and Falco (2013) found they “provide linkages between activities and outcomes as well as to serve as a framework for developing quality and purposeful activities.” For cybersecurity, there are no perfect solutions. However, “living” models like logic models could expose these new truths during a project’s life cycle, especially when the operating environment can be so dynamic and ambiguous at the same time.

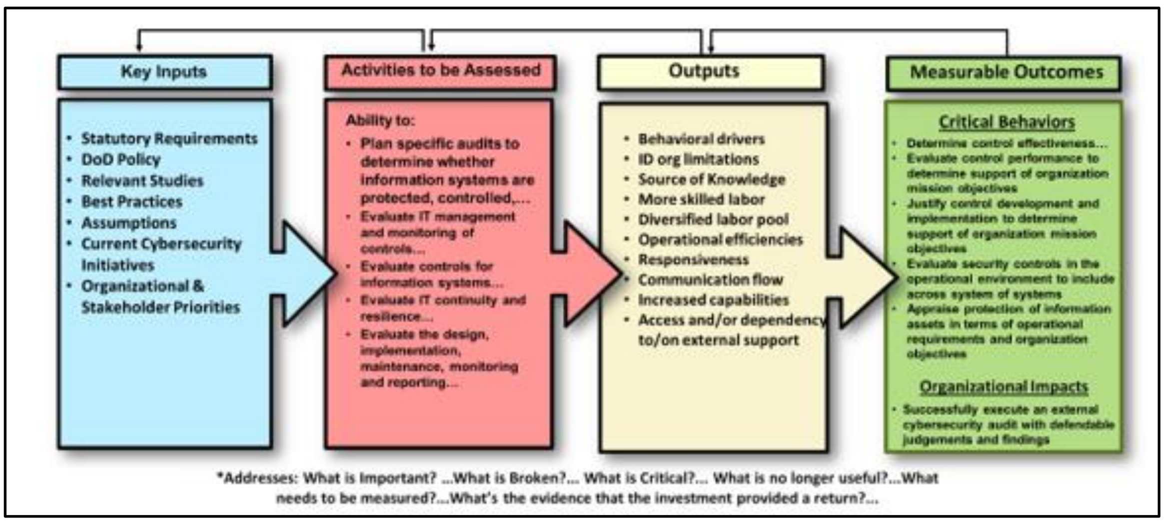


Figure 2. High Level Logic Model That Guided This Research

Kirkpatrick’s Learning Levels was the second tool selected because “Logic Models don’t show why activities are expected to produce outcomes” (Clark, 2004). The Kirkpatrick would show why and help verify if the learning stuck long enough to change the way the learners operated back on-the-job to be highly effective. The authors were especially interested in determining if what a cohort group learned in an objective-driven workshop resulted in any behavioral changes back in the workplace. Among the various learning tools available, Kirkpatrick’s four levels of learning seemed well suited to help characterize the learners’ journey to demonstrate the achievement of their indispensable “critical behaviors.” In its simplest form, Figure 3 depicts the Kirkpatrick’s learning levels.

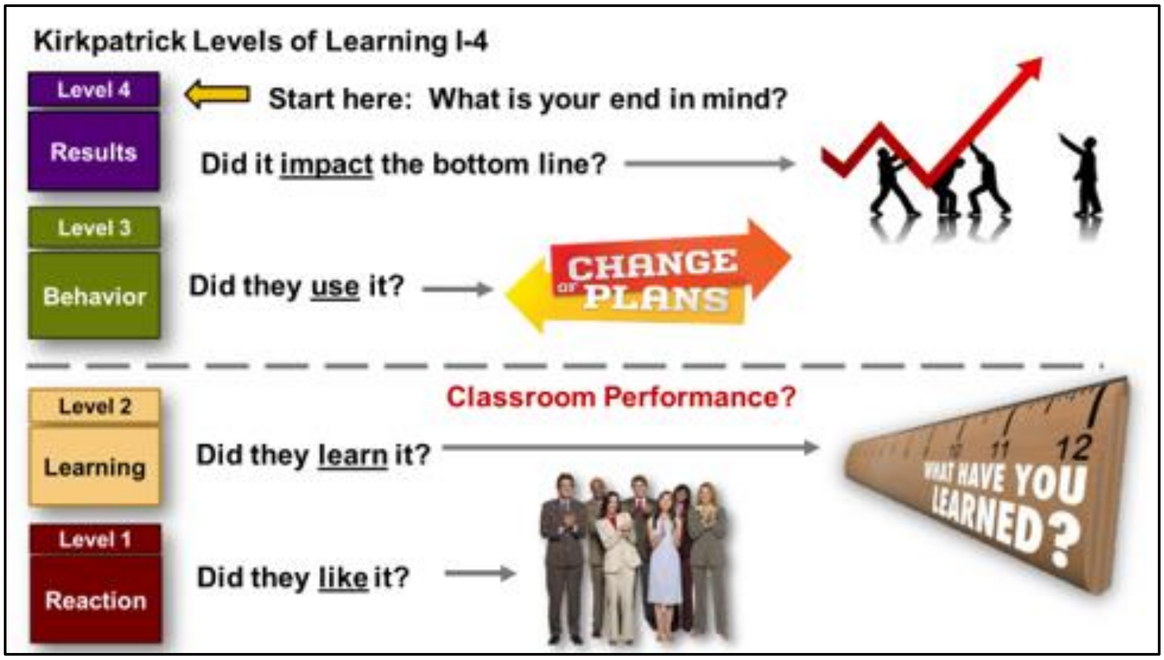


Figure 3. Kirkpatrick Learning Levels
(Kirkpatrick, 2016)

Incorporating both the Logic and Kirkpatrick Learning tools into a Performance Learning Value chain tool would provide a fully embodied visual representation (see Figure 4). It would also help show the learning dependencies leading to the learning evidence. Without the evidence, it would be hard to prove any link(s) to the initial and/or ongoing learning investment made by any organization.

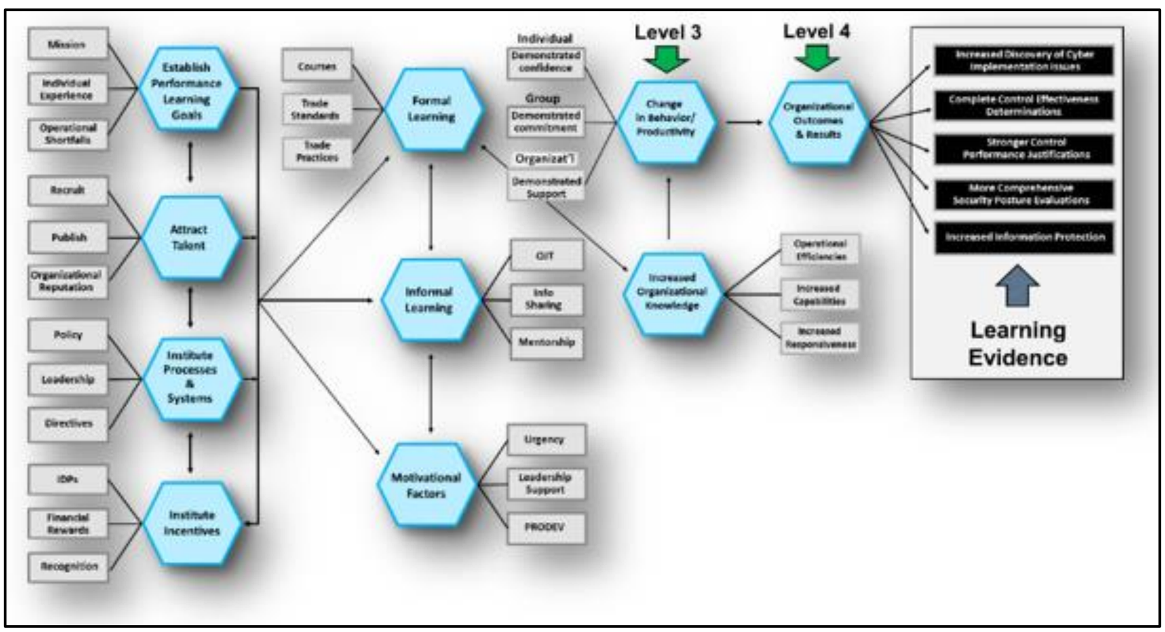


Figure 4. Performance Learning Value Chain
(Adapted from ASTD, 2004, as cited in Elkeles, Phillips, & Phillips, 2017, p. 10)

To help build greater cybersecurity knowledge and raise awareness for acquisition professionals, DAU conducted a variety of workshops—over 35 in the last three years across various DoD components and program offices. During these workshops, the following common themes surfaced from the engaging discussions:

- Current DoD cybersecurity guidance is at the strategic outcome level and generally forces program offices to take more time translating these outcomes into operation and tactical outputs.
- Enterprise cyber intelligence and warning signs can be difficult to translate into cybersecurity risk for probability and impact to their system because cyber threats are so fluid.
- Cybersecurity threats force program offices to spend more time on something that is so dynamic and sometimes difficult to translate their needs based on how they might impact their systems today.
- Risk mitigation strategies aren't tightly connected to mission assurance imperatives in the face of a hostile environment imperatives.
- Program offices may too quickly acquiesce (and accept higher risk levels) to cybersecurity design decisions because of their inability to change in their acquisition life cycle stage or to accept the perils of their inherent legacy design.
- Program offices still have to convince their resource sponsors and MDA of the needs and consequences to address potential cyber vulnerabilities.

The three models discussed previously were not used in the cybersecurity workshops. However, they set the stage for a more comprehensive case study assessment.

Case Study

For most organizations that seek to connect their learning gains in class (level II) with objective applications in their workplace afterwards (Level III), the bridge between level II and Level III can be a difficult challenge. Without it, what evidence can organizations use to confirm that the resources they allocated to Level II learning gains actual paid off? The Assistant Auditor General for Financial Management & Comptroller Audits who reports to the Naval Audit Agency graciously volunteered to participate in this case study. They wanted to ensure their auditors could apply what they learned in what they considered to be a vital functional domain—cybersecurity. Figure 5 represents the current instantiation of the Naval Audit Service Directorates. Earlier, a couple of their personnel attended DAU's cybersecurity awareness workshops. They left with a very strong feeling that their cohort group needed the same experience. Later, and after several subsequent interactions between DAU and the Naval Audit directorate's team leads as well as their leadership, the directorate welcomed a way to confirm the critical cybersecurity behaviors expected of them in the prosecution of their all audit responsibilities were met.

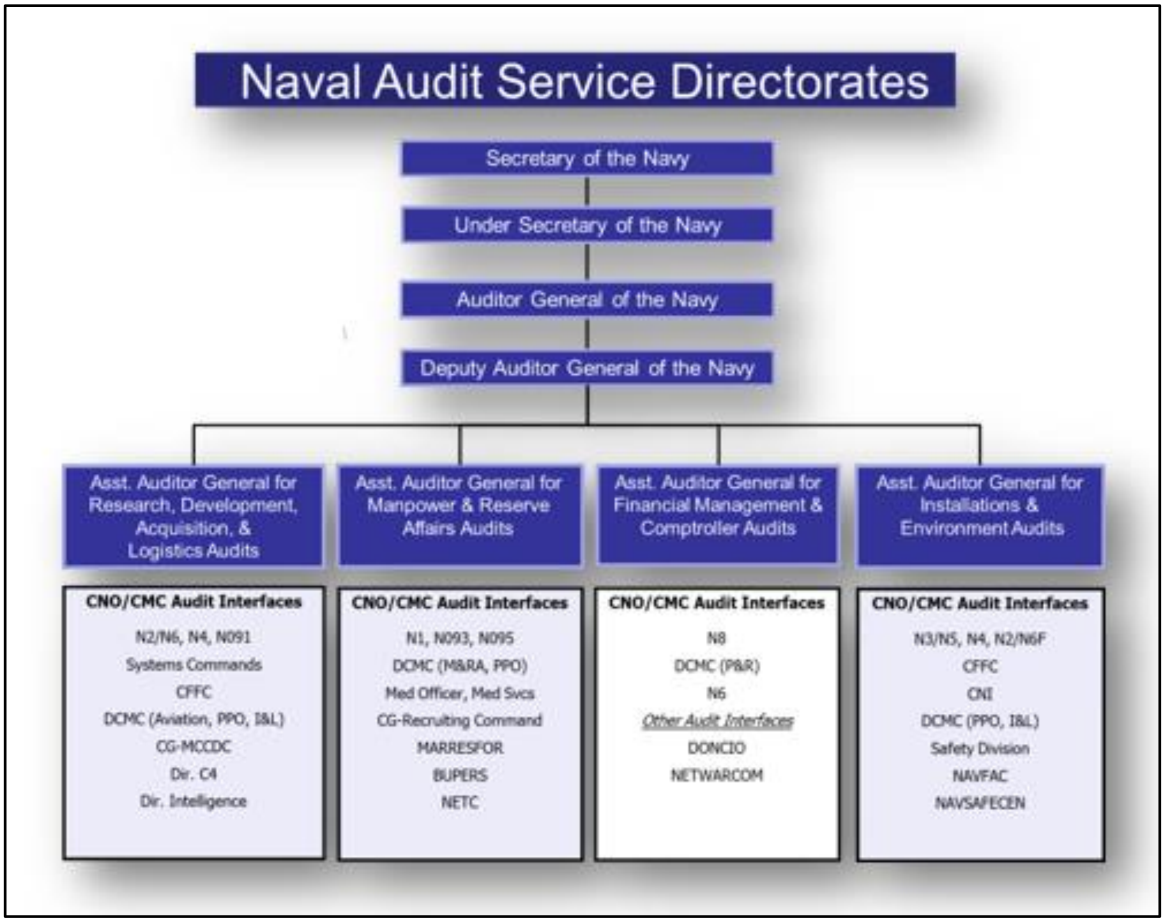


Figure 5. Naval Audit Services Directorate Structure

The cybersecurity workshop customized for the Naval Audit Services Directorate addressed the following learning objectives (i.e., Learning Level II):

- Determine the effectiveness of security controls in support of risk management.
- Evaluate the performance of security controls in support of organizational mission assurance objectives.
- Justify security control development and implementation in support of organization mission assurance objectives.
- Evaluate security controls at system interfaces and that span system of systems.
- Appraise protection of information assets in context of a threat level for protected information assets.

The learning objectives cut across the five domains that constituted the team's responsibilities (see Figure 6).

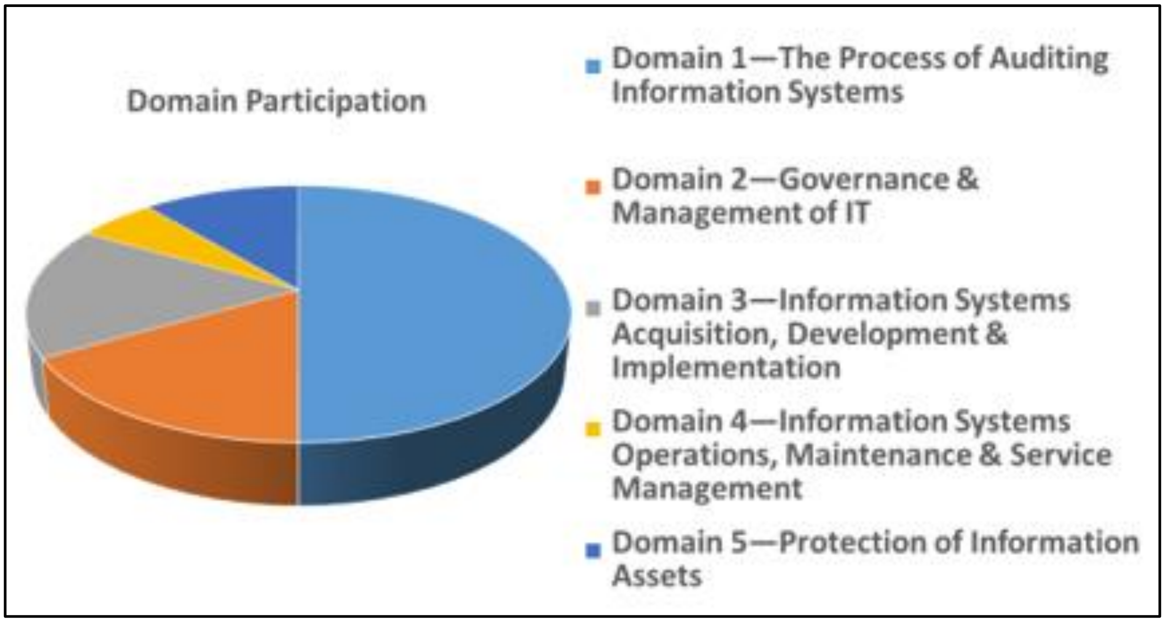


Figure 6. Domain Participation

The directorate's intact teams who attended the workshop also previously committed to connecting Level II learning objectives with the Level III critical behaviors. Just as importantly, their leadership committed to what Kirkpatrick calls its required drivers (i.e., monitor, encourage, reinforce, and reward) to assure their Level III achievements (Kirkpatrick, 2016, p. 56). Without them, a key feedback mechanism would be missing, and accountability opportunities would be lost.

Results and Findings

Figure 7 summarizes what 19 respondents had to say about their Level II learning levels "before and after" after the workshop. There were noticeable shifts and distinctions from this highly interactive and hands on event in each learning category without exceptions. Domain 2 had the most significant shift where the respondents no longer needed assistance after the workshop. Domains 1 and 2 virtually eliminated their lack of understanding for any domain afterwards.

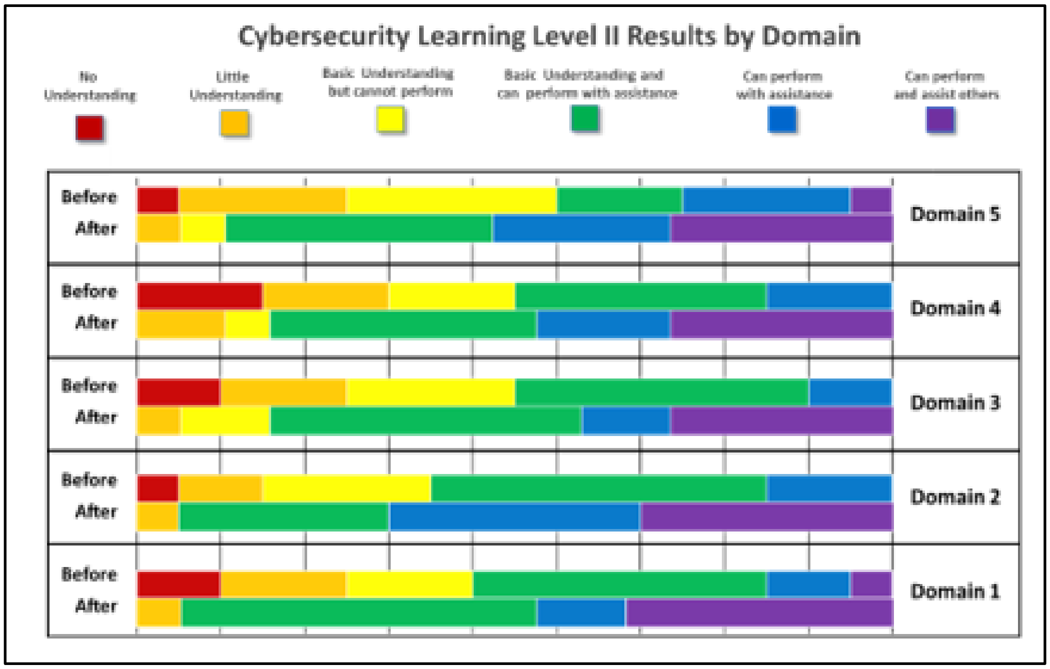


Figure 7. Cybersecurity Workshop Results

The respondents also provided a number of illuminating comments when asked, "What initial successes will likely occur as you consistently apply what you learned?" Here's what they had to say:

- We plan to incorporate all the concepts we learned in future cybersecurity audits.
- Being able to plan and execute an audit using cybersecurity and cyber resilience concepts and policies/guidance on a system or process.
- To be able to initiate an audit in cybersecurity with the training, tools, and material provided in confidence.
- I will more often consider risks concerned with access to any naval systems that are applicable to assigned future audits.
- Also, I plan to work with the audit team to develop potential audit topics that involve cybersecurity within the DoN.
- Be able to identify potential cybersecurity internal control weakness regarding people and processes.
- Cybersecurity attack vulnerability minimized.
- I will pursue more knowledge in this area to get a better understanding.
- Agencies will be better prepared to tackle cyber obstacles they may have not known existed prior to the audit.
- I think the senior Navy leadership will start seeing our capabilities and request more cybersecurity audits.

However, the more important aspect surrounding the abilities and attitudes of the learners to apply what they learned in the workshop back on-the-job (i.e., Level III) that doesn't atrophy, and what results their learning afforded. Furthermore, what will happen and what needs to happen to strengthen the bridge between Level II and Level III? Here are the Level III critical behaviors that were jointly developed up front with the team:

- Auditor determines control effectiveness and makes audit findings on controls for organization evaluation.
- Auditor evaluates control performance to determine support of organization mission objectives and makes audit findings on controls for organization evaluation.
- Auditor justifies control development and implementation to determine support of organization mission objectives and makes audit findings on controls for organization evaluation.
- Auditor evaluates security controls in the operational environment to include across system of systems and makes audit findings on controls for organization evaluation.
- Auditor appraises protection of information assets in terms of operational requirements and organization objectives to make audit findings and recommendations.
- Auditors are able to successfully execute an external cybersecurity audit with defendable judgements and findings by reviewing information, work products, or systems outputs based on a set of accepted auditing criteria.

The achievement of these Level III critical behaviors represents the litmus test. Through a suitable dose of feedback (i.e., monitor, encourage, reinforce, and reward), Level III critical behaviors and Level IV results can be achieved, later.

Extendability

The generalizability and extendibility of the claims from this research should be able to be prove validity through independent repeatability (Creswell, 2015). The NIST 800-181 National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework defines tasks, knowledge, skills, and abilities of numerous groups throughout the Cybersecurity Workforce. Using these workforce tasks, knowledge, skills, abilities, attitudes, learning objectives, and critical behaviors should be extendable to other workforce groups. Measurement issues of study constructs can easily exploit instruments like the Kirkpatrick model. However, caution should be taken to avoid use of just a single measure (Lund Research Limited, 2012). There is more than one interaction and measurement in any research project. Addressing these considerations would reduce the burden of proof for validity and broader extendibility.

Conclusion

Despite the DoD's good intentions in their policy declarations, focusing on the cybersecurity behavioral changes in the acquisition community is an equally important consideration that doesn't appear to be highly visible. The number of cyber threat actors who have the ability to exploit DoD's systems is growing at a staggering rate, while too many people involved in the acquisition community may not have fully embraced (or even understand) their role in cybersecurity. It's vitally important to elevate the acquisition community's knowledge of all cybersecurity risks in order to more carefully plan, decide, and act for the inescapable and impending cybersecurity threats. Admittedly, the danger signs are very telling, and they're not all good. No one would argue that cybersecurity is taking center stage as our dependency on the internet continues to increase.

Following a particular organization responsible for auditing the implementation of cybersecurity imperatives has been quite informative and has highlighted instrumental triggers and influencers that are so central to the achievement of desired learning outcomes.

The personnel involved in the case study is still underway. In Part 1 of a two-part research project, the authors helped reinforce which critical behaviors the participants had to embody to assess mission assurance. Time will tell if it resulted in any expected gains. In the interim, the Assistant Auditor General for Financial Management & Comptroller Directorate intends to monitor, encourage, reinforce, and reward the behaviors required by their daily duties to guide them—and convert shock into action. Part II will address their successes as well as any particular challenges they faced through ample objective evidence.

From a macro viewpoint, what steps should the DoD take now to translate their high-level outcomes into achievable acquisition behavioral changes?

- Ensure that programs don't stop cybersecurity development and testing at the interfaces, and instead compel programs to instinctively develop in a real world environment.
- Publish the critical cybersecurity competencies and proficiency levels required by all defense acquisition professionals.
- Recognize that any new policy requires a companion discussion on learning behavior implications and compel the services to report annually on their actions to address them.

What steps should YOU take to better prepare for your cybersecurity acquisition responsibilities? In many cases, it comes down to personal attitudes. Here are several that require more immediacy:

- Don't outsource your cybersecurity thinking to someone else. Take time to learn the risks and issues. Be prepared for all eventualities.
- Think critically about cybersecurity. Open your apertures, think beyond compliance, and build a more robust cybersecurity posture.
- Daily exercise the cybersecurity critical behaviors incumbent in your duties and hold your colleagues accountable to the same standards.
- Always assume compromise and set the lowest threshold for trust in all system interfaces. Never trust another system, especially if unexpected behavior occurs.

References

Barrett, M., Maron, J., Pilitteri, V., Boyens, J., Witte, G., & Feldman, L. (2017). *The cybersecurity framework: Implementation guidance for federal agencies* (Draft NISTIR 8170). Gaithersburg, MD: National Institute of Standards and Technology. Retrieved from https://csrc.nist.gov/csrc/media/publications/nistir/8170/draft/documents/nistir8170-draft.pdf

Behler, R. (2018). *Director, Operational Test and Evaluation FY 2017 annual report*. Retrieved from http://www.dote.osd.mil/pub/reports/FY2017/

Clarke, H. (2005). *Theories of change and logic models: Telling them apart*. Retrieved from http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/TOCs_and_Logic_Models_forAEA.pdf

Coates, D. (2018). *Worldwide threat assessment of the U.S. intelligence community*. Retrieved from Director, National Intelligence, website: https://www.dni.gov/files/documents/Newsroom/Testimonies/2018-ATA---Unclassified-SSCI.pdf

Creswell, J. (2014*). Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

Creswell, J. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (5th ed.). Upper Saddle River, NJ: Pearson.

Cyber Security and Information Systems Information Analysis Center. (2018). The DoD cybersecurity policy chart. Retrieved from http://iac.dtic.mil/csiac/ia_policychart.html

Defense Science Board. (2016). *Cyber defense management*. Retrieved from https://www.acq.osd.mil/dsb/reports/2010s/Cyber_Defense_Management.pdf

Director, Operational Test & Evaluation. (2017). *Cybersecurity OT&E—guidance* (Ver. 3.0). Retrieved from http://www.dote.osd.mil/docs/TempGuide3/Cybersecurity_OT&E_Guidance_3.0.pdf

DoD. (2015). *The DoD cyber strategy*. Washington, DC: Author. Retrieved from https://www.defense.gov/Portals/1/features/2015/0415_cyber-strategy/Final_2015_DoD_CYBER_STRATEGY_for_web.pdf

DoD, Chief Information Office. (2014). *Cybersecurity*. (DoDI 8500.01). Washington, DC: Author. Retrieved from http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/850001_2014.pdf

Elkeles, T., Phillips, J. J., & Phillips, P. P. (2017). *Chief talent officer: The evolving role of the chief learning officer* (2nd ed.). Routledge.

Gilmore, J. (2017). *Director, Operational Test and Evaluation FY 2016 annual report*. Washington, DC: DoD. Retrieved from http://www.dote.osd.mil/pub/reports/FY2016/

Hall, J. (2017). *Developmental Test and Evaluation FY 2016 annual report*. Washington, DC: DoD. Retrieved from https://www.acq.osd.mil/dte-trmc/docs/FY2016_DTE_AnnualReport.pdf

Helen, K. (2014). Power-up your research with diagrams and models. Retrieved from http://www.qsrinternational.com/nvivo/nvivo-community/blog/power-up-your-research-with-diagrams-and-models

Kendall, F. (2016). Next Generation Operational Control System (OCX) Nunn-McCurdy Certification Basis of Determination and Supporting Documentation, Under Secretary Secretary of Defense (AT&L), Letter to the Senate Arms Service Committee. Retrieved from https://myclass.dau.mil/bbcswebdav/institution/Courses/Deployed/ACQ/ACQ404/Archives/Student%20Materials/Student_Materials/5%20SAMC%20Class%20Prep%20Readings%20FY17-1%20Nov%2014-18/z%20-%20Additional%20Material/USD%28AT%26L%29%20Oct%2016%20Ltr%20to%20Congress%20Re%20OCX%20Certification.pdf

Kirkpatrick, J., & Kirkpatrick, W. (2016). *Four levels of training and evaluation*. Alexandria, VA: ATD Press.

Lund Research Limited. (2012). *Threats to external validity*. Retrieved from http://dissertation.laerd.com/external-validity-p3.php

Newhouse, W., Keith, S. Schribner, B., & Witte, G. (2017). *National initiative for cybersecurity education (NICE) cybersecurity workforce framework* (NIST Special Publication 800-181). Washington, DC: National Institute for Standards and Technology. Retrieved from https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-181.pdf

O'Brien, M. (2018, April). Data breach affects Saks, Lord and Taylor. *San Diego Union-Tribune*, p. A2. Retrieved from https://www.nbcsandiego.com/news/business/Data-Breach-Hits-Saks-Fifth-Avenue-Lord--Taylor-Stores--478488543.html

Office of Management and Budget (OMB). (2016). *Managing information as a strategic resource* (OMB A-130). Washington, DC: Author. Retrieved from https://www.federalregister.gov/documents/2016/07/28/2016-17872/revision-of-omb-circular-no-a-130-managing-information-as-a-strategic-resource

Office of Management and Budget (OMB). (2017). *Federal Information Security Modernization Act of 2014: Annual report to Congress.* Washington, DC: Author. Retrieved from https://www.hhs.gov/sites/default/files/fy_2016_fisma_report%20to_congress_official_release_march_10_2017.pdf

Snyder, D., Power, J., Bodine-Baron, E., Fox, B., Kendrick, L., & Powell, M. (2015). *Improving the cybersecurity of the U.S. Air Force military systems throughout their life cycles*. Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/research_reports/RR1007.html

Spaulding, D., & Falco J. (2012). *Action research for school leaders*. Pearson.

Under Secretary of Defense for Acquisition, Technology, & Logistics (USD[AT&L]). (2015). *Operation of the defense acquisition system* (Incorporating change 3, August 10, 2017). Washington, DC: DoD. Retrieved from http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002_dodi_2015.pdf

Panel 24. Implementing and Costing Rapid Prototyping

Thursday, May 10, 2018	
3:30 p.m. – 5:00 p.m.	**Chair: John Burrow**, Lecturer, University of Mary Washington ***Maturing Cost Estimation in a Rapid Acquisition Environment*** Jennifer E. Manring, The MITRE Corporation Thomas Restivo, The MITRE Corporation Natalie Faucher, The MITRE Corporation Richard Tepel, The MITRE Corporation ***Set-Based Evaluation Tool (SET): A Software Analysis Tool to Support Set-Based Decision Methods*** Stephen Hunt, SPAWAR Richard Byers, LCDR, USN (Ret.), SPAWAR Rhonda Hoeckley, New Venture Research Corp. ***Application of Set-Based Decision Methods to Accelerate Acquisition through Tactics and Technology Exploration and Experimentation (TnTE2)*** Carly Jackson, SPAWAR Systems Center Pacific Aileen Sansone, Naval Undersea Warfare Center Division Newport CAPT Christopher Mercer, USN, Assistant Secretary of the Navy for Research, Development, and Acquisition Douglas King, Col, USMC (Ret.), Marine Corps Warfighting

John Burrow—has spent more than 30 years in service to the federal government. In his prior position as a Deputy Assistant Secretary of the Navy, he managed matters associated with the Department of the Navy's Research, Development, Test, and Evaluation investments.

As the department's leader for research and development, he oversaw science, technology, and prototyping efforts across more than 15 naval warfare centers, including more than 25,000 scientists and engineers. He is the principal architect of the department's initiatives to accelerate fielding of advanced technologies and warfighting capabilities.

Author and co-author of numerous publications, Burrow has taught at the University of Mary Washington, the Massachusetts Institute of Technology Naval Postgraduate School Professional Summer, and the University of Mississippi.

Among his numerous recognitions, Burrow received the Senior Executive Service Presidential Rank Award in 2015, the Navy's Distinguished Civilian Service Award, and the department's Commendation for Superior Civilian Service in 2014.

Maturing Cost Estimation in a Rapid Acquisition Environment

Jennifer E. Manring—is a Principal Economics and Business Analyst at The MITRE Corporation. She has supported numerous DoD Sponsors, providing life cycle and independent cost estimates, cost benefit analysis, and data collection and analysis on a wide variety of DoD systems and platforms. Manring is trained and experienced on a number of commercial parametric software cost models and risk analysis tools. She has led several cost research initiatives in cloud computing, service-oriented architecture, and agile development and various independent schedule analysis assessments. Manring holds a BS in mathematics from the Virginia Polytechnic Institute and State University. [jmanring@mitre.org]

Thomas Restivo—The MITRE Corporation

Natalie Faucher—The MITRE Corporation

Richard Tepel—The MITRE Corporation

Introduction

The Department of Defense (DoD) acquisition community is increasingly pursuing means to introduce new capabilities to the warfighter as quickly as possible. When facing emerging threats, the warfighter cannot wait for a new, critical capability to work its way through the rigid and time-consuming traditional acquisition process. In an era of tightening federal budgets and increased demand for new technology to help meet mission requirements, agencies are searching for ways to deliver critical mission functionality faster and with less risk.

The traditional acquisition planning process, with its numerous maturity milestones and decision gates, was designed to reduce risk and field a mature, sustainable capability, and is not suitable for obtaining smaller, innovative technologies that may have shortened technology life cycles, or for helping users counter emerging threats. The DoD is challenged to quickly address urgent operational needs (UONs) that could endanger military personnel or lead to mission failure (Wizner, 2013). During the Global War on Terror (GWOT), the Secretary of Defense (SECDEF) promoted and implemented a decision process to shorten existing budget planning and procurement cycles, to create rapid acquisition methods to equip the warfighter, and to fulfill UONs. However, this process was implemented in an ad hoc manner across the Services. Critical review of these rapid methods provided insights regarding shortfalls of a shortened process, and revealed a need to mature and formalize a rapid process.

Circumventing the traditional acquisition process to field capability more quickly affects requirements, cost, and affordability planning. Therefore, the DoD is challenged to generate confident and credible cost estimates where programs may have less definition and/or greater uncertainty in a rapid acquisition environment. The military sector strives "to be an innovation leader in developing technology to protect troops on and off the battlefield" (DoD, 2017). When utilizing rapid methods, the DoD must ensure that it understands the total costs of a capability to make informed decisions about the capability and systems being acquired. To make effective decisions, it is essential to establish a repeatable process and assess initial costs, as well as the potential enduring impact on costs, as solutions move from rapid processes into traditional Programs of Record (PoRs).

Cost estimating plays a critical role in a rapid decision-making process by providing decision makers a deeper understanding of cost implications in rapid acquisition environments. Embedded in the traditional capability planning process are well-documented and recognized best practices for developing credible cost estimates to support DoD decision planning. However, new rapid acquisition approaches, with their short timelines, challenge the cost-estimation community. Therefore, the cost community needs to understand and mature estimation techniques to adapt and operate effectively when using rapid acquisition approaches. The DoD needs to be able to deliver accurate and credible cost estimates on rapid acquisitions to make better informed decisions.

This research provides the acquisition, cost, programmatic, and system engineering communities a deeper understanding of the impacts on cost estimating processes and cost approaches. This will help the DoD understand key areas where cost estimation should be adapted, and areas where there is increased uncertainty. This research focuses on all aspects of rapid acquisition to help program offices develop credible and confident cost estimates needed to make informed, data-driven decisions. This research will generate key insights that programs need to fully understand both near-term and long-term cost challenges of a rapid acquisition process compared with capabilities developed and acquired using traditional acquisition procedures. Results of this research on maturing cost methods in a rapid acquisition environment will improve the ability of program offices to estimate the cost of implementing rapid capability in a consistent and repeatable way.

Understanding the impact that a rapid acquisition process has on generating credible cost estimates helps to prove this research proposition and deliver outcomes that are impactful for programs and the defense acquisition community. This research will impact programs by improving the ability to make informed, data-driven decisions in a rapid acquisition environment. It will also benefit the DoD, Joint, and Services' portfolios at an enterprise level, where the research will help cost communities and program leadership assess and evaluate cost implications of capabilities acquired when using rapid acquisition methods.

Terminology

This report uses the following terminology specific to the rapid acquisition landscape:

- **Deliberate planning** is the traditional acquisition approach, which is based on three principal decision-making processes. Specifically, the DoD uses
    - As the requirement process: Joint Capabilities Integration and Development System (JCIDS)—the formal DoD procedure that defines acquisition requirements and evaluation criteria for future defense programs
    - To deliver a mature capability to an end user: Defense Acquisition System (DAS)—an event-based acquisition management process governed by milestone reviews and other decision points
    - To allocate resources to satisfy requirements: Planning, Programming, Budgeting & Execution Process (PPBE)—a cyclical process to determine DoD funding requirements and affordability
- **Rapid acquisition** is a non-traditional acquisition approach used to acquire and field urgent capability in response to adversarial threats, or to leverage new market technologies quickly.
- **Urgent need(s)** refer to capability required to remedy shortfalls that could endanger military personnel or lead to mission failure, as well as to the DoD's need to leverage emerging and relevant technology to bring innovations to the field, and protect troops. Urgent needs may be Service-specific (referred

to as UONs), or joint across multiple Services—Joint Urgent Operational Needs (JUONs) or Joint Emergent Operational Needs (JEONs).

- **PoR** is a term used in this research to describe a program that has approved funding across the defense program, achieved through the Program Objective Memorandum process, and resulting in an official line in the budget.
- **Project** is defined as a stand-alone effort outside a PoR (which might transition into a PoR); or, alternatively, a project can be a sub-set capability within a PoR.
- **Cost estimate** is defined as the summation of individual cost elements, using established methods and valid data to estimate future costs of a program, based on what is known. The cost estimate is continually updated with actual data, revised to reflect changes, and analyzed by calculating differences between estimated and actual costs (GAO, 2009).

Background

The GWOT and Iraqi and Afghanistan wars required changes within the DoD in both the requirements and acquisition planning processes to quickly address emerging capability shortfalls. The DoD promoted and implemented streamlined decision processes to shorten existing budget planning and procurement cycles to under two years. The DoD created and implemented policy in an ad hoc manner across the Services to support a top-down push for streamlined acquisition, and faster fielding of urgently needed capability. All Services established JUON and UON processes to quickly field critically needed capability.

Consequently, the United States (U.S.) Government Accountability Office (GAO) reported, “total funding for the fulfillment of urgent needs is at least $76.9 billion from fiscal years 2005 through 2010” (GAO, 2010). This included numerous programs of various sizes. Some urgent requirements grew into large, complex programs such as Mine Resistant Ambush Protected (MRAP), Joint Improvised Explosive Device Defeat Organization, and Unmanned Aerial Vehicles. Other smaller programs targeted special operations, equipping the soldier and fielding emergent technologies. One common element in these programs was the practice of rapid acquisition to meet urgent needs and field capability quickly.

The GAO (2011) criticized the ad hoc and complex processes created by the DoD, finding,

> The department does not have a comprehensive approach to manage and oversee the breadth of its activities to address capability gaps identified by warfighters in-theater. Federal internal control standards require detailed policies, procedures, and practices to help program managers achieve desired results through effective stewardship of public resources.

Further criticism from the Defense Science Task Force on the DoD fulfillment of UONs reinforced the lack of mature, repeatable rapid processes. “The department, as well as the acquisition community it depends on, has struggled in their ability to field new capabilities in a disciplined, efficient, and effective way” (Defense Science Board Task Force, 2009).

Under the 2011 National Defense Authorization Act (NDAA), Congress formally required updates in DoD policy. Interim policy was created in 2012, and by 2015, policy for rapid acquisition was recognized in Department of Defense Instruction (DoDI) 5000.2 and Chairman of the Joint Chiefs of Staff Instruction (CJCSI) 3170.01H. These updates formalized requirements and acquisition processes for rapid acquisition and fulfilling urgent needs.

The DoD implemented policy based on criticism and lessons learned, but a void remained in policy and guidance for creating confident and credible cost estimates in a rapid acquisition environment. Ad hoc processes resulting from urgent conditions can lead to poorly defined assumptions, failure to consider all costs, inadequate data collection, inappropriate estimating methodologies, inadequate assessment of risk and uncertainty, and indefensible estimates used in the budget planning process.

Generating confident and credible cost estimates is difficult when rapid acquisition programs have less definition and greater uncertainty than traditional acquisition progams. The rapid acquisition process should be robust enough to support decisions based on affordability, using mature cost methods grounded in a repeatable process, thereby bringing credibility to the estimate and decision planning.

Several studies have assessed a non-traditional planning process compared to the traditional process. Understanding differences between traditional acquisition planning and non-traditional methods provides insights into important factors enabling implementation of rapid acquisition methods. The following are differences identified in GAO audits and Defense Science Task Force recommendations:

- Traditional planning consists of three core components: (1) requirements validation, (2) acquisition planning, and (3) affordability planning required in the approval of a new military system. These three processes include maturity milestones and decision gates throughout system fielding to deliver a 99% solution in roughly three to 11 years.
- Traditional planning encompasses full life-cycle acquisition of a system owned by a Service in a formal PoR. This process is scalable for military solutions and adaptable to each individual Service. Deliberate planning is well documented, formal, and has repeatable processes to define requirements, acquisition, and affordability.
- Non-traditional planning responds to and encourages innovation for quick fielding of a capability or system. This approach does not focus on fielding a 100% solution, which is a critical factor in shortening procurement time. The decentralized process reduces the number of requirements and acquisition approvals through a shortened decision chain within the Services and Office of the Secretary of Defense (OSD).
- Non-traditional planning evolved in an ad hoc manner without a well-documented, repeatable process to address affordability and cost estimation. Historically, urgent needs waived affordability requirements during wartime operations.

Recent DoD policy updates define two key parameters for rapid acquisitions. The first parameter requires that the system be fielded in 24 months. The second requires that a disposition analysis be conducted within the first year of sustainment. The system disposition analysis determines (1) termination of the program, (2) sustainment for current contingency, or (3) transition to a PoR. This policy update creates a critical decision point for re-assessing requirements, acquisition, and affordability of the system. The recommendation then follows the steps for validation and approval as defined in JCIDS. This process identifies decision gates that form the foundation of a repeatable and defensible decision process. Understanding the path a system takes, and decision gates within rapid acquisitions, allows the DoD to better understand and anticipate the total costs of a rapidly acquired system.

Consolidating the acquisition process to make it deliver capability more quickly affects requirements, cost, and affordability planning in both the near and long terms. In a rapid acquisition environment, acquiring a partial system in the near term might change the affordability and sustainment of a complete system in the long term. Recent policy updates do not provide detailed guidance on how to identify and assess the cost impacts across an accelerated program.

Rapid Acquisition Landscape

The research team conducted a comprehensive research of open sources to characterize and understand the rapid acquisition culture and community. Literature reviews and recent policy updates demonstrate the DoD's commitment to a rapid acquisition process. Recent policy updates and implemented best practices address gaps and shortfalls from the audits. However, a gap remains regarding how rapid acquisition processes address affordability. The DoD has not provided specific guidance on how to develop cost estimates to support rapid acquisitions. Critical to understanding a program's affordability is a credible cost estimate grounded in repeatable processes. Open source literature identified factors that affect both near-term and long-term program costs. This research assessed and characterized these factors to provide a better understanding of their impact on cost-estimating methods in a rapid acquisition environment.

- Rapid acquisition process accepts a less than 100% solution for use in a limited definition and/or use case. This creates greater uncertainty around near-term cost estimates as well as long-term costs of the full program, because the program technical baseline is bounded by the use case and not by cost of a fleetwide implementation.
- The DoD created innovation forums to provide a better understanding of technology maturity for science and technology (S&T) and bring innovation to DoD programs. However, the new demonstration forums are not specific to a Service, program, or capability gap. This DoD process relies on limited seed funding by a Service or by the PoR that selects the technology and incorporates it into its program. If a PoR chooses to incorporate a showcased technology, its decision may affect both near-term and long-term costs associated with customizing a commercial solution to meet a specific military need, and overcoming additional acquisition barriers, such as funding governed by appropriation laws and industry sourcing.
- Rapidly acquired solutions may be viable in a rapid fielding scenario with limited production quantities. However, this drives uncertainty into a program by producing impacts on both quantities (expansion from initial plan) and lead time in cases of a limited supplier/parts base. These impacts may affect both near-term and long-term program cost estimates, as well as the new policy fielding constraint of less than 24 months. The DoD has no process in place to address gradations of programs deemed viable under a rapid acquisition process but not sustainable under a traditional process.
- Abbreviated testing is typically associated with a rapid acquisition. A traditional process requires a complete testing program to demonstrate full functionality of a system in the operating environment. The testing phase identifies risks that can be addressed prior to full production and fielding of a system. To meet urgent fielding requirements, some programs may consider

testing in parallel with fielding a system. Then, work-around solutions must be found to address issues found during the testing phase that could not be handled before fielding. Work-around solutions generate costs and schedule impacts to a program using rapid acquisition methods. Therefore, this should be taken into consideration, as it impacts total program cost and affordability of the system.

- Rapidly fielded solutions may include combinations of commercial off-the-shelf (COTS) and modified COTS solutions, thus limiting requirements for new development and potential platform-integration efforts. In many cases, COTS products require additional engineering to integrate with another system. Formalizing methods used to estimate rapid acquisitions will ensure that these additional integration costs will be accounted for in the cost estimating process.
- Limited deployment of a capability might require only a limited training and sustainment approach—one that is aimed at meeting the near-term fielding requirement. For example, MRAP is considered a successful program that has fielded thousands of vehicles to meet urgent needs, but early deployment suffered from the lack of an advance-planning timeline typically found in logistics and sustainment areas of a traditional program. MRAP had no process to establish common parts or logistic chains to support the vehicles' numerous variants. While the near-term implementation of the vehicle met critical warfighter needs, the program did not transition into a PoR because of long-term affordability and sustainment issues.
- Urgent schedule constraints have an impact on acquisition strategy for a near-term solution. Acquiring a capability rapidly requires a streamlined acquisition process and market sources for the product. Urgent schedule constraints might limit acquisition strategies and source to a single vendor to meet time constraints of the rapid acquisition. This may affect near-term and long-term costs, because the near-term solution may develop a reliance on a commercial proprietary system that may not be sustainable under a PoR. A traditional acquisition approach includes a mature process for assessing competition and sources in the marketplace.

Role of Cost Estimation

The open source review of literature did not provide specific guidance on DoD policy or processes for conducting cost estimates in a rapid acquisition environment. No open source documents described if or how any review of cost estimates should be conducted for UONs and rapid acquisition. Furthermore, the 2016 policy updates do not contain specific guidance on how rapid acquisitions across the Services and joint programs should address affordability. Yet, the GAO recognizes that cost estimating is a critical part of project formulation and execution.

Best practice processes for developing cost estimates are defined in the 2009 publication *GAO Cost Estimating Guide: Best Practice for Developing and Managing Capital Program Costs* to ensure proper stewardship of public funds, employ effective management practices, and provide reliable cost information to government officials and decision makers. This GAO (2009) guide provides processes, standards, and procedures for developing, implementing, and evaluating cost estimates for use across the federal government.

The GAO process represents a repeatable method that results in high-quality, comprehensive, and credible estimates. Twelve distinct steps are organized into four evolutionary phases for developing a cost estimate. The GAO observed that, when an agency did not incorporate all 12 steps, its estimates were unreliable. Furthermore, when an agency lacked an overall comprehensive process, its ability to create credible cost estimates was impaired. Therefore, the GAO recommends that each step be followed to ensure that a quality estimate is used when making decisions. The GAO describes four phases and the corresponding steps included in each phase (see Figure 1).

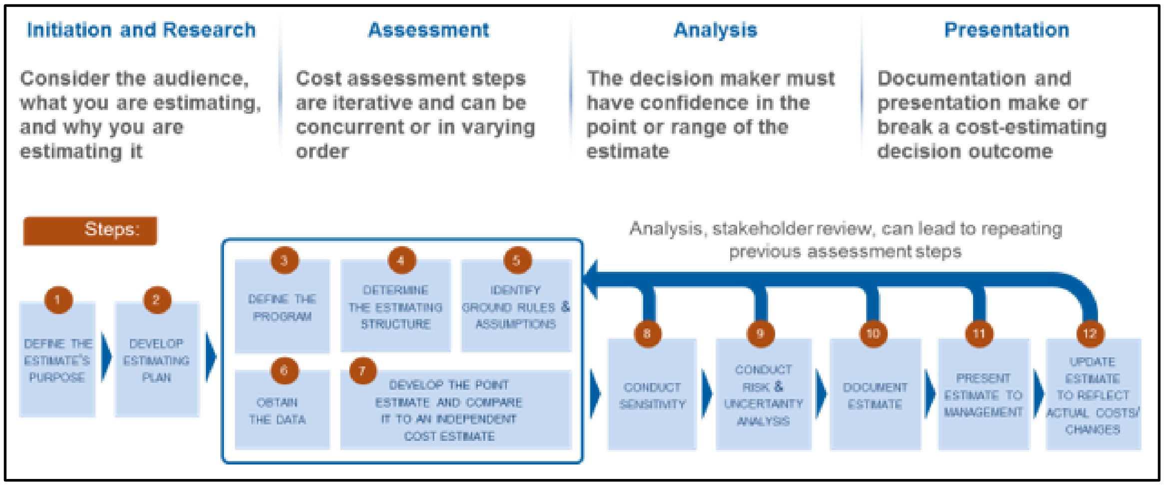


Figure 1. GAO Cost Assessment 12-Step Process
(GAO, 2009)

The GAO (2009) presents the diagram in Figure 1 to illustrate how the phases and steps are related. The GAO emphasizes that the fundamental scarcity of resources and growing budget demands make it imperative that government acquisition programs deliver as planned, "because every dollar spent on one program will mean one less available dollar to fund other efforts" (GAO, 2009).

The rapid acquisition process accepts a less than 100% solution in order to provide capability in a shortened timeline. Consolidating a traditional process to deliver capability more quickly affects requirements, cost, and affordability planning in both the near and long term, and thus requires a repeatable cost estimating process that provides reliable cost information to government officials and decision makers operating in a rapid environment. The GAO understands that "an estimate that meets all these steps may be of little use or may be overcome by events if it is not ready when needed. Timeliness is just as important as quality. In fact, the quality of a cost estimate may be hampered if the time to develop it is compressed" (GAO, 2009).

Relying on a standard process that emphasizes pinning down the technical scope of work, communicating the basis on which the estimate is built, identifying the quality of data, determining the level of risk, and thoroughly documenting the effort should result in cost estimates that are defensible, consistent, and trustworthy. In a rapid acquisition environment, it is important to have a credible cost-estimating process to ensure that limited resources are allocated effectively to meet the warfighters' need (GAO, 2009).

In the literature review and findings of this research, the research team identified several key cost processes used in rapid acquisition environments that deviate from the traditional process. These differences, reflected in the case study research, provided the foundation for recommendations to mature and improve cost estimation methods and processes in a rapid acquisition environment. From these insights and recommendations, the research team created guidance that offers the acquisition, cost, and programmatic communities a deeper understanding of the cost implications of rapid acquisition for service programs and the DoD enterprise.

Research Methodology

This research used a case study approach to characterize and classify findings on cost methods employed in rapid acquisition environments, and compared these characterizations with cost processes used in traditional processes, as defined in the *GAO Cost Estimating and Assessment Guide* (GAO, 2009), through a formal case study research design.

Fundamental to case study research is its design. It is a linear and iterative process that requires planning to conduct valid and thorough research. To model the case study research, the research team employed research and design methods described in “Case Study Research Design and Methods” by Robert K. Yin (2014). Yin’s methods provide necessary rigor by capitalizing on the strengths and compensating for limitations of case study research. Yin provides strategies and methods that the team used to construct and conduct this case study research.

Characterization of the problem, and findings from literature reviews, raised questions about the credibility of cost estimation methods in a rapid acquisition environment. An examination of the rapid acquisition community reinforced the relevance and importance of examining “how” cost estimates are conducted and “what” is different in a rapid acquisition environment compared to a traditional one. Key research questions relevant to rapid acquisition are as follows:

- What are the overarching factors and characteristics that affect cost processes and methods?
- What are the key programmatic and technical differences compared to a PoR?
- What are the impacts on cost approaches and processes?

Research questions help to form the hypothesis and boundaries of the research. The research team developed a hypothesis focused on factors that affect cost estimation of a rapid acquisition. Effects of these factors depend on solution maturity, type of program, type of system, and size of system acquired through rapid methods. The hypothesis also considers processes surrounding rapid acquisition, and characteristics like number of resources, types of resources, and variations from the traditional approach. The research hypothesis is as follows:

Cost estimates for rapid acquisition projects will improve in reliability and credibility by using a proven and repeatable approach specific to the rapid acquisition environment.

The research team collected and examined relevant evidence from case studies to better understand and support the research hypothesis. The collected data and information provided insight into how costing methods used in rapid acquisition cost estimating differ from methods used in traditional processes and PoRs.

Pre-screening of case study candidates was conducted to provide further insight into the rapid acquisition community and identify case study candidates. Interviews with 35 subject matter experts (SMEs) across 25 programs helped the research team characterize rapid acquisition in the DoD, and identified future opportunities for validating the research outcome. Additionally, the research team used data obtained during the pre-screening phase to refine their characterization of rapid acquisition environments and address these findings during the data-collection phase.

Seven case studies were conducted using an established case study protocol and framework for collecting data. Four of the case studies were conducted using a detailed questionnaire and interview. Three of the case studies were conducted using a high-level questionnaire and response. Cases varied from responses to urgent cyber threats to PoR technology insertion of hardware and software solutions through rapid acquisition methods. Program managers, engineers, and cost analysts participated in the case studies.

As part of the interview process, the research team input developed a data-collection questionnaire for analysis. The data-collection questionnaire needed to address and collect data on all steps in the GAO process. Each question was assessed and mapped either to a single step or to multiple steps. Some questions addressed all steps in the GAO process. Each question was also assessed for the type of data being collected. Almost all data collected was "qualitative" or "open-ended." Key areas on the data collection questionnaire included the following:

1. What overarching factors do you think most affect rapid acquisition costing?
2. What are the key programmatic and technical differences in a rapid acquisition (e.g., testing, training, documentation, maintenance, etc.)?
3. How is a traditional cost estimating process changed to adapt to a rapid acquisition environment?
4. What are the main rapid acquisition cost estimating challenges?
5. What rapid acquisition cost estimating process recommendations would you suggest?

To analyze qualitative data, the research team assigned an identifier to information collected during an interview. They recorded data from the interviews in a Microsoft Word file, then grouped it by question. They also assigned codes to information collected against a step in the GAO process and used additional codes to identify challenges and recommendations for a rapid acquisition environment. Once they had coded all data against the 12 steps, the research team consolidated and grouped the findings by step. Through discussion and consensus on the findings throughout the analysis process, the research team identified themes, patterns, and trends in the data. Further analysis triangulated the findings with insights from literature and pre-screening interviews.

The research team used a multilevel review process to validate results. One level involved reaching out to 25 SMEs representing a wide variety of rapid acquisition programs, approaches, and experiences. The research team asked these SMEs to review key findings and recommendations, and provide feedback on each key finding, stating whether they agreed, disagreed, or had recommended changes. Nine of the 25 SMEs provided validation feedback. For the second level of validation, the research team used detailed SME reviews. Four SMEs participated in a detailed review and discussion session, and provided comprehensive feedback on the full set of results. The research team incorporated feedback received from all SMEs into the final set of findings and recommendations.

Results

The research team analyzed qualitative data collected from open source materials, ancillary interviews, high-level case studies, and detailed case studies. Through an assessment of this information, they identified themes, patterns, and trends which they organized by “findings” and “recommendations.” They mapped and coded each finding and recommendation to one of the 12 steps in the *GAO Cost Estimating and Assessment Guide* (GAO, 2009). The GAO’s 12 steps are recognized as best practices by the cost estimating community, and served as the overarching theoretical framework for the research. The research team identified key findings and recommendations to highlight results with the greatest impact and importance.

The research team identified several rapid acquisition characteristics that impact cost methods, and provide context in which recommendations are made. These characteristics include the following:

- Rapid acquisition emphasizes delivery of a capability quickly, which causes very short acquisition timelines.
- To achieve these shortened timelines, rapid programs operate at a fast pace, and have a great concurrency of efforts.
- Schedule is the top priority; cost and capability are flexible to support desired schedule.
- There are many rapid acquisition approaches that vary in solution maturity, size, type, and timeline, as well as acquisition strategy.

Rapid acquisition, with its compressed timelines, pose unique challenges to the cost estimating process. Specific challenges identified during the findings assessment are shown in Figure 2. The research team made recommendations in this report to specifically address these challenges, while considering the constraints of rapid environments.

Figure 2. Cost Estimating Challenges in a Rapid Environment

The research team identified, organized, and aggregated key results to highlight the most important findings and recommendations among detailed results. Six major themes emerged:

1. cost estimating process
2. cost analyst
3. documentation
4. uncertainty/risk

5. trade-offs
6. scope/baseline

Overall, the research team identified 10 key findings and 15 key recommendations across these six major themes. Key findings and recommendations maintained the mapping to the GAO 12-step process, and mappings are shown in parentheses following each result. In some cases, the result is mapped to "All Steps" as the result applies across the entire 12-step process. The team selected key recommendations to specifically address the most important challenges to the cost estimating process in a rapid environment, while also considering constraints imposed by rapid environments.

Theme: Cost Estimating Process

Key Findings:

- Cost-estimating processes, approaches, and constraints vary (All Steps).
- Initial estimates are typically developed quickly to obtain funding then continue to be refined over time (All Steps).
- Cost data collection is especially challenging in shortened timeline (Step 6).

Key Recommendations:

- Initial and follow-on estimates should follow the GAO 12-step high-level process at an appropriate level, given time constraints and the demands of rapid acquisition (All Steps).
- Initial estimates should be broken out into life-cycle phases (development, production/procurement, and sustainment), and investment phases broken out to WBS level 3, where feasible (All Steps).
- Begin data collection early, and allow for as much time as possible to collect desired data; where possible, consider identifying and collecting data that can be used for future rapid acquisitions (business intelligence; Step 6).
- Identify cost drivers and conduct sensitivity analyses on them as soon as possible. Initial estimates should identify high-level cost drivers and conduct some level of sensitivity analysis on them (Step 8).
- Conduct cross-checks and cross-verification of at least major cost elements and cost drivers (Steps 7 and 8).
- Conduct high-level schedule analysis to ensure capability can be delivered as planned (Steps 7 and 8).

Theme: Cost Analyst

Key Findings:

- Trained cost analysts are often brought in after initial estimates are developed and are often not adequately resourced (All Steps).

Key Recommendations:

- Trained cost analysts should be engaged early on, and continue to be embedded within a program to handle the rapid pace of change (All Steps).
- Cost analysts need to be adequately resourced to support initial and follow-on estimates (All Steps).

Theme: Documentation

Key Findings:

- Acquisition and cost documentation are not a top priority (Step 10).

Key Recommendations:

- Cost estimate, and what is known about the programmatic approach and technical solution, should be reasonably documented within timeline constraints (Step 10).
  - Rapid timeline should allow for flexibility in documentation medium and level of detail (focus on the most important pieces of the estimate).

Theme: Uncertainty/Risk

Key Findings:

- Initial estimates have the greatest uncertainty and risk, but generally only point estimates are developed when cost analysts are not involved (Step 9).

Key Recommendations:

- Key areas of uncertainty and risk should be identified, and all estimates should be risk-adjusted (Step 9).
- Specific areas of uncertainty and risk to consider for rapid programs include: scope definition, GR&As, solution options, software development, integration, fielding, sustainment, and supply chain (Step 9).

Theme: Trade-offs

Key Findings:

- Trade-offs may be required; particularly on requirements (Steps 3 and 8).

Key Recommendations:

- Key cost, schedule, performance, and functional trade-offs that the program will need to evaluate should be incorporated upfront into the estimating plan (Steps 3 and 8).
- Requirements should be continuously prioritized early in order to justify trade-offs that may be needed to deliver capability quickly (Steps 3 and 8).

Theme: Scope/Baseline

Key Findings:

- Acquisition and cost documents are developed in parallel (Step 3).
- Acquisition and technical efforts are abbreviated or developed in parallel in certain areas to accommodate rapid timelines and field solution quickly (Step 3).
- Transition planning to PoR varies from unknown to well understood (Step 3).

Key Recommendations:

- Programs ensure a tight coupling of cost, programmatic, and technical SMEs and should hold regular GR&A discussions to ensure key program personnel agree with GR&As and to keep up with fast pace of change in rapid programs (Step 3).
- Consider cost estimate adjustments to reflect reductions in documentation, integration, testing, and training efforts in a rapid environment and plan for full efforts in these areas when programs transition to PoR (Step 3).

Overall, Step 8 had the greatest number of key recommendations, followed by Step 3 and "All Steps." Although Step 2 and Step 3 had more detailed findings (as shown in Figure 3), the most important recommendations are in Step 8, Step 3, and those that apply across "All Steps." These steps focus on the importance of defining the program's characteristics, conducting a sensitivity analysis of the cost estimate, following the overall GAO 12-step process at an appropriate level for program maturity and available time to develop the estimate, and involving cost analysts early to develop the cost estimate.

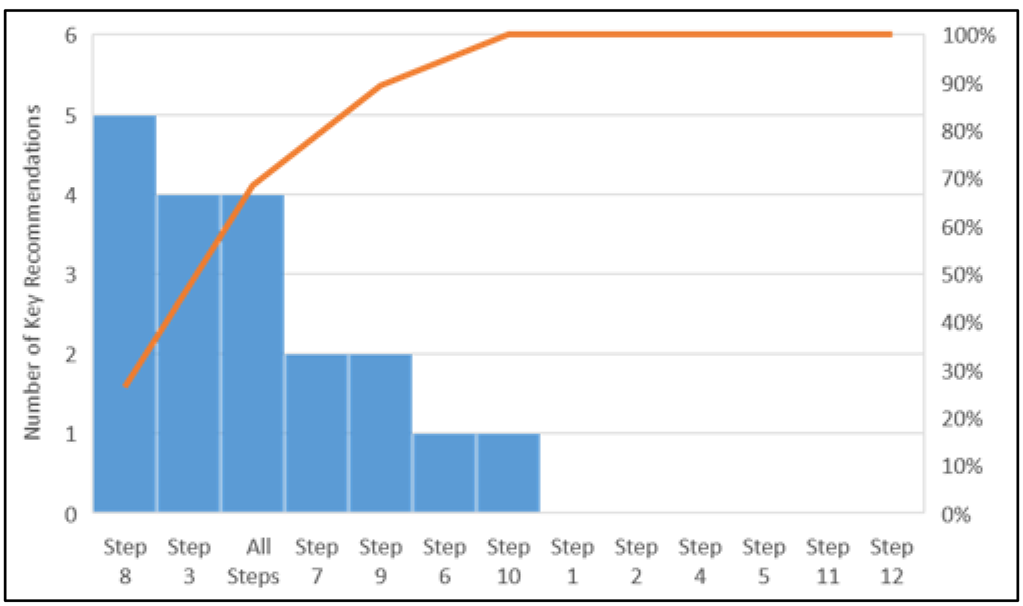


Figure 3. Numbers of Key Recommendations by GAO Step

In summary, the research team conducted case studies, organized the information they collected, and analyzed it for themes, patterns, and trends. They aligned results with the GAO's 12-step cost estimating process which represented a theoretical framework for analysis. Key results yielded 10 findings and 15 recommendations. Results were most numerous in Steps 2 and 3, although all 12 steps had some findings and recommendations. Figure 4 shows a summary of key recommendations mapped against the GAO's 12-step cost estimating process. Although Step 2 and Step 3 had the greatest number of detailed results, the most important recommendations are in Step 8, Step 3, and those that apply across "All Steps." These steps focus on the importance of defining the program's characteristics, conducting a sensitivity analysis of the cost estimate, following the overall GAO 12-step process at an appropriate level for program maturity and available time to develop the estimate, and involving cost analysts early to develop the cost estimate.

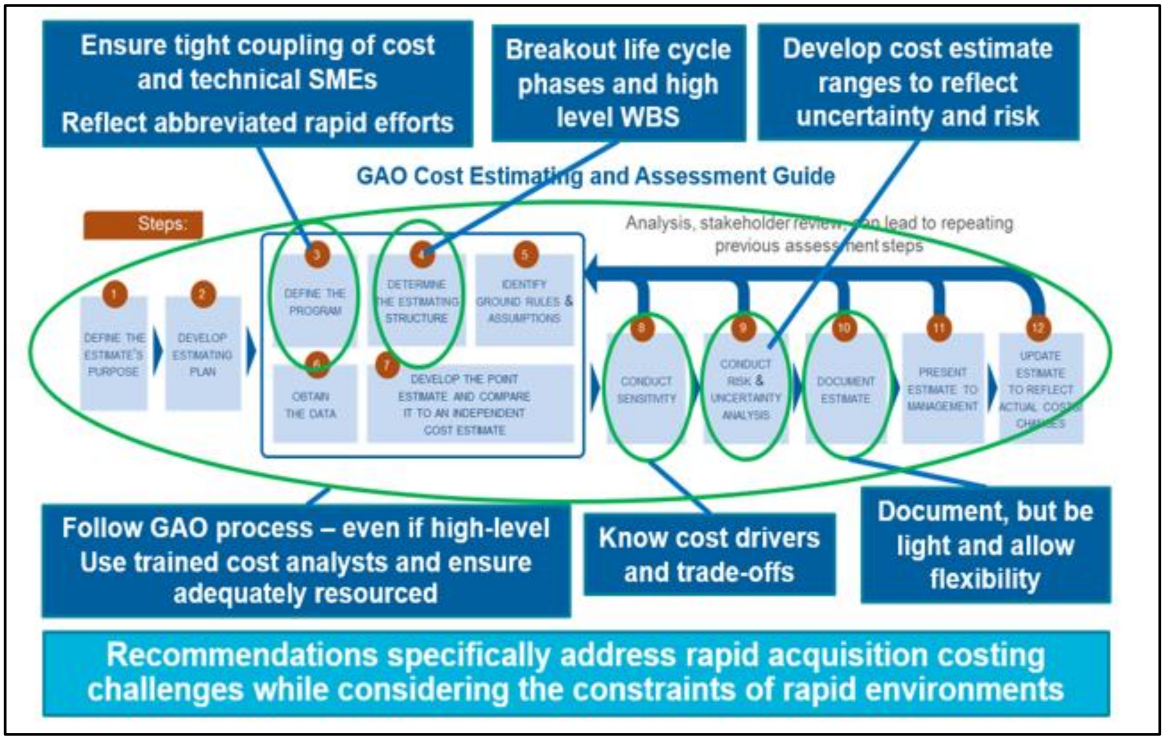


Figure 4. Summary of Key Recommendations

Recommendations can be easily applied, are flexible, and can accommodate a variety of rapid timelines and approaches. Some factors to consider when applying the recommendations include: rapid acquisition approach, size of program (in dollars), new program start versus an established program, timeline, solution maturity, and solution complexity. The research team aligned its recommendations with the GAO's best practices to establish credibility and accommodate ease of use. Recommendations allowed flexibility in implementation to ensure broad applicability across a variety of rapid acquisition methods.

Summary

Agencies are increasingly pursuing means to introduce new technology and capabilities as quickly as possible to counter emerging threats. For critical capabilities, warfighters cannot wait for new technology or capability to work its way through a rigid and time-consuming acquisition process. Agencies are using rapid acquisition methods to help achieve these goals. Rapid acquisition approaches have challenged the acquisition community's ability to generate credible cost estimates. Ad-hoc and disparate processes have been used over the past decade, leading to poorly defined assumptions, inadequate data collection, inappropriate estimating methodologies, and inadequate assessments of risk and uncertainty.

This research set out to advance the capabilities of the acquisition and cost communities by maturing cost analysis aspects of rapid acquisition processes, and developing a proven and repeatable approach to generating accurate and credible cost estimates. The research team developed guidance to aid the cost, acquisition, and system-engineering communities who face unique challenges within rapid acquisition environments. This guidance includes findings that characterize these environments and recommendations for developing credible cost estimates, given the constraints of rapid acquisition requirements.

The research team performed multiple activities during the development of guidance for cost estimating communities. They performed a comprehensive literature and landscape review, and conducted an assessment and detailed characterization of the rapid acquisition environment. Qualitative data-collection methods included interviews with 35 SMEs and seven case studies representing a broad range of programs, solution types, and rapid acquisition approaches. The research team compiled and analyzed the qualitative data for patterns and trends, and used the GAO's cost estimating best practices to develop 10 key findings and 15 key recommendations. By mapping their findings and recommendations to the GAO's 12 steps, the team ensured that their recommendations could be easily applied to a variety of rapid timelines and approaches. Finally, the research team identified key findings and recommendations to highlight the most important results.

This guidance describes common characteristics of rapid acquisition environments, and unique challenges faced by cost, acquisition, and system-engineering communities operating within rapid acquisition–compressed timelines. It recognizes that there are many rapid acquisition approaches that vary in solution maturity, size, type, timeline, and acquisition strategy and presents specific characteristics that impact cost methods in rapid environments. These include emphasis on quick delivery of capability, short acquisition timelines, fast-paced environments with a high degree of concurrency of efforts, and the need for cost and capability trade-offs to meet schedule.

This guidance provides an easily usable and adaptable set of findings and recommendations aimed at strengthening the ability of the cost and acquisition communities to produce credible cost estimates for capabilities that require rapid acquisition methods. The recommendations guide program offices that are implementing rapid acquisitions to help mature their cost estimating processes and align them with best practices; decrease the variance seen today across rapid acquisition cost estimating processes; and help the cost community establish repeatable, proven processes for operating in rapid environments. The recommendations are linked to the GAO cost estimating best practices that, when implemented, help ensure that credible, reliable, and confident cost estimates are delivered—even within the constraints of rapid acquisition timelines. Ultimately, a credible cost estimate helps improve a program office's ability to make better informed, data driven decisions (at the program and portfolio levels).

Recommendations

The research team developed several recommendations for cost analysts developing cost estimates in a rapid environment. These recommendations specifically address the challenges they face, and consider the constraints imposed by a fast-paced rapid acquisition environment. Recommendations include the following:

- **Apply the GAO 12-step process, even if it is applied at a high-level.** There are multiple levels of depth at which each step can be applied. The rapid acquisition approach, solution maturity level, and timeline will constrain the depth to which each step can be applied in a given situation. Research results indicate that every step can be applied at some level in a rapid environment. Including all steps will ensure that the full spectrum of best practices is applied.
- **Use trained cost analysts and engage them early on.** The research revealed that many of the initial estimates were not developed by trained cost analysts. Although well-intentioned and limited by time, the estimates lacked best practices that a trained cost analyst would have applied. Initial estimates are often the most important in establishing budgets and baselines, so, incorporating cost analysis best practices upfront is critical.
- **Develop cost estimate ranges to reflect uncertainty and risk in rapid acquisitions**. Rapid acquisitions often have greater uncertainty and risk than traditional acquisition programs. Much less may be known about the solution and implementation. In rapid acquisitions, it is important that uncertainty and risk be reflected in a range estimate, not a single point estimate.
- **Know cost drivers and trade-offs**. Rapid environments are fast-paced. Schedule is top priority, and cost and capability trades are often required. Knowing the critical trades that may be required upfront, and key cost drivers in the estimate, help ensure that options can be examined often, quickly, and effectively.
- **Understand and reflect abbreviated acquisition and technical efforts.** There are adjustments made to efforts in the interest of faster delivery to the field. Integration, testing, documentation, and training efforts may be reduced compared to traditional acquisition efforts. The cost estimator needs to understand which efforts may be reduced, reflect these reductions in the estimate, and allow for additional efforts that may be moved to the transitioning to a PoR later.
- **Document, but be light and flexible**. Documentation is not a top priority in rapid acquisitions, and there is minimal time in a rapid environment. Research findings uncovered very limited documentation or records of key scope and GR&As. The intent of this recommendation is not to impose an unreasonable burden in a constrained environment, but to reinforce the need for flexible media to document the most important features related to the cost estimate to establish credibility and traceability.

Recommendations can be easily applied, and are flexible enough to accommodate a variety of rapid timelines and approaches. Some factors to consider when applying these recommendations include: rapid acquisition approach, size of program (in dollars), new program start versus an established program, timeline, solution maturity, and solution complexity. These recommendations are aligned to established best practices, yet allow flexibility in implementation to accommodate ease-of-use and ensure broad applicability across a variety of rapid acquisition methods.

References

Defense Science Board Task Force. (2009). *Fulfillment of urgent operational needs*. Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics.

DoD. (2017, January 5). *Department of Defense Science and Technology*. Retrieved from https://www.defense.gov/News/Special-Reports/0715_science-tech

GAO. (2009). *GAO cost estimating and assessment guide: Best practice for developing and managing capital program costs*. Washington, DC: Author.

GAO. (2010). *Warfighter support: Improvements to DoD's urgent needs process would enhance oversight and expedite efforts to meet critical warfighter needs*. Washington, DC: Author.

GAO. (2011). *Warfighter support: DoD should have a more comprehensive approach for addressing urgent warfighter needs.* Washington, DC: Author.

Wizner, A. M. (2013). *The next acquisition challenge: Transitioning enduring capability.* Carlisle Barracks, PA: U.S. Army War College.

Yin, R. K. (2014). *Case study research designs and methods*. Washington, DC: Sage.

Acknowledgments

The research conducted for this project was supported by several external communities, including Dr. Megan Fillinich, Chief Technology Officer, Naval Surface Warfare Center, and multiple programs in the Program Executive Office for Command, Control, Communications, Computers, and Intelligence, as well as various cost communities and subject matter experts whose perspectives provided insights into the challenges surrounding maturation of cost methods in a rapid acquisition environment.

Disclaimer and Distribution Statement

The views, opinions and/or findings contained in this report are those of The MITRE Corporation and should not be construed as an official government position, policy, or decision, unless designated by other documentation.

Approved for Public Release; Distribution Unlimited. Case Number 18-0817.

Set-Based Evaluation Tool (SET): A Software Analysis Tool to Support Set-Based Decision Methods

Stephen G. Hunt—is the Senior Scientist Technical Manager, Director for Mine Warfare Rapid Prototyping at the NSWC Panama City Division. He is one of five executives responsible for collaborating across the Naval Research & Development Establishment on rapid prototyping, experimentation, and demonstration efforts. Hunt is a prior Major Program Manager for the Deployable Joint Command and Control program. He has a BS in mechanical engineering from Virginia Tech and an MS in science management from Florida Tech. [stephen.hunt@navy.mil]

Richard Byers, LCDR, USN (Ret.)—is the Technical Area Expert for Set-Based Design at the NSWC Panama City Division. He is the project development lead for SET. He was the lead systems engineer for the Smart Mining Initiative. Byers is a prior Surface Warfare Officer (CRUDES) and enlisted submarine fire-control technician (FTG). He has a BS in electronics technology engineering from Norfolk State and an MS in systems engineering from the Naval Postgraduate School. [richard.w.byers@navy.mil]

Rhonda Hoeckley—is a writer, published author, and senior analyst for the New Venture Research Corporation. She supported the NSWC Panama City Division Innovation Cell in 2014–2015, helping develop the concept for a smart mine, and then served on the Smart Mine Initiative team in 2016. She currently supports the Smart Mining Project Office and the Set-Based Design Evaluation Tool (SET) development team, both at NSWC Panama City. She has a BA in journalism and English from Texas A&M University. [rhonda.hoeckley.ctr@navy.mil]

Abstract

This paper describes a software analysis application, the Set-Based Design Evaluation Tool (SET), which supports an innovative accelerated acquisition methodology for rapidly informing prototyping investments across the Navy and Marine Corps. The tactics and technology exploration and experimentation (TnTE2) method is fostering innovation and is used to quickly respond to high-priority urgent or emerging operational needs. This methodology brings together operational and technical teams of warfighters and engineers, leveraging aspects of a systems engineering methodology called Set-Based Design (SBD) to rapidly assess emerging technologies and engineering innovations against a specific capability-based framework. The basis of this practice is rooted in the SBD systems engineering construct to enable data-based decision-making. The SET software automates the analysis by coding the configuration evaluation portion of SBD into a user-friendly application to significantly increase the speed of analysis, reduce the chance of data input error, and standardize the reporting. Specifically, SET provides a streamlined and systematic way to

- Create a Capability Concept Wheel
- Quickly process extremely large data sets (trillions)
- Integrate and process/filter data
- Produce concise visuals of data relationships and solution alternatives
- Provide reports of analysis results

Its demonstrated benefits include enabling *users* to understand and rapidly assess interdependencies between requirements, components, and variables of large and complex data sets; providing a means for *decision-makers* to explore the tradespace and perform cost versus capability trade-offs; and giving *leaders* an automated tool to maintain and manage evolving requirements.

Introduction

In today's world, new technological breakthroughs seem to occur daily. SpaceX successfully launches the heaviest rocket in history into space, and then precisely lands its two booster rockets on their designated landing pads. Robots from Boston Dynamics effortlessly scramble over rough terrain with perfect balance, and easily open closed doors and walk through. Medical technologists around the world print human hearts, kidneys, and livers in the race to produce the first viable 3-D printed organ that can be implanted in the bodies of people on long organ donor lists. Each of us read about or watch these events (often in real time) on small screens on phones in our hands, or sometimes on even smaller screens on our watches. Even those of us who work in technical fields cannot help but be awed at the technological innovation that seems to be exploding around us.

In this dynamic atmosphere, anyone paying attention would find it difficult to believe there is only one technological answer out there for any given technical problem. Instead, when faced with such a problem, we would expect to be able to choose from a plethora of technology solutions—some of which we likely didn't even know existed until we began to fully analyze our problem and start our solution search in earnest. But with this wealth of technological possibilities comes a challenge: How do we quickly and effectively evaluate, compare, and choose the "right" solution from a large pool of varying potential solutions without spending excess time and money on the search?

The DoD is facing this challenge as it works to maintain our technological edge over adversaries who are rapidly catching up. While it is crucial that we continuously explore new technologies (and enhance older ones where feasible) as rapidly as possible, there is not sufficient time, funding, or personnel for the DoD and Services to pursue every technological idea that has promise, raising the questions: Which technologies do we invest in, and how can we get them in the hands of the warfighter as soon as possible? Which capabilities are the most important to satisfy?

Defense leadership views rapid concept exploration and prototyping, the "fail fast/learn fast" mindset, as key to meeting this challenge. In response, the U.S. Navy is implementing an innovative accelerated acquisition methodology to rapidly inform prototyping investments across the Navy and Marine Corps. The tactics and technology exploration and experimentation (TnTE2) method is fostering innovation and is used to quickly respond to high-priority urgent or emerging operational needs. This new methodology brings together operational and technical teams of warfighters and engineers leveraging aspects of a systems engineering methodology called Set-Based Design (SBD) to rapidly assess emerging technologies and engineering innovations against a specific capability-based framework. The methodology expands the tradespace to assess a much larger range of options. The process eliminates options only when they are proven infeasible based on objective quality evidence (OQE), and delays making critical constraining decisions until after the requirements and the solution options are better understood.

Crucial to the success of this method is the ability to quickly and effectively perform complex data analysis on extremely large data sets and then translate the results into formats (including visualizations) that decision-makers can quickly interpret to choose the most feasible solutions to further explore for a specific problem. The information analyzed includes "the ilities," such as adaptability, durability, interoperability, portability, scalability, supportability, and stability, among other non-functional requirements, to assess the operational burden of a specific solution. Decision-makers often have to weigh technical capability against critical parameters such as performance, maturity, cost, schedule, development time, and risk, all of which, when combined and compared, create millions of options in different sets of permutations. Since no commercial or government off-the-shelf tool currently exists that can quickly fuse, analyze, and display all the required data to support the decision-making process, the engineers and data scientists at the Naval Surface Warfare Center Panama City Division (NSWC PCD) developed a new software application tailored for complex data analysis. This new tool lends itself to both the rapid prototyping process and naval system acquisition. The Set-Based Design Evaluation Tool (SET) is a desktop software application that can be run on any computer with a Windows operating system. Currently in its beta phase of development, the SET is evolving through hands-on use by the technical and operational community. Through this user operational system evaluation context, SET is becoming more capable and robust and is already successfully being used to support the rapid prototyping process, including the following:

- Problem definition
- Capability concept generation
- Fleet valuation exercises
- Data analysis from demonstration and experimentation events of individual technologies

SET is designed on the backbone of the SBD engineering methodology, and thus helps ensure a sufficient degree of engineering rigor is applied in the rapid prototyping process—something that is often missing in rapid prototyping efforts. The tool organizes and analyzes the data to help produce and document OQE that is traceable to warfighter missions, scenarios and tasks, and helps define requirements.

This paper describes the history of SET, its current scope of capability (with a recent use case), and what SET will look like in the future as it evolves in capability.

Historical Evolution of SET

In 2014, NSWC PCD's Innovation Cell (iCell) proposed bringing U.S. sea mining into the 21st century by prototyping and demonstrating a modular "smart" mine suite, which would: (1) include communications, command and control, sensors, and both kinetic and non-kinetic effector nodes; (2) launch from unmanned surface and undersea vehicles; and (3) be able to be pre-positioned in international waters to persistently influence the adversary at a time/place of our choosing. The proposed iCell concept gained early DoD and Navy leadership support, and when U.S. Pacific Command (PACOM) released a Joint Emergent Operational Needs Statement (JEONS) in 2015 seeking an asymmetric capability to address the threat of contested environments, PACOM and then-Deputy Assistant Secretary of the Navy for Research, Development, Test, and Evaluation (DASN RDT&E), Dr. John Burrow, felt that the smart mine concept could potentially offer a solution. Burrow tasked NSWC PCD to lead a cross-Naval Research and Development Establishment (NR&DE) Smart Mine Initiative (SMI) to explore innovative concepts and share technologies which could be integrated to meet the vision of a smart mine for the JEONS.

At the time of the release of the JEONS, the DASN was also exploring implementing the innovative SBD systems engineering methodology across the Navy R&D warfare centers to change the paradigm of Navy system design. The SBD methodology, which has been used successfully for years in the automotive industry by Toyota, first emerged for potential use in Navy ship design in 2008 when Vice Admiral Paul Sullivan, then-Commander of the Naval Sea Systems Command, issued a memo that expressed the need for evolving models and analysis tools to be compatible with, among other things, SBD (Singer, Doerry, & Buckley, 2009). An instantiation of the SBD methodology was successfully used by the U.S. Marine Corps in 2013 in concept exploration for an affordable, survivable, high water speed Amphibious Combat Vehicle (ACV) to replace the cancelled

Expeditionary Fighting Vehicle (EFV; Burrow et al., 2014). The first Navy instantiation of the SBD methodology was used in 2014 to develop preliminary designs for the Navy Ship to Shore Connector (SSC; Mebane et al., 2011), a new ship to replace the Landing Craft, Air Cushion (LCAC). The DASN RDT&E thought the SMI effort was a good opportunity to utilize the SBD methodology at the beginning of development of a new Navy capability to be rapidly prototyped to meet an urgent warfighter need. In 2016, NSWC PCD and its cross-warfare center SMI team began to incorporate aspects of SBD methodologies into a six-month analysis of the smart mine capability tradespace. The goal was to identify feasible configurations for potential prototyping.

In simple terms, SBD is a methodology that allows designers to fully explore a design tradespace, evaluating very large numbers of design configurations early in the design process to quickly eliminate designs which are not feasible based on OQE so that the most promising ideas can be modeled or prototyped to determine viability prior to choosing a final design. In its most basic form, SBD is design discovery by way of elimination. The delay in final design choice until superior OQE data is gathered results in better understanding of the requirements, more optimal designs, lower risk, lower cost, and increased stakeholder interaction. Using this design methodology, engineers are not selecting the best solution so much as they are eliminating the worst. It allows for convergence on a solution set that increases understanding of design decision impacts.

SBD is a significant paradigm shift from the Navy's traditional design methodology, which follows a classic design path: (1) Converge as quickly as possible on a solution (a single "point" in the identified solution space) that has acceptable risk and fits within the limitations of either budget or available time, or both; then (2) rapidly and incrementally develop and evolve that solution until it meets the requirements and can be fielded to the warfighter. This can be an effective approach if the optimal solution is selected at the start (highly unlikely). Choosing that optimal solution is challenging, however, especially when the requirements are relatively immature and not well understood (a common acquisition problem), or when the design incorporates technologies that don't perform as expected (another common acquisition problem). What happens when the solution selected proves to be less than optimal and issues arise? The process, which planned and funded solely for success, remains locked on that "point-based design" and the program manager has to spend extra time, effort, and funding to modify the design until it meets the (potentially immature, and more likely ill-conceived) requirements.

SBD takes a very different approach. As noted previously, instead of quickly converging on a single point based design solution, SBD expands the tradespace to assess a much larger range of options that includes cross domain intersections. It delays making critical decisions until later in the process when both the range of requirements and the potential solution options are better understood through OQE, data analysis, modeling and simulation (M&S), and rapid prototyping.

A good analogy for the value of SBD can be found in the story of those infamous builders "The Three Little Pigs," who classically used a point-based design approach to design a shelter to protect themselves from the Big Bad Wolf. Under their point-based design approach, they generated hard and fast requirements too early with incomplete knowledge of the problem, resulting in incorrect assumptions, such as the structure had to provide shelter, it had to stand up to strong wind, and it had to hold at least one pig. When they began looking for solutions to meet their requirements, they constrained their trade space to the design options they were familiar with rather than expanding the space to look for more innovative options. For example, the type of shelter they quickly settled on was a building (expanded trade space options may have included a tree house or a cave or an underground home or a wolf-free island that needed no shelter). They chose materials they were most familiar with, such as straw, sticks, and bricks. An expanded tradespace may have included glass and metal. They then proceeded down three independent paths and conducted their trade studies in stovepipes. Their studies for materials, floor plan, roof pitch, and so forth, were all conducted independently with no cross-domain intersection considerations. No one understood the core required capability (the critical design factor), which was not maintaining security of the pigs as much as wind resistance. As a result, the first design, the straw house (chosen because it met the basic requirements and appeared to be cost effective), failed integration and operational testing when the wolf came knocking. They then had to redesign the straw house for a better (but not fully) understood requirement for wind. Under their point-based design approach, it took multiple iterations—each with an associated cost and schedule—before they gained enough knowledge to inform the final design and realize a brick house would work. Their point designs became hindsight engineering. And with a dangerous enemy like the wolf, they could ill afford these high-risk return trips to the materials pile to get it right.

Under the SBD approach, the pigs would have designed all three houses concurrently, and then used field and lab testing, low-efficacy models, and simulation techniques to eliminate poor design configurations prior to committing resources for building the final house. They would have made design decisions based on OQE, and they would have had a better understanding of how to meet the now fully understood requirements. Bottom line: If they had used the SBD approach, they could have saved time and money and all lived out their lives safely and comfortably in brick homes (or maybe even in a cool cave or on a wolf-free tropical island).

When the wolf is at the door, SBD can show that not all feasible designs (those that look good on paper) are viable (actually work), which allows leadership an opportunity to make important decisions and understand their impact before the house is built. The SBD engineering effort has a cost and schedule that is front loaded. It may not be quicker getting started; however, the savings in cost and schedule come in the latter half of development by avoiding costly and lengthy redesign, test, and production reiteration.

Specifically, the SBD methodology does the following:

- Considers large data “sets” of candidate solution alternatives in the trade space (often containing millions of potential configurations)
- Takes advantage of modern automated analytical frameworks that leverage high-speed computing power to develop, explore, manage, and visualize large data sets
- Reduces the trade space in a progressive, deliberate manner by eliminating alternatives only when objective evidence (analysis, M&S, rapid prototyping) shows they do not meet the necessary criteria (i.e., feasibility/viability and cost/schedule/performance)
- Increases knowledge as the sets of alternatives are narrowed
- Converges to more globally optimal solutions with greater fidelity
- Reinforces confidence in final recommendations to leadership with a pattern of reproducible, defensible artifacts in support of the decision process, specifically before highly constraining decisions
- Builds a body of lessons learned on options that were eliminated, which can inform future efforts

A critical requirement of SBD implementation in system development is a means to rapidly conduct the complex data analysis and translate the results for decision-makers. For the previous Navy SBD efforts, a one-of-a-kind automated analytical framework tool called the Framework for Assessing Cost and Technology (FACT) was developed for each specific effort to provide visual simulations that integrated data and model input on design, cost, schedule, etc., to allow leadership to quickly see how the various design choices affected outcome.

Specifically, the FACT tool (whose capabilities are shown in Figure 1) enabled the efforts to

- Process large data sets through integrated M&S
- Provide data integration, processing, and concise visuals of data relationships and solution alternatives
- Allow users to understand and rapidly assess interdependencies between requirements, components, and variables of large and complex data sets
- Allow decision-makers to explore the tradespace and compare alternatives
- Allow leaders to maintain and manage an evolving requirements set

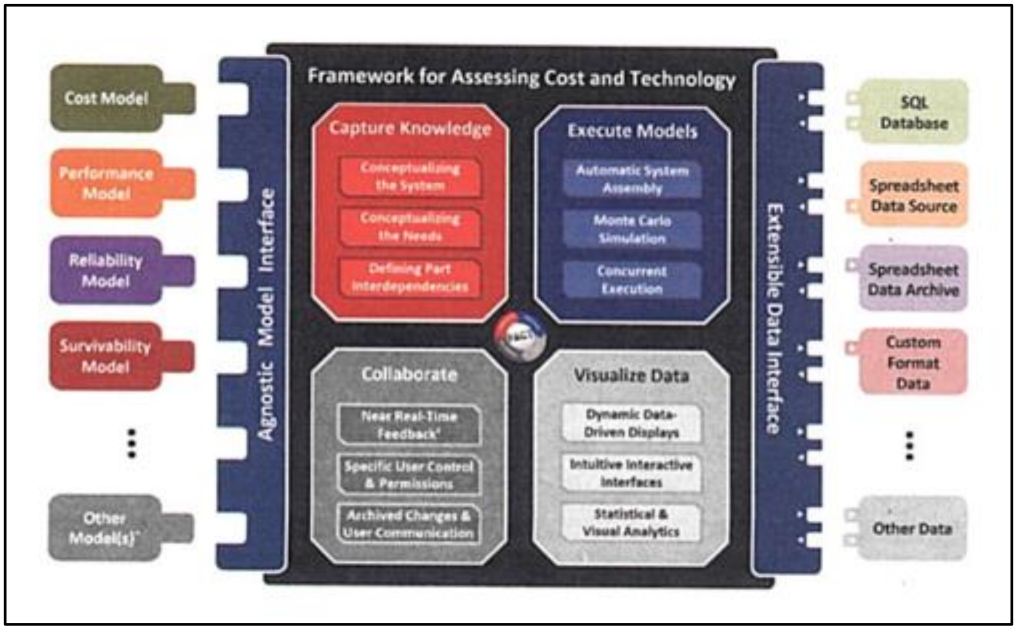


Figure 1. Framework for Assessing Cost and Technology (FACT) Structure

While the FACT tools developed and used for both previous efforts shared a common software architecture, each version required unique software coding, man-hours, and expense to tailor it to the subject matter. Since SMI did not have the time nor funding available to pursue a lengthy rework of the existing tool, when the SMI effort launched NSWC PCD and the Georgia Tech Research Institute (GTRI) (which developed the FACT software for the U.S. Marine Corps) worked together following the FACT example to create a less detailed, but more adaptable, version of the previous tools to meet the needs of the smart mining effort. Creating the SMI version of FACT took six months of focused development and coding. In the end it was a single instantiation of an SBD tool that was developed for a single purpose. It built upon and progressed previous SBD tools, as it did allow the user to quickly (~30 minutes) analyze very large data sets. At the end of its development, the SMI FACT was able to help leadership visualize and analyze a tradespace of 1.9 quadrillion possibilities that resulted in a recommendation of nine feasible configurations that included cost, schedule, risk, and performance.

Since the DASN planned to implement SBD across the entire Navy R&D community, it became clear that a new FACT-like tool was needed that was flexible and robust enough for systems engineers to apply to any SBD system development effort without the need for major recoding. After the SMI six-month effort was completed, NSWC PCD systems and software engineers began developing SET to fill that tool gap. The SET team is building on the FACT capabilities, which improved with each of the FACT instantiations (as shown in Figure 2). The SET team has been working on the tool continuously since 2016, adding more capabilities daily (sometimes "on the fly" as needs emerge while being used to support an R&D event). As noted previously, the tool is still in beta form, but it is being used to support current rapid acquisition efforts using SBD methods. The vision for SET is to reach the robustness of the previous FACT tools (including being able to ingest inputs from external models such as cost models) in an "off-the-shelf" version flexible and powerful enough to work across all problem solution efforts employing a SBD methodology.

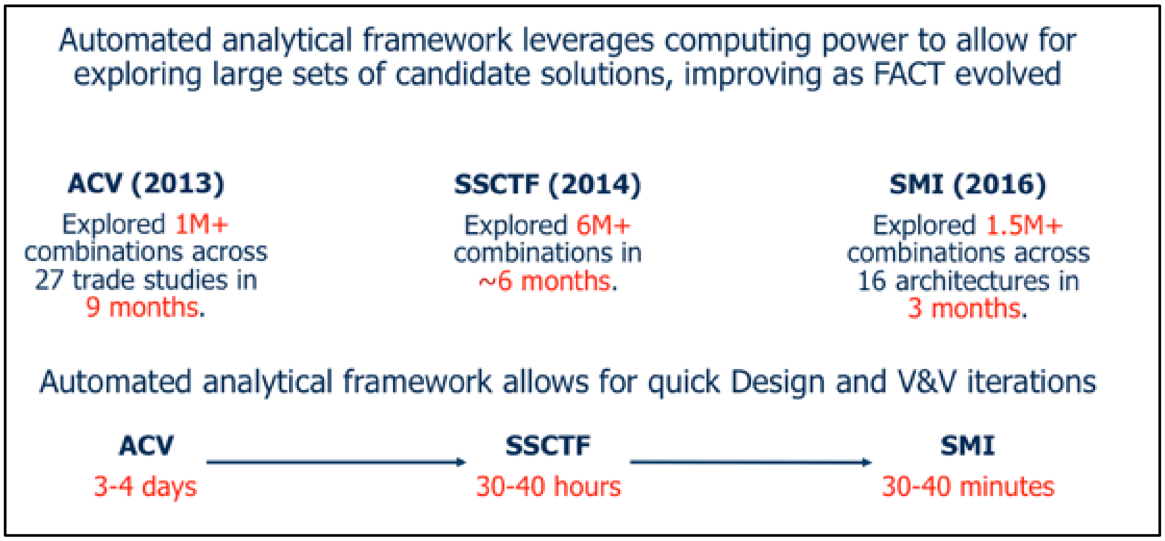


Figure 2. Improvements in FACT Performance

Current SET Capability and a Representative Case Study

SET provides an easy-to-use tool to quickly process extremely large data sets (into the trillions), integrate and process/filter the data, and provide concise visuals (such as scatter plots and histograms) of data relationships and solution alternatives. Currently it is limited to evaluation and comparison of individual technologies, but the envisioned end state will allow for evaluation and comparison of sets of configurations which can be quickly formed and re-formed using the tool to allow decision-makers to clearly see how different design choices in configurations affect system outcome.

SET is structured around the elements of a Force (e. g., the warfighter) Engagement Process framework using tailored SBD elements that can be used to support the rapid prototyping process and overall Navy system development. The framework translates a specific, emerging Fleet or Joint Force concept of operation (CONOP) into increments of capabilities. Those increments of capability, presented through a SET visualization tool that models a Capability Concept Wheel (CCW), are used during a series of scenario-based wargames with teams of warfighters to provide insight into the relative value of increments of capability to effective mission operation(s). Once the capability concepts which are the most highly valued by the warfighters are identified, a robust database of relevant technologies that can support those capability concepts is developed through data calls and calls for proposals. Subject matter experts (SMEs) then analyze and bin the submitted technologies by capability concept, and assess them operationally and technically. Technologies are eliminated from consideration only when OQE shows them to be infeasible for helping solve the problem. The resulting narrowed group of technologies are then assessed operationally (in a scenario) and technically by warfighters in an Advanced Naval Technology Exercise (ANTX). In cases where a technology is supported by such a high degree of OQE that further evaluation is superfluous, the technology may skip an ANTX and begin planning for prototyping.

The Force Engagement Process framework (shown in Figure 3) underlying SET involves three phases:

- Force Valuation
- Assessment Workshop
- Demonstration and Assessment

A fourth phase is envisioned for the future. It includes the integration of models (e. g., cost and/or performance), so that decisions on acquisition, contracting, and system prototype configurations can be completed. Data analyzed through SET will lead to recommendations for continued system design of feasible configurations that lead to execution of viable designs and solutions.

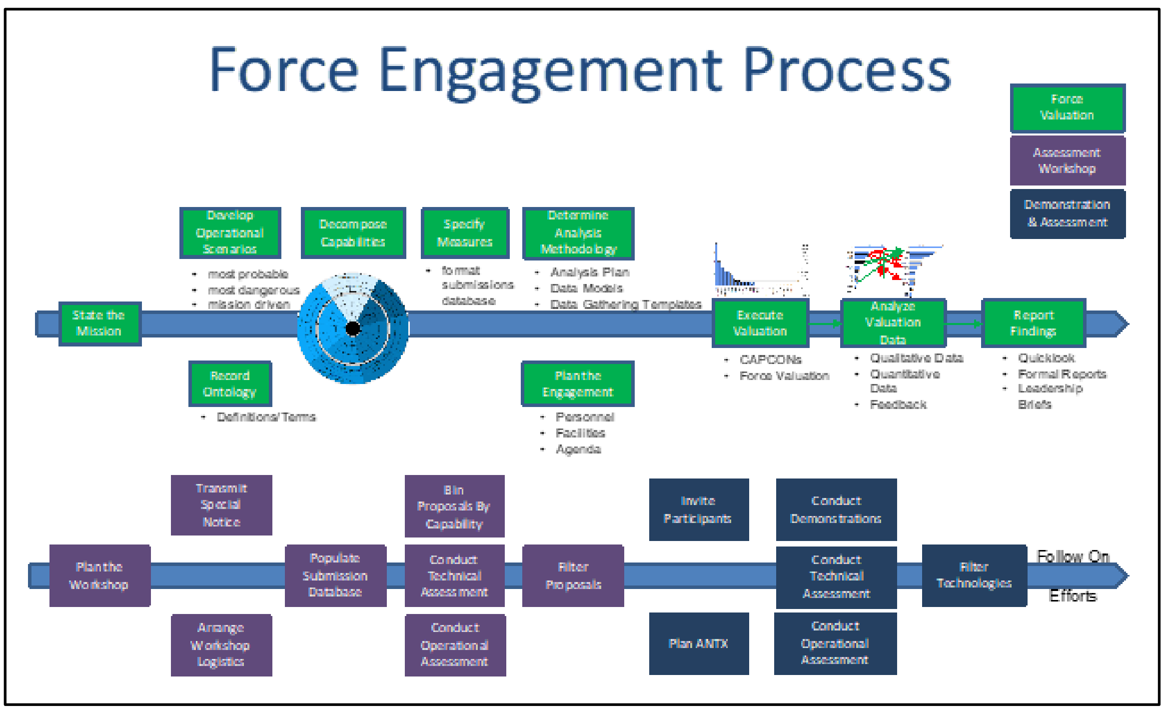


Figure 3. Force Engagement Process Underlying SET

In the Force Valuation phase, the systems engineering team makes the engineering preparations for and executes a Force Engagement Team (FET) valuation exercise in which warfighter teams with a mix of mission-appropriate skill sets use a Capability Concept Wheel in a tabletop wargame to determine which of the capability concepts they value most highly in a relevant operational scenario. In this phase, SET can be used to record the definitions of terms so all participants use the same taxonomy; automatically create the Capability Concept Wheel based on manual inputs from the capabilities decomposition; record the points assigned during the valuation wargame; analyze the valuation results; and generate reports (including visualizations such as scatter plots and histograms).

In the Assessment Workshop phase, the team issues data calls and requests for proposals to collect data on relevant technologies that potentially support the most highly valued capabilities. The collected data is then used to create a technology database. SMEs meet to bin the technologies by capability, then conduct technical and operational assessments to eliminate technologies based on initial filtering criteria, thus narrowing the number of potential technologies going into the ANTX. In this phase, SET can be used to import the technology database, bin the technologies, record the assessment data, and filter the technologies.

In the Demonstration and Assessment phase, the team plans and executes an ANTX or other demonstration in which technologists are invited to bring and demonstrate the filtered technologies so warfighters can physically view them and conduct technical and operational assessments. In this phase, SET can be used to create the electronic assessment forms, record the results (through individual networked computer tablets) from the technical and operational assessments, filter the technologies based on assessment results, and generate reports (including visualizations).

The current beta version of SET was most recently used to support a recent Urban 5th Generation Marine (U5G) effort. As outlined in the Marine Corps Operating Concept (MOC) and the Marine Corps Intelligence Activity's Future Operating Environment, the growth of crowded, poorly governed, or lawless areas (particularly in and around the world's littorals) will force future commanders to consider how to conduct operations in complex terrain. In 2017, a U5G Task Force was established to develop concepts that enable situational awareness, counter reconnaissance, maneuver, fires, and command, control, communications, computers and information (C4I) operations within and among the populations that reside in the urban littorals. The task force—comprising a core team of operational, acquisition, and technical subject matter experts from Headquarters Marine Corps Combat Development and Integration (CD&I), Marine Corps Warfighting Lab (MCWL), Marine Corps Systems Command (MCSC), and the NR&DE—is charged with executing a progressive series of ANTX that will inform emerging concepts of operations and future acquisitions.

The first end-to-end exercise of the TnTE2 method, U5G ANTX 2018, was held in March 2018 at Marine Corps Base Camp Pendleton, CA. The ANTX was structured to provide warfighters with the opportunity to assess the operational utility of emerging technologies and engineering innovations that enhance the U5G concept of operations as it applies to a Marine Air Ground Task Force (MAGTF), which is given missions to operate in an urban environment.

The ANTX explored the five domains of air, space, cyber, logistics, and intelligence, with the focus on how those areas affect the operation in the urban environment at the company level. The ANTX used two vignettes to provide context for employment: (1) A rifle company must secure a key piece of infrastructure in a hostile environment where adversaries blend with civilians requiring a high degree of urban situational awareness, precision effects, and minimal signature; and (2) A rifle company, as part of a larger operation, must conduct offensive operations to clear a complex urban area consisting of multiple city blocks, underground corridors (subway, sewer basements, etc.), and multi-story buildings. Once cleared, they must secure and defend the area while potentially providing assistance to any remaining civilians. During the course of both vignettes, the rifle company is continually conducting offense, defense, and stability.

SET was used to support each of the three phases of the Force Engagement Process in the U5G effort.

Primary goals for SET during the ANTX event included the following:

- Improving methods, processes, and tools
- Integrating assessments of operations, technology, and capability
- Leveraging technology to enhance data capture and analysis
- Providing traceability to U5G core capabilities
- Providing data longevity for future study
- Providing insight upon which solutions are ready for rapid acquisition, experimentation, or are a science/technology of interest

Prior to the start of the ANTX, Operational and Technical SMEs helped draft the criteria, scales, and weights associated with assessing the candidate technologies. The ANTX event was split into two days of Limited Technical Assessments (LTAs) and 2 days of Limited Objective Experimentation (LOE). During the LTAs, squads of Marines handled individual pieces of technology at static displays or as part of live demonstrations. During the LOEs, platoons carried out operational missions, both day and night. Three platoons were issued multiple pieces of technology while rotating between mission objectives and operational locations. Upon conclusion of each LTA or LOE event, an embedded data collector would capture both quantitative and qualitative assessment data from individual marines, and the technical assessors via the SET assessment interface. Technical assessments were captured by tailored, diverse groups vice individuals to ensure subject matter expertise was factored in and the groups could build upon separate areas of knowledge for a complete assessment: mechanical, electrical, computer science, and so forth. Over four days, SET processed 2,664 total assessments: 2,210 operational, 304 technical, and 150 scenario based.

SET provided near real time statistical observations on operational and technical performance of demonstrated technologies from the warfighter and engineering perspectives in the form of histograms and scatter plots. It translated concepts of capability assessment, technical assessment, and operational assessment into data views which enabled cross referencing of performance not only from a total score perspective, but it enabled drilling down to specific questions of interest, such as operational relevance or personnel burden. SET also highlighted those technologies that traced back to the top capabilities required for success within the proposed scenarios.

Future developments of SET may be able to provide justification criteria for DoD acquisition by providing the linkage between capability requirements and in field testing.

The next several figures are good representations of data from the U5G ANTX. Figure 4 presents the Capability Concept Wheel that the U5G team created using SET for the Force Valuation phase. Figure 5 presents a visualization of the Assessment Workshop phase binning of the 93 technologies that were evaluated. Figure 6 presents an example (non-U5G) of an assessment form created using SET (similar to what was created for the U5G ANTX). Figure 7 provides a sample of a visualization from the U5G ANTX assessment results.

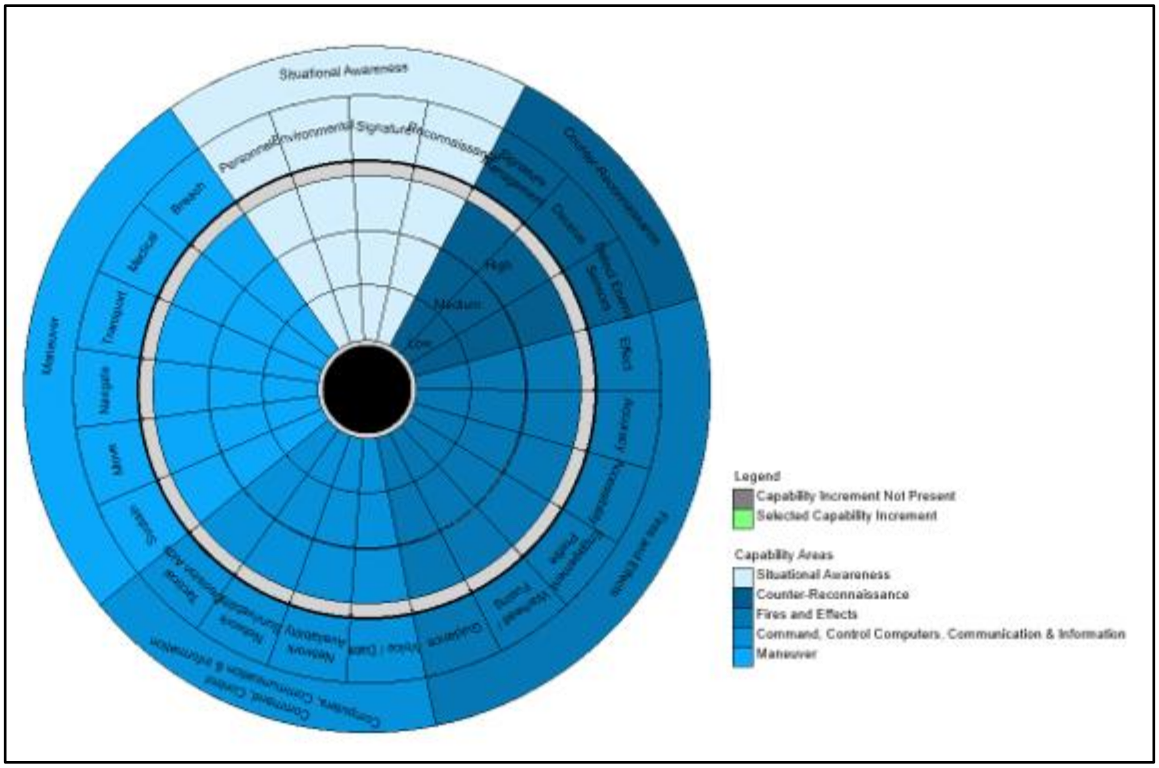


Figure 4. U5G ANTX 2018 Capability Concept Wheel

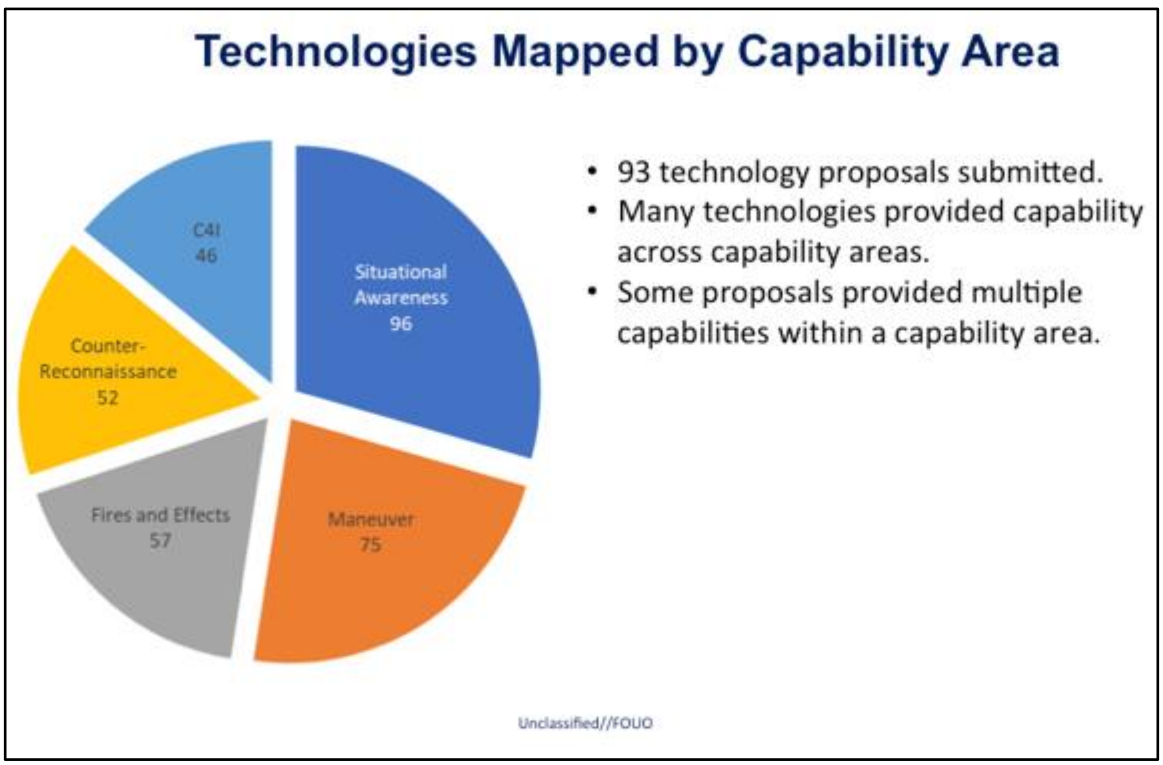


Figure 5. U5G ANTX 2018 Technologies Mapped by Capability Area in SET

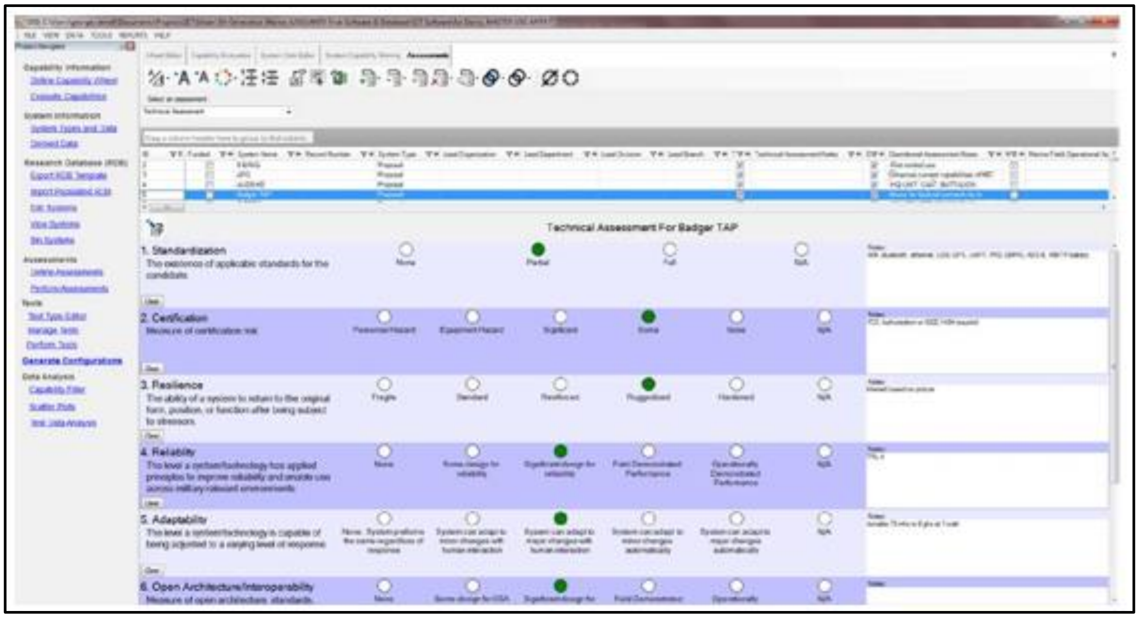


Figure 6. Sample Assessment Form Created With SET

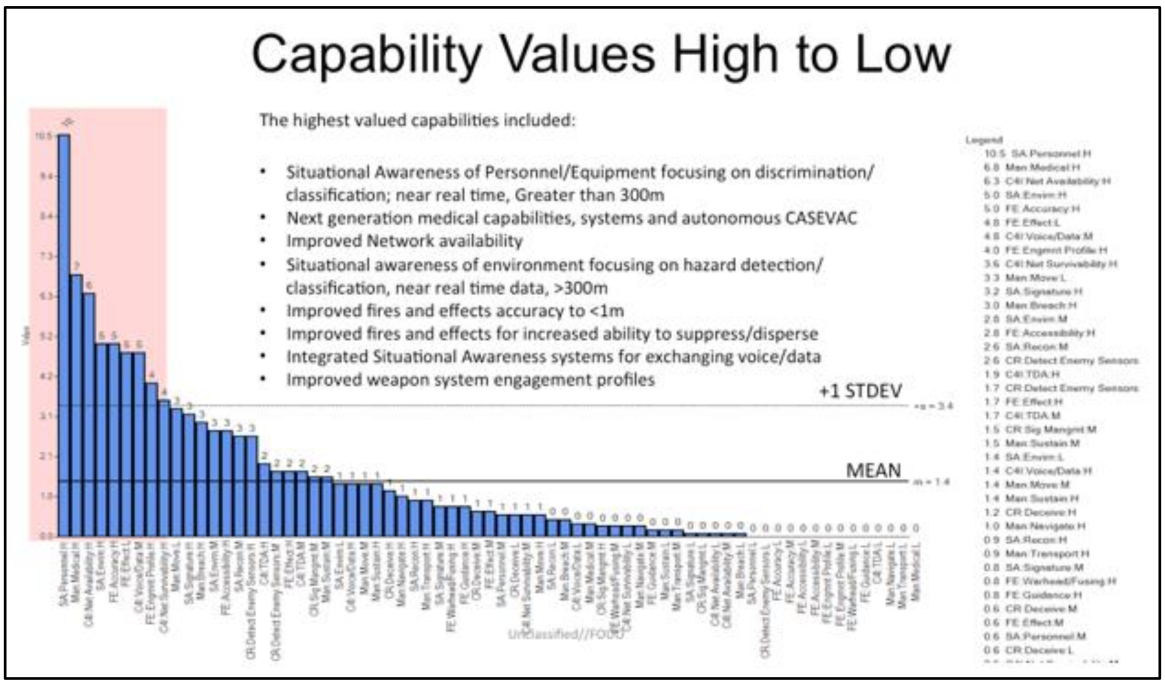


Figure 7. Example of Assessment Results Visualization From SET for U5G ANTX 2018

Next Stages of SET Development

In its current instantiation, SET successfully evaluates measures of stakeholder utility and may apply these measures against technology solutions whether they are tactics, process, or systems. However, to come to full realization as an automated analytical framework tool, current development needs to address not just stakeholder utility, but also a means to understand the relationships between sets of solutions and how they may drive the feasibility of design.

A close focus on improving the user interface will help fortify the methodology and help teams adopt the principles of SBD. Focus on SET outputs will aid in institutional learning, reuse, and knowledge retention. New designs will be able to build upon previous explorations.

Functions that are currently under development in SET include the following:

- The capability to create design sets and configurations, providing definition and constraints
- Establishing mathematical relationships to set intersections
- Providing a GUI which allows input, modification, and visualization of design-impacting attributes
- The capability to evaluate a design space for feasibility and dominance, specified by a set of constraints and given specific functional relationships
- The capability to interface with other models, such as cost and risk, to provide other interpretations of set feasibility

Ultimately, the goal of SET is to help inform future design and improve design quality by allowing greater design space exploration and providing support to a methodology which may provide solutions more resilient to requirement changes, grant early understanding of design relationships, and reduce design rework.

References

Singer, D. J., Doerry, N., & Buckley, M. E. (2009). What is Set-Based Design? *Naval Engineers Journal, 121*(4), 31–43. doi: 10.1111/j.1559-3584.2009.00226.x

Burrow, J., Doerry, N., Earnesty, M., Was, J., Myers, J., Banko, J., McConnell, J., Pepper, J., & Tafolla, T. (2014). Concept exploration of the Amphibious Combat Vehicle.

Mebane, W. L., Carlson, C. M., Dowd, C., Singer, D. J., & Buckley, M. E. (2011). Set-Based Design and the ship to shore connector. *American Society of Naval Engineers, 3,* 79–92. doi: 10.1111/j.1559-3584.2011.00332.x

Application of Set-Based Decision Methods to Accelerate Acquisition Through Tactics and Technology Exploration and Experimentation (TnTE2)

Carly Jackson—is the Senior Scientific Technical Manager responsible for rapid prototyping information warfare capabilities. As the director of prototyping, she leads diverse and highly technical teams through the development, experimentation, and fielding of advanced command, control, communications, computers, intelligence, surveillance and reconnaissance, space, and cyber technologies. Just prior to assuming this role, she was detailed to ASN (RD&A), where she played a critical role in formulating the Department of Navy's strategic initiative to accelerate acquisition through rapid prototyping. Jackson holds BS and MS degrees in mechanical engineering from UCLA and an MBA from Pepperdine University. [carly.jackson@navy.mil]

Aileen Sansone—supports DASN(RDT&E) and the Marine Corps Warfighting Laboratory in the planning and execution of S2ME2 follow-on efforts as well as co-leading the Urban 5th Generation Marine Task Force TnTE2 effort and participating in Expeditionary Advanced Base Operations TnTE2 planning. Previously she was detailed to the Deputy Assistant Secretary of the Navy for Research, Development, Test, and Evaluation (DASN[RDT&E]) Rapid Prototyping, Experimentation, and Demonstration (RPED) office. During her time at DASN(RDT&E), she co-led the planning and execution of the S2ME2 Task Force efforts. Dr. Sansone holds a PhD degree in electrical engineering from the University of Maryland. [aileen.sansone@navy.mil]

CAPT Christopher Mercer, USN—is the Department of the Navy's Director of Rapid Prototyping and Experimentation. He has spearheaded the development and implementation of agile, set-based methods as underpinnings to the Department's reforms in accelerated acquisition through rapid prototyping. CAPT Mercer is an Engineering Duty Officer and Acquisition Professional with over 20 years of research, development, and procurement experience including multiple projects using set-based design techniques. CAPT Mercer holds a BS degree in Marine Engineering from Maine Maritime Academy and an MS degree in electrical engineering from the Naval Postgraduate School. [christopher.p.mercer@navy.mil]

Douglas King, Col, USMC (Ret.)—holds a Master of Arts and Science in Operational Art and a second in Strategic Study. He is a graduate of the School of Advanced Military Studies and was designated as a Master Tactician by the United States Army. He served 28 years active duty as a Marine Corps planner, armor officer, and reconnaissance Marine, and retired as MCCDC G3/5 responsible for concepts and plans. He currently serves as the Director of the Ellis Group, responsible for continuous and progressive examination of how the USMC operates and fights. He was the Marine Corps lead author of the Army–Marine Corps Counterinsurgency Doctrine, *A Cooperative Strategy for 21st Century Seapower*, the Naval Operating Concept, and the Marine Corps Operating Concept. [douglas.king@usmc.mil]

Abstract

The tactics and technology exploration and experimentation (TnTE2) method has been shown to foster innovation and create speed in responding to high-priority urgent or emerging operational needs. The TnTE2 method rapidly shepherds a balanced team of warfighters and technologists through a series of capability-based rapid prototyping and experimentation cycles, which accelerate complex warfighting concepts and tactics development. Set-based design (SBD) methods enable full exploration of both the warfighting capability and technology trade space. SBD tools and visualizations facilitate complex data analysis and decision-making. TnTE2 methods were synthesized in the recent Ship to Shore Maneuver Exploration and Experimentation (S2ME2) Task Force (TF), where a first-ever Advanced Naval Technology Exercise (ANTX) informed a developing warfighting concept and started associated rapid prototyping projects—all within the standard program

objective memorandum (POM) cycle. The TnTE2 method is open, competitive, and merit-based. Future applications of TnTE2 will be expanded to include the deliberate use of new acquisition authorities, policies, and contracting vehicles, and has the potential to fundamentally change how systems are acquired with agility and speed.

Background

USS *JOHN C. STENNIS*, somewhere in the South China Sea.

Warfare Commanders, War Room, 0330 hours (local time).

Final strike planning …

"I can't show protection in this close with these environmentals, boss. As Air Defense Commander, I'm saying … if you launch this morning, we're gonna take losses."

"We are going to take losses, Steve. This strike needs to happen now though, this morning, with the sun at our backs. These Alert fives you've had us on are taking a toll on birds AND crews. CAG's down three Hornets as it is, so another 24 hours isn't in the cards. Look, I just need to be here, and quiet, for another hour, okay? Keep working the problem. Rollo, anything new from CHUNG HOON?"

"No sir, just the 'POSS-SUB' we told you about earlier, never confirmed. Theater ASW's barrier search will be done at 0430, but buoy data's showing some movement in the sound velocity profile down the threat axis."

"Ok. IWC, talk to me about layers, RF and acoustic. I want to be launching in an hour and striking five minutes after sunrise."

"Admiral, METOC's loading the 0330 acoustic predictions now. Surface layers have been really close for days, so I'm confident in Rollo's posture for that strike window. Steve's got a problem though. NAVOCEANO launched a new RF modelling format yesterday that is MUCH better, but our NITES system needs a new card to use it. OCEANO's paralleling the old format, but not over GBS, so I can't get it passively. Right now, my best RF prediction is the air search RADAR which is why Steve's Aegis system is being so conservative. TWC and I think we can get NITES back online by 0400 though … right, Melony?"

"That's right, Ender. Admiral, I just chatted with Linda Collins and Admiral Beckett in San Diego. They did a rapid prototyping authorization for us, and I expect a file for the new NITES card any minute. It'll take 10 minutes to print and install the card, so NITES should be good to go by 0400."

"Alright, good. Port the new RF paths directly to Aegis so Steve can update apportionment for inbounds. I want us all playing off the best sheet of music when we go on this. Everyone stay on your bricks, but keep 'em on Low. We don't need any leakage while our tattletale is enjoying his last hour above the surface. I'll be in TFCC at 0415. Ready, Break."

This vignette is completely imaginary. The scenario and all characters are fictitious, but they plausibly juxtapose today's strike force leadership and planning with future capabilities envisioned by authors of this paper.

Introduction

Over 30 years ago, the U.S. Navy knew it had to coordinate air, surface, and subsurface warfare in order to maintain superiority over what we then called near-peer competitors. The composite Warfare Commander (WFC) construct helped us achieve that coordination, establishing Air Defense Commanders, (like Steve in the vignette above), Strike Warfare Commanders (CAG), and Sea Combat Commanders (Rollo) to form a team of super teams, each dominant in their domains and each contributing to shared situational awareness of the overall battlespace. This was ahead of its time, actually, and it served us very well until the "knee" of a technology curve began pressing on our chest about five years ago.

For the past three years, we've been working to integrate information warfare (IW) into that composite model (led by IWC, or Ender, in the vignette). We hope IW will help us match our own pace of maneuver and lethality to that of our "informatized" adversary. This force-matching instinct, however, which has been bred over centuries of warfare, approaches a new, information-enabled battlespace as a more complicated but ultimately predictable environment. While IW feels revolutionary, it may have missed a fundamental change in that environment. A technology-fueled explosion of interdependency between tools, operators, and tactics now accelerates naval warfare near a distinct complexity threshold. General Stanley McChrystal found himself on the wrong side of that threshold in 2004 while fighting Al Qaeda in Iraq (McChrystal et al., 2015). There, he discovered an environment of such staggering complexity that prediction and operational efficiency were rendered obsolete by resiliency and adaptability. Since our IW work is largely about efficiently sensing and predicting highly contested battlespaces, maybe we missed the root cause of that heavy feeling in our chest five years ago. Now, as the feeling moves toward our throat, we think it is less about technology and more about interdependencies and speed, about unpredictable complexity.

In 2015, the Chief of Naval Operations (CNO) said that "core attributes" like Integrity, Accountability, Initiative and Toughness must underpin a decentralized command structure (CNO, 2015). To anyone who grew up with "Honor, Courage, & Commitment," the first two items on the CNO's new list were unsurprising. *Initiative* and *Toughness*, however, seemed new, and a little edgy. They sounded a lot, in fact, like McChrystal's resiliency and adaptability. If we mean to use these new attributes against a peer competitor, we should stop prioritizing clever tactics and exquisite tools, which have unknown but probably short shelf lives in a complex battlespace. Instead, we should prioritize a *method* for speedy, coordinated, correct capability fielding, so that we can adapt faster than our adversary in a complex battlespace.

This paper describes a methodology that has been shown to synchronize the development of tactics and technology in response to high-priority and emerging operational needs. The tactics and technology exploration and experimentation (TnTE2) method, pronounced "T-N-T-2," fosters innovation and creates speed by rapidly shepherding a balanced team of warfighters and technologists through a series of capability-based rapid prototyping and experimentation cycles. TnTE2 methods and tools were synthesized through two recent Task Forces that applied the method to rapidly identify highly-valued capabilities, inform developing warfighting concepts, and initiate rapid prototyping lines of effort—all within the standard program objective memorandum (POM) cycle. The recent Task Forces (i.e., Ship to Shore Maneuver Exploration and Experimentation [S2ME2] and Urban Fifth Generation Marine [U5G]), formulated and applied the TnTE2 method with a keen understanding of the nature and politics of major military innovation (Gardiner, 1992; Hayes & Smith, 1994), agile teaming constructs (Rubin, 2013), war room processes

(Burrow, 1997), and set-based design principles, as expanded by the Department of the Navy (DoN) for the Small Surface Combatant Task Force (Garner et al., 2015) and Advanced Combat Vehicle (Burrow et al., 2014). The TnTE2 method and the Task Force strategy collectively enabled the rapid and full exploration of highly complex technical and operational solution spaces and promoted in major naval innovation. The TnTE2 method "bring(s) together technology, doctrine, and policy objectives" (Gardiner, 1992, pp. 10–11) to encourage innovation from within the Naval bureaucracy, promoting "diversity of potential sources of innovation" (Hayes & Smith, 1994, p. 75), and provides a forum for "prototype(s) … (to) demonstrate feasibility … at critical junctures" (Hayes & Smith, 1994, p. 100).

Throughout S2ME2 and U5G Task Force execution, the Assistant Secretary of the Navy for Research, Development and Acquisition (ASN[RD&A]) worked across Congress, the Office of the Secretary of the Defense (OSD), the DoN, and Naval Systems Commands to explore underutilized financing, acquisition, and contracting authorities. This "Fast Lane" effort, further described below, is aimed at effectively scaling up the successful prototyping effort and applying new accelerated acquisition authorities and tools to speed capability to the field.

Unless we think the pace of technology, or that of our peers, will relax, the "knee on our windpipe" must be addressed. Strong efforts to accelerate capability fielding have existed in both the Fleet and the Naval Research and Development Enterprise (NR&DE) for years. In a few recent cases, these efforts have collided like neutrons, releasing tremendous energy in the form of a special, new kind of rapid prototype. These rapid prototypes are distinguished by the fact that their specific capability contribution is valued and fully characterized from both a tactics and technical perspective, allowing one to imagine a future acquisition system where our "technology bench" is expansive, fully characterized, and standing ready to field. Such rapid fieldings of emergent, modified, or adapted technologies to U.S. forces at sea (hypothesized below) might be coordinated, afloat, by a technology warfare commander (TWC, or Melony, in the vignette).

Today we face a multi-domain warfighting problem defined by unprecedented complexity. Recent, deployed experience with that complexity tells us that a successful capability-based concept exploration method rooted in Set-Based Design (SBD) can provide a simple but powerful mechanism to prepare for high-end conflict. Only through synchronized discovery and injection of tactics and technology, tailored in iterative and progressive cycles toward specific warfighting capability, can we pace technology and outpace our adversaries.

Prior Work

Our most senior leaders recognize the potential of a deliberate and continuous partnership between the engineering, acquisition, and operational communities in addressing high-end conflict. The Chief of Naval Operations' (CNO) *The Future Navy* was heavily influenced by the pace of technology, and that of our adversaries, speaking often to notions like "rapidly iterative approach[es]" and encouraging us to "simultaneously build and innovate" (Richardson, 2017, p. 7). The CNO told us to optimize requirements via "meaningful discussion[s] [between] industry leaders, technologists, our defense labs, the requirement officers, and our budget people" (Richardson, 2017, p. 7).

On January 7, 2016, Assistant Secretary of the Navy for Research, Development, and Acquisition (ASN [RD&A]), the Honorable Sean Stackley, laid down a marker with Congress, calling for "active and continuous engagement by our Naval Research and Development Establishment (NR&DE) with Fleet forces" (Stackley, 2016, p. 4). The elevation by Stackley of this imperative concept seeded a revolution in acquisition methods, which has grown strongly in the two years since his call. The 2018 National Defense Strategy (DoD, 2018) also encourages developing operational concepts and technologies together:

> Evolve innovative operational concepts. Modernization is not defined solely by hardware; it requires change in the ways we organize and employ forces. We must anticipate the implications of new technologies on the battlefield, rigorously define the military problems anticipated in future conflict, and foster a culture of experimentation and calculated risk-taking. We must anticipate how competitors and adversaries will employ new operational concepts and technologies to attempt to defeat us, while developing operational concepts to sharpen our competitive advantages and enhance our lethality. (p. 7)

Excellent and recent articles may also be found on experimentation and prototyping. Simmons (2017) compared deliberate practice with deliberate experimentation through an operational lens in his November 2017 article, “Forget the 10,000-Hour Rule; Edison, Bezos, & Zuckerberg Follow the 10,000-Experiment Rule.” By drawing his comparisons in operational, albeit commercial, contexts, Simmons offers insight on accelerating technology insertion into capability. He notes that practice alone, while useful in static fields of endeavor, is “next to useless in areas that change rapidly, such as technology” (Simmons, 2017). By contrast, experimentation in large numbers brings the power of odds and large data sets to bear on the problem of non-linear technology acceleration. In fact, the undeniable resemblance of Simmons’ experimental success graphic, shown in Figure 1 (Simmons, 2017), to recent portrayals of adversary capability is thought-provoking.

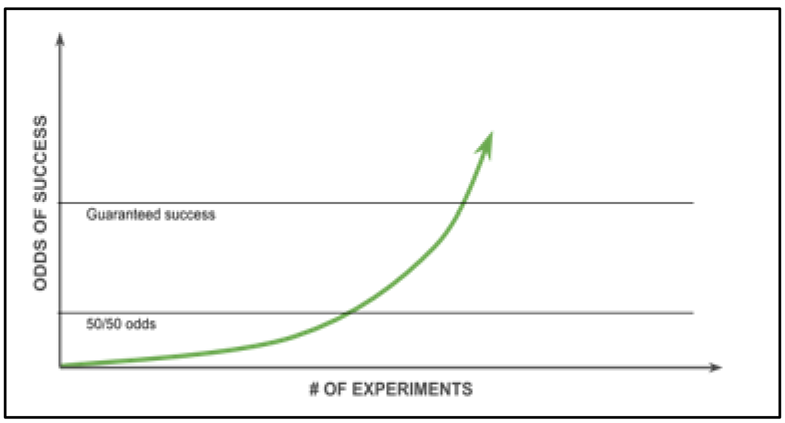


Figure 1. Success Curve
(Simmons, 2017)

The Path to Prototype Warfare (Kozloski, 2017) examines how Cold War industrial and geopolitical environments shaped the DoD’s acquisition system in the 1960s and how those environments differ today. Concluding that technology and international relations have invalidated the DoD’s acquisition model, Kozloski explores an alternative one based on rapid, mission-targeted, research and prototypes rather than monolithic, unwieldy programs of record. Whereas Cold War strategy intentionally avoided surprise by making large fielding decisions based on assured technology predictions and intelligence, today’s strategy must do the opposite. “[Quickly] … equip … with weapons custom designed for a specific … mission,” says Kozloski; increase “the number and type of threats a defending force must consider during battle”; and deliver “promising weapon systems … quickly to the operational forces in limited quantities.” Kozloski then asks pointed questions about why the military does not act on these ideas and proposes that his questions “be part of a debate on reforming the acquisition process and designing the future force" (Kozloski, 2017). He even nearly described the TnTE2 method by asserting that "…rapid deployment of unique weapons would demand that the… military… quickly develop… tactics…" (Kozloski, 2017). Even without the final mental leap to integrate TnTE2, this article is bluntly insightful. The assertion that strategic surprise "is viewed often as a tool of the weaker state" (Kozloski, 2017) should not delay our adoption of it, formally, as a defense strategy. After all, the only thing worse than *becoming* the weaker state would be not recognizing it.

Historical Perspective

We recognized the complex, multidomain warfighting problem presented to our nation in World War II (WWII), and we acted on it with a national fervor not approximated in peace or war since. The 1941 attack on Pearl Harbor galvanized our national will, sacrifice, and, innovative spirit, but Naval Forces had been eagerly learning from the embattled British, and from industry, for years before that. The amphibious craft used during America's march across the Pacific evolved from Floridian inventor Daniel Roebling's hurricane floodwater recovery "amphibious tractor" (amtrac), built less than a decade earlier. Historian Henry Shaw, Jr., writes, "When the first production [Landing Vehicle Tracked] (LVT) rolled off Roebling's assembly line … in July 1941, there was already a detachment of Marines at nearby Dunedin learning to … develop tactics for their effective use" (Shaw, 1991). Shaw exactly captured TnTE2's exploratory spirit here, adding that "no sooner did the LVTs make their appearance in significant numbers than the thought occurred that the tractors could be armed and that they could have a role as an assault vehicle, leading assault waves" (Shaw, 1991).

Air power saw tremendous innovation as well. Marine observers from captains to colonels visited British air stations throughout 1941, studying "the weapons and equipment being used and the tactics and techniques being practiced." What we learned about air control and weapons and RADAR, to say nothing of carrier-based aviation in the Pacific, expanded air power's mission from strictly surveillance into an offensive game changer.

Let's not ignore lessons from the other side either. German Blitzkrieg tactics, developed only a generation earlier, had become one of the most effective maneuver methods in history. Small, nimble, lethal units wreaked havoc on Allied troops by moving quickly, aggregating and disaggregating in precise time and space to inflict specific damage. They stayed in touch, and alive, using radios that were simplistic but faster than Allied trucks and horses.

The atomic bomb, infamous for its destructive power, was also an inspiring example of operating force–scientist teaming on a just-in-time, rapid prototyping effort which set the conditions for Japan's surrender and, ultimately, for Allied victory. Richard Rhodes' *Making of the Atomic Bomb* documents this historic operational-technical collaboration in an enjoyable read (Rhodes, 1987; see Figure 2).

Figure 2. Manhattan Project Scientists and Military Personnel Gathered Around the Bomb Pit, Ready to Watch the Little Boy Bomb Being Loaded Into the *Enola Gay*
(Rhodes, 1987)

In the decades following WWII, the U.S. Navy matured into a highly trained, precision firepower, sea-based force. Innovation continued more slowly but no less impressively. Submarine-launched Polaris missiles, nuclear propulsion, modern aircraft carriers, space-based communications, and the Aegis weapon system all capitalized on late 20th-century technology with (Cold) war-winning effect.

Between 1989 and about 2010, we practiced and refined the methods and technology that put us on top despite an increasingly deliberate and laborious defense acquisition system (DAS). Meanwhile, factors like the post–Cold War “Peace Dividend” and a shift in focus from peer competitors to combatant commander hotspots stretched U.S. Naval forces thin. Figure 3, taken from a recent *Strategic Readiness Review*, starkly depicts the reduced force and increased operation tempo of the current Fleet (DoN, 2017).

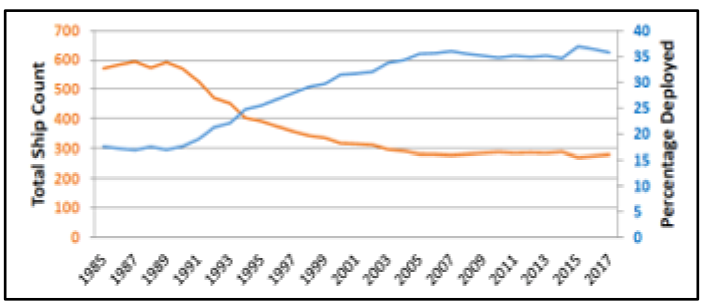


Figure 3. The Dramatic Rise in Percentage of the Total Force Deployed: 1985–Present
(DoN, 2017)

Very recently, in the last three to five years, honest self-assessment and glide slope comparisons with China and Russia have created the sense of urgency captured in the CNO’s *The Future Navy* (Richardson, 2017) and the *Marine Corps Operating Concept* (USMC, 2017). Naval component commanders are revamping Concepts of Operation (CONOP), experimenting with new warfighting constructs like Distributed Maritime

Operations (DMO) and Expeditionary Advanced Base Operations (EABO), and prioritizing underutilized mission areas like Information Warfare (IW).

During this time of renewed *operational* innovation, Secretary Stackley's aforementioned Congressional testimony brought clarity and focus to the Navy's renewed focus on *technical* innovation. In that January 2016 testimony, he identified four key enablers to accelerate acquisition (Stackley, 2016). The complete list shown here became the four cornerstones of acquisition reform as we know it today. TnTE2, described in full detail later in this paper, now emerges as the executable method to build on those cornerstones.

"We want to 'learn fast' through prototyping—completing projects as rapidly as possible and certainly within 24 months of project selection—to improve follow-on system acquisition decisions before incurring significant costs ...

Key Enablers:

1. *leveraging the breadth and depth of technical talent and facilities from across the Naval Research & Development Establishment (NR&DE)*
2. *active and continuous engagement by our NR&DE with fleet forces including the Warfighting Development Centers*
3. *designing our major weapon systems for rapid technology insertion*
4. *funding expressly for rapid prototyping, experimentation, and demonstration"*

ASN(RD&A) Testimony on Acquisition Reform: Experimentation and Agility, January 7, 2016

Figure 4. ASN(RD&A) Testimony on Acquisition Reform
(Stackley, 2016)

Previous acquisition reform efforts, such as the Better Buying Power initiatives, all had the effect of tinkering with the current DAS, adding complexity, rigor, work content, schedule, and cost. In 2016, however, the threat from peer adversaries motivated leaders Capitol Hill, acquisition, and operational communities to pursue "skip-gen" approaches. Prototyping how we do prototyping, the DoN launched a series of dedicated, task-force-like initiatives focused on Smart Mining (SMI), Counter small Unmanned Aerial Systems (CsUAS), Unmanned Systems (UxS), and Advanced Combat Systems Technology (ACST). The creation and branding of the Naval Research and Development Establishment (NR&DE) marshaled the collective and collaborative efforts of the 15 naval labs and warfare centers and the Office of Naval Research (ONR) behind these initiatives in support of Secretary Stackley's accelerated acquisition strategy. Led by the Deputy Assistant Secretary of the Navy for Research, Development, Test and Evaluation (DASN[RDTE]), the NR&DE designed and developed the methodologies and committed dedicated experts to rapidly form the Fleet/Force Engagement Teams (FET), which explored tactics and technology with operational forces.

In parallel with these DoN efforts, the United States Marine Corps (USMC) released the Marine Corps Operating Concept (MOC; USMC, 2017), which articulated a new central warfighting problem: the future operating environment was changing exponentially. Even in 2016, said the Marines, the confluence of a technology explosion, conflicted littorals, complex terrain, adversary use of the information environment and exploitation of intelligence, surveillance, and reconnaissance systems created increased challenges to naval and littoral maneuver. Recognizing the USMC was at a significant, strategic, inflection point as a service, leadership focused increasingly on challenges and opportunities that would define the future, rethinking the methods, tools, and processes that inform investment decisions. They wanted to accelerate the speed at which USMC develops, integrates, and deploys future systems, and they wanted to co-evolve tactics development and integration. Understanding the complexities of amphibious warfare operating spaces was a daunting task. USMC leadership was searching for new approaches and technologies that would enable real world experimentation and revolutionize warfighting approaches to peer threats. These approaches included artificial intelligence, manned and unmanned teaming, 21st-century combined arms leveraging the information environment, enhanced littoral maneuver and enhanced intelligence, surveillance, and reconnaissance (ISR) and counter ISR. These approaches were applied to five capability concepts:

- Ship to Shore Maneuver: clandestine, overt maneuver from varied distance to support reconnaissance, assault, and heavy material
- Amphibious Fire Support and Effects: finding targets and engaging targets; emphasis on long range precision against moving targets
- Clear Amphibious Assault Lanes: identify and clear mines and obstacles; beach reconnaissance and survey
- Amphibious C4: secure resilient and robust communications; common operational picture; precision navigation and targeting
- Amphibious Information Warfare: corrupt the enemy perception with decoys, electronic support, and attack

S2ME2 Task Force

Having both recognized a strategic transition point, USMC and DoN leaders empowered a task force to accelerate the exploration and experimentation of advanced tactics and technology for 21st-Century Ship to Shore Maneuver warfare. The USMC deputy commandant, Combat Development and Integration (DC CD&I), and DASN(RDT&E) chartered the Ship to Shore Maneuver Exploration and Experimentation (S2ME2) Task Force on August 24, 2016 (DoN, 2016a). This task force applied the TnTE2 methodology, rapidly shepherding a balanced team of warfighters and technologists through an iterative series of capability-focused rapid prototyping and experimentation cycles. Chartering members provided the operational imperative and mission focus, and they empowered an experienced team of warfighters and technologists to set new standards for speed, scale, and rigor.

The establishment and structure of a task force was critical to the overall strategy. The S2ME2 Task Force, simply referred to as the “Task Force” for the remainder of the paper, primarily consisted of operational and technical subject matter experts from DASN(RDT&E), Headquarters Marine Corps Combat Development and Integration (CD&I), Marine Corps Systems Command (MCSC), Marine Corps Warfighting Lab (MCWL), and the labs and warfare centers of the NR&DE.

Naval Technology Exercises

The S2ME2 Task Force expanded upon the concept of an Annual Naval Technology Exercise, which had been pioneered by the Naval Undersea Warfare Center (NUWC) Division Newport in August 2015. NUWC Division Newport leadership leverages their Naval Innovative Science and Engineering (NISE) program to host an annual event that provides government and industry participants access to a collaborative, low-risk environment to demonstrate technologies across NUWC Division Newport labs and ranges. Recognizing the tremendous, unique, intrinsic potential of NR&DE-wide resources and facilities, DASN(RDT&E) subsequently promoted Advanced Naval Technology Exercises (ANTX) as key events in the early exploration phases of TnTE2. Operational Force–championed ANTXs are distinguished by mission focus, team structure, and full employment of TnTE2 methods to evaluate a highly complex tactics and technology trade space. A detailed description of the TnTE2 methods is the focus of the section titled "TnTE2 Within S2ME2." A summary of Annual Naval Technology Exercises and ANTXs that have been executed, or are in planning phases, across the NR&DE labs and warfare centers are included as the appendix.

In just seven months, more than 130 emerging technologies were evaluated, and 52 highly-valued capability prototypes were demonstrated for naval operators, scientists, and engineers during a two-week exploration event at Marine Corps Base Camp Pendleton. This was the first-ever Advanced Naval Technology Exercise (ANTX), and it set benchmarks and best practices for implementing the TnTE2 methodology and ANTX constructs across the entire department and throughout the NR&DE. Unexpected findings included novel concepts such as Proximity: Unmanned Systems can achieve tactically relevant proximity with persistence and low signature. They can effectively maneuver far forward of manned units with very low risk in order to sense and provide effects. A few promising "skip-gen" type technologies were left with operational forces for extended user evaluations. The most extensive integration resulting from S2ME2 ANTX has been the collaboration between ONR's Information Support to Operations office and the 1st Light Armored Reconnaissance Battalion. By October 2017, 14 prototypes, organized around six highly-valued mission threads, were integrated into Fleet/Force Experimentation venues including BOLD ALLIGATOR 17 and DAWN BLITZ 17. S2ME2 Task Force efforts directly initiated more than 30 follow-on rapid prototypes and S&T accelerations, many of which are scheduled to culminate with Fleet/Force experimentation events in fiscal year 2018.

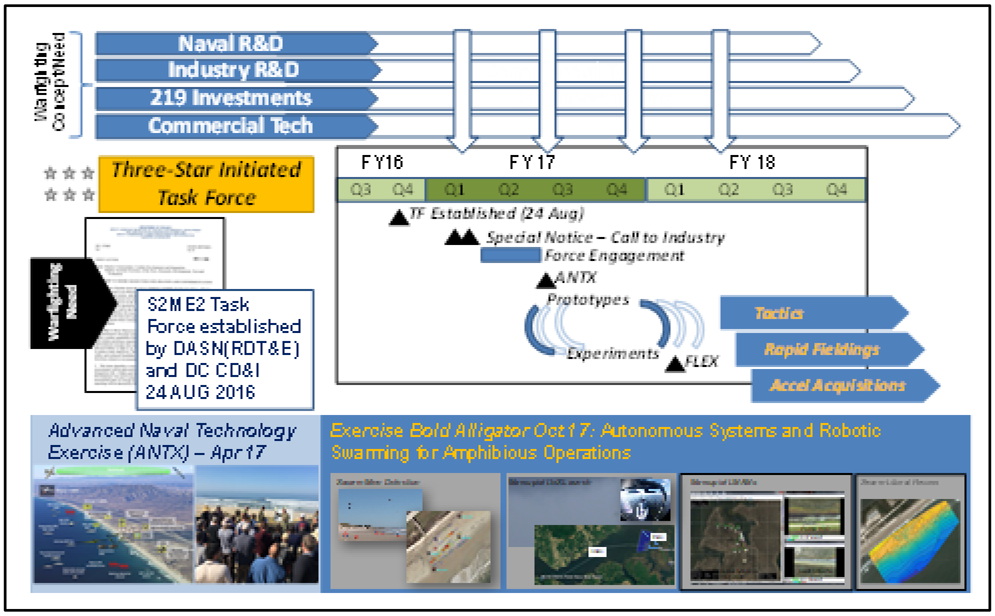


Figure 5. S2ME2 Task Force Strategy

The complexity of the solution space and the speed at which leadership demanded execution and recommendations required a synchronized effort from warfighters *and* technologists. The complexity of the solution space was generally defined by the following:

- **Tightly coupled and multi-domain operating environments—**requires our most innovative operators and planners to explore warfighting capabilities and alternate tactics
- **Technology options are significant in number and diversity—**requires our most innovative scientists and engineers to explore the applicability and limitations of emerging technologies
- **Technology and tactics pairings are significant in number and diversity—**requires BOTH our most innovative operators and technologists to rapidly explore, iterate, and assess technology and tactics pairings

The S2ME2 Task Force employed TnTE2 methods through a series of capability-based workshops, which led up to field demonstrations, exercises, and experiments. Throughout the course of these rapid iterations, the operational and technical experts explored the solution space together, assessing emerging capabilities from their respective areas of expertise. The method effectively unleashed the intrinsic ingenuity of our practicing operators, planners, scientists, engineers, and industry partners, and it leveraged the full potential of our Naval laboratory infrastructure to "reestablish agile experimentation at appropriate levels (seen as) critical to achieving a strategic innovation advantage" (DoN, 2015).

S2ME2's operational team contributed expertise in the specific emerging concept of operations, scenarios, and related tactics of interest. "Delivering capability at the speed of relevance" (DoD, 2018, p.10) demanded mission focus and expertise from the operators. These operators were "current" and empowered by leadership to advance tactics for the high end fight. For S2ME2, the Operational Champion was a senior general officer with sufficient experience and authority to set priorities. This is considered a critical element for all TnTE2 implementation.

S2ME2's technical team of practicing scientists and engineers provided expertise related to the specific technologies and their potential applicability and limitations in the relevant environment. While the S2ME2 Task Force did not require long-term reassignment, it established a core team and prioritized the efforts of an extended team of experts for the duration of the Task Force. For example, expertise was resourced from across the NR&DE and USMC commands for dedicated periods of time for Force Valuation workshops, planning workshops, ANTXs, and Fleet/Force experiments.

The underlying thesis for S2ME2 was that mature and emerging technologies from the DoD, the DoN, and industry had been largely untapped. Many mature prototypes inspired better and/or alternate tactics for amphibious ship to shore maneuver. By studying and in many cases, leveraging significant prior S&T and R&D investments, the Task Force quickly and cost-effectively characterized the tactics and technology trade space and informed:

- S&T, R&D, and rapid prototyping investments
- Industry and NR&DE internal investments
- Concept and requirements development
- Acquisition plans

Extensive participation by small, large, traditional, and non-traditional DoD and industry partners was encouraged by a Special Notice promulgated on the Federal Business Opportunities website (https://www.fbo.gov/) and socialized through various public websites and industry forums. Overall, more than 48 R&D organizations participated in S2ME2 ANTX through their current contracts or new cooperative research and development agreements (CRADA) with the appropriate NR&DE lab or center. More than 75 operational assessors from over 10 Fleet and Force organizations participated in the Force Valuation workshops, planning, and ANTX and FLEX events, providing more than 760 technical and operational assessments that informed leadership decisions.

The S2ME2 Task Force has fostered innovation and accelerated development of complex warfighting capability in ways that informed the tactics as much as the systems that were prototyped. The Task Force directly informed the development of the Marine Corps Operating Concept (MOC; USMC, 2017) and emerging Ship to Shore Maneuver concept of operations. USMC leadership recognized the potential of TnTE2 as an innovative process to integrate tactics and technology evaluation, initiate follow-on prototyping for experimentation, further technology development, and inform decisions to move into rapid fielding. USMC leadership also considers S2ME2 as a rapid prototyping, experimentation, and demonstration (RPED) project and will continue employing these methods to accelerate capability development and acquisitions (DoN, 2016b).

Throughout S2ME2 execution, DASN (RDT&E) worked with Congress, the Office of the Secretary of the Defense (OSD), the DoN, and Naval Systems Commands to explore underutilized financing, acquisition, and contracting authorities. This "Fast Lane" effort, further described below, is aimed at effectively scaling up the successful prototyping effort and applying new accelerated acquisition authorities and tools to speed capability to the field. Another iteration of TnTE2, focused on the "Urban 5th Generation (U5G) Marine" is in progress at the time of this writing, and planning has begun for a third iteration focused on Expeditionary Advanced Base Operations (EABO) and Information Warfare.

At present, in early 2018, the authors sense a tipping point, either just ahead or just behind us, beyond which our most senior leadership view accelerated acquisition, based on TnTE2, as an executable method of preparing Naval and Joint forces for high-end conflict.

Recent guidance from the 2018 National Defense Strategy reinforces the authors' belief that such combined action between the Fleet/Force and the NR&DE is not only desired, but essential to delivering "a more lethal, resilient, and rapidly innovating Joint Force" (DoD, 2018, p. 1).

It is believed the strategy and methods synthesized by the S2ME2 Task Force constitute the first comprehensive response to recent DoD and DoN calls for faster capability fielding. The remainder of this paper will describe TnTE2 in detail, including specific S2ME2 implementations, how it has evolved since 2017, and remaining work. We will also describe additional, "Fast Lane," Accelerated Acquisition (AA) work, both planned and in progress, and suggest ways for readers to get involved.

TnTE2 Within S2ME2

TnTE2 methods, as applied during the S2ME2 and U5G Task Forces, have proven successful in rapidly identifying highly-valued capabilities, informing developing warfighting concepts, and initiating rapid prototyping lines of effort. Set-based design (SBD) principles and decision methodologies, as expanded by the DoN for the Small Surface Combatant Task Force (Garner et al., 2015) and Advanced Combat Vehicle (ACV; Burrow et al., 2014), allowed for the full exploration of complex technical *and* operational solution spaces.

The TnTE2 method can be generalized into four interrelated major elements, as shown in Figure 6.

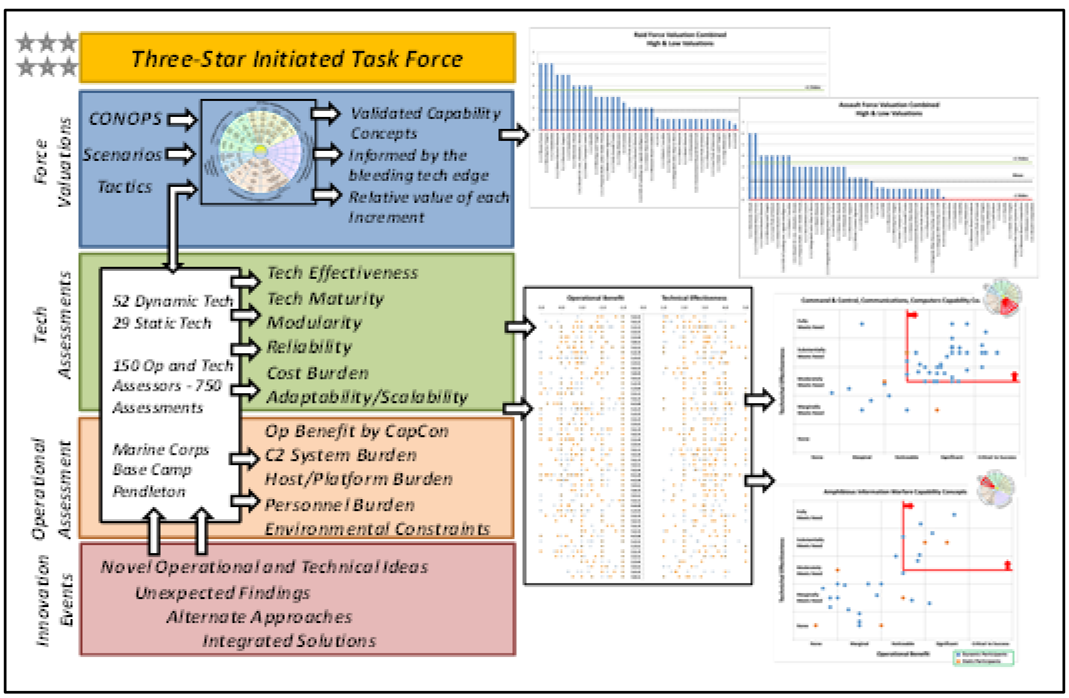


Figure 6. Tactics and Technology Exploration and Experimentation (TnTE2) Method

It is important to note that the metrics and data depicted in Figure 6 were captured at a specific point during the S2ME2 ANTX. The method and major elements are applied iteratively throughout the TnTE2 process. Leadership is continuously informed by additional and more quantitative data gathered throughout the assessment process, (i.e., capability increments, “tech bank,” assessment results, databases, systems engineering artifacts, architectures, cost estimates), which increases fidelity and confidence in decision-making.

Major Element: Force Valuations and the Capability-Based Framework

The development of a capability-based framework is foundational to the TnTE2 method. The Capability Concept Wheel (CCW) is the capability-based framework that translates concepts of operations into capability areas and then into increments of capability. Increments of capability include definitions and sufficient parameterization such that an operator can assess the relative value of discrete capability increments and a technologist can identify specific technology solutions. The CCW is a shared framework that must be developed, acknowledged, and maintained by both the operational and engineering community. It becomes the primary communications tool used throughout wargaming, tactics development, and the series of iterative and progressive rapid prototyping, experimentation, and demonstrations that ensue.

The CCW is essentially a table that includes underlying definitions and parameters. It is typically presented as a wheel with increments of capability of increasing complexity along the radial axes. Capability increments can be cumulative or exclusive, and differences are typically denoted with asterisks to guide the user. Joint publications are typically referenced, but in cases of multi-domain or joint efforts where lexicon differs among warfighter communities, the framework establishes a common lexicon and definitions that are agreed upon for the duration of the Task Force’s efforts. The S2ME2 CCW was developed over the course of two four-day workshops and validated during a Force Valuation dry run. Helpful hints are provided to operational and technical subject matter experts throughout the efforts of a Task Force. Hints specific to CCW development are shown in Table 1.

Table 1. Hints for Capability Concept Wheel Development

Transform Emerging Warfighting Concepts *into* Capability Areas *and then into* Increments of Capabilities
Helpful Hints: • Thinking dominated by needs and capabilities • Stay clear of discussing proposed solutions • Any solution discussions should be limited to refining increments of capabilities • Multiple solutions and combinations of solutions will be developed and analyzed

The major capability areas of the S2ME2 CCW are shown in Figure 7.

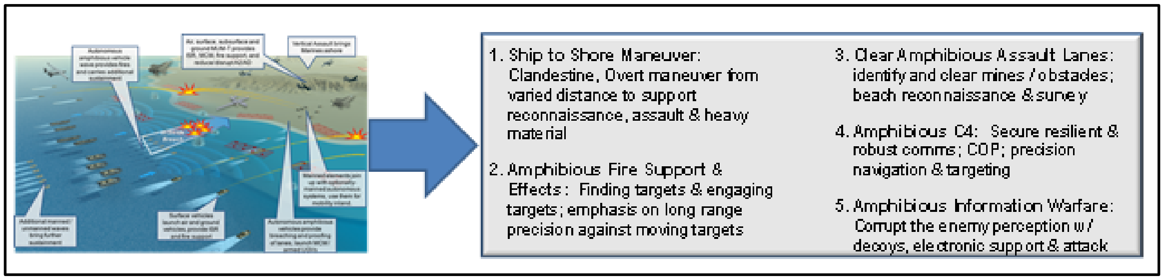


Figure 7. S2ME2 Capability Areas

It is imperative that a CCW provide coverage for the entire trade space of tactics and technology options. Initial workshops to develop the CCW are a mix of facilitated brainstorming and wargaming sessions where operators are asked to articulate capabilities a planner or operator would employ during specific scenarios. In the context of TnTE2, a capability must be technology or solution agnostic and *strictly* defined as the ability to perform or achieve certain actions or outcomes. Capability examples might include abilities like: sensing the spectrum, where increments of capability span frequencies and/or ranges; or engaging kinetically, where increments of capability span ranges and/or effects. Allowing capabilities and/or increments to be limited to a specific technology or solution will stifle innovation, as there may be more than one technical approach to achieving a specific capability.

Force Valuation workshops are typically executed over the course of a week where teams of planners and operators play a series of scenario-based tabletop wargames that provide insights into which capabilities the operators value most for a given scenario and mission. Care must be taken when scripting scenarios, so as to not limit tactical creativity or preordain solutions. Multiple scenarios may be “played” to learn more about capabilities unique to specific areas of responsibility (AORs) and/or phases of conflict.

Each team “plays” a given scenario multiple times (i.e., Game 1, 2, … n) and with decreasing resources, as shown in Figure 8. Operators are asked to place their resources on the specific increment of capability they value most in executing the mission. “Pulling rank” is discouraged, and teams must collaborate to converge on a single set of capabilities required to execute a mission plan. Each team is carefully manned with planners and operators of various experience and expertise. Discussions regarding specific technologies may aid in developing the CCW and during the Force Valuations, but the fact that the CCW is entirely capability-based is critically important. Teams may be shuffled and scenarios and games repeated to provide more fidelity to the capability valuations. Plays are captured with “chips” or “thumbtacks” on the CCW, and statistical analysis provides a summary capture of the results for leadership review and validation.

Flag or general officers should validate the results from these valuations because these results directly inform the ensuing prototyping and experimentation campaign and S&T investments. Data gathered from Force Valuation workshops provide decision-makers with insight into the relative value of each increment of capability for a given mission and scenario. If results are anticipated to vary significantly between theaters, then multiple Force Valuation Workshops may be conducted. Results may be normalized, but retaining the ability to visualize unique insights from different phases, scenarios, and AORs is important.

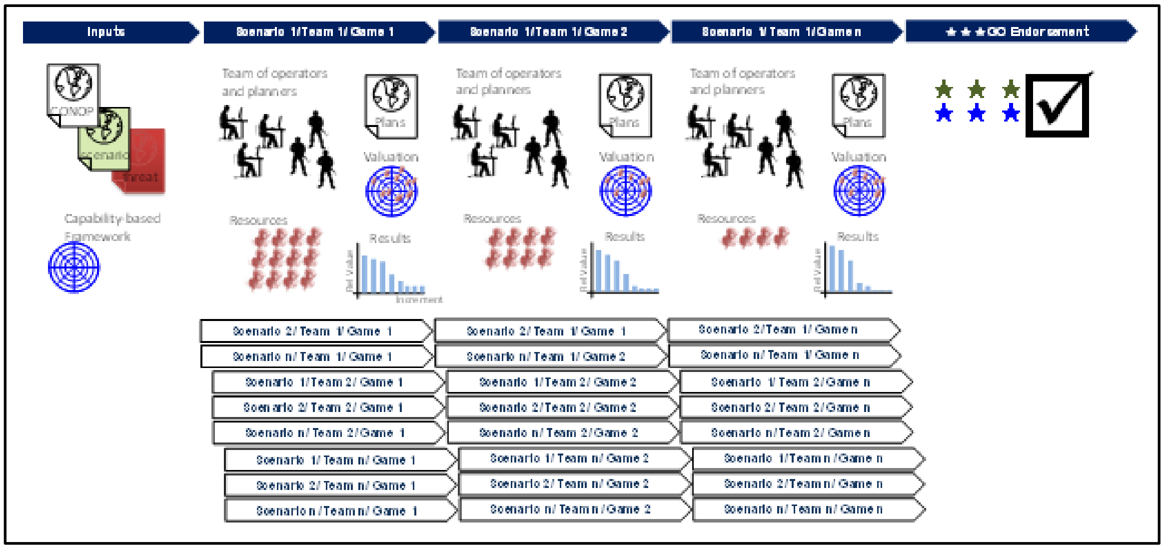


Figure 8. Force Valuation Process

Force Valuation workshop outputs are typically visualized as histograms that show Capability Increments ranked according to their relative value in the relevant mission scenarios. A typical set of mission scenarios, selected by the operational leadership, includes "most stressing" and "most likely" scenarios at a minimum. For S2ME2, these two scenarios were "amphibious raid" and "amphibious assault." In general, S2ME2 Force Valuation findings demonstrated that capabilities required to plan and execute an amphibious raid differed greatly from those required in amphibious assault. This prompted a leadership decision to focus the majority of S2ME2 ANTX mission-threads on evaluating alternate tactics and technology pairings to help advance the raid forces. Force valuation results also demonstrated that operators and planners employed many diverse capability sets. No single capability would completely enable the emerging tactics. These discoveries further justified the need for a deliberate, low-barrier-to-entry, exploration phase that was loosely scripted to allow for rapid iterations of tactics and technology pairings.

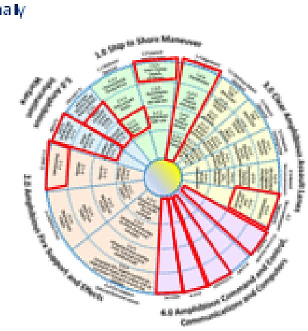
Major Element: Technical and Operational Assessments

Simultaneous to the execution of the Force Valuation workshop, a "Special Notice" was posted to the Federal Business Opportunities website and distributed through other various public websites and industry forums. The S2ME2 CCW formed the basis of a call to industry, academia, and government labs for emerging technologies and/or engineering innovations. The capability areas, increments, and sample metrics were fully articulated in Section III of the Special Notice (SPAWAR Systems Command, 2016), which emphasized a desire for responses offering mature and emerging technologies that could be fielded within 18 to 24 months. Submitters were asked to identify the specific increment, or increments, of capability that a particular technology was capable of providing. Submissions were maintained in a government-owned relational database known as the "Tech Bank." Scientists, engineers, and subject matter experts then binned the technologies into the CCW increments and conducted technical assessments against categories that included technical maturity, integration readiness, reliability, standardization, etc. Operators, planners, and subject matter experts then assessed the technologies against categories, which included operational relevance, personnel burden, environmental constraints, etc. It should be noted that assessments and/or weighting criteria used to evaluate offerors' technologies will vary depending on the urgency and/or specifics of the operational imperative, which will vary with different study areas.

More than 130 emerging technologies or engineering innovations were submitted in response to the S2ME2 Special Notice, providing a variety of technology solutions for the majority of the capability increments. Ensuring the CCW was wholly capability-based reduced the urge to presuppose technical areas and solution spaces and preserved the entire trade space. For example, both swarming UUVs and post-processed surface search radar returns were able to provide bathymetry predictions in an amphibious environment. In fact, at the time of the Force Engagement Team (FET) workshop, only one of the highly-valued capability increments had less than two technology options. In general, highly valued increments of capability that garner few to zero submissions are typically fed back into the S&T and/or R&D communities for action.

Technical and operational assessments during S2ME2 were iterative and repeated to gather more detailed and quantitative data as the Task Force progressed through the series of workshops, exercises, and experimentation events. Leadership was continuously informed by additional quantitative data gathered throughout the assessment process. Sample outputs from more than 50 dynamic and almost 30 static technologies assessed at S2ME2 ANTX are depicted above in Figure 6. During the ANTX, each technology was grouped into one or more of six exercise teams. Teams Shield, Spear, Dagger, Cutlass, Broadsword, and Battleaxe were formed around mission threads shown in Figure 9. Team Battleaxe, the sixth team, focused on the persistent aspects of amphibious command, control, communications, and ISR that could be applicable to both raid and assault scenarios.

- Early Intelligence Preparation of the Battlespace, Reconnaissance – Team SHIELD
  - Unmanned surface vehicles, deliver swarming unmanned surface and underwater vehicles
  - SIGINT systems, COMINT payloads
  - Tailored processing, exploit, and dissemination from commercial satellites; high-resolution elevation models
  - GPS denied position, navigation, and timing
- Threat Identification – Team SPEAR
  - Cooperative UAVs and USVs; cross-UxV command and control; UUV with high-resolution camera; optionally manned jet-ski with communications relay payload
  - High-bandwidth LPI/LPD line of sight E-band communications
  - Real-time wave characterization for improved unmanned vehicle launch and recovery ops at-sea
- Reconnaissance and Threat Elimination – Team DAGGER
  - Manned Jet-Ski, UUVs for recon and mine detection
  - Diver sonars, underwater comms, and Multi-Utility Tactical Transports for obstacle marking and removal
- Maneuver Ashore – Team CUTLASS
  - High Speed and swarming USVs for ISR and fires support
  - Remotely operated AAVs; direct fire weapons; UAVs and 3D printed kinetic quadcopters
  - UAVs; Mesh datalinks and LPI/LPD communications nodes
  - ISR and Multi-INT data fusion tools
- Combat Power Ashore – Team BROADSWORD
  - UGVs for ISR and fires support; UAVs and UGVs for combat resupply
  - Expeditionary 3D manufacturing module
- Amphibious Command, Control, Communications and ISR – Team BATTLEAXE
  - High bandwidth LOS comms; 4G LTE tactical networks; man-portable and expeditionary common operating pictures



Note. Team Broadsword's "capability concept" is highlighted in red.

Figure 9. S2ME2 Teams by Mission-Threads

A "capability concept" is the combination of a valued set of capability increments *and* an associated concept of operations. A capability concept is solution agnostic and may include many different technology options, and permutations thereof. Team Broadsword's associated capability concept is highlighted in red in Figure 9. Force Valuations, described above, provide insight into the highest valued increments of capability. Technical and operational assessments provide insight into the most promising emerging technologies. The combination of this data is used to prioritize exploration efforts and inform near-term rapid prototyping investments. It should be noted that a capability concept is not intended to fully define a system, or systems, as in the form of a "detailed specification." Once technologies are identified for follow-on rapid prototyping and experimentation efforts, systems engineering effort and expertise are required to fully decompose the system of systems required for end-to-end integration of a given capability concept.

Major Element: Innovation Events

Innovation is fostered throughout the life of the Task Force, enabled primarily through the expansive characterization of a capability-based trade space where technology and tactics are fully characterized *and* the direct and continuous interaction between our Fleet/Force operators and planners and practicing scientists and engineers. This second element is so foundational to our strategy and to our methods that it was highlighted as the second key enabler of acquisition reform through experimentation and agility (Stackley, 2016).

Fleet/Force operators and planners, like those from our warfighting development centers and current or recent deployers, must be "current" on advanced tactics, and they must be empowered by an Operational Champion to develop them within specific missions and scenarios. Practicing scientists and engineers should be experts in the "state of the art" with a solid grasp of technical capabilities and limitations of the "bleeding" technological edge. Helpful hints specific to promoting innovation throughout the various phases of effort are provided in Table 2. For example, during a FET workshop, experts are asked to articulate new and innovative capabilities and technological solutions. During an ANTX, experts are provided a loosely scripted event where new and innovative technology/tactics pairings are captured and assessed. Facilitating these "direct and continuous interactions between our scientists and engineers and the Sailors and Marines they support will drive innovation and more importantly, align technical ideas with operational needs at the earliest stages in prototype development and experimentation" (Stackley, 2015, p. 5).

Table 2. Hints for Promoting Innovation

Helpful Hints:
▪ Thinking dominated by new and innovative views of needs, capabilities and solutions
▪ Think deep throws through incremental experimentation plans
▪ Seek vignettes that apply to many CapCons
▪ Develop and analyze multiple solutions and combinations of solutions

SBD principles are applied to the decision-making that occurs throughout the TnTE2 continuum.

- Full exploration of the trade space is enabled by the capability-based framework and full and open calls to government and industry partners.
- Progressive and iterative series of prototyping and experimentation events are used to gather more qualitative and quantitative information.
- Expansive consideration of technology/tactical pairings are characterized and evaluated prior to making decisions to proceed to the next phase.

SBD principles and decision-making methodologies are applied to ensure that trade spaces are fully characterized with dense and diverse sets of tactics and technology pairings. Force valuations, technical and operational assessments, and visualization tools provide leadership with the ability to make rapid data-driven decisions. The TnTE2 continuum is shown Figure 10. The TnTE2 continuum includes exploration, rapid prototyping, Fleet/Force experimentation, and rapid fielding phases. The CCW, or the "framework," is continuously matured and developed, informed throughout the continuum by innovations in tactics and/or disruptive technologies. Even as initial decisions are made, the framework and solution trade space can be updated and reemployed as technologies, threats, and budgets change.

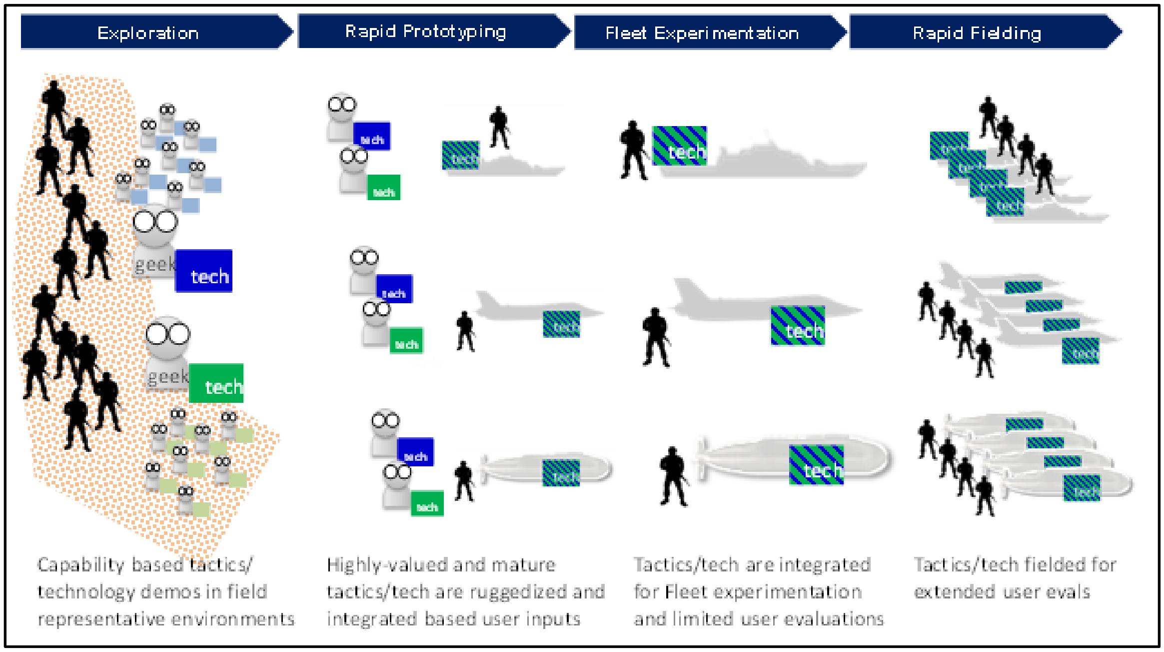


Figure 10. TnTE2 Continuum

For S2ME2, the exploration phase was realized through the first-ever TnTE2 ANTX event. As described above, ANTXs are a very low-barrier-to-entry event where technologists demonstrate their technologies with operator participation and observation in a field or simulated environment. Speed and scope are prioritized; full integration is of lower priority. Technologies are demonstrated to operators in choreographed mission threads to help them assess the potential impact of the technology on advanced tactics. In many cases, advanced sensors or algorithms are demonstrated on COTS platforms that are not intended for fielding or acquisition. In these cases, the analysis and assessment must remain focused on the contribution of the specific emerging technology, not the platform. Even throughout these early stages, balanced and focused partnership with operational experts is critical. Emerging concepts of operations and tactics must be informed by the technological art of the possible, and technological solutions must be informed by emerging concepts of operations. S2ME2 Task Force findings confirmed observations from *The Politics of Naval Innovation* that a single, "new technology has not revolutionized naval warfare. … It was the final integration of several technologies," most already existing that were simply applied in new ways that included: "1) synthesis—new combinations of existing technologies, 2) a keystone—a missing link for a new ensemble of technologies, or 3) tactical innovation—new uses for existing forces" (Lautenschlager, 1983, p. 50).

TnTE2 methods and an exploration phase of significant scope and scale, such as realized through ANTXs, are the catalysts that ignite an extensive campaign of rapid prototyping, rapid fielding, and accelerated acquisitions. Highly-valued technologies are selected to proceed to rapid prototyping, experimentation, and fielding phases for extended user evaluations. Concepts of employment are continuously evolved throughout the process and tested during these phases. Operators refine the concepts of employment and inform technology refinement while technologists are documenting top-level requirements, interface requirements, architectures, and cost data. Tight integration of operational and technical experts, documentation, and learning is paramount. The S2ME2 Task Force focused "on testing out ideas and concepts first and getting fast feedback through rapid prototyping" (DoN, 2016). USMC and ASN(RD&A) are embracing these methods as the new paradigm for rapidly identifying and integrating new technologies into military tactics, informing Naval Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, and Facilities (DOTMLPF) solutions, and guiding future Navy and/or Marine Corps acquisitions.

Remaining Work

TnTE2 tools and methods continue to be refined, and planning has begun for future initiatives and Task Forces focused on other Navy and Marine Corps mission priorities. The areas provided below have been identified as opportunities to expand the impact of TnTE2 methods and to ensure more confident transitions.

- Funding expressly identified to support rapid prototyping, experimentation, and demonstration efforts which may be allocated within budget execution years based on data provided by open, merit-based methods such as TnTE2
- Direct and deliberate implementation of alternate financing vehicles, such as prize challenges, multi-award contracts, Other Transactions (OTs; DAU, 2017), throughout the TnTE2 continuum to enable rapid financing actions at the various decision points
- Improvements in NR&DE and DoN processes and databases such that the framework and solution trade space, aka the "Tech Bank," can be updated and reemployed as technologies, threats, and budgets change.
- Automation and improved visualization of the technical and operational assessments made during ANTXs to enable even greater decision speed
- Strategies and methods for rapid security classification of many-to-many technology and tactics pairings
- Direct and deliberate use of TnTE2 methods in other wargame series, such as USN and USMC Title 10 wargames
- Improvements to the NR&DE live, virtual, and constructive (LVC) ranges for concept exploration through experimentation events
- Efforts to expedite integration and installation of prototypes into Fleet/Force tactical systems
- Direct and deliberate integration of TnTE2 with acquisition and POM planning efforts

The authors continue to seek opportunities to expand and apply TnTE2 methods to high-priority, emerging, and complex warfighting areas, where the technology and tactics trade space are inherently complex. Planning efforts are underway for the first-ever tri-chartered ANTX with direct involvement of an operational Fleet command. Planned for December 2018, the next ANTX will focus on Information Warfare as it applies to the emerging concept of Expeditionary Advanced Base Operations (EABO).

Evolving Accelerated Acquisition

Since 2015, NR&DE's FET-led efforts like SMI, CsUAS, UxS, and ACST and more recent S2ME2 and U5G Task Forces have laid important groundwork for TnTE2. S2ME2 and U5G Task Forces embraced the full potential of a unified partnership between operators and technologists. As a result of these recent Task Forces, TnTE2 methods identified impactful, mature technology candidates and initiated several rapid fieldings. While the DoN was leaning forward with TnTE2 methods and agile teaming constructs, Congress initiated the largest acquisition reform since the 1986 Goldwater–Nichols Act. Although our traditional defense acquisition system will continue to guide most major defense platform acquisition, the 2016 and 2017 NDAAs authorize alternate pathways to accelerate urgent and component-level acquisition (USG, 2016, 2017).

Middle-Tier Acquisition (Section 804) and Acquisition Agility (Section 806) pathways are depicted in Figure 11. These new authorities recognize that increased use of rapid prototype development and experimentation early in a program's formulation are fundamental to improving acquisition outcomes (Dougherty, 2018; GAO, 2017). Additionally, they allow for agile methods and approaches, such as TnTE2, to be applied throughout the problem identification and definition phases, limited trials, and decision points.

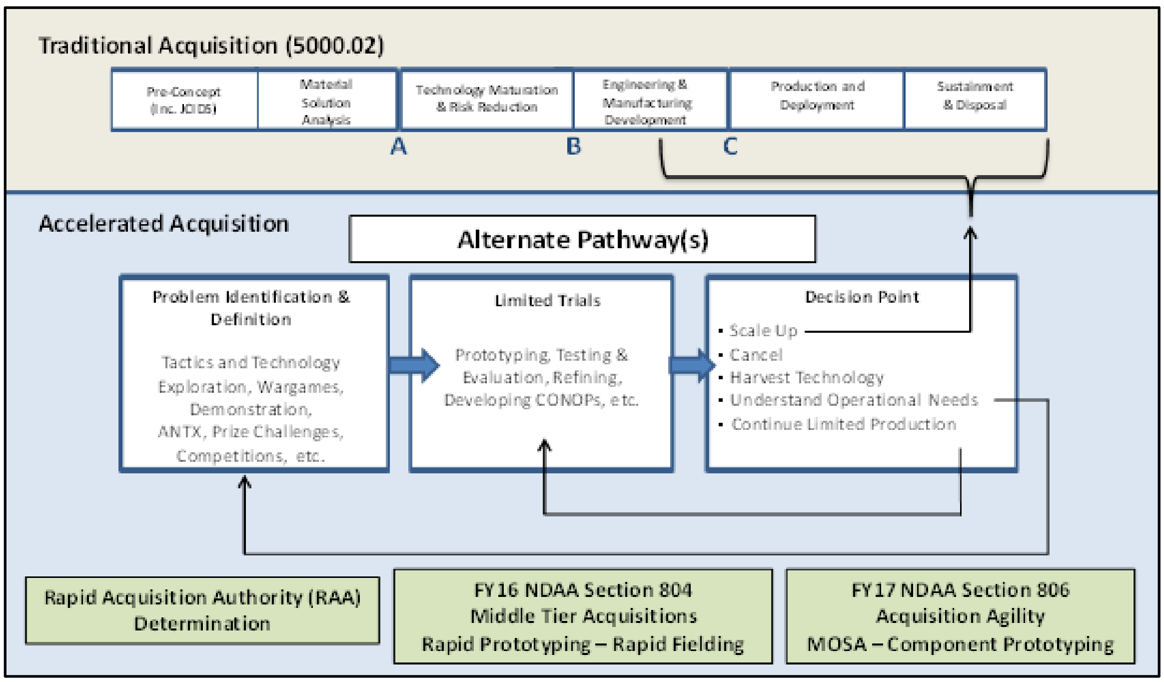


Figure 11. Alternate Acquisition Pathways

Specifically, Middle Tier Acquisition introduced authorities that release certain service urgent programs from constraints in the Department of Defense (DoD) Instructions and the Joint Capabilities Integration and Development System (JCIDS) process. In addition, Middle Tier Acquisitions allow the services to establish a flexible rapid prototyping fund to accelerate programs within the Planning, Programming, Budgeting, and Execution (PPBE) cycles. Acquisition Agility introduced authorities to acquire incremental capabilities by separating "component /technology acquisition" from "platform acquisition" for Major Defense Acquisition Programs (MDAP). This component acquisition allows components and their underlying technologies to be developed through an agile framework. Their host platform(s) requiring substantial investment will remain in the inventory for decades and will be acquired through the established, deliberate, acquisition process. Using these new pathways depends on the services' ability to prototype, experiment, and integrate components seamlessly into host platforms. These authorities could fundamentally alter how we execute MDAPs and provide alternative acquisition strategies to programs leveraging new methods, like TnTE2, which embrace open, competitive, merit-based, and fast acquisition as the rule vice the exception.

The implementing policies for these alternate acquisition authorities are being developed by DoN leadership and experts in the Accelerated Acquisition "War Room," located in Crystal City, VA. The following have been identified as the key enablers to accelerating acquisitions:

- **Organizational Constructs**, which recognize and are comfortable employing new acquisition authorities, pathways, and tools aimed at agility and responsiveness
- **Technical Authority** culture and process change to embrace speed to capability, risk, and uncertainty as sources of resiliency
- Innovative **Contract and Agreement Strategies** to leverage inherent authorities and employ new tools applied to the appropriate level of service
- Innovative **Financing Constructs**, which leverage DoD, national, and global investments in technology
- Modern **Systems Development and Engineering Methods**, such as "agile" and "DevOps" that accelerate and streamline development, test, and deployment
- **Financial Management Methods and Benchmarks**, which encourage agility and nimble financial decisions, obligations, and expenditures within a budget cycle
- **Cost Accounting Regulations and Practices** aligned to the spirit of the Acquisition Agility Act, which are capable of monitoring Platform affordability without limiting incremental increases in capability
- Expanded availability and access to **Fleet/Force Experiments** for extended user evaluation periods
- Strategies and organizational alignment for rapid **Security Classification** determinations
- **Delegation of Authority** to the lowest possible level
- Flexible and agile **Requirements Development** methods that provide acquisition professionals the trade space to deliver technically and tactically relevant capability

Acquisition professionals with expertise in one, or more, of these transition-enabling subject areas are strongly encouraged to visit the Accelerated Acquisition War Room to explore opportunities to contribute to this DoN Strategic Initiative.

A Future Perspective

"Deliver performance at the speed of relevance. Success no longer goes to the country that develops a new technology first, but rather to the one that better integrates it and adapts its way of fighting." (DoD, 2018, p. 10)

While this paper is about a method proven for rapid prototyping and capability development, we must acknowledge the inextricable linkage between operators and the technical and acquisition workforce necessary to institutionalize our ability to deliver capability at the speed of relevance. General McChrystal's experience in Iraq, described in his excellent book, *Team of Teams: New Rules of Engagement for a Complex World* (McChrystal et al., 2015), describes a transformation of linear thinking and organizational charts into a new interwoven structure enabling parallel plans and execution, not at the squad level but across the enterprise. The general's foe bore little resemblance to our peer naval competitors, but the Internet-enabled, shape-shifting environment he encountered has clear parallels to the contested maritime and uncertain fiscal environments we face. If the fight goes to the side who best understands and exploits the environment, and it usually does, then the environment is our "Rosetta Stone." In increasingly complex warfighting environments, adaptability, not control, reigns supreme. It follows that an adaptable structure should form the basis of our acquisition system and strategy. OPNAV strategic planner CDR Frank Goertner recently observed that Russia and China "both appear intent on being first to learn early and learn fast in the operating environment" (Goertner, 2018). Second place has little value in that race.

As technology, peer capability, and the rapidly changing interdependency between them proliferate, how will naval research, development, and acquisition change its structure to pace technology and outpace our adversaries? Taking full advantage of Congressional authorities is just the first step. If accelerating new system fieldings from 10 years to one year satisfies us today, what should satisfy us in the future, and how will that change over time? It takes both experience and imagination to answer that question. It means blending enduring ideas, like deterrence and combined arms maneuver, with new trans-regional complexities like Information Warfare. It requires conceptually refreshing and refitting how we deploy and sustain highly agile, lethal, expeditionary formations, and *that* requires faster decision-making. Faster decisions, at the Enterprise level, require flat, dynamic reporting chains, shared situational awareness, and trust, which can only be developed deliberately and systemically by the *many*, versus accidentally or episodically by the *few*.

Looking ahead, we imagine the ability to download, print, or otherwise create new warfighting capability in the thick of a fight, in minutes or seconds, or fractions of sections. That kind of responsiveness certainly requires some new technology, but it may also require methods like TnTE2, alternate acquisition mechanisms, and agile teaming constructs. Scientists and engineers from across the NR&DE have already begun forming that team with our warfighters and acquisition professionals. Our methods unify developers and operators in the same way that the DevOps culture unified software developers and software operators in commercial industry. The only difference is that our developers know that a mistake, or a delay, can cost the life of an operator, a crew, or a fleet. Scientists and engineers will be critical shipmates in future conflicts because of the rapid pace of technological development.

We are not guided here by what we can prove or plan exquisitely, but simply by what is necessary. Put another way,

> Until one is committed, there is hesitancy, the chance to draw back. Concerning all acts of initiative (and creation), there is one elementary truth, the ignorance of which kills countless ideas and splendid plans: that the moment one definitely commits oneself, then Providence moves too. All sorts of things occur to help one that would never otherwise have occurred. A whole stream of events issues from the decision, raising in one's favor all manner of unforeseen incidents and meetings and material assistance, which no man could have dreamed would have come his way. … "Whatever you can do, or dream you can do, begin it. Boldness has genius, power, and magic in it" [Goethe]. (Hutchinson, 1951)

References

Bernstein, J. I. (1998, May 18). *Design methods in the aerospace industry: Looking for evidence of set-based practices* (Master's thesis). Cambridge, MA: Massachusetts Institute of Technology.

Burrow, J. (1997). *The 21st century surface combatant war room: A review and case study assessment of the war room process*. Blacksburg, VA: Virginia Polytechnic Institute and State University.

Burrow, J., Doerry, N., Earnesty, M., Was, J., Myers, J., Banko, J., … Tafolla, T. (2014). *Concept exploration of the Amphibious Combat Vehicle.*

Defense Acquisition University (DAU). (2017, December 15). *Other transaction authority for prototype projects 10 U.S.C. 2371b.* Retrieved from https://www.dau.mil/acquipedia/Pages/ArticleDetails.aspx?aid=dd749bb2-2ed0-49ca-868e-83f02962ee3e

DoD. (2015, January 7). *Operation of the defense acquisition system* (DoDI No. 5000.02). Washington, DC: Author.

DoD. (2018). *National Defense Strategy of the United States of America*. Washington, DC: Author.

DoN. (2015, October 15). *Increase resources and opportunities for experimentation* [Memorandum]. Washington, DC: Author.

DoN. (2016a, August 24). *Joint letter: Ship to shore maneuver exploration and experimentation.* Washington, DC: Author.

DoN. (2016b, December 22). *Department of the Navy accelerated acquisition for the rapid development, demonstration and fielding of capability* (SECNAVINST 5000.42). Washington, DC: Author.

DoN. (2017, December 11). *Strategic readiness review*. Retrieved from http://s3.amazonaws.com/CHINFO/SRR+Final+12112017.pdf

DoN, Headquarters United States Marine Corps (USMC). (2017). Marine Corps operating concept (MOC): *How an expeditionary force operates in the 21st Century*. Washington, DC: Author.

Dougherty, G. (2018). Promoting disruptive military innovation: Best practices for DoD experimentation and prototyping programs. *Defense Acquisition Research Journal, 25*(1), 2–29.

Future Force Staff (FFS). (2017, March 1). *Set-based design ushers in a modern approach to shipbuilding.* Retrieved from http://futureforce.navylive.dodlive.mil/2017/03/set-based-design-ushers-in-a-modern-approach-to-shipbuilding/

GAO. (2017). *GAO report weapon systems: Prototyping has benefited acquisition programs, but more can be done to support innovation initiatives* (GAO-17-309). Washington, DC: Author.

Gardiner, S. (1992, August). The military-technical revolution: More than military and more than technical. *RSAS Newsletter, 4*(3).

Garner, M., Doerry, N., MacKenna, A., Pearce, F., Bassler, C., Hannapel, S., & McCauley, P. (2015, November). *Concept exploration methods for the Small Surface Combatant.* Paper presented at the World Maritime Technology Conference 2015, Providence, RI.

Giachetti, R. (2016). *Rethinking the systems engineering process in light of design theory*. Monterey, CA: Naval Postgraduate School.

Goertner, F. (2018, February 27). The Navy needs to do more than rebuild for the future; it needs to reinvent itself. Retrieved from http://cimsec.org/navy-needs-rebuild-future-needs-reinvent/35561

Hayes, B., & Smith, D. (1994). *The politics of naval innovation* (Naval War College paper).

Hutchinson, W. M. (1951). *The Scottish Himalayan expedition*. London, England: J.M. Dent & Co.

Kozloski, R. (2017, July 17). *The path to prototype warfare*. Retrieved from https://warontherocks.com/2017/07/the-path-to-prototype-warfare/

Lautenschlager, K. (1983). The technology and the evolution of naval warfare. *International Security, 8*(2).

McAfee, A., & Brynjolfsson, E. (2017). *Machine platform, crowd: Harnessing our digital future*. New York, NY: W. W. Norton & Company.

McChrystal, S., et al. (2015). *Team of teams: New rules of engagement for a complex world*. England: Portfolio Publishing.

Rhodes, R. (1987). *The making of the atomic bomb.* New York, NY: Simon & Schuster.

Richardson, J. (2015, December 6). CNO identifies four core attributes to guide Navy leaders (Story Number: NNS151206-02). Retrieved from http://www.navy.mil/submit/display.asp?story_id=92293

Richardson, J. (2017, May 17). *The future Navy*. Retrieved from http://www.navy.mil/navydata/people/cno/Richardson/Resource/TheFutureNavy.pdf

Rubin, K. (2013). *Essential scrum: A practical guide to the most popular agile process*. Addison-Wesley.

Shaw, H. I. (1991). Opening moves: Marines gear up for war. Retrieved from https://www.nps.gov/parkhistory/online_books/npswapa/extContent/usmc/pcn-190-003115-00/sec1a.htm#roebling

Simmons, M. (2017, October 26). Forget the 10,000-hour rule; Edison, Bezos, and Zuckerberg follow the 10,000-experiment rule. Retrieved from https://medium.com/the-mission/forget-about-the-10-000-hour-rule-7b7a39343523

Singer, P., & Cole, A. (2015). *Ghost fleet*. Boston, MA: Houghton Mifflin Harcourt.

SPAWAR Systems Command. (2016, October 7). *Special Notice: Advanced naval technology exercise (ANTX)—2017 ship to shore maneuver exploration and experimentation exercises*. San Diego, CA.

Stackley, S. (2016, January 7). *Statement of the Honorable Sean J. Stackley, Assistant Secretary of the Navy (Research, Development, and Acquisition) Before the House Armed Services Committee on Acquisition Reform: Experimentation and Agility.* Retrieved from http://docs.house.gov/Committee/Calendar/ByEvent.aspx?EventID=104314

United States Government (USG). (2016). The National Defense Authorization Act for Fiscal Year 2016.

United States Government (USG). (2017). The National Defense Authorization Act for Fiscal Year 2017.

Appendix: Annual Naval Technology Exercises and ANTX Summary

- 13 August 2015. NUWC Division Newport hosted the inaugural Annual Naval Technology Exercise. Undersea Constellation technology was demonstrated.
- 16–18 August 2016. Cross Domain Communications and Command & Control Above, On, and Under the Sea Annual Naval Technology Exercise hosted by NUWC Division Newport. Exercise focused on cross-domain collaboration for maritime superiority.
- ★ **15–29 April 2017. DASN RDT&E/Marine Corps DC, CD&I Task Force led Ship to Shore Maneuver Exploration and Experimentation (S2ME2). ANTX focused on evaluating technology and tactics parings that enable high speed and agility in order to rapidly deploy combat capabilities from the sea.**
- 15–16 August 2017. Battlespace Preparation in a Contested Environment hosted by NUWC Newport/Keyport & NSWC Panama City. Exercise focused on battlespace preparation in a contested environment.
- 21 August–1 September 2017. ANTX Innovation and Sensor Fusion Experimentation Exhibit (ISFEE) hosted by NSWC Crane. Experiment focused on networked sensors, weapons, data fusion, data diffusion, processing, exploitation and dissemination and a Counter-UAS scenario.
- 13–14 September 2017. Surface Warfare Distributed Lethality in the Littoral hosted by NSWC Dahlgren. Demonstration focused on improving air/surface warfighting through unmanned system integration.
- ★ **15—23 March 2018. DASN RDT&E/Marine Corps DC, CD&I Task Force led Urban 5th Generation (U5G) Marine ANTX. ANTX focused on assessing technology and tactics pairings that enable situational awareness, counter-reconnaissance, maneuver, fires, and C4I operations within and among the populations resident in the urban littorals.**
- 21–25 May 2018. Coastal Trident 2018 to be hosted by NSWC Port Hueneme Division. Demonstration will focus on leveraging new and emerging technology to address operational and technical challenges presented by asymmetric threats in port and coastal regions.
- 29–31 August 2018. Human-Machine Optimization and Integrated Targeting in the Maritime Environment to be hosted by NUWC Division Newport and Commander, Naval Meteorology and Oceanography Command (CNMOC). Exercise will focus on human-machine Optimization and Integrated Targeting in Maritime Environments.
- ★ **TBD December 2018. Information Warfare ANTX to be hosted by SSC Pacific and TBD Operational Force/Fleet champion.**

Note: Operational Force championed ANTXs denoted by ★.

Panel 25. Why Causal Learning Must Be Adopted Within Acquisition Research

Thursday, May 10, 2018	
3:30 p.m. – 5:00 p.m.	**Chair: Robert Stoddard,** Principal Researcher, Software Engineering Institute, Carnegie Mellon University **Discussant: Ricardo Valerdi,** Associate Professor, University of Arizona **Discussant: Raymond J. Madachy,** Associate Professor, Naval Postgraduate School ***Experience Searching for Causal Factors in Personal Process Student Data*** William R. Nichols, Jr., Carnegie Mellon University Michael Konrad, Carnegie Mellon University ***Further Causal Search Analyses With UCC's Effort Estimation Data*** Anandi Hira, University of Southern California Robert Stoddard, Carnegie Mellon University Mike Konrad, Carnegie Mellon University Barry Boehm, University of Southern California

Robert Stoddard—is a Software Engineering Institute Principal Researcher at Carnegie Mellon University. His research includes machine/causal learning, applied statistics, Bayesian probabilistic modeling, Six Sigma, and quality/reliability engineering. Stoddard achieved an MS in Systems Management and significant doctoral progress in reliability and quality management. He is a Fellow of the American Society for Quality and a senior member of the IEEE. Stoddard holds five ASQ certifications and is a Motorola-certified Six Sigma Master Black Belt. [rws@sei.cmu.edu]

Ricardo Valerdi—is an Associate Professor in the Systems & Industrial Engineering Department at the University of Arizona. Previously he was a Research Associate at the Massachusetts Institute of Technology. He is a Visiting Associate in the Center for Systems and Software Engineering at the University of Southern California and a Visiting Scientist at the Software Engineering Institute at Carnegie Mellon University. Previously he was a Visiting Fellow at the Defence & Systems Institute at the University of South Australia.

Dr. Valerdi is a two-time recipient of the Best Thesis Advisor Award in the MIT Technology & Policy Program, the Best Article of the Year Awards in the *Systems Engineering* journal and *Defense Acquisition Journal*, and Best Paper Awards at the Conference on Systems Engineering Research (2009 and 2013), International Society of Parametric Analysts (2005), and Conference on Predictive Models in Software Engineering (2011). In 2014–2015 he was Visiting Fellow of the Royal Academy of Engineering (UK) hosted by the University of Bath. He received the Frank Freiman Award, the top honor given by the International Cost Estimating and Analysis Association, and is a member of the Mexican Academy of Engineering.

His research focuses on cost estimation, test & evaluation, cybersecurity, and sports analytics. His research has been funded by the Army Test & Evaluation, Navy Acquisition Research Program, Air Force Office of the Surgeon General, Air Force Acquisition Chief Process Office, Air Force Office of Scientific Research, BAE Systems, Lockheed Martin, Raytheon, IBM Center for the Business of Government, the National Collegiate Athletic Association, the California Department of Education, and the Arizona Diamondbacks. He teaches courses in cost estimation, systems engineering, sports analytics, and the science of baseball.

Dr. Valerdi is the Editor-in-Chief of the *Journal of Cost Analysis and Parametrics* and from 2009–2014 was the founding Editor-in-Chief of the *Journal of Enterprise Transformation*. He served on the Board of Directors of the International Council on Systems Engineering, and is a Senior Member of the Institute of Electrical and Electronics Engineers. He received his BS/BA in Electrical Engineering from the University of San Diego in 1999, and his MS and PhD degrees in Systems Architecture and Engineering from the University of Southern California in 2002 and 2005. Between 1999 and 2002, he worked as a systems engineer at Motorola and has been affiliated with the Aerospace Corporation's Economic and Market Analysis Center. He served as Program Chair of the 20th and 24th Forum on COCOMO and Software Cost Modeling.

His contributions to the field include the Constructive Systems Engineering Cost Model (COSYSMO), a model for estimating systems engineering effort, which has been calibrated with data provided by BAE Systems, Boeing, General Dynamics, L-3 Communications, Lockheed Martin, Northrop Grumman, Raytheon, and SAIC.

He is also founder and Chief Scientist of the Science of Sport and a consultant to the Arizona Diamondbacks, Los Angeles Angels of Anaheim, San Diego Padres, Colorado Rockies, Washington Nationals, LA Galaxy, Seattle Sounders, Dallas Mavericks, and Orlando Magic. His work has been featured on ESPN, Fox Sports Arizona, and LA Times. In collaboration with faculty in the UA College of Medicine, he developed the first-ever concussion simulator for football for the NCAA.

Raymond J. Madachy—is an Associate Professor in the Systems Engineering Department at the Naval Postgraduate School. He has 30 years of experience working in industry, academia, and consulting in technical and management positions. His research interests include modeling and simulation of processes for architecting and engineering of complex systems; system total ownership cost modeling and tradespace analysis; systems and software measurement; integrating systems engineering and software engineering disciplines; and integrating empirical research with process simulation. He has over 100 publications with three books including *Software Process Dynamics*.

Experience Searching for Causal Factors in Personal Process Student Data

William R. Nichols, Jr.—joined the Software Engineering Institute (SEI) in 2006. He serves as Team Software Process (TSP) Mentor Coach and maintains PSP and TSP Software Engineering Data. Prior to joining the SEI, he led a team at the Bettis Laboratory near Pittsburgh, PA, developing and maintaining nuclear engineering and scientific software. Dr. Nichols's research interests include software engineering project planning, process modeling, and economics. He received a PhD in physics at Carenegie Mellon University. [wrn@sei.cmu.edu]

Michael Konrad—is a Principal Researcher at the SEI, providing analytic support to various projects at the SEI using statistics, machine learning, and most recently, causal learning. Since 2013, Dr. Konrad has contributed to research in requirements engineering, software architecture, and system complexity measurement. From 1998 to 2013, he contributed to CMMI in many technical leadership roles. Prior to 1998, Dr. Konrad was a member of the teams that developed the original Software CMM and ISO 15504. He is coauthor of the Capability Maturity Model Integration for Development (CMMI-DEV). Dr. Konrad received his PhD in mathematics from Ohio University in 1978. [mdk@sei.cmu.edu]

Abstract

The objective of this study is to apply recently developed techniques to infer causality from observational software engineering data. Determining causation rather than just correlation is fundamental to selecting factors that control outcomes such as cost, schedule, and quality. The Tetrad tool's PC and FGES causal search algorithms were applied to software engineering data from 4940 programs written in the C programming language collected during Personal Software Process (PSP) training. PSP programs have previously been used in empirical research quantitative relationships between developer and project factors. Both algorithms successfully identified the expected relationships and did not find contradictory or implausible associations. Many of the available causal inference search algorithms require Gaussian distributional families with linear effects. The linear relationship may be especially important for software engineering research and may require prior knowledge and data transformation. Because software engineering has depended on small-scale, low-power experiments, often using non-representative students, inferring causal relationships would expand the insight available to researchers. Inferring causation from observational software engineering data shows much promise, but is currently limited by researcher understanding of the capability and limits of causal inference, the quality of the underlying data, and the general requirement for linear effects.

Introduction

Despite repeated calls for empirical studies in software engineering (Perry, Porter, & Votta, 2000) and guidelines for their conduct (Kitchenham & Dybå, 2004) it is usually impractical to run controlled software development experiments. Thus, most data in software engineering are observational, presenting challenges to causal inference. Without causation, selection or control of factors will not have the desired effect on outcomes. Understanding causation is fundamental to the forward-looking control of the software development process.

The epistemological problems of inferring causation from observational data are now being overcome (Pearl, Glymour, & Jewell, 2016; Spirtes, 2010) and accepted in research (Fedak et al., 2015). This study aims to apply causal search techniques to a previously-studied software engineering data set to validate the overall approach and gain experience with the capabilities and limitations of these methods.

Tools for Causal Inference

The University of Pittsburgh, Carnegie Mellon University (CMU), and the Pittsburgh Supercomputing Center serve as founding members of the Center for Causal Discovery (CCD). The CCD develops and maintains causal algorithms, software, and tools, including Tetrad. Tetrad enables users not only to search for causal graphs from a dataset, but also to estimate and evaluate parametric models. We applied the PC and FGES algorithms to selected data from PSP training. These two algorithms were chosen to exercise two different search approaches.

The PC algorithm, named after its creators Peter Spirtes and Clark Glymour, sets a Fisher Z-based *p*-value cutoff for conditional independence testing. The FGES algorithm uses a fast greedy approach to search the space of causal Bayesian networks to return the most probable model(s) based on the Bayesian information coefficient (BIC) score (Sanchez-Romero et al., 2018). Both algorithms assume that each node is a linear function of its parents plus a Gaussian noise term.

One advantage of score-based search algorithms over constraint-based search algorithms is that they can obtain quite accurate adjacencies within the causal graph equivalence class. Also, score-based algorithms typically output only directed or undirected edges. Because equivalence class scoring almost always favors one orientation over the other, bi-directed edges are rare. A limitation of score-based search algorithms is that they can be slow and might not scale as well as constraint-based searches.

Generally in our analyses, we don’t have full knowledge of how well the assumptions of the various search algorithms are met, so we usually employ more than one algorithm. Also, because they rely on different mathematical mechanisms to construct the output graph (MEC), we favor applying one or more constraint-based searches and one or more score-based searches, and comparing for commonalities in the direct causal relationships that are identified. This provides some protection against the uncertainties about how well the various assumptions are met by the dataset analyzed. Employing two or more search algorithms based on different mathematical approaches for inferring causal structure also allows us to take advantage of their respective strengths (e.g., there is less ambiguity in the direction of causality with score-based searches).

Personal Software Process (PSP) Data

Dataset Summary

The Personal Software Process was developed by Watts Humphrey at the Software Engineering Institute to demonstrate how an individual can apply the process principles underlying the Capability Maturity Model for personal work. The PSP contains coherent frameworks for defining the development process and measurements for process and products. A progressive development process with activity steps, measurements, and a sequence of training assignment exercises is described *in A Discipline for Software Engineering* (Humphrey, 1995).

PSP classes are taught by trained and authorized instructors who submit resulting data to the Software Engineering Institute for use in research. Several versions of the course have been taught over the years; this study uses the 10-assignment course taught primarily through 2006 because it contains a large sample and consists primarily of professional software developers rather than university students. Additional descriptions of the data and prior analyses can be found in Rombach et al. (2008) and Vallespir and Nichols (2012). For this study, because we wanted to reduce the number of potential hidden confounding factors, we selected only programs using the C programming language and students who completed the entire 10-program sequence. Although using only a subset of the data introduces some risk for bias, lack of hidden confounders is a key assumption for both search algorithms.

Data Attributes

When using PSP for the program exercises, students record the direct (i.e., stopwatch) time engaged in particular activities (e.g., planning, design, design review, code, code review, compile, test, post mortem analysis) along with information on the defects injected and removed in each phase. For the more focused analysis that is the subject of this report, we used only the total effort (sum of all activities), the construction activity effort (design and code), and the total defects for the 494 students using the C programming language and implementing all 10 assignments. The data variables we examined were as follows:

1. Assignment Average Minutes (abbreviated AsgAveMin)—for each of the 10 program assignments, the average of the log-transformed effort required by the 494 students using the C programming language to complete that assignment. AsgAveMin can be thought of as a proxy for the requirements size of each program. (AsgAveMin is defined for 10 assignments.)
2. Student Size Factor (abbreviated StuSizeFactor)—for each of the 494 students, the ratio of total new lines of code written by the student for the 10 program assignments compared to the total new lines of code written for the 10 program assignments averaged across all 494 students. (StuSizeFactor is defined for 494 students.)
3. Student Effort Factor (abbreviated StuEffFactor)—for each of the 494 students, the ratio of the student's total effort for the 10 program assignments to the overall student average. Thus, StuEffFactor is very similar to StuSizeFactor, but focuses on a student's total effort rather than the total new lines of code written. (StuEffFactor is defined for 494 students.)
4. Student Defect Arrival Rate (abbreviated StuDAR)—for each of the 494 students, the ratio of defects introduced during program design and code activities (that is, during construction to the construction effort). This factor characterizes a student's specific tendency to introduce defects while developing software, using "defects per hour" as the unit. Using hours as the unit instead of the more common minutes should not affect causal inference and provides a better scale for the log-transform. (StuDAR is defined for 494 students.)
5. Construction Minutes (abbreviated ConstMin)—for each of the 494 students and 10 program assignments, the effort expended in construction (design and code) activities, measured in minutes. Thus, ConstMin does not include effort expended in planning, reviews, compile, and test. (ConstMin is defined for 494 students and 10 program assignments.)
6. Lines of Code for the product (abbreviated LOC)—for each of the 494 students and 10 program assignments, the sum of added and modified logical lines of code written. Thus, LOC does not include lines of code that were reused without modification. (LOC is defined for 494 students and 10 program assignments.)
7. Total Development Effort (abbreviated MinTot)—for each of the 494 students and 10 program assignments, the student's total effort expended on the

program assignment, measured in minutes. (MinTot is defined for 494 students and 10 program assignments.)

8. Total Number of Defects Injected and Discovered (abbreviated DefTot)—the number of defects discovered after the step in which they were injected is completed. For example, any design defect discovered during coding would be counted, but not a defect both introduced and discovered in the same design phase. Also, a typo corrected during normal coding would not count, but a typo discovered by a compiler would. We asked students using interactive development environments to disable automated checking so that these defects would be visible and counted in compile. (DefTot is defined for 494 students and 10 program assignments.)

Note that AsgAveMin can be thought of as a proxy for the requirements size because each student must perform the same exact 10 exercises. The exercises vary in difficulty as does the amount of code required to implement those requirements. Thus, rather than using an estimate based on function points as a measure of requirements size, we estimate the relative size of specific exercises from the arithmetic average of the log-transformed effort expended on an exercise across all developers.

By taking the log transformation, we accomplish several purposes. First, we reduce the effect of outliers, helping to make the search algorithms we utilize less sensitive to outliers, and thus the model(s) returned from causal search more stable. Second, transformed distribution is approximately Gaussian, which the theorems for consistent convergence of the chosen algorithms require. Third, the log transformation will later help to linearize factor effects.

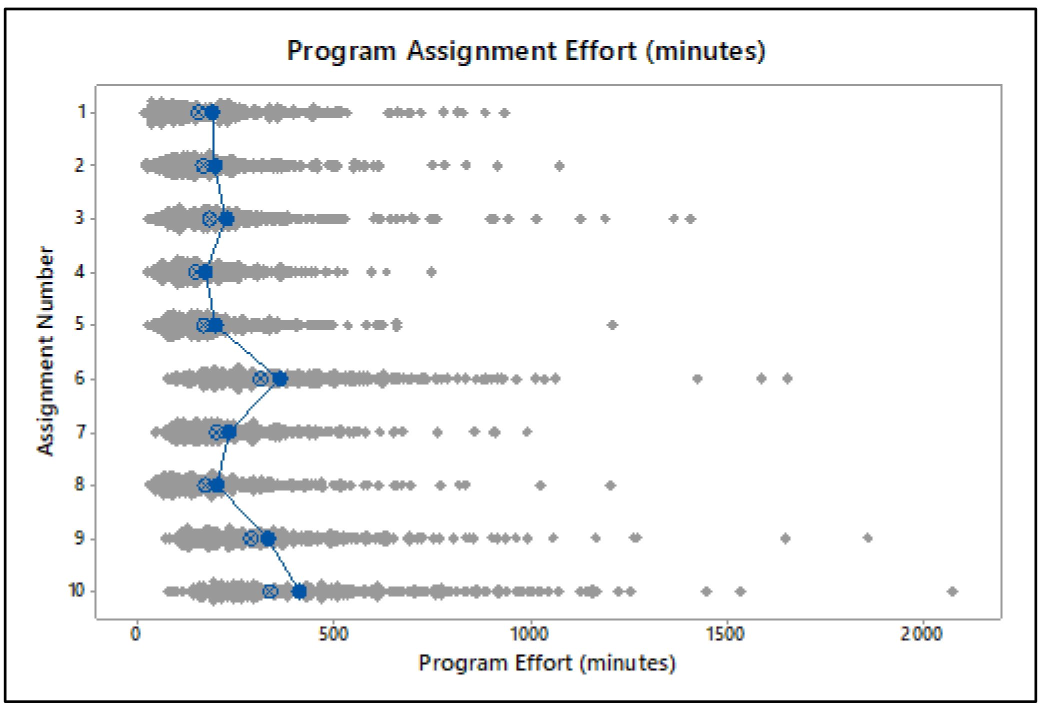


Figure 1. Individual Values Plot of Actual Effort for Each Programming Assignment

Figure 1 shows that individual effort distributes widely for each programming exercise. The bulls-eye symbols indicate the median and the circles (connected by a line) are the mean. The differences are statistically significant differences between program assignments. The mean values are also shown in a scatterplot of size versus effort in Figure 2 , which also shows that the size varies by nearly a factor of three, while effort varies by roughly a factor of two. We interpret slope as code production rate and the intercept as a start-up cost. We decided to use only the effort factor for this study because for the aggregate, the correlation between size and effort is very strong, suggesting that adding size does not add much information. In this figure, the averages were computed from the log transformed data, then retransformed into the natural units of LOC and minutes.

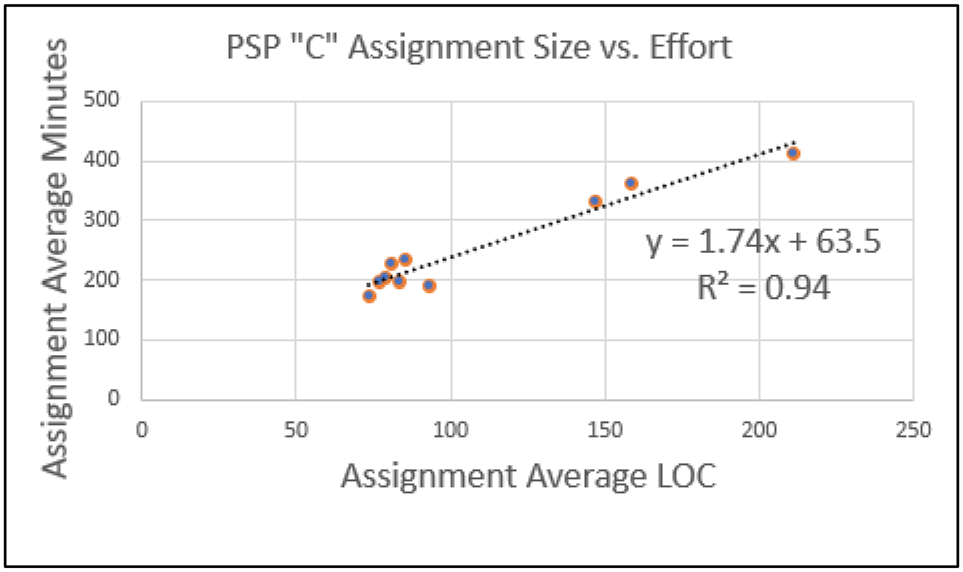


Figure 2. Correlation Between Program Size and Effort for 10 PSP "C" Programming Language Exercises

We also included factors that account for programmer variability. The software engineering literature contains numerous studies reporting variation in programmer productivity (Sackman, Erikson, & Grant, 1968; Curtis, 1981; Valett & McGarry, 1989; DeMarco & Lister, 1999; Card, 1987; DeMarco & Lister, 1985; Sheil, 1981). Few studies, however, explicitly report individual differences in defects or size of solutions to similar problems. Nonetheless, a more recent work (Caliskan et al., 2018) reports that individual programmer characteristics can be identified from the compiled (and even optimized) binary. We decided to explicitly account for programmer idiosyncrasies in coding style, line counting standards, and solution approach with programmer-specific factors that affect product size, defect counts, and production rates. This is supported by a separate ANOVA analysis of the data that finds that such programmer factors are statistically significant and approximately doubles the amount of variance accounted for by the coefficient of determination (from $R^2 \approx 0.3$ to $R^2 > 0.6$). The untransformed distributions of student factors (StuSizeFactor and StuEffFactor) seem to follow lognormal distributions as shown in Figure 3 (note the heavy skew to the right); and the log-transformed data (lnStuSizeFactor and lnStuEffFactor) thus approximately follow normal distributions as shown in Figure 4. Defect arrival rates (StuDAR), untransformed and transformed, are shown in Figure 5 and Figure 6.

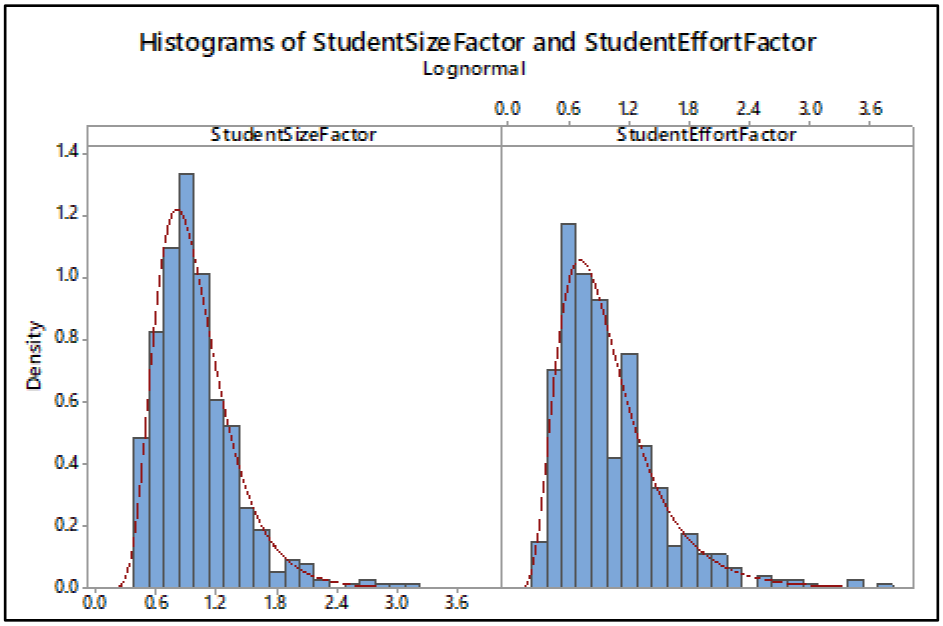


Figure 3. Histogram of Student Size and Effort Factors (494 Points)

The untransformed student-dependent size and effort factors are shown in Figure 3. The log-transformed distributions approximate a Gaussian as shown in Figure 4. Please note that the scales in the paired plots may differ and ordinates are displayed below the left hand plot but above the right hand plot.

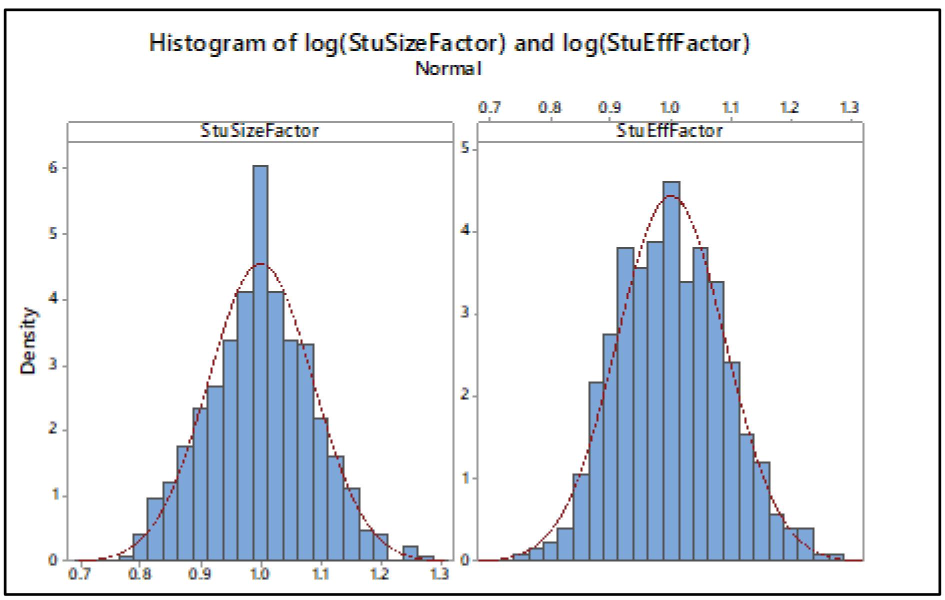


Figure 4. Histograms of Log-Transformed Student Size and Effort Factors

The defect arrival rate is the student's rate of injecting defects during design and coding activities. Because this can be zero, we have added an offset of 1.0 in the transformation to prevent negative rates. This small offset will not affect the search provided that the distribution is approximately Gaussian. The StuDAR at zero are likely an artifact of data gathering practices. We are, however, reluctant to clean the data because of the threat of introducing bias.

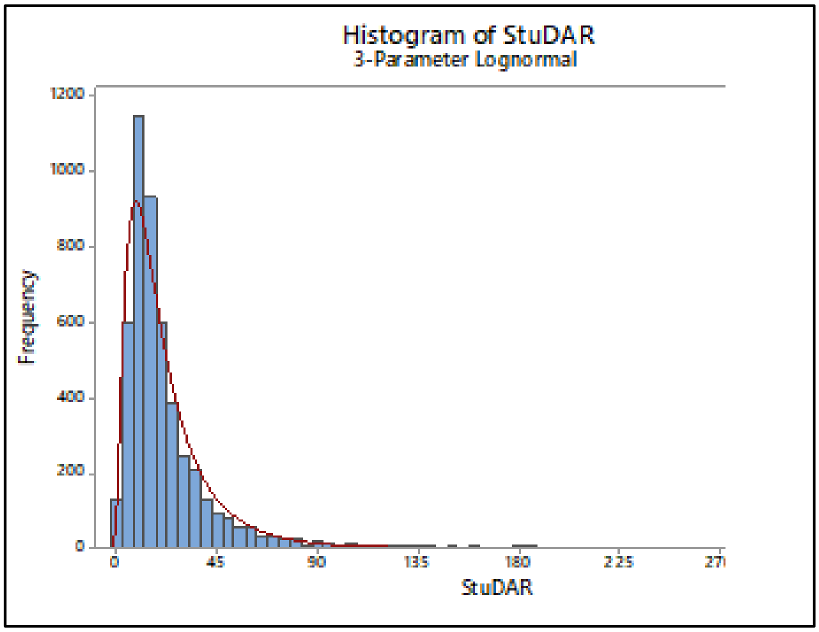


Figure 5. Student Defect Arrival Rate (494 points)

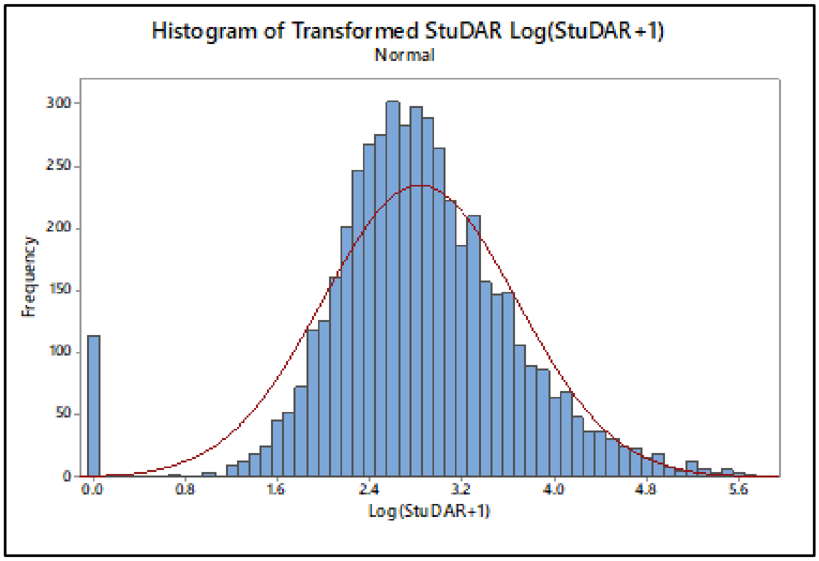


Figure 6. Log of the Student Defect Arrival Rate (494 Points)

For our initial model, we examine the construction time (ConstMin) as a candidate causal contributor to lines of code and defects, and total effort (respectively LOC, DefTot, and MinTot). See Figure 7 for a histogram of ConstMin. Mathematically, construction time is a product of the construction rate times and the product size. Because this rate may already be implicit in other programmer factors, we avoided using it in this analysis. Moreover, that rate may be causal, but uncontrollable. Other factors, including estimation accuracy, effort in review, and review rates, will be considered in future work. For this analysis we use the construction effort because the expected model described in a later section is simple to construct and interpret, thus helping to validate the overall approach.

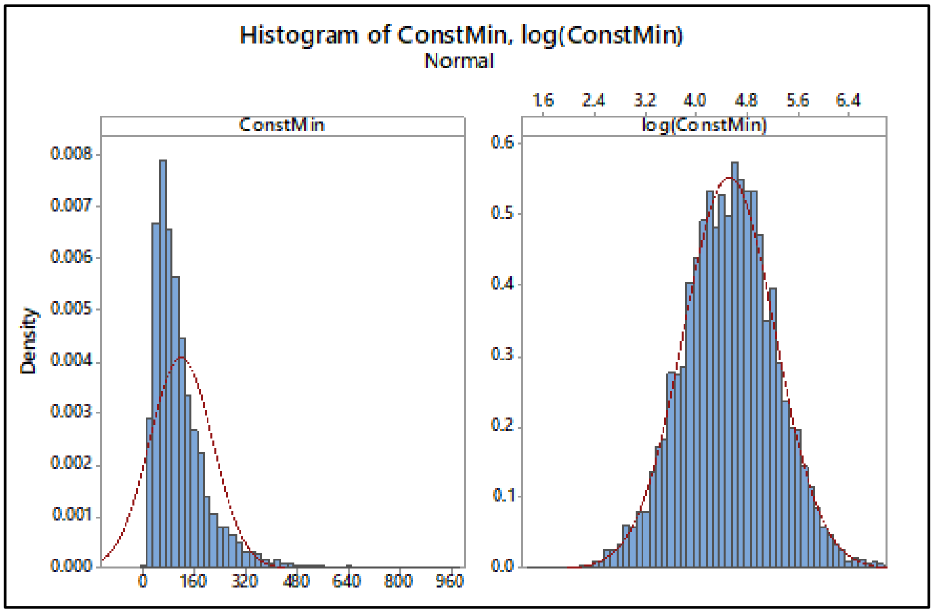


Figure 7. Distribution of Construction Effort and Log Transform of Construction Minutes of Effort (4940 Points)

The outcomes of interest are the program size measured in Lines of Code (LOC; shown in Figure 8), total effort measured in minutes (MinTot; shown in Figure 9) and total defects (DefTot; shown in Figure 10).

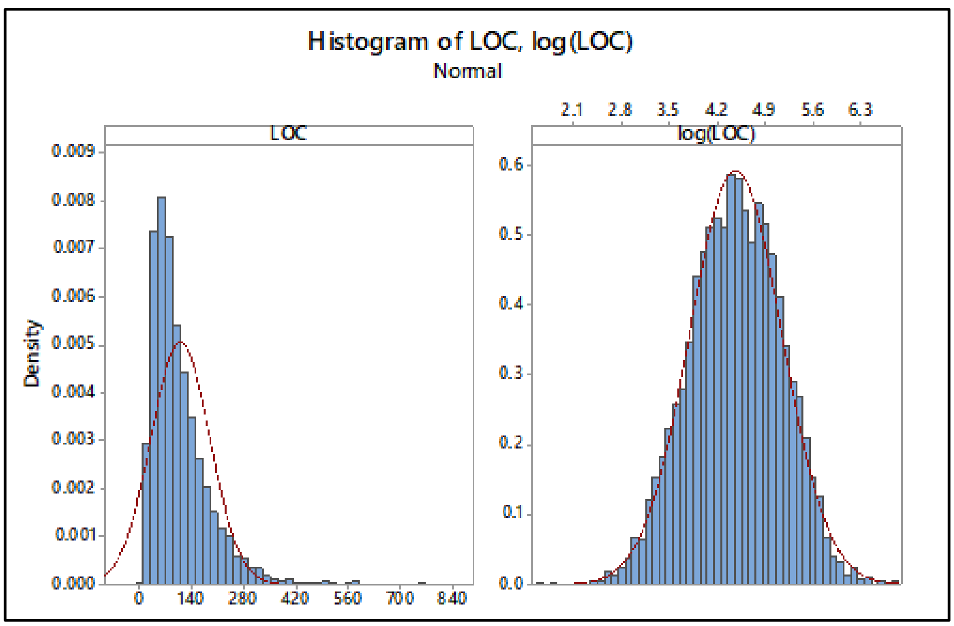


Figure 8. Students' Sum of Added and Modified Logical Lines of Code (LOC) and Log-Transformed LOC for Each Student Exercise Pair (4940 points)

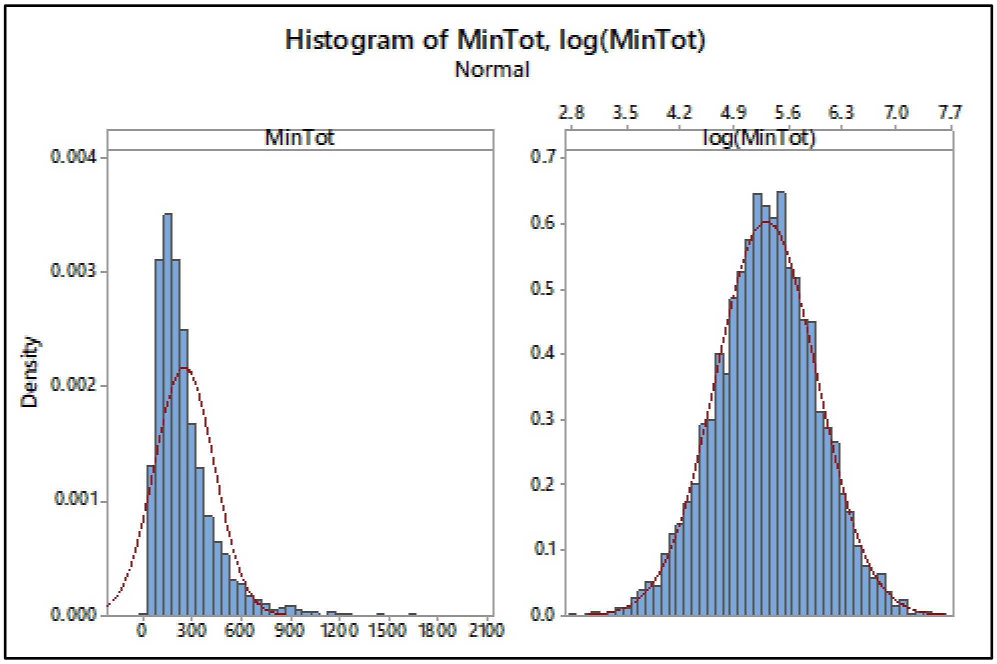


Figure 9. Students' Total Effort in Minutes (MinTot) and Log-Transformed MinTot for Each Student Exercise Pair (4940 Points)

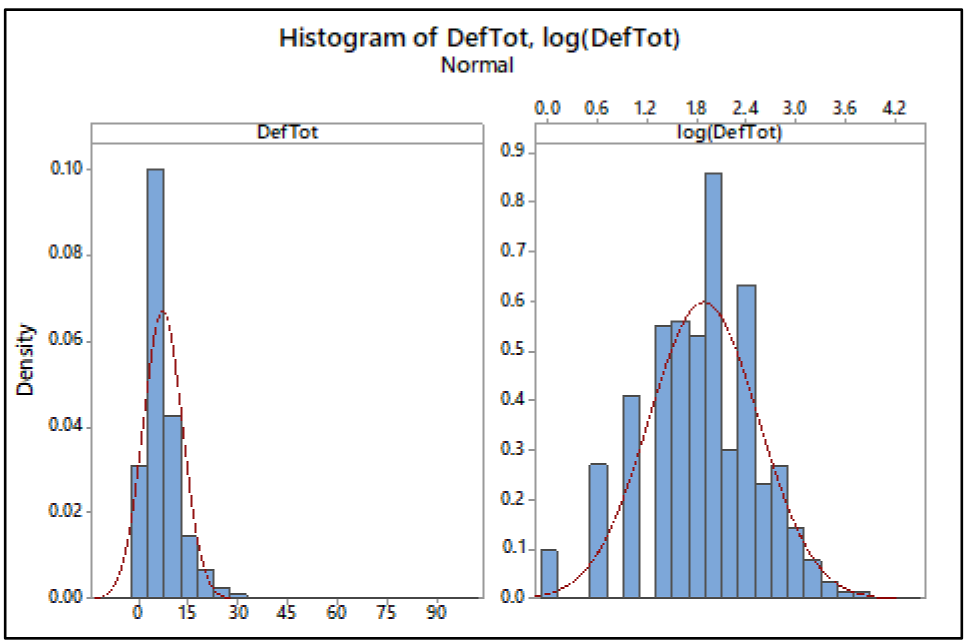


Figure 10. Students' Total Number of Defects Injected and Discovered (DefTot) and Log-Transformed DefTot for Each Student Exercise Pair (4940 Points)

Data Analysis Approach

Research Objective

For this study, we wanted to

- evaluate the effectiveness of applying the causal search algorithms to the PSP software engineering data.
- determine effective data transformations that facilitate correct use of the selected causal search algorithms.

The PSP data is useful for this research because

1. All students develop programs from the same specifications.
2. The measurements framework is required and reinforced by instructors.
3. The data has been analyzed before and is familiar (Rombach et al., 2008; Grazioli, Nichols, & Vallespir, 2014; Vallespir & Nichols, 2011; Vallespir & Nichols, 2012).

Based on prior analysis and experience, we expected some correlating factors to exhibit a causal relationship.

Expected Models

PSP data is used in planning and tracking projects that are run according to the Team Software Process (TSP). In TSP planning, estimates of component size, conversion factors to lines of code, overall production rates, activity effort distributions, historical defect injection rates, and activity defect removal yield are used to predict likely outcomes (Nichols, 2012).

For a PSP programming assignment (i) and student (j) pair, we would expect—without examining any process data—that the program size can be estimated by the untransformed values:

$$LOC_{ij} = ReqSize_i \times SSF_j \quad (1)$$

In the above, we can use the average effort (*AsgAveMin*) as a proxy for *ReqSize*. The actual size will likely vary based on factors such as design versus code effort. This is not yet modeled.

Likewise, we expect the student effort (j) for each program should be related to the assignment size and student dependent factor:

$$MinTot_{ij} = ReqSize_i \times SEF_j \quad (2)$$

Total development time should also be influenced by other factors, including design-specific effort review time and review effectiveness (not included in this model) and the actual defect arrivals. The defect arrivals should be related to the student's (j) defect tendencies, the program size (i), and the actual effort in construction (ij). It is during construction (design and coding) that most defects are injected.

$$DefTot_{ij} = ConstMin_{ij} \times StuDAR_j \quad (3)$$

That these relationships are products is a problem for the search algorithms. We will, therefore, take advantage of the observation that the values from the left hand side of these equations distribute approximately lognormally, by using log transforms. The resulting transformed equations and data thus consists of a linear sum of normally-distributed data, making them more suitable for the search algorithms we intend to employ.

If the causal search is successful in finding consistent causal models involving these four introduced factors for problem requirements size and three developer traits (respectively, $ReqSize_i$, SSF_j, SEF_j, and $StuDAR_j$), we will have added support to the case that these factors can be useful for prediction and mitigation in software development, at least at the individual and team levels for planning and tracking.

Causal Discovery

We ran the PC search algorithm with alpha = .05 (a hyper-parameter defining the *p*-value cutoff for inferring conditional independence) and the domain-knowledge constraints described below.

Tetrad allows users to add constraints regarding the required presence or absence of a particularly-oriented direct causal relationship between two nodes (i.e., parameters). For example, a causal link may be required or forbidden or the direction restricted. A known temporal order can be enforced by placing the nodes into knowledge-box tiers such that causality is only permitted forward, not backward. Adjacencies between nodes appearing within the same tier may be allowed or forbidden. We chose to structure tiers as follows:

- Tier 1: Assignment Average Minutes (AsgAveMin), Student Size Factor (StuSizeFactor), Student Effort Factor (StuEffFactor), and Student Defect Arrival Rate (StudDAR)
- Tier 2: Construction Minutes (ConstMin)
- Tier 3: Lines of Code (LOC), Total Effort (MinTot), and Total Number of Defects Injected and Discovered (DefTot)

Essentially the tiers correspond to pre-development inputs (characteristics of the problem or developers), in-process data (the construction effort in minutes), and process outputs (size, effort, and total defects). The default setting is that nodes in the same tier can have direct causal relationships between them, but a node in a lower tier (assigned a higher number) cannot have an oriented edge pointing to a node in a higher tier (assigned a lower number).

Results

Node Links

The search results are shown in Figure 11 and Figure 12. There are some differences but no explicit contradictions.

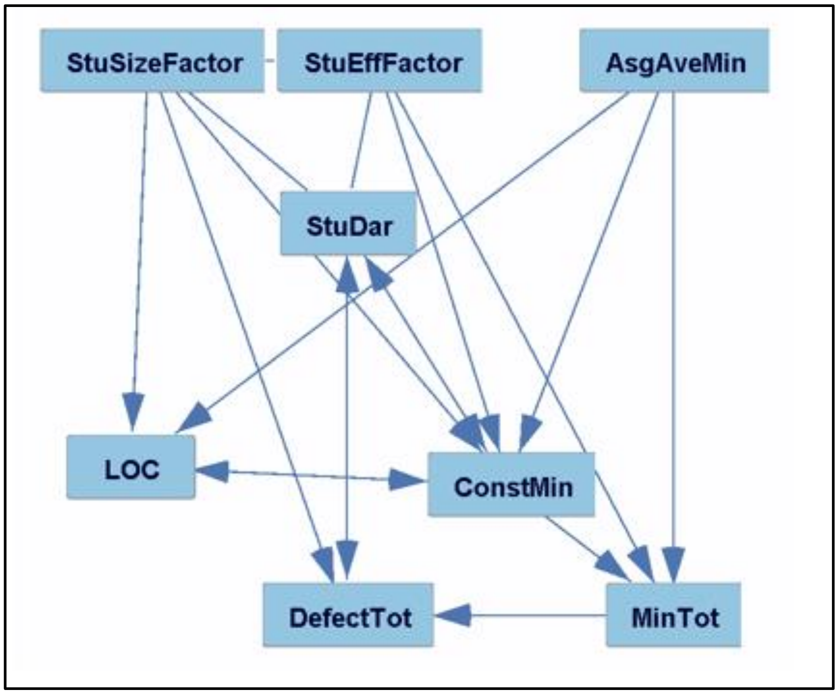


Figure 11. Resulting DAG From Tetrads PC Search Algorithm for Data From Programs Written in C

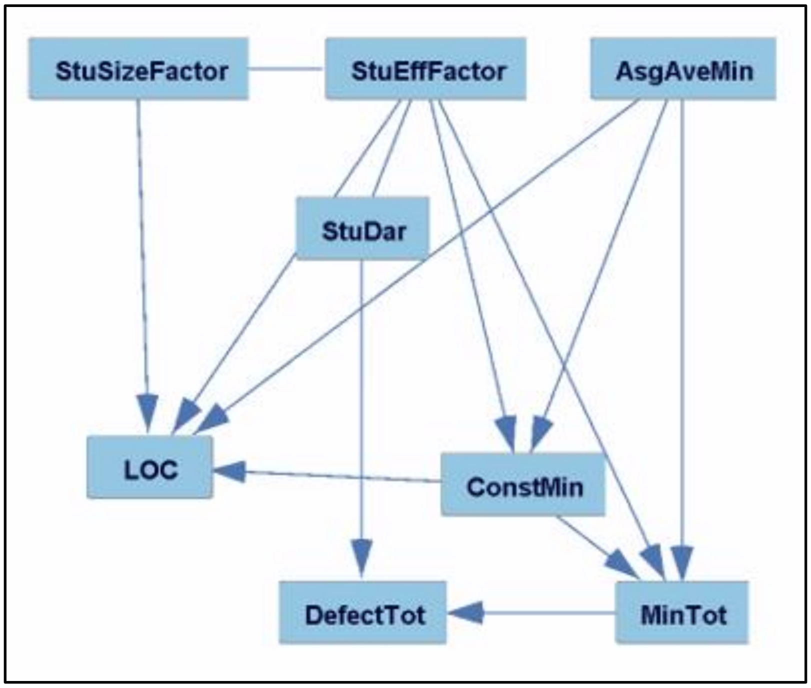


Figure 12. Resulting DAG From Tetrads FGES Search Algorithm for Data From Programs Written in C

The following direct causal edges occur in both graphs:

1. AsgAveMin --> ConstMin
2. AsgAveMin --> LOC
3. AsgAveMin --> MinTot
4. ConstMin --> MinTot
5. MinTot --> DefectTot
6. StuEffFactor --> ConstMin
7. StuEffFactor --> MinTot
8. StuSizeFactor --> LOC

The following undirected causal edge occurs in both graphs:

1. StuEffFactor --- StuDAR (This means there's evidence of a direct causal relationship between the two nodes, but there's insufficient information to determine the direction.)

The following directed causal edges occur in only one graph:

1. ConstMin --> LOC (The graphs returned by PC shows as a bidirected edge, meaning there's evidence for a hidden confounder of the two nodes.)
2. StuDAR --> DefectTot (The graphs returned by PC shows as a bidirected edge, meaning there's evidence for a hidden confounder of the two nodes.)
3. StuEffFactor --> LOC
4. StuSizeFactor --> ConstMin
5. StuSizeFactor --> DefectTot

The following undirected causal edge occurs in only one graph:

1. StuSizeFactor --- StuDAR

Threats to Validity

Internal Validity

That the PSP data collection forms filled out and collected during PSP training were consistent and completely filled in did not necessarily mean that students recorded complete and accurate data. Others (Johnson et al., 2003) found significant errors, though these seem primarily to be in hand calculations prior to automated tool support. Our analysis relies only upon direct measures for which all further calculations are performed by the data-gathering spreadsheet used during training. Nonetheless, data correctness validation relies upon the diligence of the student and instructor.

The difficulty of the exercises or rigor required by the course may lead to a survival bias. Excepting the LOC counter, which is the very first exercise, the series of exercises consists of writing programs that must perform statistical and floating-point math calculations. Some are more complicated and challenging to program than others, and many students failed to complete the course, perhaps in part due to challenges they encountered while working on these problems. Our analysis uses only data from students who completed the course.

There may also be a bias in the experience of the students regarding domain experience with statistics or programming experience in general. The data used for the study includes 494 software practitioners who took part in the training at the Software Engineering Institute (SEI) or at external locations. The data provided limited information that could help us measure the programmers' experience. We have indications of years of programming and the number of lines of code in the target language, but no information about domain experience. It is likely that more domain-experienced programmers (i.e., those with strong elementary knowledge of statistics and linear algebra) performed better.

There is some risk of maturation bias. The PSP course deliberately teaches estimation, design techniques, and review of both code and designs, while the developers may gain experience within the numeric programming domain. The maturation effect on total defects and defects escaping into test is evident. Although total code production rate appears to be unchanged, the overall rate does initially slow, then return to near its original level. The risks with respect to inference on the effects of PSP training were addressed in Rombach et al. (2008) and Hayes and Over (1997); however, there could also be process-drift effects.

External Validity

We report only on results from the C programming language. By selecting only data from students using C, we have mitigated issues arising from the use of different programming languages between subjects (e.g., in determining StuSizeFactor) at the expense of some generality. Different languages can be more or less suited to solving certain programming problems, and this could affect the assignment normalization factors. Future analyses will compare results with other programming languages.

Because the work was performed in class settings rather than under normal industrial conditions, there is a risk that the results might not generalize beyond the academic setting.

Construct Validity

The program assignment requirements were deliberately left vague on some key points, mostly related to the specifics of input and output formats, error checking, and so forth. Individual interpretations of the requirements could vary somewhat, adding some variation to implemented program size and effort.

Our proxy for an independent measure of program size, the average effort in minutes (AsgAveMin), aggregated these individual programmer choices and thus might be subject to some systematic bias.

Discussion

All the identified causal relationships have face plausibility. Larger assignments cause more effort and larger amounts of code. The expected relationships for size and effort both appear reasonable. The relationship with defects is less clear. The FGES search algorithm finds the StuDAR and total effort (MinTot) rather than construction effort causing defects, while PC-Stable has a bidirected edge (double arrow) between DefectTot and StuDAR (indicating a hidden confounder). PC also has a connection from StuSizeFactor to DefectTot.

It is possible that the algorithms lead to slightly different models because the data is from mixed-causal systems or is insufficient. Another possibility is that there is overlapping information in these specific variable constructs (creating dependencies with hidden variables). To avoid this problem it is often preferable to use only direct rather than derived values; however, that approach runs into the problem that both PC and FGES assume linear relationships between a child node and its parents, plus a Gaussian noise term. Additional work will be needed to find better variable selections with minimal variable overlap, paired with the most appropriately-selected causal search algorithms. In particular, there are search algorithms that were designed to search for non-linear causal relationships among variables having skewed noise distributions; and there are still variants, as well, that endeavor to take into account hidden confounders.

We offer several observations of lessons learned during this exploratory work.

First, it is imperative to visualize the data. We are not yet certain of the sensitivity to deviations from Gaussian, but single peaks and lack of outliers are surely important. Moreover, the distribution characteristics affect the available transforms. We have focused on simple relationships, but these will not get us from planning through production. Much work remains to model more complex and stepwise systems.

Second, mechanistic relationships with direct measures can become complicated. Using derived measures risks including the same factor multiple times and can lead to mathematical artifacts because of ratios. The requirement for linear effects constrains the available choices in ways that are not immediately obvious. Moreover, it is not yet clear that the natural mechanistic models can be successfully transformed for analysis by the algorithms currently available.

Third, while transformations can sometimes simplify the problem, they make the data relationships less intuitive.

Forth, count data, such as defects, can be poorly behaved with low numbers. We counter this first by counting all defects, not just test, and by using an offset of “+1.” Nonetheless, the distributions can become noticeably discrete on the left-hand side of the peak.

Fifth, the data quality and consistency are a concern. Long lists of problems with software engineering data are available (Shull, Singer, & Sjøberg, 2008). The analyst must be keenly aware of the strengths and weaknesses of particular datasets.

Ultimately, the causal mechanisms we expected to see do appear in the resulting models and implausible causal mechanisms do not. Which specific models are better predictors of future performance is left for future work.

Conclusion

The PC and FGES search algorithms returned results that are generally consistent with each other and with overall expectations.

Causal inference methods should be applied in software engineering, but with caution. We have made only initial steps toward assessing the degree to which different search algorithms are sensitive to deviations from the assumptions about shape (Gaussian for some, skewed for others), outliers, linear effects, or homoscedasticity. Real datasets are likely to be subject to problems of construct validity, measurement inconsistency, determinism, and process drift. Guidelines on reporting data characteristics and the sensitivity of different algorithms will be included in future work.

References

Caliskan, A., Yamaguchi, F., Dauber, E., Harang, R., Rieck, K., Greenstadt, R., & Narayanan, A. (2018). When coding style survives compilation: De-anonymizing programmers from executable binaries. *Network and Distributed Systems Security (NDSS) Symposium 2018.* doi:10.14722/ndss.2018.23304

Card, D. N. (1987). A software technology evaluation program*. Information and Software Technology, 29*(6), 291–300. doi:10.1016/0950-5849(87)90028-0

Curtis, B. (1981). Substantiating programmer variability. *Proceedings of the IEEE, 69*(7), 846.

DeMarco, T., & Lister, T. (1985). Programmer performance and the effects of the workplace. *Proceedings of the 8th International Conference on Software* (pp. 268–272).

DeMarco, T., & Lister, T. (1999). *Peopleware: Productive projects and teams*. New York, NY: Dorset House.

Fedak, K. M., Bernal, A., Capshaw, Z. A., & Gross, S. (2015). Applying the Bradford Hill criteria in the 21st century: How data integration has changed causal inference in molecular epidemiology. *Emerging Themes in Epidemiology*, *12*. doi:10.1186/s12982-015-0037-4

Grazioli, F., Nichols, W., & Vallespir, D. (2014, January). An analysis of student performance during the introduciton of the PSP: An empirical cross-course comparison. In *TSP Symposium 2013 Proceedings* (CMU/SEI-2013-SR-022; pp. 11–21). Retrieved from http://repository.cmu.edu/cgi/viewcontent.cgi?article=1772&context=sei

Hayes, W., & Over, J. W. (1997, December). *The Personal Software Process (PSP): An empirical study of the impact of PSP on individual engineers* (Report No. CMU/SEI-97-TR-001). Retrieved from https://resources.sei.cmu.edu/asset_files/TechnicalReport/1997_005_001_16565.pdf

Humphrey, W. S. (1995). *A discipline for software engineering*. Boston, MA: Addison-Wesley Longman Publishing.

Johnson, P. M., Agustin, J., Chan, C., Moore, C., Miglani, J., & Doane, W. E. J. (2003). Beyond the Personal Software Process: Metrics collection and analysis for the differently disciplined. In *Proceedings of the 25th International Conference on Software Engineering* (pp. 641–646). doi:10.1109/ICSE.2003.1201249

Kitchenham, B. A., & Dybå, T. (2004). Evidence-based software engineering. In *Proceedings of the 26th International Conference on Software Engineering (ICSE'04).* Edinburgh, Scotland. doi:10.1109/MS.2005.6

Nichols, W. R. (2012). Plan for success, model the cost of quality. *Software Quality Professional, 14*(2), 4–11.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics : A primer.* Hoboken, NJ: Wiley.

Perry, D. E., Porter, A. A., & Votta, L. G. (2000). Empirical studies of software engineering: A roadmap. In *Proceedings of the Conference on the Future of Software Engineering* (pp. 345–355). doi:10.1145/336512.336586

Rombach, D., Münch, J., Ocampo, A., Humphrey, W. S., & Burton, D. (2008). Teaching disciplined software development. *Journal of Systems and Software, 81*(5), 747–763. doi:10.1016/j.jss.2007.06.004

Sackman, H., Erikson, W. J., & Grant, E. E. (1968). Exploratory experimental studies comparing online and offline programming performance. *Communications of the ACM, 11*(1), 3–11.

Sanchez-Romero, R., Ramsey, J. D., Zhang, K., Glymour, M. R. K., Huang, B., & Glymour, C. (2018). Causal discovery of feedback networks with functional magnetic resonance imaging. *Preprint*, 1–54.

Sheil, B. A. (1981). The psychological study of programming. *Computing Surveys, 13*(1).

Shull, F., Singer, J., & Sjøberg, D. I. K. (Eds.). (2008). *Guide to advanced empirical software engineering.* London, England: Springer. doi:10.1007/978-1-84800-044-5

Spirtes, P. (2010). Introduction to causal inference. *Journal of Machine Learning Research, 11,* 1643–1662. Retrieved from https://dl.acm.org/citation.cfm?id=1859905

Valett, J., & McGarry, F. (1989). A summary of software measurement experiences in the software engineering laboratory. *Journal of Systems and Software, 148*(2), 137–148. Retrieved from http://www.sciencedirect.com/science/article/pii/0164121289900162

Vallespir, D., & Nichols, W. (2011). Analysis of design defect injection and removal in PSP. In *TSP Symposium 2011 Proceedings* (pp. 19–25). Retreived from https://www.fing.edu.uy/sites/default/files/biblio/22573/designdefectspsp.pdf

Vallespir, D., & Nichols, W. (2012). An analysis of code defect injection and removal in PSP. In *TSP Symposium 2012 Proceedings* (CMU/SEI-2012-SR-015; pp. 3–19). Retrieved from https://resources.sei.cmu.edu/asset_files/SpecialReport/2012_003_001_34121.pdf

Acknowledgments

This material is based upon work supported in part by Cyber Security and Information Systems Information Analysis Center (CSIAC). We would also like to thank David Zubrow and Robert Stoddard of the SEI for encouragement, support, and sharing insights for the work in this paper. Additionally, we thank David Danks, Kun Zhang, Madelyn Glymour, and Joe Ramsey for their help in understanding causal discovery, the algorithms, and the tools.

Disclaimer and Distribution Statement

Further Causal Search Analyses With UCC's Effort Estimation Data

Anandi Hira—is currently a PhD student under Dr. Barry Boehm at the University of Southern California's (USC's) Computer Science Department. Her research interests lie in cost estimation and models, software metrics in relation to project management, and process improvement. Hira has been a part of the Unified Code Count (UCC) development effort at USC's Center for Systems and Software Engineering (CSSE) for the past six years and has been collecting and analyzing the data to compare the effectiveness of functional size metrics. [a.hira@usc.edu]

Barry Boehm—is the TRW Professor in USC's Computer Sciences, Industrial and Systems Engineering, and Astronautics Departments. He is also the Director of Research of the DoD-Stevens-USC Systems Engineering Research Center and the founding Director of the USC Center for Systems and Software Engineering. He was a Director at DARPA-ISTO, TRW, Rand Corporation, and General Dynamics. His contributions include the COCOMO family of cost models and the Spiral family of process models. He is a Fellow of the primary professional societies in computing (ACM), aerospace (AIAA), electronics (IEEE), and systems engineering (INCOSE), and a member of the U.S. National Academy of Engineering. [boehm@usc.edu

Robert Stoddard—is a Software Engineering Institute Principal Researcher at Carnegie Mellon University. His research includes machine/causal learning, applied statistics, Bayesian probabilistic modeling, Six Sigma, and quality/reliability engineering. Stoddard achieved an MS in Systems Management and significant doctoral progress in reliability and quality management. He is a Fellow of the American Society for Quality and a senior member of the IEEE. Stoddard holds five ASQ certifications and is a Motorola-certified Six Sigma Master Black Belt. [rws@sei.cmu.edu]

Michael Konrad—is a Principal Researcher at the SEI, providing analytic support to various projects using statistics, machine learning, and most recently, causal learning. Since 2013, Konrad has contributed to research in requirements engineering, software architecture, and system complexity measurement. From 1998 to 2013, he contributed to CMMI in many technical leadership roles. Prior to 1998, Konrad was a member of the teams that developed the original Software CMM and ISO 15504. He is coauthor of the main Capability Maturity Model Integration for Development (CMMI-DEV) books. Konrad received his PhD in mathematics from Ohio University in 1978. [mdk@sei.cmu.edu]

Abstract

Correlation does not imply causation. Though this is a well-known fact, most analyses depend on correlation as proof of relationships that are often treated as causal. Causal search, also referred to as causal discovery, involves the application of statistical methods to identify causal relationships using conditional independences (and/or other statistical relationships) within data. Though software cost estimation models use both domain knowledge and statistics, to date, there has yet to be a published report describing the evaluation of a software dataset using causal search. In a previous paper, the authors ran a PC causal search algorithm on Unified Code Count's (UCC's)[1] dataset of maintenance tasks and compared them to correlation test results. This paper builds on the previous paper to introduce causal discovery to software engineering research by exploring additional causal search algorithms (PC-Stable, fast greedy equivalent search [FGES], and fast

[1] http://ucc.usc.edu

adjacency skewness [FASK]) and comparing their results to the traditional multi-step regression analysis.

Introduction

Though analysts seek causation, "data, on their own, only communicate associations" (Elwert, 2013). Two variables with high correlations, or associations, may have causal or non-causal relationships (Elwert, 2013). Correlations, by themselves, cannot generally determine which relationships are causal (Cook, Campbell, & Shadish, 2002). Hence, statisticians emphasize correlation does not imply causation. Judea Pearl (2001) stated, "Behind every causal conclusion there must lie some causal assumption that is not testable in observational studies," suggesting that one cannot gain complete causal knowledge through only observational studies or data alone. Experiments that manipulate causal variables can help identify causal relationships (Cook et al., 2002). However, there are "practical and ethical considerations that limit the application of controlled experiments in many cases" (Spirtes, 2010).

In software engineering, it is impractical to run software development effort experiments. Developing the same software with various personnel attributes or manipulating the project's size and product attributes with the same team would become very expensive for software development teams. Hence, most data in software engineering are observational versus from controlled experiments. Though causal inference (which includes causal search) is characterized as "finding answers to questions about the mechanisms by which variables come to take on values, or predicting the value of a variable after some other variable has been manipulated" (Spirtes, 2010), estimation model developers have not used such statistical methods to confirm or reject causal assumptions. Until recently, estimation model developers lacked the tools to systematically evaluate whether the size and effort drivers identified by software project experts do, in fact, have significant (both in the statistical and effect size sense) causal effects on effort and, therefore, should be preferred for selection over other candidate drivers for effort estimation.

Using the theory of causal inference, one can now perform a causal search "based on unmanipulated data" (Spirtes, 2010). According to Spirtes, causal inference consists of two parts: "search for a causal graph, and estimation of the free parameters from sample data and the causal graph." In this paper, the authors run causal search algorithms that return causal graphs (the first part defined by Spirtes) with the intent to use the causal discovery results to estimate the parameters (the second part defined by Spirtes).

Previously, the authors reported the differences between correlation test results and commonly-used causal search algorithm results on UCC's dataset of maintenance tasks (Hira et al., 2018). This paper builds upon the first paper in several ways. First, the authors take the opportunity to discuss the differences between constraint-based versus score-based search algorithms as both are used in this paper. Second, the authors run additional causal search algorithms (of both types) in order to better illustrate some of the other search algorithms and their capabilities given a relatively small dataset, and to further confirm or modify findings of the first paper in terms of factors having direct or indirect causal effects on software development efforts. Third and last, the authors more clearly contrast the traditional approach of correlation and multiple regression with that of the causal search approach. Interestingly, not all causal search algorithms will necessarily result in the same set of causal relationships.

Unified Code Count

Development Environment and Dataset Summary

The University of Southern California (USC) maintains Unified Code Count (UCC), a small- to medium-sized tool that provides source lines of code (SLOC) counting metrics for about 30 programming languages, such as logical SLOC (Park, 1992) and cyclomatic complexity (McCabe, 1976). UCC is an object-oriented project written in C++, and each year, development teams work on enhancements or extensions that range in size from 45 to 1,425 equivalent source lines of code (ESLOC, defined in the subsection titled Dataset Attributes). USC releases an updated UCC annually with new language parsers, additional features, and/or additional metrics. UCC's current dataset covers recent projects consisting of six new language parser projects, five new features projects (such as GUI interface or additional input options), and 19 projects researching and adding cyclomatic complexity metrics to UCC's outputs. Data for analysis came from the developed code, weekly timesheets, test case documentation with corresponding test data, and explanatory reports summarizing the steps taken and the results of projects that began and completed between 2010 and 2014.

Dataset Attributes

Along with size and effort, UCC's dataset contains project and personnel characteristics as defined by Constructive Cost Model® (COCOMO®) II. COCOMO® II is a parametric software cost estimation model that requires size, product, and personnel attributes as input, and outputs the estimated effort in Person-Months (PM). A summary of the attributes included in UCC's dataset of maintenance tasks are as follows, where items numbered 6 to 13 are effort factors defined by COCOMO® II:

1. Equivalent Logical Source Lines of Code (ESLOC): Logical SLOC was developed and defined by the Software Engineering Institute (SEI) to standardize SLOC measurement (Park, 1992). Equivalent logical SLOC (ESLOC) makes modifications to reused code equivalent to new code, which has been calculated using Nguyen's modification to COCOMO® II's reuse model (Nguyen, 2010; Boehm, Madachy, & Steece, 2000).
2. IFPUG Function Points (FPs): Each project in the dataset has been sized using version 4.3.1 of IFPUG's FPs method.
3. IFPUG Software Non-Functional Assessment Process (SNAP) Points: Each project in the dataset has been sized using version 2.3 of IFPUG's SNAP method.
4. COSMIC Function Points (CFPs): Each project in the dataset has been sized using version 4.0 of COSMIC's FPs (CFPs) method.
5. Total Effort: Effort in terms of hours, including time spent on training, requirements gathering, coding, testing, and documenting.
6. Applications Experience (APEX): Most of the personnel that join UCC's development team do not have prior industry experience in similar application types, though they have sufficient computer science education. Therefore, the Low rating is used for APEX on all data points (as opposed to Very Low or Nominal, etc.).
7. Platform Experience (PLEX): The development personnel have little experience in the graphical interface platform and building cross-platform applications. Hence, a Low rating for PLEX best describes the development teams for all data points.

8. Use of Software Tools (TOOL): Currently, the UCC development environment only uses tools corresponding to the Very Low rating.
9. Personnel Continuity (PCON): COCOMO® II's highest personnel turnover rating is 48% per year. On average, UCC faces a 90% turnover over four months. Two of the authors had to adjust the rating value for this parameter in a previous study (Hira, Sharma, & Boehm, 2016).
10. Documentation Match to Life-Cycle Needs (DOCU): All teams are required to document the requirements of the project and summarize the work completed and decisions made. However, a couple of projects had substantially more documentation with respect to the requirements and earned a High rating for DOCU, and one project had less than the required documentation and earned a Low rating.
11. Analyst Capability (ACAP): "Analysts are personnel who work on requirements, high-level design, and detailed design" (Boehm et al., 2000). Some teams showed high and very high analyst capability, whereas very few teams showed low capability in analysis and design.
12. Programmer Capability (PCAP): Though most of the programming personnel were sufficiently capable, some developers had especially proficient programming skills.
13. Product Complexity (CPLX): Although most of the UCC maintenance tasks fall within the Nominal or Average rating for CPLX, some were rated Low based on the types of control operations and computation operations implemented.

The applicable COCOMO® II effort factors and their corresponding values for each of the projects were evaluated by reviewing the source code, deliverables, and Hira's weekly notes on teams' progress at the time of data collection. Since items 6–9 are rated the same across all data points, they are not included in the causal search algorithm runs.

Causal Search Algorithms Further Explained

Causal search algorithms typically take a dataset and hyper-parameters governing the search and output a set of graphs whose nodes are the variables appearing in the dataset (and depending on the algorithm, may include latent variables) and whose edges indicate some kind of direct causal connection between the pair of nodes they join. (Optionally, the algorithms also take sets of required and prohibited direct causal relationships between pairs of variables, which the user can use to encode the results of experiments or elicited domain knowledge.) There are many variations on this simple theme among the dozens of search algorithms, but in terms of understanding how they function, and thus something of their relative strengths and limitations, it will help to organize them into two broad categories: constraint-based and score-based search algorithms (Spirtes, 2010).

For both categories of searches, pointwise-consistent convergence has been proven. In other words, with increasing datasets drawn from the same population, the algorithm will eventually find the correct causal graph(s). Unfortunately, uniformly-consistent convergence has not been proven, which could provide the rate of convergence and level of confidence for particular causal relationships (Spirtes, 2010).

Constraint-Based Search Algorithms

The first practical constraint-based search algorithm developed was the PC search algorithm, the algorithm used in the previous paper by the authors (Hira et al., 2018). In its simplest form, constraint-based search involves two stages: Adjacency Search and Edge Orientation. Starting with a complete undirected graph, edges are iteratively removed by testing for the conditional independence of joined nodes given a subset of neighboring nodes. If conditional independence is found, the edge is removed and the conditioning set employed is noted for later use in the Edge Orientation stage. This process is continued until all edges have been evaluated in this way. The result of this first stage, Adjacency Search, is thus an undirected graph. Edge Orientation starts with an undirected graph and iteratively orients edges according to a few rules that make use of the conditioning sets noted during the Adjacency Search stage. The result is an equivalence class of graphs, called a Markov Equivalence Class (MEC), rather than a single graph, because it is often impossible to determine the orientation of all the edges in the undirected graph that is output from the Adjacency Search stage (Spirtes, 2010).

For example, suppose we have a dataset featuring three variables, X1, X2, and X3, and the only independence discovered among them is X1 is independent of X3 conditioned on X2. We also suppose we have no additional knowledge to encode about X1, X2, and X3, only the dataset. Then the Adjacency Search stage will output the undirected graph X1 – X2 – X3 (as well as some kind of note that the conditioning set that made X1 and X3 independent is {X2}). Then, given that particular independence, it necessarily follows that during the Edge Orientation stage, the direction of orientations for the edges of this undirected graph will not be able to be determined uniquely. Indeed, any of the following three pairs of orientations are valid, constituting the MEC: { X1 → X2 → X3, X1 ← X2 ← X3, X1 ← X2 → X3 }. Note that the following sequence of orientations is not part of the MEC: X1 → X2 ← X3. This type of relationship among variables is referred to as a collider. In a collider, the independence conditioning set is the empty set, because X1 is independent of X3 unconditionally. Hence, if the only independence found among X1, X2, and X3 is that X1 and X3 are unconditionally independent, then the MEC would consist of exactly one graph: X1 → X2 ← X3. Thus, colliders provide important clues for orienting edges during the Edge Orientation Stage (Spirtes, 2010).

While the idea of a set of graphs being the output of a causal search may disappoint, it is important to note that all graphs in an MEC have the same set of colliders and are built on top of the same undirected graph. Thus, all graphs in an MEC manifest the same set of correlations present in the dataset but may vary as to the causal orientations for some edges.

The hyper-parameters of a constraint-based search algorithm typically include but are not limited to

1. type of independence test used (e.g., Fisher Z Test, Conditional Correlation Test)
2. confidence level for conditional independence testing (cutoff for p values)
3. maximum size of condition set

Constraint-based search makes very significant use of independence tests, and what type of independence test to use for what purpose is an ongoing area of research to help achieve both accuracy and speed across a range of different assumptions (e.g., non-Gaussian univariate distributions). Another area of research is how to achieve both accuracy and speed in determining edge orientations. The need to conduct search on enormous datasets and a very large number of variables (e.g., one million cases and tens of thousands of variables) motivates research for better algorithms.

Score-Based Search Algorithms

To those readers more familiar with machine learning, score-based search algorithms employ a familiar mechanism: a maximum likelihood-based score (such as Bayesian information criterion [BIC]). Like constraint-based search, there are two stages, both are iterative, and in each iteration of each stage there is both a currently-considered MEC (see the previous section for an explanation of this term, but it is important to note that all graphs in an MEC share the same underlying undirected graph and the same colliders) and a set of neighboring MECs and that each either possesses an additional edge (first stage of search) or has one edge removed (second stage of search; Spirtes, 2010).

In each iteration of the first stage, from the currently-considered MEC, the algorithm scores all neighboring MECs that have one additional edge. The best-scoring neighboring MEC then becomes the currently-considered MEC in the next iteration. The algorithm continues to iterate, building graphs one edge at a time, until a better score cannot be attained. In the second stage, the algorithm proceeds similarly but in reverse, considering only those MECs having one edge removed. Again, the algorithm halts when no better score can be attained, and the resulting MEC is returned as the output (Spirtes, 2010).

The advantages of score-based search algorithms over constraint-based search algorithms are as follows:

- They can obtain more accurate adjacencies.
- They will typically output only directed or undirected edges. There are no bi-directed edges because equivalence class scoring will almost always favor one orientation over the other (or make the rarely-required arbitrary selection). A bi-directed edge signifies that there may be an unmeasured variable affecting the two variables.

A limitation of score-based search algorithms is that they can be slow and might not scale as well as constraint-based searches.

Applying Three Additional Search Algorithms (Beyond the First Paper)

As mentioned earlier, the authors ran the PC search algorithms earlier (Hira et al., 2018). PC is often cited by data analysts experienced with performing causal search as the “go to” causal search algorithm, given its generally high accuracy (generally better at getting adjacencies right) and high scalability. In this paper, the authors run three additional search algorithms on the same dataset: PC-Stable (constraint-based), FGES, and FASK (both score-based), in order to better illustrate some of the other search algorithms and their capabilities given a relatively small dataset.

Here is a short description of the particular niche where each algorithm has some relative strengths over the others:

1. PC-Stable is a variant of PC that addresses the problem that the causal graphs output by many search algorithms depend on the order of the variables within the dataset (Colombo & Maathuis, 2014).
2. FGES is a score-based search algorithm and perhaps best qualifies as the causal-search data analysts’ favorite “go to” search algorithm after PC (particularly if they do not like to deal with bi-directed edges; Center for Causal Discovery, 2017).

3. FASK, a score-based search algorithm, addresses the problem that oftentimes variables have asymmetric distributions. The previously listed algorithms assume that direct causal relationships are linear up to an error term that is Gaussian, whereas FASK actually exploits skew in error terms' distributions to determine how best to orient edges (Sanchez-Romero et al., 2018).

Tetrad

As part of a National Institutes of Health (NIH) Big Data initiative, the University of Pittsburgh, Carnegie Mellon University (CMU), and Pittsburgh Supercomputing Center serve as founding members of the Center for Causal Discovery (CCD). The CCD develops and maintains causal algorithms, software, and tools, including the Tetrad[2] program with its GUI, API, and command-line interfaces (referred to as Tetrad in this paper). Tetrad allows users to run causal search algorithms on a dataset as well as estimate and evaluate parametric models. The authors loaded the data from the UCC's dataset (described in the section titled Unified Code Count) and ran causal search algorithms, which returned causal graphs representing the cause-effect relationships discovered in the dataset.

PC-Stable Causal Search Results

The PC-Stable causal search is run on UCC's dataset first with all size-related variables (ESLOC, CFPs, FPs, and SNAP), and second, individually with only one size-related variable at a time. The authors run the algorithm with the simultaneous inclusion of all size-related variables to determine whether the size metrics might work together to characterize total effort, as represented by the variable TotalEffort.

The graphical search results from PC-Stable are displayed in Figures 1–5, which conclude that product complexity (CPLX) is consistently identified as the single causal factor of TotalEffort except for the search with all size metrics included. When all size metrics are included, the size metric COSMIC Function Points (CFPs) is identified as the sole causal factor on TotalEffort. Interestingly, ESLOC, FPs, and SNAP have an undirected edge with CPLX, which thus has a potential causal effect on TotalEffort through CPLX (Figures 2–4), while CFPs' role with CPLX is flipped (see Figure 5). Lastly, each of the figures shows an undirected edge between analyst capability (ACAP) and programmer capability (PCAP).

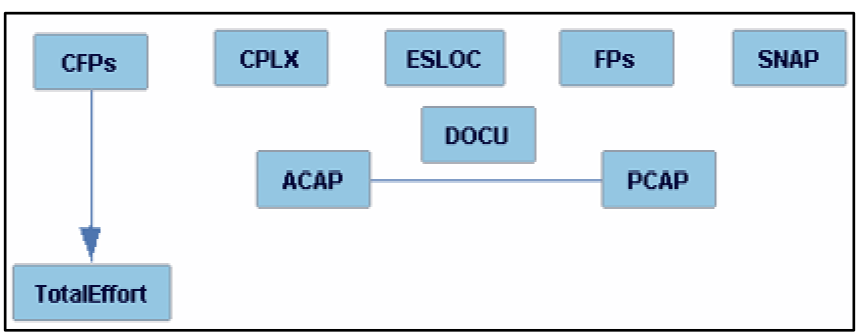


Figure 1. PC-Stable Result When All Size Metrics Are Included

[2] https://github.com/cmu-phil/tetrad

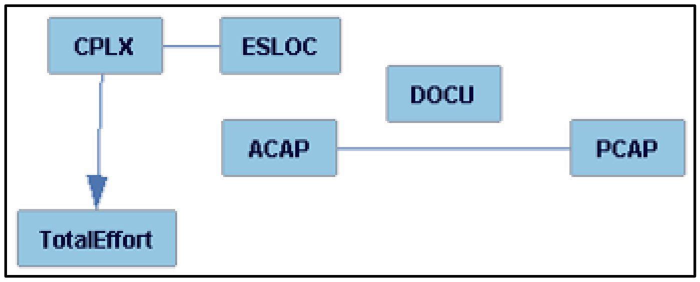


Figure 2. PC-Stable Result When Only ESLOC Is Included as a Size Metric

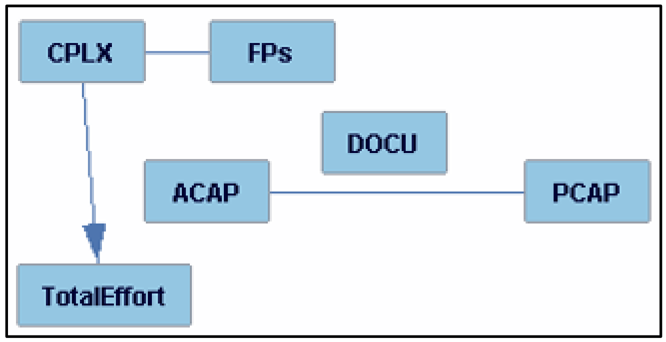


Figure 3. PC-Stable Result When Only IFPUG FPs Is Included as a Size Metric

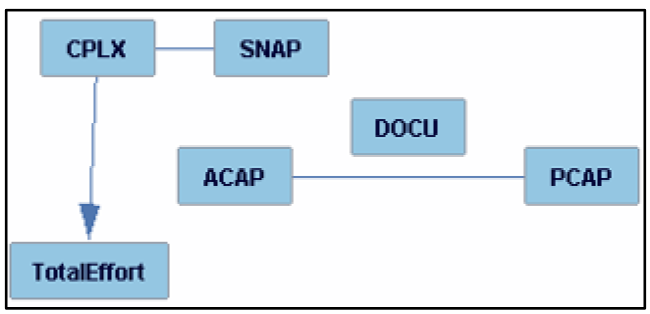


Figure 4. PC-Stable Result When Only IFPUG SNAP Is Included as a Size Metric

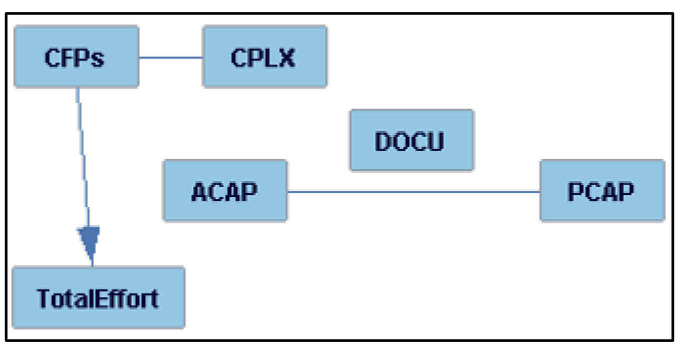


Figure 5. PC-Stable Result When Only CFPs Is Included as a Size Metric

FGES Causal Search Results

The graphical search results from running FGES may be seen in Figures 6–10. The causal graphs returned by FGES have some different conclusions than those returned by PC-Stable. The PCAP factor shows up in all cases as having a causal effect on TotalEffort (and an undirected edge with ACAP), while CFPs also have a causal factor on TotalEffort (consistent with PC-Stable results). Only with ESLOC is there an undirected edge between the size metric and CPLX (see Figure 7; consistent with PC-stable result).

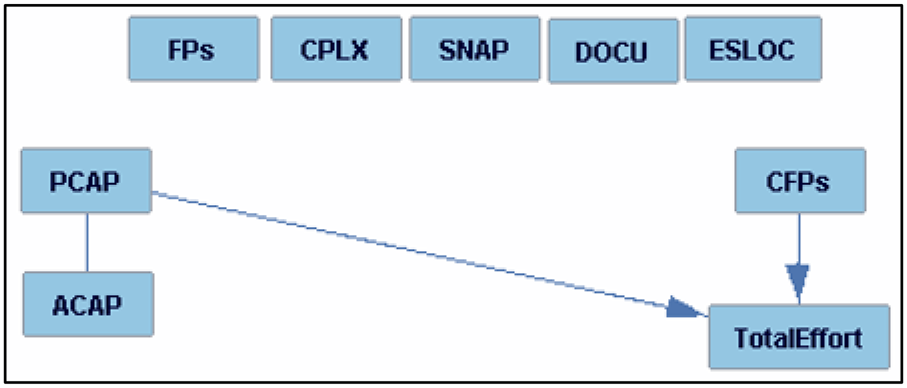


Figure 6. FGES Result When All Size Metrics Are Included

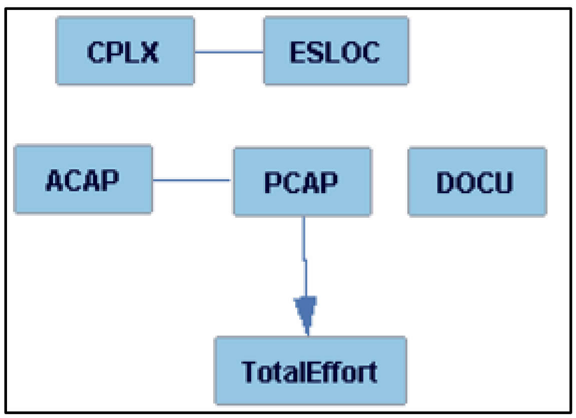


Figure 7. FGES Result When Only ESLOC Is Included as a Size Metric

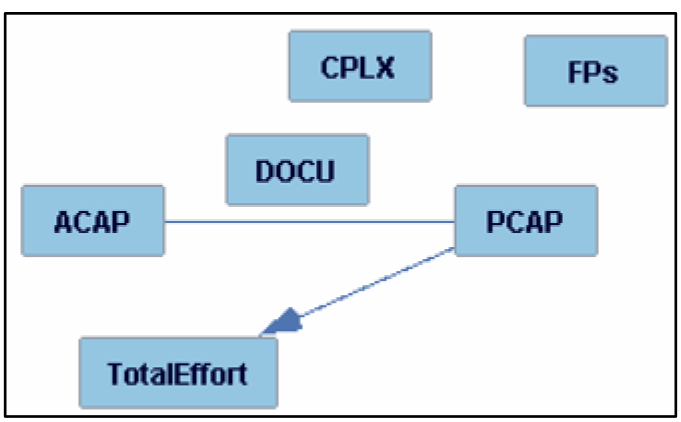


Figure 8. FGES Result When Only IFPUG FPs Is Included as a Size Metric

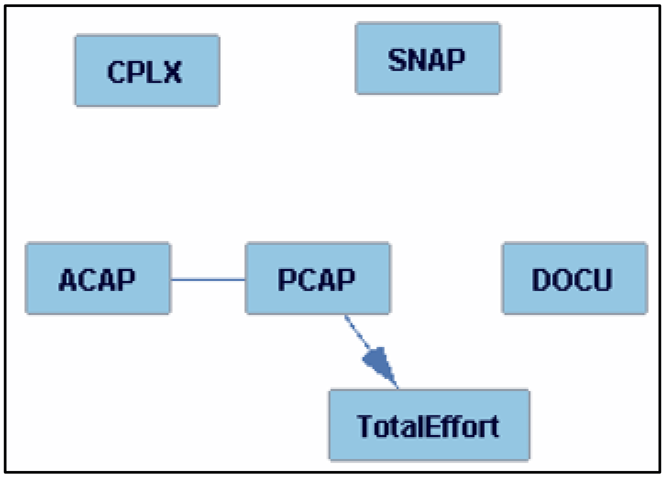


Figure 9. FGES Result When Only IFPUG SNAP Is Included as a Size Metric

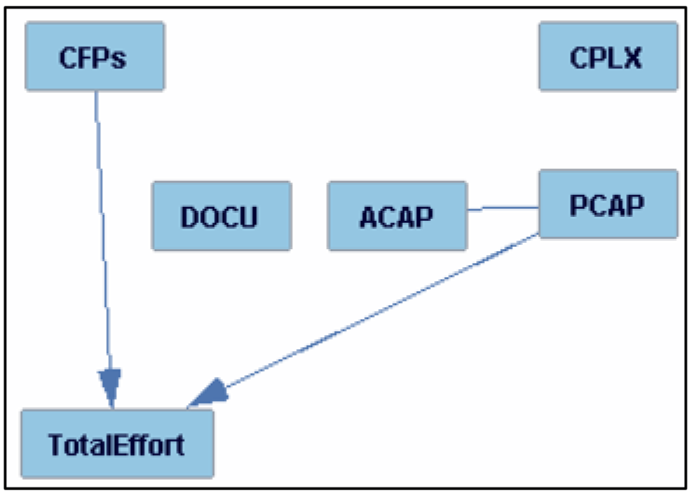


Figure 10. FGES Result When Only CFPs Is Included as a Size Metric

FASK Causal Search Results

Before conducting the FASK causal search, the authors revisited the distribution of each factor as FASK achieves improved edge orientation when distributions are skewed. Table 1 summarizes the results of checking for skewness. Since most variables are skewed, the authors proceeded to run FASK, aware that the lack of full skewness might render some edge orientations incorrectly.

Table 1. Summary of Variables' Skewness

Variable	Skewed
CPLX	YES
CFP	YES
FP	NO
SNAP	YES
ESLOC	YES
ACAP	NO
PCAP	NO
DOCU	YES
Total Effort	YES

The graphical search results from FASK, Figures 11–15, show that the algorithm returned some similar and some very different results compared to both PC-Stable and FGES. Most interestingly, the algorithm returns the factor Documentation Match to Lifecycle Needs (DOCU) as having a causal effect on TotalEffort, along with some of the size metrics. Except when all size metrics are included in the analysis, DOCU is identified as having a causal effect on ACAP, and there is an undirected edge between DOCU and PCAP. Additionally, SNAP is identified as having causal effects on ACAP and DOCU, and FPs as having a causal effect on DOCU.

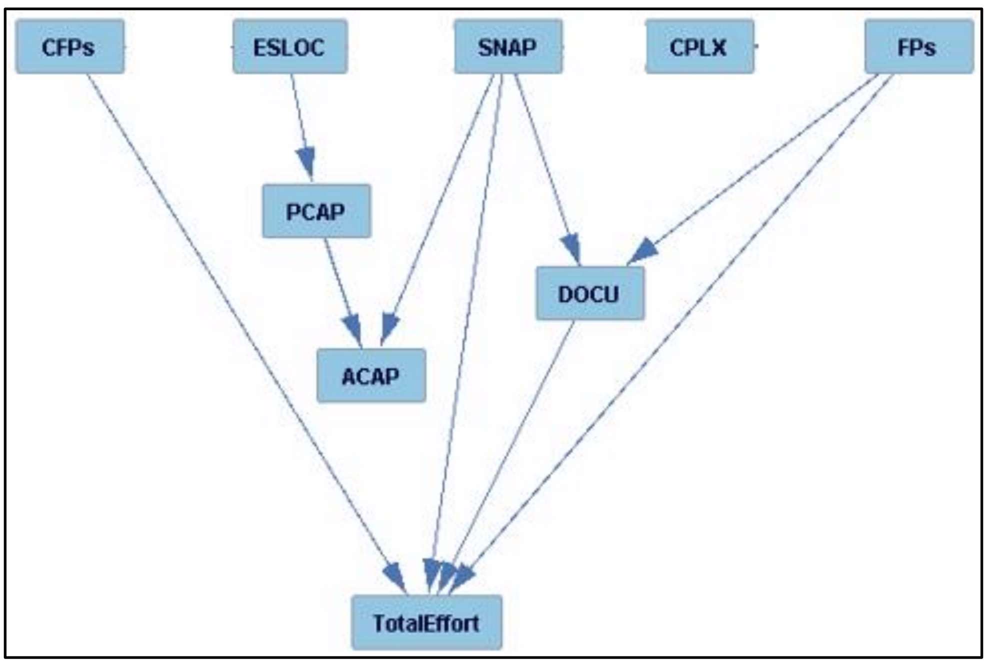


Figure 11. FASK Result When All Size Metrics Are Included

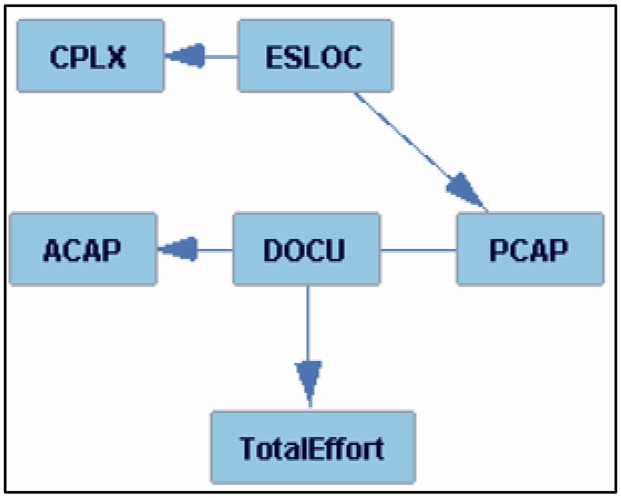


Figure 12. FASK Result When All Size Metrics Are Included

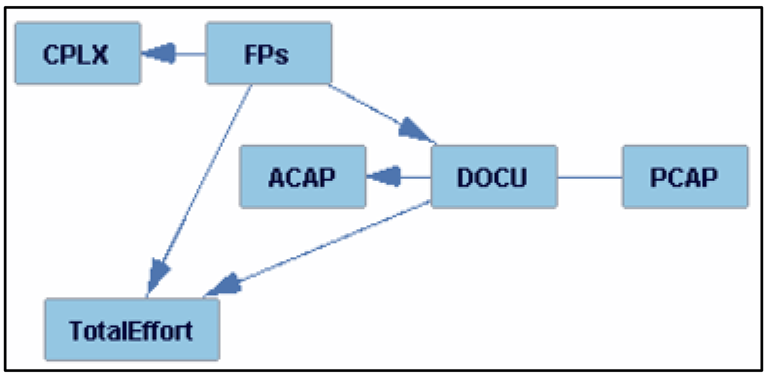


Figure 13. FASK Result When All Size Metrics Are Included

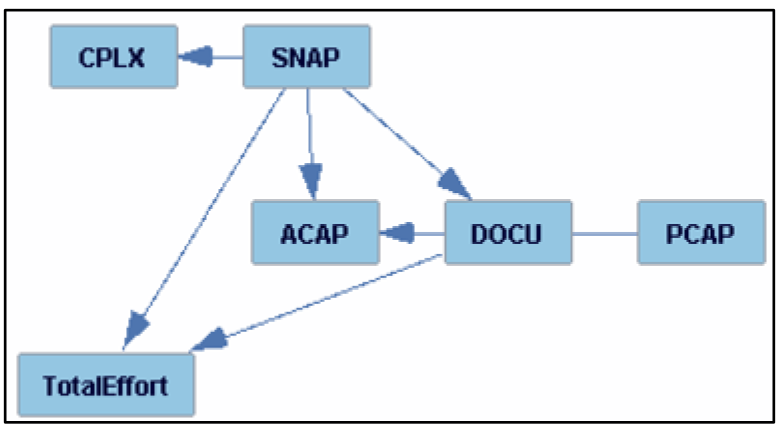


Figure 14. FASK Result When All Size Metrics Are Included

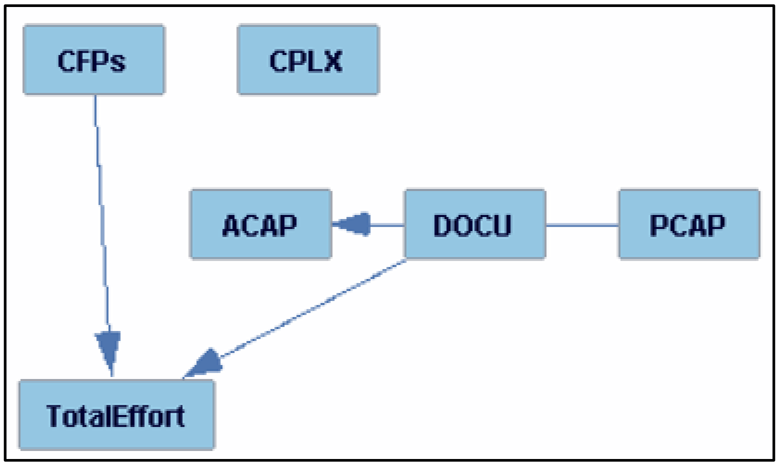


Figure 15. FASK Result When All Size Metrics Are Included

Traditional Stepwise Multiple Regression

In order to determine the most influential factors and an acceptable prediction model, data analysts typically use the stepwise multiple regression approach. As displayed in Table 2, stepwise regression produces results indicating that PCAP and CFPs remain significant in predicting TotalEffort. Although models with three or more factors achieve a reasonably high Adjusted R-Squared, the simplified, two-factor model with PCAP and CFPs produces similar Adjusted R-Squared results. The details of the resulting prediction model are detailed in Table 3, and Figure 16 displays that the residuals of the model are normally distributed and do not have multi-collinearity.

Table 2. Stepwise Regression Results Summary With Variables Considered During Each Run

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	CFPs	ESLOC	FPs	SNAP	ACAP	PCAP	DOCU	CPLX
1	57.6	56.1	51.0	35.0	231.50						X		
1	54.6	53.0	47.9	39.3	239.65	X							
2	84.9	83.8	78.3	-2.3	140.54	X					X		
2	78.5	76.9	67.4	6.9	167.87				X		X		
3	85.3	83.6	76.1	-0.9	141.35	X					X		X
3	85.1	83.4	77.2	-0.6	142.30	X		X			X		
4	85.4	83.0	73.0	1.1	143.98	X			X		X		X
4	85.3	83.0	68.6	1.1	144.13	X				X	X		X
5	85.4	82.3	61.9	3.0	146.83	X			X	X	X		X
5	85.4	82.3	67.0	3.0	146.89	X	X		X		X		X
6	85.4	81.6	55.2	5.0	149.92	X	X		X	X	X		X
6	85.4	81.6	59.9	5.0	149.99	X			X	X	X	X	X
7	85.4	80.8	53.0	7.0	153.29	X	X		X	X	X	X	X
7	85.4	80.8	52.0	7.0	153.29	X	X	X	X	X	X		X
8	85.4	79.8	49.5	9.0	156.90	X	X	X	X	X	X	X	X

Table 3. Coefficients and Corresponding P-Values for the Regression Model Predicting TotalEffort

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-1791	244	-7.33	0.000	
PCAP	1948	264	7.38	0.000	1.12
CFPs	76.3	10.9	7.00	0.000	1.12

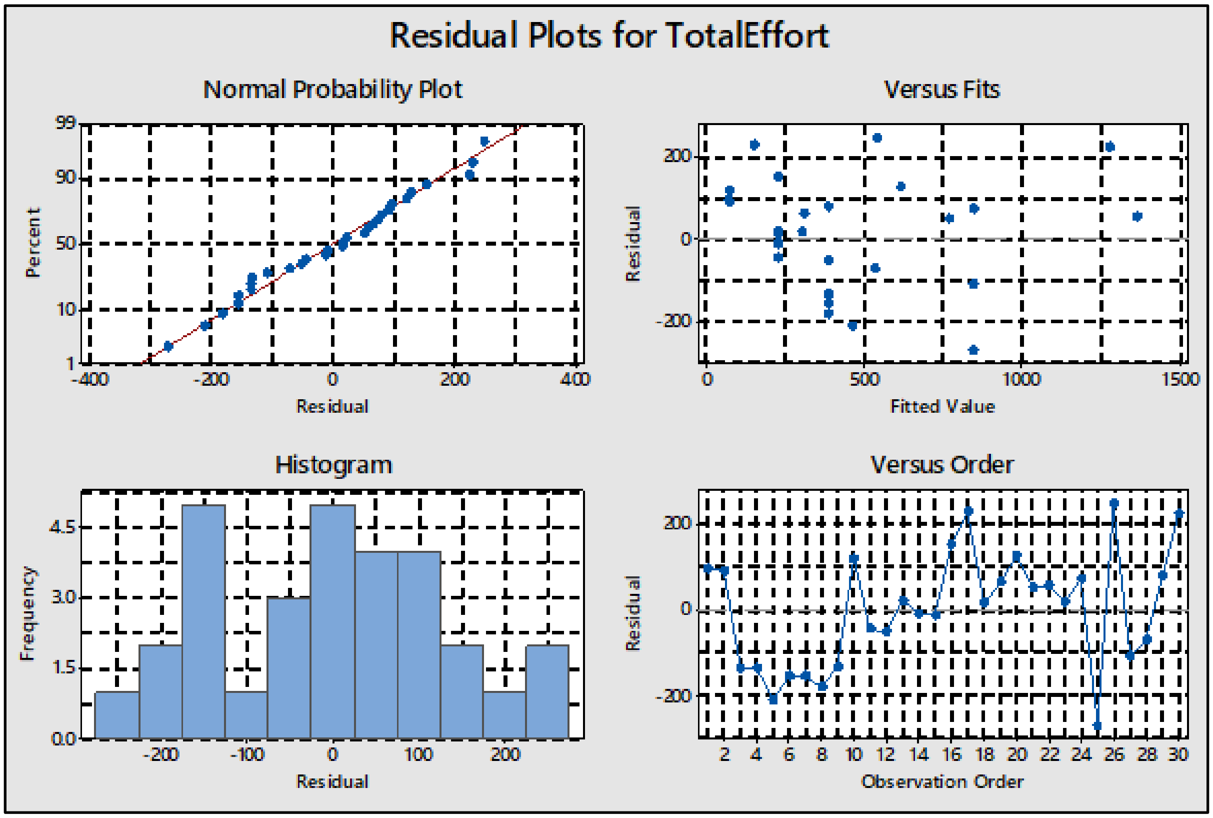


Figure 16. Residual Plots for the Regression Model Predicting TotalEffort Displaying the Fulfillment of Linear Regression Assumptions

Summary of Results

Table 4 depicts a summary of the results of the PC (Hira, Sharma, & Boehm, 2018), PC-Stable, FGES, and FASK causal searches in which the direct and indirect causal factors are noted. Contrary to the traditional stepwise multiple regression approach which identified PCAP and CFPs as significant factors of TotalEffort, the causal search has uncovered direct causal evidence, depending on the algorithm, of FP, SNAP, CFP, CPLX, PCAP, and DOCU. The algorithms also identified ESLOC, FP, SNAP, ACAP, and PCAP as indirect causal drivers of TotalEffort.

There remains a number of ways in which this summary may be interpreted. One approach would be to look for the complete absence or presence of causal relationships as noteworthy new knowledge. Another approach would be to look for direct and indirect causal factors showing up across a majority of causal algorithms. Lastly, one could choose to be much more inclusive and look at all direct and indirect factors that show up at least once.

Acknowledging the small size of the data set and the impact of the small sample size on the causal search results, one can still see that causal search can inform a researcher of alternative independent factors that may actually be the causal force on the dependent factor. As such, given the small sample size, there is evidence that CPLX, PCAP, DOCU, and several size measures (CFP, FP, and SNAP) may be considered as direct causal factors. It is interesting to also note that PCAP only showed up as a direct causal factor on TotalEffort using the FGES score-based algorithm, while DOCU showed up as a direct causal factor using FASK. Among the size measures, the strongest candidate for a direct causal factor for TotalEffort would seem to be CFP. In conclusion, these causal search results should motivate the researcher to collect additional data and continue analyzing beyond what a traditional regression approach might offer toward the ultimate goal of estimating a quantified actionable model for TotalEffort—an actionable model that is suitable for use in deciding how to intervene and change a project's course toward achieving desired target outcomes. Correlation alone really doesn't provide sufficient insight.

Table 4. Identified Direct and Indirect Causes of TotalEffort by Algorithm and Scenario

	Direct Cause of Total Effort								Indirect Cause of Total Effort							
Algorithm	ESLOC	FP	SNAP	CFP	CPLX	ACAP	PCAP	DOCU	ESLOC	FP	SNAP	CFP	CPLX	ACAP	PCAP	DOCU
PC-AllSizeMetrics				YES			YES				YES			YES		
PC-ESLOC							YES							YES		
PC-FP							YES							YES		
PC-SNAP			YES				YES							YES		
PC-CFP				YES			YES									
PC-Stable-AllSizeMetrics				YES			YES									
PC-Stable-ESLOC					YES				YES							
PC-Stable-FP					YES					YES						
PC-Stable-SNAP					YES						YES					
PC-Stable-CFP				YES									YES			
FGES-AllSizeMetrics				YES			YES							YES		
FGES-ESLOC							YES							YES		
FGES-FP							YES							YES		
FGES-SNAP							YES							YES		
FGES-CFP				YES			YES							YES		
FASK-AllSizeMetrics		YES	YES	YES				YES								
FASK-ESLOC								YES	YES						YES	
FASK-FP		YES						YES							YES	
FASK-SNAP			YES					YES							YES	
FASK-CFP				YES				YES							YES	

Note. Black cells are Not Applicable as factor not present in model.

Implications for Acquisition Research

The authors believe the implications of adding causal searches and ultimately, causal estimations of causal influence, could transform acquisition research in the following ways:

1. The need to pursue expensive and, more likely, prohibitive experiments of acquisition factors could be obviated by use of causal methods appropriate for observational data. (Or at a minimum, causal research findings should be used to more efficiently focus such experiments.)
2. Revisiting research data affiliated with the acquisition research arena could quickly help filter likely causal factors from the merely correlated factors, thereby reducing researcher distraction and accelerating progress in acquisition research.
3. Different acquisition researchers could more easily begin to unite causal conclusions into a more holistic causal model of acquisition research outcomes.
4. The acquisition research community could identify and prioritize constrained research funding towards causal research outcomes worthy of investment in repeatability and reproducibility studies.

5. Causal research findings would be more confidently tested by real-world acquisition program interventions and risk less in wasteful interventions.
6. SEI's acquisition experts indicate specific acquisition outcomes and associated factors worthy of causal learning include
    a. delivering required warfighter capability at an affordable price and on schedule, driven by potential factors such as lack of competencies in the workforce, greater system complexity, and underestimation of cost and schedule,
    b. unrealistic assumptions made by decision-makers very early in the acquisition lifecycle,
    c. program acquisition strategy (and changes in the strategy), the program structure, and the technical challenges facing the program,
    d. leadership incentives to conduct critical thinking and apply evidence-based knowledge and practice early in a program,
    e. people, process, requirements, and incentives in general,
    f. the overall contracting process in context of changes that are often difficult to make and take a lot of time, leading to issues on the contractor side,
    g. the technical strength of the government team,
    h. measurement of agile versus more traditional development, and
    i. measured effectiveness of the Department of Defense (DoD) 5000 mandated reviews.

Next Steps

The authors welcome research collaboration making use of causal search and estimation, especially acquisition research focused on impacts to acquisition cost and schedule. Research collaboration of this nature would complement existing research funded through the SEI as part of a three-year (FY 2018–2020) DoD research project titled "Investigating the Feasibility of an Integrated Causal Model for Software Cost Control (SCOPE)." Many forms of acquisition research collaboration exist, including (1) providing subjective and objective research data, (2) connecting our research team with acquisition managers and organizations who might provide research data, (3) helping to identify the nature of acquisition outcome measures worthy of study along with insights to the potential causal factors of those outcomes, and (4) learning how to conduct causal search and estimation and working with our research team as a direct contributing member or as a reviewer of causal results.

References

Boehm, B. W., Madachy, R., & Steece, B. (2000). *Software cost estimation with Cocomo II with Cdrom.* Prentice Hall.

Center for Causal Discovery. (2017). Fast greedy equivalence search (FGES) algorithm for continuous variables. Retrieved from http://www.ccd.pitt.edu/wiki/index.php?title=Fast_Greedy_Equivalence_Search_(FGES)_Algorithm_for_Continuous_Variables

Colombo, D., & Maathuis, M. H. (2014). Order-independent constraint-based causal structure learning. *Journal of Machine Learning Research, 15*(1), 3741–3782.

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Elwert, F. (2013). Graphical causal models. In *Handbook of causal analysis for social research* (pp. 245–273). Dordrecht: Springer.

Hira, A., Sharma, S., & Boehm, B. (2016, May). Calibrating COCOMO® II for projects with high personnel turnover. In *Proceedings of the International Conference on Software and Systems Process* (pp. 51–55). ACM.

Hira, A., et al. (2018). Preliminary causal discovery results with software effort estimation data. In *Proceedings of the 11th Innovations in Software Engineering Conference* (pp. 1–11). ACM.

McCabe, T. J. (1976). A complexity measure*. IEEE Transactions on Software Engineering, 4,* 308–320.

Nguyen, V. (2010, September). Improved size and effort estimation models for software maintenance. In Pro*ceedings of the 2010 IEEE International Conference on Software Maintenance* (ICSM; pp. 1–2). IEEE.

Park, R. E. (1992). *Software size measurement: A framework for counting source statements* (No. CMU/SEI/92-TR-20). Pittsburgh, PA: Carnegie-Mellon University, Software Engineering Institute.

Pearl, J. (2001). Causal inference in the health sciences: A conceptual introduction. Health *Services and Outcomes Research Methodology, 2*(3–4), 189–220.

Sanchez-Romero, R., Ramsey, J. D., Zhang, K., Glymour M. R. K., Huang, B., & Glymour, C. (2018). Causal discovery of feedback networks with functional magnetic resonance imaging. Retrieved from https://www.biorxiv.org/content/early/2018/01/10/245936

Spirtes, P. (2010). Introduction to causal inference. *Journal of Machine Learning Research, 11*(May), 1643–1662.

Acknowledgments

This material is based upon work supported in part by Cyber Security and Information Systems Information Analysis Center (CSIAC) and in part upon work funded and supported by the DoD under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the SEI, a federally-funded research and development center. The authors would like to thank Michael McClendon and Julie Cohen (both of the SEI) for their acquisition insight to high value acquisition targets of causal learning described at the end of the paper. The authors would also like to thank David Zubrow (SEI) for his encouragement and support and for sharing his insights for the work in this paper. Additionally, the authors thank David Danks, Kun Zhang, Madelyn Glymour, and Joe Ramsey for their help in understanding causal search, the search algorithms, and the Tetrad tool.

Panel 26. Enhancing Supply Chain Effectiveness

Thursday, May 10, 2018	
3:30 p.m. – 5:00 p.m.	**Chair: Lorna Estep,** Director of Resource Integration United States Air Force ***Literature Review: Metrics for Naval Humanitarian Assistance and Disaster Relief (HADR) Operations*** Aruna Apte, Naval Postgraduate School ***Navy Expeditionary Logistics*** Uday Apte, Naval Postgraduate School

Lorna Estep—is a member of the Senior Executive Service, is the Director of Resource Integration, Deputy Chief of Staff for Logistics, Engineering and Force Protection, Headquarters U.S. Air Force, Washington, D.C. She is responsible for the planning, programming and budgeting of weapons systems sustainment, equipment, and logistics and installations resource requirements. As part of the Air Force corporate structure, she monitors performance of operations and maintenance, working capital funds and investment programs; participates in program and financial review groups; and advocates for financial adjustments to optimize force readiness. She oversees preparation and defense of these Air Force programs to the Office of the Secretary of Defense, Office of Management and Budget, and Congress.

Ms. Estep started her career as a Navy logistics management intern. She has directed the Joint Center for Flexible Computer Integrated Manufacturing, was the first program manager for Rapid Acquisition of Manufactured Parts, and has served as Technical Director of Information Technology Initiatives at the Naval Supply Systems Command. In these positions she has developed logistics programs for the Department of Defense, implemented one of the first integrated and agile data-driven manufacturing systems, and directed the development of complex technical data systems for the Navy.

As the Director of Joint Logistics Systems Center, Ms. Estep had the duties of a commanding officer for a major subordinate command. In addition, she acted as the Logistics Community Manager, an emerging organization to coordinate and implement the revised Defense Department logistics strategy for achieving Joint Vision 2010 through modern information techniques and processes. She has also served as Chief Information Officer for the Naval Sea Systems Command in Arlington, Va.; Executive Director of Headquarters Materiel Systems Group at Wright-Patterson AFB; Deputy Director for Logistics Readiness at the Pentagon; and Executive Director, Air Force Global Logistics Support Center. Prior to her current assignment she was Deputy Director, Logistics, Air Force Material Command.

Literature Review: Metrics for Naval Humanitarian Assistance and Disaster Relief (HADR) Operation

Aruna Apte—is tenured Associate Professor in the Graduate School of Business and Public Policy (GSBPP), NPS. Dr. Apte received her PhD in Operations Research from Southern Methodist University. Her accomplishments include

- Over 50 research articles including 26 peer-reviewed journal articles (12 out of these are in HADR) and one patent
- Excellence in Research award in GSBPP, NPS, for her defense-focused research
- Selection to present her research on NPS day
- Briefing several admirals, diplomats, and senior governmental officials in HADR
- Outstanding Teaching award at the Cox School of Business, SMU
- Liskin award for outstanding teaching at GSBPP
- Hamming award for excellence in teaching at NPS

Abstract

We perform a literature review to develop a set of readiness metrics for humanitarian assistance and disaster relief (HADR) to be used by commanding officers to assess HADR readiness in the same way they are now able to assess combat readiness. The scope is at the strategic level with tactical and operational inputs.

Introduction

Humanitarian assistance and disaster relief (HADR) operations are part of the Cooperative Strategy for 21st Century Seapower of the United States (U.S.). In this research, we further investigate, through literature survey, whether any metrics can be defined and developed to enhance the efficacy and efficiency of HADR operations. Such measurement will be instrumental in successfully following a fundamental principle: “If we are going to do HADR anyway, then why not do it smartly.”

In the past two to three decades, the United States Navy (USN) has been the active and principal supplier of disaster relief due to its many unique and critical capabilities (Apte et al., 2013; Apte, Goncalves, & Yoho, 2016). Whether this effort will continue and be sustained in an environment of fiscal austerity and budget cuts is not given. Therefore, it is critical to identify resources the USN possesses, due to its core competencies and capabilities, that support humanitarian logistics, and to understand the USN’s readiness level to utilize these resources in the best possible way.

The United States Marine Corps (USMC) can rapidly respond to disasters because it maintains high levels of readiness on a constant basis. The USMC provides critical resources for these missions through their Marine Expeditionary Units (MEUs), which are flexible and adaptable enough to accomplish a wide range of operations, including non-combat missions (Apte & Yoho, 2014). Given the recent frequency of disasters around the world, it is probable that the occurrence of these events will continue, thus creating a demand for the relief capabilities. The MEUs have flexible and adept forces that can be deployed to austere environments while meeting urgent timelines (USMC, 2009).

Background

There have recently been significant suffering and casualties due to natural disasters across the world. Some governments offer humanitarian assistance. Figure 1 shows which donors provided the most humanitarian assistance in 2012. The United States spent almost twice as much as the next highest donor, the European Union (EU). Since 2008, the United States has spent substantial capital on humanitarian aid, as shown in Figure 2. When the USN steps in to help, the naval combatant commands, such as Pacific Command (PACOM) and Southern Command (SOUTHCOM), are the organizations that have to act.

After the 2010 Haiti earthquake, the functional organization and staffing of the COCOM had significant gaps in the ability to provide an effective and efficient response. A strategic plans officer for the UN said, “The military’s planning capability is not the most expensive part, but it is probably the most valuable. The international coordination structure would not have stood up if they weren’t there—we tapped into the Joint Task Force (JTF) planning capacity” (Joint Center for Operational Analysis [JCOA], 2010, slide 77). Given vast AORs, the number of disasters in the last decade, and the lack of lead time to prepare for relief for certain types of disasters (Apte, 2009), organizations such as the USN and the USMC need to have a playbook with readiness metrics. Therefore, three questions need to be answered:

1. How does an organization know when it is ready to respond to a disaster and whether it is capable of delivering relief?
2. What core competencies are these organizations exploiting to be ready for humanitarian missions?
3. What are the resources that can deploy the capabilities that support these core competencies?

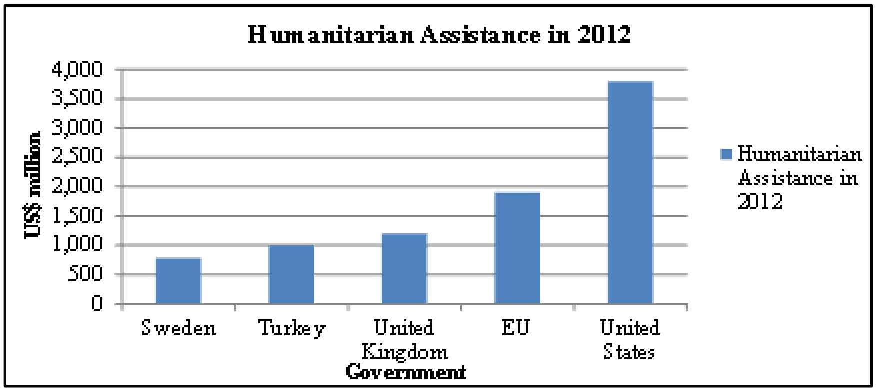


Figure 1. Humanitarian Assistance in 2012, Top Five Government Donors
(Global Humanitarian Assistance, 2013)

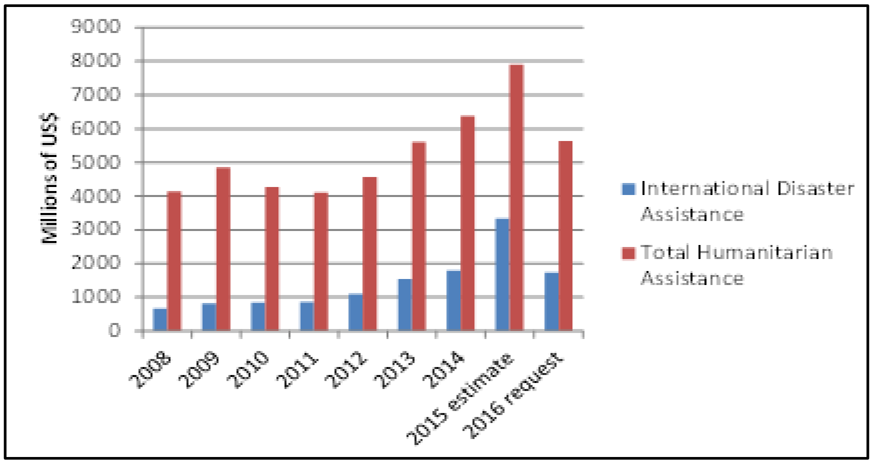


Figure 2. Humanitarian Aid by United States
(Margesson, 2015)

We studied over 80 documents, including peer-reviewed scholarly articles, government documents, white papers, research papers, and DoD briefings. This review helped us understand the definitions and descriptions of post-disaster performance indicators and pre-disaster readiness metrics. During the process, the literature was divided into four categories: disasters and lessons learned, civil and military collaboration, core competencies and capabilities, and challenges in humanitarian operations. These topics assist us in developing the path for recognizing readiness in humanitarian organizations. We follow the path to formulate a conceptual model for readiness assessment. We studied four disasters in detail for the lessons learned: the 2010 earthquake in Haiti, the 2011 earthquake and tsunami in Japan, 2013's Typhoon Haiyan (Yolanda) in the Philippines, and the 2014 earthquake in Nepal. The literature review process helped us identify a framework for readiness metrics in naval humanitarian operations based on the core competencies of the USN and the USMC.

Literature Review

Many humanitarian organizations (HOs) respond to the disasters around the globe. In this research, we define HOs as those organizations that provide humanitarian relief, whether military or non-military (NMO) and whether government or non-government (NGO). These organizations have core capabilities and competencies (Apte et al., 2016) from which they provide humanitarian assistance. When a disaster strikes, the host nation requests outside assistance, if needed. When requested, the USN and the USMC, under the guidance of USAID, get deployed for HADR. Other HOs also provide assistance based on their core competencies and capabilities. Many times, the relief falls short of meeting the demand. The reasons why this happens will help us understand how to measure the readiness that is embedded into the core capabilities and competencies of the organizations. Relief falls short for many possible reasons:

- The disaster was massive in scope and scale.
- The distribution and transportation of critical supplies and services was not well-managed; hence, the affected region did not receive necessary supplies.
- Adequate needs assessment was not possible, resulting in mismatching of delivered commodities.
- Information and knowledge was not managed from previous humanitarian missions to identify lessons learned.

- In some cases lessons were learned, but no after action reports were generated, and as a consequence, no metrics were formulated to mitigate the next disaster.

We study the literature to explore, define, and develop these reasons. Some areas are endogenous to the organization and some are exogenous. The Endogenous Factors section is further divided into the following subsections:

- Performance Indicators and Readiness Metrics,
- Core Competencies and Capabilities, and
- Issues and Challenges in Humanitarian Operations.

The Exogenous Factors section is further divided into

- Disasters and Lessons Learned and
- The "Three Cs" of Civil–Military Organizations.

Endogenous Factors

Performance Indicators and Readiness Metrics

The absence of clear performance indicators and/or readiness metrics in humanitarian organizations (HOs) has been recognized by the humanitarian community. Davidson (2006) says that, due to the incapability and lack of time, HOs do not measure the performance indicators. The organizations lack any fundamental framework to understand the readiness metrics since they do not have a good measure of performance indicators after the disaster. There are several factors that contribute to the difficulty of defining and measuring either the performance indicators or readiness metrics in HOs (Davidson, 2006).

In the U.S. Department of Defense (DoD) literature, there are discussions about military readiness metrics. However, these are predominantly about conflict readiness (GAO, 2016). In a broad sense, the DoD defines readiness as the ability of the forces to combat and meet the demands to achieve security objectives and the needs of the national strategy. One observation is that the DoD's rebuilding efforts for readiness may not work if there is not a comprehensive plan in place. A framework is necessary for combat readiness (GAO, 2016). This observation further accentuates the lack of any specific framework for readiness metrics for missions other than war, and it demands that such a framework be developed. Vast amounts of money, to the tune of $350 billion, indicate the importance that the DoD places on the readiness of its services for current and future operations (Trunkey, 2013). The readiness is assessed at the individual service level and at the joint forces level. Typically, the DoD reports readiness through the Status of Resources and Training System (SORTS). Recently SORTS, due to its limitations, was transitioned into the Defense Readiness Reporting System (DRRS) which uses a dashboard style display. DRRS is a major improvement.

There are many more aspects to a supply chain (such as material and information flow, players of the supply chain), and one prevalent issue, especially in the commercial supply chain, is the *last-mile delivery problem*. With this particular issue, efficiency or minimizing the cost is the objective. However, in response supply chains, the goal is more than these objectives due to humanitarian concerns. Huang, Smilowitz, and Balcik (2011) focus on meeting the need through quick and sufficient but equitable distribution. The authors measure the performance of the supply chains based on these three criteria. Their observations about number of vehicles, routes, and impact of demand offer practical insight into relief operations. The performance measures suggest possible readiness metrics, such as maintaining a larger number of small vehicles for effective and equitable distribution of critical supplies and services, and they suggest some rules of thumb for quick decisions.

Van der Laan, de Brito, and Vergunst (2009) offer a review of literature identifying the necessary conditions for performance measures for humanitarian supply chains. The authors present a framework that involves two phases, design and implementation. The first phase depends on strategically important functions and the will of the organization to measure operational performance and implementation of an information system to do so. The second phase, which depends on implementation, includes the principles that the framework be future-oriented, that it be aligned with the selected strategy, and that it strike a balance between financial versus non-financial as well as quantitative and qualitative indicators.

A more focused approach, focused specifically on the rapid needs assessment that is defined as a core competency for HOs (Apte et al., 2016), is discussed by Benini and Chataigner (2014). Needs assessment being the key objective for determining the affected region and population, the authors describe a particular tool "prioritization matrix," recently a prevalent tool in determining demand. They offer expansion of this tool based on logic behind it. The matrix is based on composite indicators that are managed through spreadsheets and is the intersection of decision science and humanitarian operations. The authors use the data from the 2013 Typhoon Haiyan (Yolanda) in the Philippines to substantiate their analysis.

Norio et al. (2011) review the causes and impacts of the 2010 Japan earthquake and tsunami. The management of the expanded capacity and capability after the 1995 Hanshin-Awaji earthquake in Japan significantly helped provide disaster relief for the 2010 earthquake. However, the authors believe more can be done. When there is potential for a disaster to turn into a crisis (as in the 2010 earthquake and tsunami in Japan), it is necessary to deploy a collaborative framework based on available resources. Such a framework should take into account the geographic scope of the disaster, thus enabling different governance approaches and mutual assistance and recovery systems. The authors believe that centralized power for sudden and dispersed disasters is vital, existence of a new international platform for joint management is essential, further research of such frameworks is needed, and the lessons learned from the 2010 Japan earthquake and tsunami mandate that infrastructure around the nuclear power plants be robustly planned and designed.

Figure 2 showed the extent of humanitarian aid provided by the United States. The DoD executes humanitarian operations with the budget granted by the State Department, since the DoD does not have its own budget for HADR. These humanitarian activities are rendered through the Overseas Humanitarian Disaster and Civic Aid (OHDACA) program. All HOs, including the DoD, currently face the challenge of measuring the impact of their work (Bonventre, 2006). Bonventre (2006) lists at least three reasons why the DoD should measure the impact of humanitarian assistance programs:

- First, measuring the impact of HOs offers opportunities for future and mid-course corrections in the projects through feedback loops enabling planners to underscore activities that are cost-effective.
- Second, collection and sharing of data prevents the duplication of activities performed by all HOs. Not duplicating activities helps us understand the core competencies and capabilities of HOs.
- Third, analysis based on collected data offers transparency and quantifiable results that do not leave any ambiguity.

The key point here is understanding core competencies and capabilities of all the organizations involved so duplication of efforts is reduced.

Core Competencies and Capabilities

Apte et al. (2016) identify the competencies and capabilities that are core to U.S. military and non-military organizations (NMOs) for HADR. The authors' motivation is that both military organizations and NMOs bring assets, skills, and capabilities to a humanitarian crisis; however, their competencies and capabilities are very diverse. Identification of the specific competencies and capabilities that are core to these types of organizations can enable better planning by both military and NMOs, allowing them to achieve greater effectiveness and efficiency in their humanitarian responses. Apte et al. (2016) build on existing literature on the core competency of the corporations in the private sector. In their research, Apte et al. (2016) extend the concept of identifying, cultivating, and exploiting the core capabilities of the private sector to other organizations that seek to respond efficiently and effectively to disasters. They develop a Core Competencies Test for such organizations. The authors list the top five essential services and capabilities for disaster relief as Information and Knowledge Management, Needs Assessment, Supply, Distribution and Deployment, and Health Services Support.

One of the substantial players in humanitarian assistance and disaster relief around the globe is the U.S. Navy (USN). Roughead et al. (2013) offer an in-depth analysis of the USN's humanitarian assistance, especially in the face of budget cuts and austerity. Their research does not focus on a specific disaster, but rather studies the proactive engagement or strategic pre-positioning (Apte, 2014) of humanitarian assistance. The authors describe the principal benefits of their research: strengthening relations in critical geographic areas through greater cultural understanding, improving the capabilities and readiness of the USN humanitarian assistance, and reinforcing other capabilities such as health systems of host nations.

HADR by the USN is evaluated by Apte et al. (2013) using a structured, qualitative evaluation schema complemented by expert ratings. The authors evaluate the capabilities and utility of ships in the USN. They find that there are specific types of vessels with significant disaster response utility and recommend a flotilla type that would be best suited for the humanitarian operations. Utilizing an exploratory framework that evaluates three diverse disaster cases, they scale the utility of each vessel through subject matter expertize. They find the type of ships most useful for contributing to effective disaster response.

Apte and Yoho (2014) study the USMC resources, including the Marine Expeditionary Unit (MEU), that are primarily responsible for the response. They study recent HADR events to determine how demands were met by the USMC. They identify the supplies that can meet these demands by examining both assets and capabilities of the USMC. By exploring significant gaps, if any, that can be improved by the MEU, they suggest ways to improve the effectiveness of the USMC's response to HADR. A primary take-away from their work is the challenge faced by the USMC to match the capabilities of the USMC to the demand created by future disasters. More issues and challenges in humanitarian operations that deliver disaster relief are described in the next subsection.

Issues and Challenges in Humanitarian Operations

Roughead et al. (2013) list the operational challenges for the USN, such as short-term or discontinuous engagement in HADR lacking enduring coordination and development, insufficient integration with host nations and NGO operatives, dependence on sole assets of vessels that may not serve the necessary demand, inadequate and irregular funding, and most notably, difficulty in measuring alignment of humanitarian efforts with strategic goals. The authors recommend that the USN clarify and focus on the motivation behind the humanitarian assistance to fund the operations sufficiently and without rigidity, and increase the scope and scale of the planning process of HADR allowing coordination with NGOs and host nations. But most importantly, they point out that the USN needs to develop and implement a robust set of metrics for readiness in humanitarian missions.

A major challenge in any supply chain management is measurement of the performance of that supply chain. In the commercial sector, the focus is on resources for optimizing the input (cost) or output (profit). However, for a supply chain established to respond to a disaster, a response supply chain (RSC), the focus is on the time required to respond or the ability to meet the demand. An RSC is defined to be efficient based on the amount or number of resources used to meet the goal of that organization and to be effective based on the level at which it meets the preset goal (Beamon, 2004). Developing such a system for measurement is one of the issues associated with RSCs. The author lists the issues as structure of the RSC, distribution network, inventory control, type of measuring system, coordination with other organizations involved in HA, acquisition of supplies, and finally, the actual measurement.

Beamon and Kotleba (2006) describe the stochasticity of the demand of the disaster, and if the disaster is large-scale, the strain that it creates on the physical distribution.

Other issues that are challenging are the inadequate or incorrect estimation of demands that yield both further casualties and further suffering in the affected area (United Nations, 2007; Duran, Gutierrez, & Keskinocak, 2011; Apte & Yoho, 2011; World Meteorological Organization, 2009). Estimating where and when such demand is needed (McCoy, 2008; Apte, 2009; Apte et al., 2013) is even harder. Demand after a disaster strike in the host nation is external to the organization providing relief. We now focus on such exogenous factors.

Exogenous Factors

Disasters and Lessons Learned: Haiti

On January 10, 2010, a 7.0 magnitude earthquake struck Haiti near Port-au-Prince. The earthquake caused 316,000 casualties. In addition to the Haitian losses, the earthquake also claimed the lives of members of the United Nations Stabilization Mission in Haiti. It injured 300,000 people, made one million people homeless, collapsed 100,000 houses, and damaged 200,000 structures. This earthquake decimated Haiti’s infrastructure. Air and sea transportation was reduced to unworkable. Key access roads were impassable. The medical facilities also became practically nonexistent. The most critical shortage was fuel. The utility infrastructure, including electricity and telecommunications, fell apart. The paralyzed Haitian government was overwhelmed and requested immediate assistance from all over the world. The Dominican Republic received thousands of refugees, but being a small country, it had limitations. Haiti’s urgent request to the U.S. government prompted an immediate response.

Even before the disaster, Haiti had fuel and water shortages as an underdeveloped country. Medical support was scarce. So after the disaster, conditions worsened quickly. Poor infrastructure and inadequate disaster preparedness limited the delivery of relief

(McCunn et al., 2010). Though access to the airport was limited, the U.S. Air Force (USAF) stepped in to maintain security and air traffic control. After this rapid and successful transition, medical support was delivered by many HOs including the U.S. military and other military organizations, NGOs, and government organizations.

The earthquake damaged the Port of Haiti, and it was not operable. The bulk of supplies for immediate sustainment had to be delivered by sea. The lack of a designated logistics team within the Global Response Team at the Joint Task Force (JTF) headquarters meant that deployment planning had to be done at short notice, thus increasing obstacles to an already challenged supply chain. Many in the JTF team were not in the contingency status and, hence, were deployed with suboptimal preparation. This resulted in pushing the supplies quickly but in an ad-hoc way without formal planning, sourcing, and tracking processes. A substantial part of the bureaucracy was eliminated, which enabled a quick response.

The accomplishments of the JTF during Haiti HADR can be divided into the following areas: Air Port, Sea Port, DoD medical support, shelters, overall support, and a secure environment for the operations (JCOA, 2010, slide 196). Best practices emerged from these activities:

- Deployment and support from strategic level liaisons to tactical level (National response)
- Use of unclassified operation environment for information sharing and collaboration between all stakeholders (COCOM)
- Establishing JTF Force Flow working group (Force projection)
- Interface between Humanitarian Assistance Coordination Center and Joint Operational Task Center and NGOs, private voluntary organizations (PVO)s, and UN systems (Coordination)
- Establishing Joint Interagency Information Cell

After the response to the 2010 Haiti earthquake, there were many lessons learned. One of the important findings from the HADR provided by the Joint Center for Operational Analysis (JCOA) was the swift establishment of response structure (JCOA, 2010). Also, civilian and military resources were pushed not only to resolve but to overcome the problem. This was done by (1) a pre-established Response Management Team (RMT) that is dependent on the classification of the disaster (Apte, 2009) performed in five functional areas: management, planning, logistics, administrative, and communications; (2) a Joint Staff Team plugging in with RMT which turned out to be the best practice despite not having connectivity other than commercial internet. However, USAID had visibility for the movement of DoD resources.

As for long-term planning for future disaster relief, the following are some of the implications:

- The president’s declaration about making the disaster relief a priority would help the administration and the country focus on the effort.
- Civilian and military resources may be pushed to mitigate the disaster by establishing the national response structure rapidly.
- Roles, responsibilities, authorities, and essential capabilities need to be clarified at the outset.
- Division of labor within the DoD should be clearly defined.
- Integration of HOs may raise many policy issues that need to be resolved.

- Incomplete data on the ground at the onset of the disaster is a challenge for logistics requirements and priorities.

LTG Keen, who was in charge of Operation Unified Response had the following observations:

1. Respond quickly and effectively,
2. Protect the people always,
3. Build partnerships with key players,
4. Coordinate and Collaborate (C2) to achieve unity of effort,
5. Communicate—Communicate—Communicate,
6. Support the lead Federal Agency within clearly defined roles,
7. Pull from all available resources to form the Joint Task Force,
8. Include the Host Nation Government as much as possible,
9. Work closely with the UN Humanitarian Community, and
10. Anticipate challenges with Internally Displaced Persons (IDPs). (JCOA, 2010, slide 208)

Haiti also taught a few lessons to COCOM. They had to overcome internal organizational issues, gain situational awareness, and satisfy an extraordinary demand for information. Another lesson was that the use of “open” communications and unclassified information-sharing over BlackBerry devices allowed for expanded coordination and collaboration with DoD organizations. Personal and professional relationships among key leaders permeated all levels of interaction and engagement within organizations. And lastly, quick establishment of land-based headquarters reassured the affected population and enhanced the coordination with the host country, state government, USAID, UN, and NGOs.

Disasters and Lessons Learned: Japan

On March 11, 2011, a 9.0 magnitude earthquake struck Japan. A tsunami followed soon after and the losses incurred were extremely severe. By April 13, there were 13,392 casualties, 15,133 missing people, and more than 335,000 people without food, water, shelter, and medical help (Norio et al., 2011). Several nuclear power plants were heavily damaged, resulting in rolling blackouts. The earthquake also affected the transportation system, and for a short time, all the ports were closed. Part of the high speed rail line was shut down, and the Sendai airport suffered intensive damage due to the tsunami. But the devastating blow that pushed this disaster into a crisis was the meltdown of the Fukushima nuclear power plant.

Carafano (2011) assesses the response to the 2011 earthquake in Japan and outlines the lessons for the U.S. to evaluate its own capacity to deal with a future crisis. The author studies critical areas and the corresponding key findings and resources in the United States (see Table 1).

Wilson (2012) has a focused view based on the response from the U.S., titled Operation Tomodachi, to the 2011 earthquake and tsunami in Japan. The response efforts and the collective use of the military stationed abroad offers a model for further U.S. efforts across the globe. The author identifies the activities that worked well, such as the value of maintaining U.S. forces abroad, the use and capabilities of remotely piloted aircrafts, the voluntary evacuation of the U.S. dependents, bilateral coordination, and the benefit of social media through the disaster response. However, the lessons learned, such as improving bilateral coordination, removing control and command confusion, and preparing for large-scale decontamination are also critical for handling future disasters. The author concludes that describing the success of Operation Tomodachi will induce lesser cuts in the DoD's budget since it will bring humanitarian assistance to the forefront as opposed to combat operations in Iraq and Afghanistan.

Table 1. Critical Areas and Key Findings

(Carafano, 2011)

Preparedness and Response	• Effective planning, preparedness, and mitigation measures with possible decentralization for execution of this plan • Need to nurture a national culture of preparedness by concentrating on self-reliance in communities as well as individuals
Communicating the Risk	• Community awareness and understanding risk through communication fetches better cost-effective results than protection measures such as building seawalls. • Communicating risk of low-dose radiation and building confidence for that risk
International Assistance	• The United States and, based on history, Japan have difficulty receiving aid. The United States needs to bolster its capacity to accept and apply international aid efficiently.
Critical Infrastructure	• Need to focus on the most "vital" infrastructure (United States–Canada grid) to maintain resilient infrastructure that can recover quickly in case of disaster. • Industry and federal regulators need to work together to understand lessons from Fukushima and how they can be adapted for nuclear disasters in the United States.

Terada (2012) notes that during the assistance and relief following the 2011 earthquake in Japan, information should have been shared and appropriate tasking should have been implemented among the participants. There should be more training and exercises for USJF as the DoD support for HADR increases so that professionalism is enhanced and roles are clarified (Liaison Staff, 2012).

Japan is a developed nation and fairly self-sufficient in disaster relief. However, it did not have much experience in receiving aid from across the world. Thus, one of the lessons learned was to institute training for international guidelines (Smart, 2012). It is also imperative to establish an effective media strategy for controlling and dissipating information when there is a need of receiving real-time facts.

Katoch (2012) stresses that no silos should be permitted. Clear protocols should be set with chain of control at all levels of the departments involved of the host government, military organizations, and NGOs. Organizational structures and processes, in compliance with humanitarian and military doctrines, must be pre-established at local, national, and international levels. Only close ties with such organizations is not adequate for productive civil–military coordination. This was evident during the 2011 earthquake in Japan in the coordinating pains experienced by the United States and Japan even though they are allies (Katoch, 2012).

Wanlach (2012) emphasizes establishing relationships before a disaster to share information. The author also claims that agreements have to be in place for practical methods of coordination, and the relief needs to be planned so that the strengths of the responding organization are exploited. Finally, better preparation by the host country will always help mitigate suffering.

The 2011 earthquake in Japan also taught lessons about the geographical perspective. Developing a tsunami response system using inundation maps helps disaster managers to model the potential effects of a tsunami so that the most suitable shelter locations and optional evacuation routes can be planned (Hong, 2012). Such lessons were also taught by Super Typhoon Haiyan (Yolanda) in the Philippines in 2014. Shallow draft adds to the destruction due to the fact that it produces more surges. Therefore, to understand threats, warnings must be accompanied by analysis of the impacts on the ground (Center for Excellence in Disaster Management and Humanitarian Assistance [CFE-DM], 2015).

Disasters and Lessons Learned: Philippines

On November 8, 2013, Typhoon Haiyan (Yolanda) made landfall in the Philippines, causing extensive damage. More than 1.1 million houses were damaged, and 14.1 million people were affected. The confirmed death toll was 6,183. Though the Philippines is one of the most disaster-prone countries, this typhoon was among the strongest ever to strike the country.

The extensive damage to the internal infrastructure made transportation of goods extremely difficult to the point that signs of assistance and relief were only visible three to five days after the typhoon struck the Philippines (CFE-DM, 2014). Among the international community, both military and non-military, the U.S. DoD, supporting the Armed Forces of the Philippines, and USAID played a significant role in HADR. UN agencies also responded immediately with teams for initial rapid assessment.

The heavy vertical lift capabilities of the U.S. DoD and other military organizations helped in the face of infrastructure destruction. Their capabilities also helped in scouring the thousands of affected islands that were remote and almost impossible to access. The tactical military forces provided support immediately. There were many assets of the U.S. DoD stationed in Japan and Okinawa. These included USS *George Washington* naval task force and 31st MEU to form JTF 505. Approximately 1,000 U.S. DoD personnel were deployed. Military aircraft provided support in needs assessments of remote areas, brought aid workers and supplies to these remote areas, and evacuated the affected population to other locations. The Marines helped clear roads and distribute supplies and services (Lum & Margesson, 2014).

One of the lessons learned (CFE-DM, 2014) during the Super Typhoon Haiyan assistance and relief was that civil–military collaboration needs to happen far faster than it did. It is also important to have trust among participating organizations, and this could be achieved through informal networks formed during training and exercises. It was also noted that the affected people from the most dangerous areas have to be evacuated. But two concepts that are important and applicable in any disaster are the pre-positioning of supplies and the resilience of the local population.

In addition, visual messaging in the form of accurate scenario-based storm surge inundation maps facilitated a shared framework of the operating environment. Every foreign disaster response is a bilateral agreement between the assisting state and the affected state. The response in Super Typhoon Haiyan showed that the optimal use of defense assets is best coordinated through the Multinational Coordination Center (MNCC). Recognizing the need for the MNCC to operate at strategic and operational levels simultaneously, the MNCC in Camp Aguinaldo became fully operational 48 hours before Super Typhoon Hagupit the following year made landfall (CFE-DM, 2015). Recognizing the need to augment the government’s response capabilities, private sector-led organizations, as demonstrated by the Philippine Disaster Resilience Foundation (PDRF) 88, began putting mechanisms in place for a disaster operations center aimed at coordinating and collaborating disaster risk management initiatives of businesses across all industrial sectors.

Disasters and Lessons Learned: Nepal

On April 25, 2015, a 7.8 magnitude earthquake struck Nepal, followed by 20 aftershocks. On May 12, 2015, a 7.3 magnitude earthquake with five aftershocks struck near Mount Everest. Within a week, there were 7,000 casualties, 70,000 structures damaged, and over eight million people affected (Sanderson & Ramalingam, 2015). The earthquakes and their aftershocks resulted in over 5,000 landslides, flooding many streams with sediments, and causing floods in low lying areas. This made the task of transporting supplies and services nearly impossible.

The U.S. DoD deployed soft and hard assets for HADR. The 3rd Marine Expeditionary Brigade (MEB) and other forces formed the JTF 505 to respond to this disaster under the guidance of USAID. There was substantial support for evacuation by JTF aircraft, transportation of local ambulances by JTF 505 medical personnel, including squadron flight surgeons and DART physicians. However, being a landlocked country at a high elevation presented its own set of unsurmountable issues. This tested rotary wing and tilt rotor aircraft endurance. Another unique obstacle in providing relief was complications due to diplomatic requirements of coordination in overflight and clearances from multiple countries surrounding the affected area.

In addition to the substantial HADR delivered by the U.S. DoD, the Government of India responded within four hours due to the proximity with open borders, close cultural ties with Nepal, relationship with the Armed Forces, and bilateral pre-disaster planning and training. The Chinese government also responded at the request of the Nepalese government with search and rescue teams, helicopters, and 900 personnel. The World Health Organization, the UN Cluster System, international military forces, and other HOs added their support to the disaster relief.

One of the dreadful challenges was properly caring for children whose parents were missing. Urgent repair of the roads for immediate transportation was also a formidable challenge that could have been mitigated through helipads in rural areas. The inadequate collection of field information and dissemination of the same turned out to be a major handicap. Establishing call centers in each village would help overcome this difficulty. Due to damaged government structures, the basic problem of lack of office space, though not life-threatening, was a deterrent. This meant the building codes had not been followed and strict monitoring should have been implemented. Inadequate search and rescue capabilities turned out to be devastating, so one lesson learned was to strengthen the overall search and rescue capability through security forces and international support.

Wendelbo et al. (2016) outline the challenges in executing disaster relief and the lessons learned after the Nepal earthquake, as described in Table 2.

Table 2. Challenges and Lessons Learned in Nepal
(Wendelbo et al., 2016)

Planning:	In spite of sound planning for disasters, the efforts fell short. The framework with rules and regulations were not fully funded and therefore not enforced.
Building codes:	The scientifically strong building codes that exist in Nepal were not enforced.
Household damages:	Though the damage to the infrastructure and public facilities was mitigated through inside as well as outside help, the rural households remained damaged.
Logistical challenges:	Being a poor and underdeveloped country, the infrastructure in Nepal was inadequate. The country has a single airport, which turned out to be the bottleneck. The relief efforts could not be utilized in spite of sufficiently available supply, and some teams had to return without delivering the aid.
Communication:	Nepal's communication networks physically and virtually collapsed, so the local responders could not convey the existing conditions and needs to the authorities.
Coordination:	The inadequate physical infrastructure, before and after the disaster, intensified the lack of coordination between HOs delivering support.
Misdirected focus:	Trendy methodologies were used by some HOs that are costly for locals to sustain, such as K9 teams for search and rescue instead of more efficient methods.
Funding:	Though about US $4 billion was pledged within a month, when Nepalese government launched the recovery efforts, not all the funds came through. Perhaps it was due to lack of fulfilling the promises on the donors' part or not having faith in utilization of the funds by the host nation.

The overwhelming support from HOs across the globe complicated relief efforts in Nepal. Nepal had only one runway airport and very few helicopters to transport relief workers to the inaccessible mountainous areas. Unfortunately, the lessons learned in the 2004 Indian Ocean tsunami were not well understood or implemented (Salmeron & Apte, 2010). After the tsunami, the donated supplies that could have mitigated needs to a large extent could not be distributed due to a single airstrip and a single forklift in Banda Ache (Apte, 2009).

Summary of Lessons Learned

No amount of planning for disasters can prevent casualties, suffering, and damages. But "good" planning, based on lessons learned from past disasters, can mitigate the effects of the disaster. However, a significant theme that emerges from the literature review is articulated by Markus (2012)—the sharing of information among stakeholders in terms of their mandates, activity scope, capacity, technical expertise, and funding capital has to happen before a disaster strikes.

The U.S. DoD is one of the organizations providing HADR in the Asia-Pacific region with other government organizations, NGOs. Moroney et al. (2013) claim that the following changes need to be made to spread goodwill through HADR:

- Improve the DoD's efficiency in HADR
- Enhance interagency coordination
- Develop coordination with the host nation
- Increase work with the UN and NGOs
- Align security activities and regional HADR capabilities

Another organization that plays a major role in humanitarian operations is the Logistics Cluster of the United Nations (UN). Global Logistics Cluster (2016) has extensively studied the relief provided in the past disasters to understand the lessons from these experiences. The lessons learned are tabulated in Table 3.

Table 3. Lessons Learned by Global Logistics Cluster

Coordination	Mechanisms such as meetings for unifying response, reduction of duplication in logistics operations, and identification of common needs
Information Management	Website managed by the cluster providing maps, Geographic Information System (GIS), situation reports, and consistent and timely meetings notes
Logistics Service Delivery	Humanitarian staging areas need to be pre-established so they can be activated immediately, thus reducing delays in delivery. Existing support services such as pre-positioned equipment and their handlers, fleet of vehicles with smaller secondary vehicles, air ambulances operated by local staff added significantly to the success of certain occurrences. The shortcomings, however, were delays in air transport capacity that led to reduction in cost-effectiveness.

The authors recommend that investments should be made in pre-preparedness activities that have turned out to be invaluable in certain instances. They also comment on the information management tools used, such as having an accessible system to enrich the competency further. In terms of accountability, the authors suggest that there should be clarification of roles and responsibilities associated with them in addition to pre-established tracking system. Most importantly, coordination efforts between the strategic partners in preparedness planning and advisory board for decision-makers should be done with priority given to logistics.

Evans (2016) outlines necessity of interagency training as the lesson learned. The author describes the lessons as (1) a Mobile Training Team traveling to disaster-prone areas and offering training to country teams, (2) adding courses at the end of annually held conferences at USPACOM, and (3) incorporating a specific and significant disaster management content into existing preparatory courses.

Advantages from these lessons are that each member of the DoD will go through the training so that participants will learn about

- available resources,
- utilization of the same,
- lessons from previous disasters,
- relevant topics they may face such as basic search and rescue, medical first responder, and
- appreciation of options available during the lifecycle of the disaster.

Issues and Challenges in the "Three Cs" of Civil–Military Organizations

Civil–military organizations are needed to establish, maintain, influence, and exploit relations between military, government, and non-government organizations, including the host country of the disaster. The "three Cs" for civil military organizations are communication, coordination, and collaboration. With complimentary capabilities and competencies, other government and non-government organizations participate with the U.S. in HADR. Therefore it is essential that coordination and communication among all these organizations be explored and enhanced. The premise is that such processes will enable the DoD to respond efficiently and effectively with the unique capabilities that they possess in the future of limited budgets (Apte et al., 2016; Moroney et al., 2013). The type of collaboration between military and non-military organizations is predominantly determined by the disaster classification. Logistical support and delivery of supplies continues irrespective of the alliance (Pettit & Beresford, 2005). The authors propose a model for logistical requirements in the affected regions.

The authors also present issues and challenges for measuring disaster preparedness and response. These factors can help in developing the framework for readiness metrics. More importantly, the authors describe the possible conflicts arising from military involvement in humanitarian crises. Table 4 describes these conflicts.

Table 4. Conflicts Arising From Military Involvement

(Pettit & Beresford, 2005)

Medical care	Military medicine is not necessarily appropriate for humanitarian crises. Supplies readily available to military forces may be inappropriate for refugees and disaster victims, although at the outset of a crisis they may be all that is available.
Conflict resolution	Military forces are not well suited to aid long-term redevelopment efforts. The imposition of security by outside military forces may also impede negotiation and conflict resolution.
Interaction with other organizations	Military commanders may be unfamiliar with the roles of major international organizations, and, conversely, civilians will have little experience of military organizations. There will be differences in strategy, objectives, and tactics.
Conflict with humanitarian agenda	Using military resources to achieve humanitarian goals creates tension and can undermine the appearance of neutrality of relief organizations.
Adequacy of training	Few military officers receive training in disaster relief or humanitarian assistance. There is also likely to be ambiguity over the role of military physicians in complex emergencies in international humanitarian law.
Limited commitment to disaster response	The principal mission of the military is to resolve military conflicts, and, generally, less effort and fewer resources are devoted to humanitarian aid unless an HA-specific mission is being conducted.

In November 2005, a DoD directive defined stability operations as a "core U.S. military mission" with a Priority comparable to combat operations" (DoD, 2005). This directive recommends the use of outcome-based performance measure and installing process for transparency of information. Reaves, Schor, and Burkle (2008) describe the gaps in the DoD's ability to measure the effects of HADR operations when compared with international standards. The authors' analysis reveals that only 0.7% of the 1,000 after action reports studied refer to performance measures. The authors conclude that most of the humanitarian operations performed by the DoD did not have records to identify the activities that could be quantified for most contribution to the HADR. In a focused study, Reineck (2004) estimates a readiness and deployability index for emergency center registered nurses to prepare for disaster relief.

U.S. Forces in Japan (USFJ) maintained necessary coordination and daily workings with the State Department (Embassy in Japan) and Japan Self-Defense Forces (JSDF; Terrada, 2012). This was informally done without any structured support at operational level of command and control.

Yoshitomi et al. (2012) describe the bilateral coordination between JSDF and USFJ. They suggest that the solution to preparedness issues may be establishing a standing bilateral coordination center that is staffed with people from both the forces so they could share information and plan before the disaster strikes. They also recommend that for effective coordination, more activities and exercises are needed. This will enable clarification of communications, roles, missions, and capabilities with the counterparts of other nations. Acquisition and interagency agreements are necessary to pre-position supplies and services. For successful coordination, it is also essential to understand the capabilities and equipment of the host nation counterparts.

Japan is one of the best prepared countries for earthquake in the world but had limited experience in receiving international assistance (Katoch, 2012). Absence of institutionalized civil–military coordination is a significant void that is exacerbated when a country is facing a super disaster or crisis. In spite of this, the Great East Japan Earthquake (GEJE) of 2012 is a great example of coordination between JSDF, USFJ, Swiss Humanitarian Aid Unit (SHA), and German Federal Agency for Technical Relief (THW; Terada, 2012; Smart, 2012; Fichter, 2012).

At a national level, cooperation between the Red Cross and Red Crescent Movement (RCRC) and military is common, but this cooperation gets complex when military assets are involved in an international context in the case of natural disaster (Markus, 2012). Guidelines from RCRC state that “while maintaining a dialogue with armed forces at all levels, the components of Movement preserve their independence of decision-making and action, in order to ensure adequate access to all people in need of humanitarian assistance” (Counsel of Delegates, 2005).

Super Typhoon Haiyan (Yolanda) was notably one of the best instances of civil–military coordination (CFE-DM, 2014). There were many previous experiences from the disasters in the Philippines that contributed to the disaster relief. However, connections between personnel involved in the relief and other players helped expedite the collaboration between civil and military organizations.

There were 57 countries contributing to the relief operations in Super Typhoon Haiyan. Multinational Coordinate Center (MNCC) was set up for this purpose with 29 foreign militaries that responded to the disaster. The coordination predominantly revolved around warehousing, transportation, and distribution—that is, logistics. However, a lack of framework for a common operating process and a lack of consensus on needs assessment ended up causing a duplication of efforts in the face of scarce resources. The study by Center for Excellence in Disaster Management (CFE-DM, 2015) shows the following best practices:

> Best Practice 1: A commonly understood “end-to end warning system” prepares a nation for crises
>
> Best Practice 2: Bilateral commitment executed multilaterally on the ground through the Multinational Coordination Center (MNCC) promotes optimal civilian use of foreign defense assets.
>
> Best Practice 3: When closely coordinated with the government, the private sector multiplies a nation’s surge capacity to meet the life-saving needs of the affected population. (p. 5)

The authors of CFE-DM (2015) conclude that advances in civil–military coordination occur when (1) consensus in the operating environment paves the way for unity of effort; (2) systemic changes through an inclusive multi-sectoral approach streamlines disparate efforts on emergency response preparedness; (3) a convergence in concepts, frameworks, protocols, and procedures maintains a clear distinction of responsibilities and national sovereignty; and (4) institutionalized internal and external partnerships augment a country’s latent ability to surge.

The U.S. Operational Detachment–Alpha (ODA) served in Philippines during Super Typhoon Haiyan. This was not unique to the Philippines; ODA also served in Nepal. In the aftermath of the 2015 Nepal earthquake, two teams of the ODA, 1121 and 1126, happened to be in Khatmandu, Nepal (Elwood, 2016). They stayed on to help with the HADR mission since U.S. Special Forces Green Berets are known for their capability in diverse tasks of special warfare during combat missions and in training with partner forces in coordinating exercises. This came in critical use in Nepal. The beneficial aspect of ODA can be exploited methodically if the team can be incorporated in a contingency plan for military–military collaboration. The competencies of Special Forces to react instantly with pre-established relationships and resources, critical language skills, and flexibility could then be utilized.

There was significant anticipation for a catastrophic earthquake in Nepal among many international governments and military organizations. This projection helped in a broad response from all the organizations when the actual disaster occurred. The United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) led the effort for civil–military coordination through the Humanitarian Military Coordination Center (HuMOCC; Tarantino, Suter, & Cooper, 2016). In Nepal, the military participation came in the areas of logistics and transportation, in addition to health and medical support.

The model for civil–military cooperation in disaster relief is the support provided by Joint Task Force (JTF) 505 and USAID to the 2015 Nepal earthquake (Bock, 2016). The author credits the success to the Mission Tasking Matrix (MITAM) Process. The major contributions of this tool are as follows:

- Transparency in information about needs, number of response participants, requirements, and coordination challenges
- Military planners' ability to expedite the planning process and analyze if JTF has the resources and authority to fill it
- The DoD's doctrine of supporting USAID
- Maintaining cost efficiency
- Constraining focus on specific requirements to avoid mission creep

Framework Based on Literature Survey

Unfortunately, the absence of quantifiable or measurable performance indicators or readiness metrics in humanitarian organizations (HOs) has been acknowledged by organizations that are involved in HADR whether they be military or not. Literature on critical best practices for performance measurement describe that the metrics should be aligned with the objective. The objective in the case of the U.S. DoD according to the Cooperative Strategy for 21st Century Seapower is to provide HADR with capabilities that complement the capabilities of other HOs in such operations. Learning from the lessons discussed previously in this article will help the U.S. DoD be effective and efficient in HADR and at the same time spread goodwill through the world by HADR.

Readiness is defined by the DoD as the ability of the U.S. military to fight for and meet the needs of the national strategy. No comprehensive plan exists, thus emphasizing that a framework is necessary readiness. There exist marked gaps between the way the DoD measures the performance of HADR and international standards. It has been noted that out of 1,000 after action reports studied, only 0.7% even mention performance (Reaves et al., 2008).

Some of the reasons, as expressed before, that the DoD should measure the impact of humanitarian assistance programs are as follows: (1) Measuring them offers opportunities for future and mid-course corrections in the projects through feedback loops enabling planners to underscore activities that are cost-effective, (2) the collection and sharing of data decreases the likelihood that HOs duplicate activities , and (3) analysis based on collected data offers transparency and quantifiable results that do not leave any ambiguity.

However, operational challenges exist. Current naval HADR responses are mostly reactive, not proactive or preplanned and sustainable engagements. Such activities do not necessarily align with the strategic goal. Lessons learned point to deficient integration with host nations and other HOs. More importantly, the reliance of the USN on vessels alone may not provide adequate HADR due to the complete dependence on the deployment of ships irrespective of their capabilities.

Figure 3 shows the endogenous and exogenous processes of the organizations providing relief to the affected region. Core competencies that are based on the assets and resources of the organization are endogenous to the organization. Originating from the core competencies and capabilities, the organizations establish response supply chains (RSC) for products, services, and information. This step is endogenous to the organization. The response supply chains have their own issues and challenges inherent to the organization. The HADR delivered is at the intersection of establishment of RSC, an endogenous process of an organization and the demand due to disaster in the affected region, exogenous to the organization. The demand from the disaster in the affected region dictates the relief needed that is exogenous to the organization. However, the actual relief delivered is endogenous to the organization. The consequences of the delivery of HADR result in a gap of pain that originates when needs are not met by the organization. The gap of pain experienced by the regions affected by disasters forces the question of "why" to the players of the response supply chains. All these consequential steps are exogenous to the organization. However, the resulting playbook or a set of readiness metrics in answering of "why" is endogenous to the organization itself. The objective of our research in this project was to study existing literature to understand the process and ultimately to conceptualize a framework for readiness metrics.

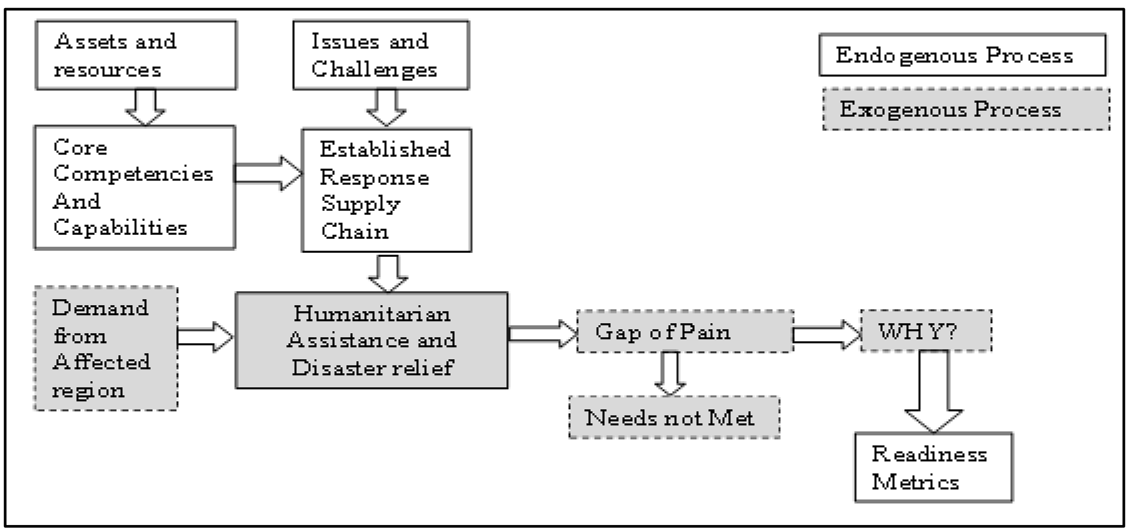


Figure 3. Process for Identifying Readiness Metrics in Organizations

The Case for Naval HADR Operations

The essential services and capabilities for disaster response as outlined by Apte et al. (2016) for military and non-military organizations are Information and Knowledge Management, Needs Assessment, Supply, Deployment and Distribution, and Health Service Support. If one focuses on military organizations and on the U.S. DoD, the capability of Information and Knowledge Management, for example, can be transformed into what is needed to be ready, an awareness of being ready, and a metric for readiness.

The disasters of the past few decades and the lessons learned from them offer insight into the needs of the countries affected, whether their needs were met by the U.S. DoD, and the effectiveness of the DoD's response. After Super Typhoon Haiyan, there was massive damage to the infrastructure. The U.S. DoD's principal capability of heavy vertical lift capabilities was critical in delivering disaster relief. This capability helped transport goods and people despite infrastructure destruction. This capability also helped rescue the affected population stranded on the many islands of the Philippines that are remote and difficult to access. USS *George Washington* naval task force and 31st MEU formed JTF 505 with about 1,000 U.S. DoD personnel and were deployed to the Philippines. Military aircraft helped in understanding the demand through intelligence gathering. Without the Marines, it would have been impossible to clear the roads and distribute supplies and services.

After the Nepal earthquake, the U.S. DoD service that could be used was the 3rd MEB located in Okinawa since Nepal is a landlocked country. Adding to the MEB, JTF 505 was formed for deployment to help with HADR under the guidance of USAID. The terrain in Nepal tested the DoD equipment and the staff. Though rotary wing and tilt rotor aircraft supported the mission, the team casualties tried the resolve of the teams.

The naval missions conducted for HADR in the past, as described in this research, help develop the Readiness Assessment Model. The output from the model must answer questions such as what is needed, is it there, what must be done, how can it be done, how can the gap between demand and supply be closed before a disaster strikes. There are many more variables that play a role in the model such as the type of the disaster (manmade or natural), onset of disaster (sudden or slow), relations with the host county, category of host country, and so forth.

In future research, we plan to dig deeper and build on the actual experiences of USN and USMC officers involved in HADR. The objective is to articulate the strategic readiness through operational details and answer the questions posed previously.

References

Adams, S. M., Sarkis, J., & Liles, D. (1995). The development of strategic performance metrics. *Engineering Management Journal, 7*(1), 24–32.

Altay, N., & Green, W. G., III. (2006). OR/MS research in disaster operations management. *European Journal of Operational Research, 175*, 475–493. doi:10.1016/j.ejor.2005.05.016

Appleby, A. (2013). Connecting the last mile: The role of communications in the Great East Japan Earthquake. Retrieved from http://reliefweb.int/report/world/connecting-last-mile-role-communications-great-east-japan-earthquake

Apte, A. (2009). Humanitarian logistics: A new field of research and action. *Foundations and Trends© in Technology, Information and Operations Management, 3*(1), 1–100. doi:10.1561/0200000014

Apte, A. (2014). Strategic and operational prepositioning in case of seasonal natural disasters: A perspective. In *Wiley Encyclopedia of Operations Research and Management Science* (pp. 1–13). Hoboken, NJ: Wiley. Retrieved from https://onlinelibrary.wiley.com/doi/10.1002/9780470400531.eorms1095/pdf

Apte, A., Goncalves, P., & Yoho, K. (2016). Capabilities and competencies in humanitarian operations. *Journal of Humanitarian Logistics and Supply Chain Management, 6*(2), 240–258.

Apte, A., & Yoho, K. (2012, September). *Capabilities and competencies in humanitarian operations* (Acquisition Research Sponsored Report Series). Monterey, CA: Naval Postgraduate School, Graduate School of Business and Public Policy.

Apte, A., & Yoho, K. (2014, September). *Analyzing resources of United States Marine Corps for humanitarian operations* (Acquisition Research Sponsored Report Series). Monterey, CA: Naval Postgraduate School, Graduate School of Business and Public Policy.

Apte, A., Yoho, K., Greenfield, C., & Ingram, C. (2013). Selecting maritime disaster response capabilities. *Journal of Operations and Supply Chain Management, 6*(2), 40–58.

Balcik, B., Iravani, S., & Smiloewith, K. (2010). A review of equity in nonprofit and public sector: A vehicle routing perspectives. In *Wiley Encyclopedia of Operations Research and Management Science*. Hoboken, NJ: Wiley.

Benini, A. (2016, August). *Severity measures in humanitarian needs assessments: Purpose, measurement, integration* (Technical note). Geneva, Switzerland: Assessment Capacities Project (ACAPS). Retrieved from https://www.acaps.org/sites/acaps/files/resources/files/acaps_technical_note_severity_measures_aug_2016_0.pdf

Benini, A., & Chataigner, P. (2014, May). *Composite measures of local disaster impact: Lessons from Typhoon Yolanda, Philippines*. Retrieved from http://aldo-benini.org/Level2/HumanitData/Acaps_140527_CompositeMeasures_Philippines.pdf

Benini, A., Chataigner, P., Noumri, N., Tax, L., & Wilkins, M. C. (2016, February). *Information gaps in multiple needs assessments in disaster and conflict areas.* Retrieved from http://reliefweb.int/report/world/information-gaps-multiple-needs-assessments-disaster-and-conflict-areas-guidelines

Bock, Y. (2016, Spring). Form with content: Ensuring mission success in HADR. *Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 13*, 43–46.

Bonventre, E. V. (2008, June). Monitoring and evaluation of Department of Defense: Humanitarian assistance programs. *Military Review,* 123.

Bourne, M., Mills, J., Wilcox, M., Neely, A., & Platts, K. (2000). Designing, implementing and updating performance measurement systems. *International Journal of Operations & Production Management, 20*(7), 754–771. doi:10.1108/01443570010330739

Buckmaster, N. (1999). Associations between outcome measurement, accountability and learning for non-profit organizations. *International Journal of Public Sector Management, 12*(2), 186–197.

Carafano, J. J. (2011). The Great Eastern Japan Earthquake: Assessing disaster response and lessons for the U.S. Retrieved from http://www.heritage.org/asia/report/the-great-eastern-japan-earthquake-assessing-disaster-response-and-lessons-the-us

Cecchine, G., Morgan, F. E., Wermuth, M. A., Jackson, M., Schaefer, A. G., & Stafford, M. (2010). *The U.S. military response to the 2010 Haiti Earthquake: Considerations for Army leaders*. Santa Monica, CA: RAND. Retrieved from https://www.rand.org/content/dam/rand/pubs/research_reports/RR300/RR304/RAND_RR304.sum.pdf

Center for Excellence in Disaster Management and Humanitarian Assistance (CFE-DM). (2014). *Lessons from civil-military disaster management and humanitarian response to Typhoon Haiyan (Yolanda).* Retrieved from http://www.alnap.org/resource/21103

Center for Excellence in Disaster Management and Humanitarian Assistance (CFE-DM). (2015). *Advances in civil-military coordination in catastrophes: How the Philippines turned lessons learned from Super Typhoon Haiyan (Yolanda) into best practices for disaster preparedness and response.* Retrieved from https://www.cfe-dmha.org/LinkClick.aspx?fileticket=QEDCMlMtTc4%3D&portalid=0

International Federation of Red Cross and Red Crescent Societies. (2005). *Council of Delegates: Seoul, 16–18 November 2005* (Resolution 7). Retrieved from https://www.icrc.org/eng/assets/files/other/cod-resolutions_2005_en.pdf

Davidson, A. L. (2006). *Key performance indicators in humanitarian logistics* (Master's thesis). Cambridge, MA: Massachusetts Institute of Technology.

DiOrio, D. R. (2010). *Operation Unified Response Haiti Earthquake 2010* (Technical report). Retrieved from http://jfsc.ndu.edu/Portals/72/Documents/JC2IOS/Additional_Reading/4A_Haiti_HADR_Case_Study_revNov10.pdf

DoD. (2005). *Military support for stability, security, transition, and reconstruction (SSTR) operations* (DoD Directive 3000.05). Retrieved from http://www.dtic.mil/whs/directives/corres/pdf/300005p.pdf

Dremsa, T. L., Resnick, B., Braun, R. F., Derogatis, L. R., McEntee, M., Turner, M., & Reineck, C. (2004). Pilot testing the Readiness Estimate and Deployability Index Revised for Air Force nurses. *Military Medicine, 169*(1), 11–15.

Duran, S., Gutierrez, M. A., & Keskinocak, P. (2011). Pre-positioning of emergency items for CARE International. Interfaces, 41(3), 223–237.

Elwood, N. (2016, Spring). The Swiss Search and Rescue Team in Japan. *Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 13*, 8–12.

Fichter, M. (2012). U.S. Army Green Berets save lives in Nepal. *Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 5,* 67–74.

Folan, P., & Browne, J. (2005). A review of performance measurement: Towards performance management. *Computers in Industry, 56*(7), 663–680.

GAO. (2016). *Military readiness: DOD's readiness rebuilding efforts may be at risk without a comprehensive plan* (GAO-16-841). Retrieved from http://www.gao.gov/assets/680/679556.pdf

Global Humanitarian Assistance. (2013). *Global humanitarian assistance report 2013.* Retrieved from http://www.globalhumanitarianassistance.org/wp-content/uploads/2013/07/GHA-Report-2013.pdf

Global Logistics Cluster. (2016, Spring). The backbone of response: Lessons from the Logistics Cluster. *Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 13,* 39–42.

Guha-Sapir, D., Hoyois, P., & Below, R. (2014). *Annual disaster statistical review 2013: The numbers and trends.* Brussels, Belgium: Centre for Research on the Epidemiology of Disasters (CRED).

Harvey, J. C., Jr. (2008). *U. S. Navy Language Skills, Regional Expertise and Cultural Awareness Strategy*. Retrieved from https://www.researchgate.net/publication/235011977_US_Navy_Language_Skills_Regional_Expertise_and_Cultural_Awareness_Strategy

Haavisto, I., & Goentzel, J. (2015). Measuring humanitarian supply chain performance in a multi-goal context. *Journal of Humanitarian Logistics and Supply Chain Management, 5*(3), 300–324.

Hong, J. (2012). Creation of a tsunami disaster response system in the Republic of Korea using inundation maps. *Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 5*, 75–78.

Huang, M., Smilowitz, K., & Balcik, B. (2011). Models for relief routing: Equity, efficiency and efficacy. *Transportation Research Part E: Logistics and Transportation Review, 48*(1), 2–18. doi:10.1016/j.tre.2011.05.004

Joint Center for Operational Analysis (JCOA). (2010). Operation Unified Response Haiti earthquake response. Retrieved from https://community.apan.org/.../Operation-United-Respose-JOCA-Report_5D00_.pdf

Kaplan, R., & Norton, D. (1992). The balanced scorecard: Measures that drive performance. *Harvard Business Review, 70*(1), 71–79.

Kaplan, R. S. (2001). Strategic performance measurement and management in nonprofit organizations. *Nonprofit Management and Leadership, 11*(3), 353–370.

Katoch, A. (2012). International response to the Great East Japan Earthquake. *Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 5,* 85–88.

Kovács, G., & Spens, K. M. (2009). Identifying challenges in humanitarian logistics. *International Journal of Physical Distribution and Logistics Management, 39*(6), 506–528.

Liason Staff. (2012). US military support: Operation Tomodachi. *Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 5*, 55–60.

Lum, T., & Margesson, R. (2014). Typhoon Haiyan (Yolanda): U.S. and international response to Philippines disaster. *Current Politics and Economics of South, Southeastern, and Central Asia, 23*(2), 209–246.

Maghsoudi, A., & Pazirandeh, A. (2016). Visibility, resource sharing and performance in supply chain relationships: *Insights from humanitarian practitioners. Supply Chain Management: An International Journal, 21*(1), 125–139.

Margesson, R. (2015). *International crises and disasters: U.S. humanitarian assistance response mechanisms* (CRS Report RL33769). Washington, DC: Congressional Research Service. Retrieved from https://www.fas.org/sgp/crs/row/RL33769.pdf

Markus, F. (2012). Japanese Red Cross cooperation with civilian and military responders in the Great East Japan Earthquake. *Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 5*, 81–84.

McCoy, J. H. (2008). Humanitarian response: Improving logistics to save lives. *American Journal of Disaster Medicine, 3*(5), 283–293.

McCunn, M., Ashburn, M. A., Floyd, T. F., Schwab, C. W., Harrington, P., Hanson, C. W., … Fleisher, L. A. (2010). An organized, comprehensive, and security-enabled strategic response to the Haiti earthquake: A description of pre-deployment readiness preparation and preliminary experience from an academic anesthesiology department with no preexisting international disaster response program. *International Anesthesia Research Society, 111*(6), 1438–1444. doi:10.1213/ANE.0b013e3181f42fd3

MEAL Humanitarian Technical Working Group, Save the Children. (2013). Key lessons to learn for Typhoon Haiyan response. Retrieved from http://www.alnap.org/resource/9736

Medori, D., & Steeple, D. (2000). A framework for auditing and enhancing performance measurement systems. *International Journal of Operations & Production Management, 20*(5/6), 520–533.

Micheli, P., & Kennerly, M. (2005). Performance measurement frameworks in public and non-profit sectors. *Production Planning and Control, 16*(2), 125–134.

Moroney, J. D. P., Pezard, S., Miller, L. E., Engstrom, J., & Doll, A. (2013). *Lessons from Department of Defense disaster relief efforts in the Asia-Pacific Region* (Rand Corporation report). Retrieved from https://www.rand.org/pubs/research_reports/RR146.readonline.html

Neely, A. (2005). The evolution of performance measurement research: Developments in the last decade and a research agenda for the next. International Journal of Operations & Production Management, 25(12), 1264–1277.

Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design: A literature review and research agenda. International Journal of Operations & Production Management, 15(4), 80–116. Reprinted in *International Journal of Operations & Production Management, 25*(12), 2005, 1228–1263.

Norio, O., Ye, T., Kajitani, Y., Shi, P., & Tatano, H. (2011). The 2011 Eastern Japan Great Earthquake disaster: Overview and comments. *International Journal of Disaster Risk Science, 2*(1), 34–42. doi:10.1007/s13753-011-0004-9

Parker, T., Carroll, S. P., Sanders, G., King, J. E., & Chiu, I. (2015). *An inside look into USPACOM response to Super Typhoon Haiyan.* Retrieved from https://www.cfe-dmha.org/LinkClick.aspx?fileticket=eF0gtnF5_iQ=

Pettit, S. J., & Beresford, A. K. C. (2005). Emergency relief logistic: An evaluation of military, non-military and composite response models. *International Journal of Logistics: Research and Applications, 8*(4), 313–331.

Reaves, E. J., Schor, K. W., & Burkle, F. M., Jr. (2008). Implementation of evidence-based humanitarian programs in military-led missions: Part I. Qualitative gap analysis of current military and international aid programs. *Disaster Medicine and Public Health Preparedness, 2*(4), 230–236.

Reineck, C. A. (2004). The Readiness Estimate and Deployability Index: A self-assessment tool for emergency center RNs in preparation for disaster care. *Topics in Emergency Medicine, 26*(4), 349–356.

Roughead, G., Morrison, S. J., Cullison, T., & Gannon, S. (2013, March). *U.S. Navy humanitarian assistance in an era of austerity*. Washington, DC: Center for Strategic and International Studies.

Rouse, P., & Putterill, M. (2003). An integral framework for performance measurement. *Management Decision, 41*(8), 791–805.

Salmeron, J. & Apte, A. (2010). Stochastic optimization for natural disaster asset prepositioning. *Production and Operations Management, 19*(5), 2010, 561–575.

Sanderson, D., & Ramalingam, B. (2015). *Nepal Earthquake response: Lessons for operational agencies.* London: ALNAP/ODI.

Sawhill, J. C., & Williamson, D. (2001). Mission impossible? Measuring success in nonprofit organizations. *Nonprofit Management and Leadership, 11*(3), 371–386.

Smart, B. (2012). Responding in a new environment: Lessons learned by an Australian urban search and rescue team in Japan. *Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 5*, 61–66.

Tangen, S. (2004). Performance measurement: From philosophy to practice. *International Journal of Productivity and Performance Management, 53*(8), 726–737. doi:10.1108/17410400410569134

Tarantino, D. A., Suter, K., & Cooper, J. (2016, Spring). Nepal Earthquake: Anatomy of a non-government response. Liaison: *A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 13,* 67–74.

Terada, H. (2012). Japan Self-Defense Force’s response to the Great East Japan Earthquake. *Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 5*, 39–44.

Trunkey, R. D. (2013). *Implications of the Department of Defense Readiness Reporting System* (CBO working paper). Retrieved from https://www.cbo.gov/sites/default/files/113th-congress-2013-2014/workingpaper/44127_DefenseReadiness_1.pdf

United Nations. (2007). *United Nations Human Development Report 2007/2008–Fighting climate change: Human solidarity in a divided world.* New York, NY: Palgrave Macmillan.

United States Marine Corps (USMC). (2009, August). *Policy for Marine expeditionary units (MEU) and expeditionary units (special operations capable) (MEU SOC)* (MCO 3120.9C). Retrieved from http://www.marines.mil/News/Publications/ELECTRONICLIBRARY/ElectronicLibraryDisplay/tabid/13082/Article/126201/mco-31209c-final.aspx

van der Laan, E. A., de Brito, M. P., & Vergunst, D. A. (2009). Performance measurement in humanitarian supply chains. *International Journal of Risk Assessment and Management, 13*(1), 22–45.

Weinberger, S. N. (2010, November). *Who should set airlift priorities during foreign humanitarian assistance/disaster relief operations and on what basis?* Retrieved from http://www.dtic.mil/docs/citations/ADA535096

Wanlach, J. (2012). Putting lessons into practice. *Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 5*, 89–94.

Wendelbo, M., La China, F., Dekeyser, H., Taccetti, L., Mori, S., Aggarwal, V., Alam, O., Savoldi, A., & Zielonka, R. (2016, May). The crisis response to the Nepal earthquake: Lessons learned. *European Institute for Asian Studies*. Retrieved from http://www.eias.org/wp-content/uploads/2016/02/The-Crisis-Response-to-the-Nepal-Earthquake-_-Lessons-Learned-colour-1.pdf

Wilson, R. K. (2012). *Operation Tomodachi: A model for American disaster response efforts and the collective use of military forces abroad*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a567991.pdf

Wisner, J. D., & Fawcett, S. E. (1991). Linking firm strategy to operating decisions through performance measurement. *Production and Inventory Management Journal, 32*(3), 5–11.

World Meteorological Organization. (2009). Fact sheet: Climate information for reducing disaster risk. Retrieved from http://www.wmo.int/wcc3/documents/WCC3_factsheet1_disaster_EN.pdf

Yoshitami, N., Arie, K., Tamura, S., Hirose, R., Imamura, E., Sakaguchi, D., & Saito, D. (2012). *Military-public-private cooperation in disaster relief. Liaison: A Journal of Civil-Military Disaster Management & Humanitarian Relief Collaborations, 5*, 21–37.

Navy Expeditionary Logistics

Uday Apte—is a Distinguished Professor of Operations Management at the Graduate School of Business and Public Policy (GSBPP), Naval Postgraduate School, Monterey, CA. He also serves as the Associate Dean of Research and Development at GSBPP. Before joining NPS, Dr. Apte taught at the Wharton School, University of Pennsylvania, Philadelphia, and at the Cox School of Business, Southern Methodist University, Dallas. He is experienced in teaching a range of operations management and management science courses in the executive and full-time MBA programs. Prior to his career in academia, Dr. Apte worked for over 10 years in managing operations and information systems in the financial services and utility industries. Since then he has consulted with several major U.S. corporations and international organizations.

Areas of Dr. Apte's research interests include service operations management, supply chain management, technology management, and globalization of information-intensive services. He has completed over 10 sponsored research projects for the U.S. DoD and has published over 60 articles, five of which have won awards from professional societies. His research articles have been published in prestigious journals including *Management Science, Interfaces, Production and Operations Management (POM), Journal of Operations Management (JOM), Decision Sciences Journal (DSJ), IIE Transactions,* and *MIS Quarterly*. He has co-authored two books, *Manufacturing Automation* and *Managing in the Information Economy*. Dr. Apte has served as a vice president of colleges at Production and Operations Management Society (POMS), as a founder and president of the POMS College of Service Operations, and as guest editor of POM journal. Currently he serves as the senior editor of POM and as associate editor of DSJ.

Dr. Apte holds a PhD in Decision Sciences from the Wharton School, University of Pennsylvania. His earlier academic background includes an MBA from the Asian Institute of Management, Manila, Philippines, and Bachelor of Technology (chemical engineering) from the Indian Institute of Technology, Bombay, India.

Abstract

The U.S. Navy, with its expeditionary warfare and logistics capabilities, is increasingly playing a critical role in conflicts involving non-state actors. Given the difficulties faced in expeditionary environments, managing expeditionary logistics is particularly challenging yet critically important today. In this research, we use case study methodology to better understand the current practices and challenges of expeditionary logistics at Explosive Ordnance Disposal (EOD), a subordinate entity of Naval Expeditionary Combat Command (NECC), and to develop a set of concrete recommendations for improving expeditionary logistics processes at EOD. We also study the current definitions of expeditionary logistics to propose a definition better suited for today's challenges, analyze similarities and differences between expeditionary logistics and commercial logistics, and develop concepts for successfully managing expeditionary logistics operations.

Introduction

In recent decades, the United States and its coalition partner countries have increasingly engaged in conflicts involving non-state actors, and it appears that these conflicts are unlikely to subside in the foreseeable future. As a powerful maritime force, the U.S. Navy plays a critical role at sea and on land in these conflicts through its expeditionary warfare and logistics capabilities. The critical requirement here is to enable operational units to carry out a short-duration mission autonomously without the routine support of a base network. Given the difficulties faced in non-permissive expeditionary environments, managing expeditionary logistics is particularly challenging yet critically important today.

However, our prior research shows that there exists scarcity of research literature or DoD documentation and guidance available on this important topic. The proposed research project will therefore begin to address this gap by developing conceptual frameworks and concrete recommendations on designing and successfully managing expeditionary logistics in non-permissive environments.

In this research, we adopt a two-phased approach. In the first phase we study a specifc instance of current expeditionary logistics (ExLog) operations in practice to (1) better understand the key elements and critical success factors of ExLog, and (2) develop recommendations for improving logistical processes being studied. In the second phase we build on the results of the first phase to develop concepts useful for optimally designing ExLog processes and successfully managing them. The specific research questions we address in the second phase are: What is expeditionary logistics and what are its key components? What are the similarities and differences between expeditionary logistics and the traditional commercial logistics? What are some of the best practices of the traditional commercial logistics that Exlog can benefit from? How to successfully manage Exlog operations?

As a starting point in this research we conducted an extensive literature survey of relevant research published in journals and books as well as documents published by the DoD on the topic. To limit the length of this symposium paper, however, we do not provide results of that literature survey, and provide instead only an overview of our own past relevant research in this area. In FY2014 we worked on an exploratory research project sponsored by the Office of Naval Research (ONR) on the topic of expeditionary logistics (Apte & Kang, 2015). A team of MBA students assisted us in that research by undertaking their MBA project to study the logistical challenges faced by the Explosive Ordnance Disposal (EOD) and Naval Special Warfare (NSW) communities (Kundra, Brown, & Donaldson, 2014). The study indicated that the main shortcoming of logistical processes was in the information systems support, and the capture and analysis of information regarding the supplies, materiel, and equipment used in expeditionary logistics. In a follow-up project funded by IMET, we developed a case based on our earlier research for use in an advanced logistics course (Yoho & Apte, 2018). This case is currently used in the capstone course of the logistics curriculum at GBBPP. Finally, continuing with our research on the same topic, two MBA students studied under our guidance the details of logistical processes supporting the deployment cycle of a Mine Countermeasure (MCM) Platoon at EOD (Reeves & Baker, 2017).

We should mention that as of the writing of this symposium paper we have yet to complete our research on the topic and the work is ongoing. Specifically, we have completed the first phase, but the work in the second phase regarding the developments of concepts and theory about expeditionary logistics is ongoing.

This research report is organized into five sections. This introductory section is followed by the second section describing the organization and the mission of Explosive Ordnance Disposa (EOD), the specific instance of ExLog proceses we studied. The third section presents details of the logistical processes supporting the deployment cycle of a (hypothetical) Mine Countermeasure (MCM) Platoon at EOD, while the fourth section provides the analysis, conclusions, and recommendations concerning the management of those logistical processes at EOD. We complete the paper in the fifth and final section of this working paper with our initial thoughts on the characterization and definition of expeditionary logistics as well as the comparison of expeditionary and commercial logistics.

Explosive Ordnance Disposal (EOD)

The Explosive Ordnance Disposal (EOD) is a subordinate entity of the Navy Expeditionary Combat Command (NECC) which is the Navy's expert command regarding expeditionary operations and logistics. The NECC exists to man, train, equip, and sustain the Naval Expeditionary Forces (NEF) for bridging the gap from operations at sea to sea–land joint operations. While the NECC command is relatively new, stood up by the Chief of Naval Operations (CNO) in January 2006, NEF is old. The NECC is composed of eight subordinate entities that are their own respective commands which deliver the unique capabilities to the U.S. and its allied forces in the expeditionary realm: Costal Riverine, Explosive Ordnance Disposal, Naval Construction (Seabees), Expeditionary Intelligence, Combat Camera, Expeditionary Logistics, Maritime Civil Affairs and Training, and Expeditionary Combat Readiness.

The Explosive Ordnance Disposal (EOD) is the Navy's technical expert in locating, identifying, rendering safe, and explosively detonating foreign and domestic ordnance. Ordnance includes conventional, nuclear, biological, chemical, underwater, and improvised types of devices. The ability to control and dispose of these various types of dangerous devices enables theater access for Carrier Strike Groups (CSGs), Expeditionary Strike Groups (ESGs), Naval Special Warfare, and Army Special Forces (SF).

EOD is a history-rich, proud community that serves alongside many SPECOPS forces, as well as traditional Navy mission communities such as ships and submarines. The EOD technicians risk their lives to perform complex, technical defusing of mines, bombs, and improvised explosive devices (IED) and, of necessity, are required to be physically fit, superior swimmers, and athletes. EOD technicians undergo rigorous schoolhouse training prior to arriving at their commands and then complete operationally challenging tours filled with deployments and stressful workups due to the high operational tempo (OPTEMPO).

EOD Group ONE, based in San Diego, CA, and EOD Group TWO, based in Little Creek, VA, are the two U.S.-based EOD elements. Each EOD group has five battalions and various shore detachments, platoons, and companies within them. The groups provide specially trained, combat ready, highly mobile EOD forces to support CSGs, amphibious ready group (ARG)/Marine Expeditionary Units (MEUs), MCM task forces and groups, NSW forces, Army SF, Military Sealift Command, unified theater commanders, CONUS Navy Region commander, and Homeland Defense and Contingency Operations.

EOD Expeditionary Support Unit (EODESU) ONE and TWO follow the same geographical structure as their fellow expeditionary forces. EODESUs provide total logistics support to the EOD forces through financial, supply chain, and logistics management as well as operational planning and global force support. Prior to formation of EODESU, ExLog was performed by the EOD teams while simultaneously experiencing stressful OPTEMPOs and very dangerous deployments stacked one after the other. The purpose of the EODESU was to relieve the EOD mobile units of logistics and maintenance duties so they could focus on their demanding operational duties. Also, the EODESUs are staffed with logistics and maintenance experts in order to perform those functions more efficiently and with increased precision, ultimately adding greater value to the Navy and improving the result provided to the warfighter.

While ESU commands are not tasked with executing any of the highly technical and versatile missions the EOD teams are tasked with, they are tasked with equipping those teams with the proper gear and equipment to successfully execute the mission. In order to know what is required and understand the details necessary to complete these difficult missions, the ESU must be knowledgeable about the EOD missions and the gear and equipment EOD teams need.

Mine Countermeasure Platoon at EOD: A Case Study

This case study focuses on EOD Mobile Unit (EODMU); in particular on one of its Mine Countermeasures (MCM) Platoons. The case begins with an overview of the logistical processes and information systems used by EODESU TWO to provide the necessary to support a (hypothetical) MCM Platoon 1201. That is followed by a description of the MK-16 equipment—an underwater breathing apparatus—since it is a heavily utilized piece of gear by the MCM Platoon. The case then tracks the logistical processes used to support the activities of the MCM Platoon throughout its deployment cycle. The case thus provides a realistic insight into the operations of the EODMU MCM Platoon and the logistical support operations of EODESU TWO.

EODESU Supply

The ESU units supply their teams as part of the services they provide. While the process by which the individual units perform this function may be slightly different, relatively speaking the same outcome is delivered. The EOD units supported by EODESU receive a variety of supplies and equipment, including expeditionary logistics overhaul (ELO) and general logistics and supply chain support. ELO is similar to the integrated logistics overhaul (ILO) process aboard ships but is specifically designed for expeditionary forces where they identify the gear needed to be repaired, reconditioned, or replaced.

ESU teams issue the following types of gear to the EOD teams they support:

- PGI (Personal Gear Issue) includes items such as uniforms, undershirts, socks, and other items that require some specificity to a member's body and measurement.
- TOA (Table of Allowances) consists of specific gear, equipment, systems, and materiel related to expeditionary missions. TOA includes items such as inflatable boats, generators, and specific wetsuits. The EOD teams keep this gear with them from the start of the Fleet Readiness Training Plan (FRTP) to the post-deployment return. TOA gear represents a challenge to the cost savings efforts due to constantly changing and non-standard nature of the gear allowed or allotted to the expeditionary teams based on their specific missions.
- COSAL (Consolidated Shipboard Allowance Listing) includes items that the ship normally carries on board. The COSAL contains nomenclature, operating characteristics, technical manuals, and equipment descriptions as described in allowance parts lists (APL) and allowance equipage lists (AEL).
- Other non-COSAL material.

The expeditionary requisitions processes are unique when compared to the standard fleet requisition processes. In the standard fleet requisition processes, over 95% of the requisitions are filled through the Navy supply system using National Stock Number (NSN) items, while less than 5% are open purchases. In comparison, in expeditionary logistics, approximately 70% of the requisitions are open purchases and only 30% are NSN requisitions (Kundra, Brown, & Donaldson, 2014). The expeditionary environment and mission add unique variables, such as distinctive operating environment, the need to stay current with technology, and the greater need for speed. Because of its availability, these situations force expeditionary units to rely heavily upon open purchases for commercial off-the-shelf (COTS) or local procurement products.

To track and store information regarding the above gear, the ESU units use multiple information systems, including the following:

- WASP: A warehouse and inventory management system
- RCRP: Readiness and Cost Reporting System
- R-Supply: A system that provides the Navy with online inventory, logistics, and financial magement tools
- DPAS: A DoD-required system that tracks property valued greater than $5,000.

The WASP, RCRP, R-Supply, and DPAS are distinctly different IT systems that are used to organize about the same type of information. In some cases, the information is actually the same, and duplicate efforts are being made to track and store transactions in different systems because the systems are not able to automatically share information with each other. For example, ESU tracks a transaction first in WASP and then manually enters the same information in RCRP. Another example is when supply parts are received from vendors at the ESUs and are automatically confirmed in R-Supply. Subsequently, the ESU members manually enter the same information that was just confirmed in R-Supply into WASP because the ESU teams use WASP as their internal inventory management system, and because there exists no interface to automatically share information between the two systems. A factor that further complicates the matter is a requirement to store information on an Accountable Property System of Record (APSR) system and since WASP is not an APSR system.

Upon completion of the mission, training, or cycle, the gear that is not meant to be kept by a member is returned and inventoried. The gear return process is more than simply stacking and counting specific clothing articles or ammunition boxes. Given the nature of EOD missions, many times the gear gets returned, but in a heavily damaged and potentially unusable state. Therefore, ESU inspectors must know what separates returned, quality gear, from gear requiring minor maintenance or depot-level repair (DLR).

In addition to managing the inventory of existing gear, ESU is required to properly document missing and damaged gear that is beyond repair. DD Form 200 is the Navy's form for financial liability investigation, required in the process that is initiated by submitting a DD Form 200. The Navy must determine, based on DD Form 200, the reason the equipment was lost or damaged and who should be responsible, if anyone, for the cost to repair or replace. DD Form 200 is required as per DoD Directive 7200.11 for lost DoD-controlled property. It is a form that is filled out electronically, but ultimately it is also kept as a hard copy and entered into the ESU IT systems manually. ESU members are required to physically search archived DD Form 200s when they need to find information.

DD Form 1149 is another DoD directive form that is required when shipping through certain seaports or airports. The DD 1149 is specifically known as the Requisition and Invoice/Shipping Document to verify what was issued against the electronic records in WASP. This hard-copy document is also manually entered into systems and kept hard-copy for storage or later use when searching for information. There is a large collection of files at EODESU TWO of forms that are necessary to conduct business but are only stored as a hard copy.

The MK-16 Underwater Breathing Apparatus

Navy EOD is the only service manned, trained, and equipped to perform underwater render safe procedures and conduct EOD dive operations. Typical EOD mission sets include Mine Countermeasure (MCM), salvage diving, ship's hull diving, search and rescue (SAR) operations, and other necessary diving missions. With such a variety of technically challenging and highly dangerous diving missions, EOD technicians are trained to perform and be successful at nearly any diving mission. The MK-16, therefore, is a common piece of equipment used in the EOD teams, and all EOD technicians are well-versed in its use and capabilities.

The MK-16 was developed to reduce magnetic and acoustic signatures emitted by diving EOD technicians. The mission of EOD technicians is one that is highly technical, diverse, and dangerous. Under such tense work conditions, a superior diving suit is required that allows full range of motion but still provides protection from the natural and enemy hazards present in the area of operation (AO). The MK-16 breathing medium is maintained at a predetermined partial pressure of oxygen (PO2) that is monitored by sensors and controls to ensure diver safety. The reason divers are required to maintain a safe level of oxygen and are monitored so heavily is that depending on the mission, they may use more or less oxygen and cannot follow a standard timetable for bottom time.

Along with MK-16, a diver's other essential equipment includes knife, hook knife, strobe, smoke or flare, thermal protection, fins, and potentially a weapon as required. The knife has many uses, but one of its main uses is to help free a trapped diver from any number of hazards. The MK-16 equipment must withstand these conditions and not puncture, disconnect, or break easily. Strobes, smoke, and flares are essential safety gear for EOD technicians because at the depths required of some of the EOD missions, there is absolutely no natural visibility and those pieces of equipment could prove to be life-saving. A weapon is a necessity depending on the mission and AO in which the dive will take place; this is a harsh reminder that the mission is not a recreational dive but is highly important and dangerous.

EODESU TWO has a team of maintainers as well as a GS civilian employee who accounts for and maintains the MK-16 system inventory. The GS civilian employee is known as the resident expert on the system. The benefit to having a civilian expert versus a military member is that ideally, the civilian remains the expert point of contact for a longer period of time, providing a long-term persistent presence as opposed to the routine rotations of assigned active duty personnel. This ensures retention of critical corporate knowledge regarding program supply and maintenance history.

The Logistical Support of MCM Platoon

This case study focuses on the Mine Countermeasures (MCM) Platoon 1201, which is a primary end-user of the MK-16. The case follows the supported unit through its training cycle, deployment, and ultimate return to the home base.

Pre-Deployment

Preparation for any deployment begins with a Fleet Readiness Training Plan (FRTP), a codified training cycle. Concurrent with the assigned deployment schedule, the MCM Platoon undergoes a FRTP cycle like most other Navy units. The purpose of this process is to train, equip, and certify unit mission preparedness. FRTP consists of various milestones, including inspections, evaluations, training, and exercises. Each one of these events helps to build unit skill and cohesion, starting with basic, individualized training, and working toward more advanced, integrated training with external units. The process is designed to prepare the unit for the upcoming deployment based on available intelligence data (intel) gathered prior to heading into theater. This same intel is what EOSESU TWO uses to prepare supply and logistics support. EODESU TWO outfits the units during expeditionary logistics overhaul (ELO) and issues all of the required gear aside from what has already been issued for the team to be successful on deployment.

The FRTP for an MCM Platoon begins with a tightly packed schedule of training events (also referred to as "workups") lasting roughly 11 months from the start. Upon completion of the workup cycle, the platoon stays in a six-month sustainment phase, when they are certified for operations, and thus may be deployed early if necessary. Otherwise, they maintain their availability status until departing on a six-month deployment, which completes the 24-month deployment cycle.

Prior to FRTP, the platoon receives expeditionary logistics overhaul (ELO) from EODESU TWO and begins workups. Part of the workups include successful completion of the requirements of the Training and Evaluation Unit (TEU). TEU does not completely oversee the FRTP process for the platoon, but provides training, classes, study materials and equipment, and some evaluation for how the unit is able to perform against the various elements of the deployment they are likely to face. At times, TEU directly issues some duplicate equipment that is required during the training. This prevents the platoon from utilizing primary issue equipment, and thereby avoiding any potential damage or loss to mission-essential gear, which in some cases can delay deployment or reduce mission capabilities of the unit. The TEU has its own supply of gear that it accounts for and purchases via EODESU TWO to support the unit training and evaluation process. ESU controls the budget used by TEU to purchase their course gear, which they acquire via DoD e-mall, GSA Advantage, GSA Leasing Support vendors, prime vendors, or other government sources of acquisition. The gear issued by TEU is generally the same as what is issued by ESU but a slight variation is possible.

ELO/Gear Issue

At the start of the deployment cycle, platoon undergoes ELO to get outfitted with the gear required for training and subsequent deployment tasking. This ELO process facilitates the issuing of a baseline of standard gear that EODESU TWO has developed over time based on coordination with the EODMUs and their historical tasking. Scheduled six to 12 months in advance, and based on long-term deployment rotations that are often available two years prior, the platoon's ELO takes approximately three weeks to fully transfer the ownership of thousands of required pieces of gear from the ESU to the platoon. The process starts with coordination between EODESU TWO and EODMU, to deconflict an appropriate start date, based on all units that may need similar support.

To start preparing for the ELO, EODESU TWO typically designates four Internal Airlift/Helicopter Slingable Container Unit 90 (ISU 90), along with a mini flyaway dive locker (FADL), for storage of all ELO gear issued to the platoon. At the completion of ELO, the ownership of these storage units will be transferred to the platoon. Before the gear is moved from the warehouse to the storage containers, EODESU supply personnel generate a DD 1149 listing all of the items required for transfer. Each commodity manager is responsible for populating a DD 1149 with the appropriate items under his purview. These documents serve as the official inventory record for equipment ownership, and in the interim, also serve as an inventory checklist utilized by both ESU personnel and the platoon commander, for verifying all items transferred.

The DD 1149 information must be entered in two separate systems. First, all items must be properly accounted for in the warehouse. The IT system utilized in maintaining an accurate warehouse accounting is Wedge Advanced Software Product (WASP). WASP is a standalone warehouse management system. ESU personnel must go into WASP to update the ownership/location status of each item, as it is transferred to the storage containers. Additionally, this same supply/inventory information must be entered into the Navy's Readiness and Cost Reporting Program (RCRP), which is the approved system of record for use in official reporting up the Navy chain of command, and which is not connected to WASP. Though WASP is not an approved system of record, it is used locally for the convenience and simplicity it provides in managing the local inventory.

DPAS warehouse is another inventory management system that is available to the supply community that satisfies about the same requirements as WASP, but adds date entry efficiencies such as bar code scanners. EODESUs have yet to implement the new system. WASP is utilized for the majority of ELO transfer items, but not for underwater items. Due to the much smaller inventory of underwater items, the dive locker works primarily with RCRP (for ownership transfer), OMMS (for repair/maintenance), and spreadsheets (for ad hoc local tracking). Once the containers have been filled, and ESU and the platoon commander have verified the transfer, the platoon commander signs the DD 1149, accepting ownership of the containers and their contents.

While the platoon usually receives the entire complement of gear required for deployment, at times, adjustments to the process are made based on supply availability and community demand for limited equipment, such as the MK-16. For example, the dive locker may delay issuance of the MK-16 if there is excess demand for use at the TEU in preparing other units for their own deployment schedule. Additionally, since mine countermeasures is a primary mission of MCM platoons, they are typically outfitted with MK-16 at the start of workups, regardless of needs of others. However, other platoons that treat mine countermeasures as a secondary mission may experience a delay in isuing the MK-16 during workups. However, in case of delay, they are provided equipment on a short-term basis as they commence specific MK-16 training evolutions during the workup cycle, and receive the full issue prior to deployment.

After about 18 months of training and sustainment, MCM platoon is deployed. During sustainment and deployment, the process for acquiring repair and replacement equipment is essentially an à la carte version of the ELO process, which is discussed in the following section.

Deployment

Upon completion of the training cycle, any training-specific gear issued by the dive locker is returned, and any outstanding ELO gear requirement is fulfilled by the ESU prior to departure. The unit then embarks on the deployment to support real-time tasking from theater commanders, execute pre-planned missions, or operate independently, depending on theater demands. The MCM Platoon 1201 is tasked with conducting a dive mission to clear a port in the Persian Gulf. This is a routine anti-terrorism/force protection (ATFP) mission to ensure safe passage for a naval surface action group (SAG), scheduled to arrive soon.

This EOD MCM Platoon is made up of eight EOD technicians. As part of their standard complement of gear, they are issued five MK-16 units and one operational support kit (OSK), which should be enough to handle the job. After four days of dive operations, two of the MK-16 units are in need of servicing. Several O-rings need replacement and one of the units needs an oxygen addition valve replaced. Until they are serviced, these MK-16s are not safe for use. In order to meet the necessary pace of operations and to avoid any extended time on station, they need to get the equipment repaired. Fortunately, these items are available within the OSK. After a quick repair evolution, all MK-16 units are fully operational. This allows the platoon to meet the mission requirements as scheduled, and more importantly, this allows a follow-on naval SAG to pull into port safely and on time.

The use of parts from the OSK, along with a subsequent replenishment request from the platoon, create a demand signal for execution back at EODESU TWO. The goal is to maintain a fully-stocked OSK, to provide some maintenance capacity on-site. With other commodities, the platoon typically coordinates with the Expeditionary Support Element (ESE), based in the theater. The ESE routes these requests through the appropriate commodity manager at ESU TWO for processing. However, in the case of underwater commodity items such as the O-rings and oxygen addition valve, they typically send e-mail to the dive locker personnel directly to request the necessary items. From a supply standpoint, this current request can be fulfilled in two different ways. The routine expendable items (the O-rings), are available immediately from the supply warehouse. The commodity manager enters the request in OMMS, which routes the request through the chain of command for approval. Once approved, the request goes to the warehouse to tag the O-rings for distribution to Platoon 1201. The oxygen addition valve, however, is considered a depot-level repair (DLR) item, and therefore is handled somewhat differently. DLR basically means that the item cannot be locally serviced, and must be sent to a dedicated repair facility. The oxygen addition valve is requested in similar fashion as the O-rings, using e-mail and an OMMS job order. However, the platoon must also send the failed part back to the ESU for exchange. The exchanged part is turned in to the depot repair facility, where it is refurbished or discarded as unserviceable. The repair facility provides a replacement part to ESU, likely a refurbished item from a previous repair. The dive locker at the ESU then generates a DD 1149 to document the parts delivery, make any necessary updates required in RCRP, and ship the O-rings and oxygen addition valve out to Platoon 1201. Upon receipt, the platoon has a DD 1149 for their records, and the OSK is back to full operational status.

This process repeats throughout the deployment, to facilitate repair and replacement activity on the MK 16.

Post-Deploymnt

Upon return, the Platoon 1201 follows up with EODESU TWO, to conduct all necessary equipment turn-in, along with associated documentation processes. The purpose of this effort is to reconcile supply-related activity that occurred throughout deployment and close out any outstanding logistics support requirements. While the Platoon 1201 is able to turn in its equipment with a materiel loss of about 5%, the record of the past turn-ins show that it is not uncommon to experience a materiel loss of as much as 30%.

Just as when the platoon received initial gear issue, the primary process for gear return is also ELO. This involves presenting any remaining gear to the supply warehouse for reconciliation. ESU personnel receive the gear, accept functional or repairable gear into inventory, and properly account for other equipment that is either unusable or lost. Functional gear may be cleaned and prepped for immediate redeployment, while repairable gear will be processed for repair or refurbishment before being returned to mission-capable status. ELO and associated data reconciliations to RCRP are important steps in the process for ensuring accountability for inventory levels. These steps support the ongoing financial improvement and audit readiness (FIAR) initiative across the DoD.

For TOA and PGI gear, Platoon 1201 returns to the supply warehouse at EODESU TWO to transfer ownership of the preponderance of ELO. Again, the process takes approximately three weeks to complete. Using the original DD 1149 document from ELO issue, along with accumulated DD 1149s generated throughout deployment for parts orders, the Platoon 1201 commander works with ESU personnel to inventory all returned items. All equipment is designated as mission-capable, serviceable, unserviceable, or missing. After accounting for all items, ESU personnel return to WASP and RCRP for appropriate electronic transfer of ownership. In the case of unserviceable or missing items, a form DD 200 must be generated to account for the loss. It is the responsibility of Platoon 1201 to generate the DD 200 and route it through their chain of command for review. A copy is provided to EODESU TWO to facilitate record keeping and to ensure inventory items are appropriately removed in WASP and RCRP, to avoid overstating the value and quantity of existing inventory.

Occasionally, due to operationally constrained deployment timelines, there is pressure to expedite the ELO process between deploying and returning platoons. A solution employed by EODESU TWO is a modified ELO. Requiring a surge of personnel and a tightly coordinated schedule, this allows a returning platoon to transfer inventory directly to another platoon starting workups. This also requires coordinated commitment from both platoons and ESU, and can reduce the typical three-week process down to one week.

Analysis, Conclusions and Recommendations Regarding the Logistical Processes at EOD

The problems and shortcoming of logistical processes described in the previous section were analyzed using selected tools of Lean Six Sigma (LSS), a process improvement methodology. Specifically, we used tools such as the Process Flowchart and the Cause and Effect analysis (resulting in a Fishbone Diagram). The Fishbone Diagram is provided in Figure 1, while a discussion of the cause and effect analysis follows. It should be noted that the analysis is organized as per the major causes shown in the Fishbone Diagram. As a sample, the flowchart of ELO/Gear Issue process is provided in the appendix.

After preparing flowcharts of all processes, they were analyzed to identify root cause(s) of various problems facing the expeditionary logistics operations.

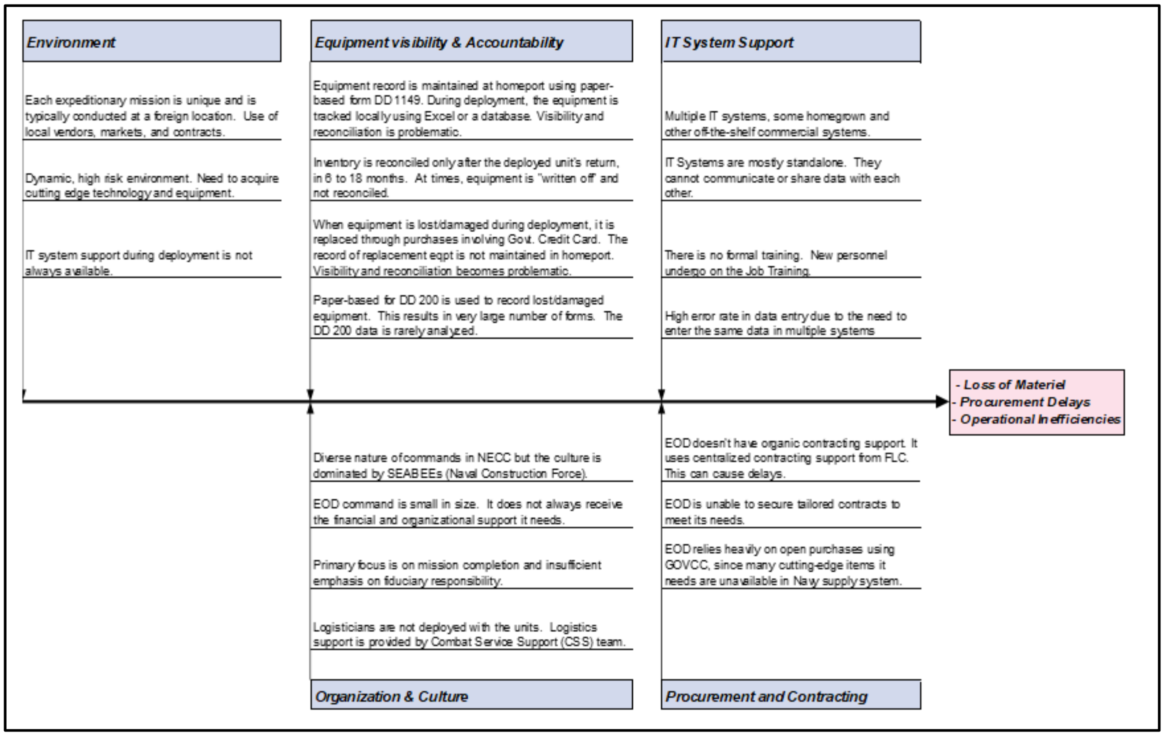


Figure 1. Fishbone Chart

Information Technology (IT) System Support

As previously mentioned, EOD represents a very small portion of the Navy's overall manning. Consequently, they are unable to claim funding that would permit them to have a written contract tailored with an inventory management program to meet their needs. As a result, EOD "makes do" with the systems it has: the commonly available commercial products such as WASP for inventory management and Navy-approved systems such as R-supply for financial management. WASP was implemented as an inventory management system several years ago. It is an improvement over their previous methods of using Microsoft Excel spreadsheets. The use of multiple systems to perform inventory management results in an ad hoc inventory management system that requires double entries to maintain duplicate databases, as well as extensive, lengthy periods of on-the-job training to master the systems. Multiple systems are required to maintain equipment inventories, and specific programs are required to be used for certain categories of equipment. Moreover, none of the databases for these inventory management programs are able to share information with the software used to track finances (R-Supply).

The effect this is having on the EOD logistics operations is reduced efficiency and effectiveness. For a typical Logistics Specialist, a tour in an expeditionary unit is unusual. The requirement to use multiple computer systems negates a key benefit of computer technology by multiplying the work required by the user. The need to perform repetitive data entries is also an invitation for natural human error. These inevitable errors introduce inaccuracies into the inventory and usage data, contribute to a loss of accountability and an inability to optimally manage inventory, and reduce buying power for the taxpayer.

Most who serve in such a unit do so for only one tour, and the majority in the Navy spend an entire career on sea-going ships. As a result, most systems that a logistician encounters in an expeditionary command are highly unfamiliar. These programs have no formal Navy training available and the Sailor must learn through on-the-job training for up to 18 months. However, as we understand, the USMC utilizes logistics programs along with the associated schools. We recommend further research be conducted to determine if similar programs and schools could be adapted for use by Navy.

Procurement and Contracting

This study showed that EODESU relies heavily on open purchases using contracts or government credit card through the commercial sources. The goal is to have the gear fixed or replaced through the fastest means possible. Relying on the readily available product allows them to procure and stay current with technology advancement at a much faster rate in an effort to always stay ahead of the next potential threat.

During our research we found that Naval Special Warfare (NSW) possess an organic ability to write and administer contracts. This greatly increases the speed with which equipment and services are obtained at NSW while reducing the workload on the unit's logisticians. In contrast, EOD is required to use the contracting services of the Fleet Logistics Center (FLC). This increases delays and administrative workload. As mentioned earlier, EOD represents a very small portion of the Navy's overall manning. Consequently, they are unable to claim funding that would permit them to have a written contract tailored with an inventory management program to meet their needs.

There is a demand for cutting-edge equipment among members of the EOD community. This compels their logistics support units to rely heavily on the use of their government commercial purchase card (GCPC.) The Navy's supply system is best suited to providing parts and equipment to traditional ships and submarines. Relative to EOD, these platforms face threats and challenges that change slowly, and as a result, the equipment and supplies they need are slow to change. This is not the case with expeditionary units. They operate in a much more dynamic environment. While a ship may have a service life of up to 50 years, much of the equipment used by EOD has a service life that is measurable in months. This time frame does not permit economical parts support. Incorporating this into the traditional maintenance model of a sea-going ship could be meaningful. The GCPC permits these commands to obtain the required equipment quickly, but this does not come without consequence. The process of purchasing with Government Purchase Card records purchase information in a form that is not readily accessible to external organizations. For example, you record the total dollar amount spent on a purchase in one system; however, the list of purchased items gets recorded individually into a different system. The two systems are not compatible with each other. The amount spent and the list of items on that purchase can only be reconciled manually by reviewing the original receipt. As a result, demand history is lost along with the ability to easily audit expenditures. Without any accurate demand history, the task of procurement and inventory management becomes significantly more difficult.

Equipment Visibility and Accountability

The process used to issue and maintain accountability of equipment is inadequate, particularly during a unit's deployment. Based on the preliminary information available at the time of this research and the interviews of subject matter experts, this study's researchers estimated that loss of materiel accountability is, at times, as high as 30% for EOD. This is mainly attributable to the methods used to assign and record accountability for equipment, the relative ease with which equipment can be replaced, and the inability to detect trends in purchases and/or surveys.

Prior to deployment, accountability for the equipment is assigned to an individual or team using a paper DD form 1149. During the EOD deployment, no supply or logistics personnel from the team's unit are deployed with them to provide support to deployed equipment and, as a result, the responsibility for maintaining custody falls to the Expeditionary Support Element (ESE) in theater. The ESE is required to do these using possibly suboptimal methods such as Excel spreadsheets or a locally maintained database. It should be noted that even when the database is a familiar program, if the EOD unit is being deployed with another service, the unit is required to use that service's program. This introduces inefficiencies and reduces effectiveness in a manner similar to that previously described.

While deployed, the teams' focus understandably shifts to the successful accomplishment of the mission. Equipment, however, can be damaged or lost and replacements are obtained from respective in-theater logistics support units to ensure maximum readiness. This use of locally deployed support personnel helps the team maintain its capability to accomplish assigned tasks. It is likely, however, that separating the functions that maintain accountability from those that use the equipment helps create a culture in which containing materiel costs are not a major.

The paper DD 1149 records are maintained at the team's homeport and are unable to be updated when equipment is lost or destroyed and subsequently replaced. Additionally, because gear that is deployed with a unit is by definition "mission essential," replacing it is a high priority. Consequently, a given piece of equipment may be replaced several times during a deployment, but it is only upon the team's return to home port that its equipment and equipment inventory records are reconciled. Some gear is deployed and returns with an individual Sailor or unit and discrepancies will be detected after the six-month deployment is concluded. As described previously, however, because these purchases are likely to have been made using a GCPC, the record of any replacements purchased during this time is largely obscured.

Additionally, a significant amount of equipment will only be reconciled after 18 months or may never be reconciled at all. Certain pieces of equipment are too costly to warrant purchasing in quantities sufficient to provide to each unit or too large to economically deploy and redeploy with a designated unit. This equipment is designated RIP/TOA and is turned over in theatre as units are relieved. Although this equipment may have a high value, because it may be more than a year since accountability was first assigned and procuring replacements for deployed equipment is relatively easy, this equipment may never be reconciled but simply "written off."

Because the method required to document and track equipment loss/damage relies upon hard copy paper documentation, it is likely that there is no effective means to accurately determine the cause of the loss/damage. Also, because of the nature of the control systems in place and the culture and attitudes it may engender, it is also unlikely that individuals with assigned accountability will be held accountable in the event of loss or damaged equipment. Lost, destroyed, and unserviceable equipment is properly recorded using the DD form 200. These forms, however, are produced at a rapid rate and the logistics units require several large binders to maintain a record of these forms. The documentation process is methodical. It is likely, however, that the sheer volume of paperwork makes it very difficult to assure accuracy in individual cases and to discern long-term patterns. Instead, the skill and memories of the unit's leaders and Sailors become the primary means for detecting trends. The reliance on paper forms and the volume with which they are produced places a significant administrative burden on the EOD logistics support commands while simultaneously obscuring trends in the information these forms record. It is probable that these factors make it unlikely that an individual Sailor will suffer any consequences in the event of a loss of accountability. This is because the same factors also make it difficult to detect a loss due to negligence or theft. The systems may also create the perception that the forms are a "paperwork drill." With this perception, it is likely Sailors prioritize their core mission responsibilities above any fiduciary accountability they may be assigned. In such an environment, it would also be inappropriate to punish the Sailor for responding to the incentives which he has been given.

Organization and Culture

The case study identifies a number of areas that offer the possibility of improved financial and operational efficiency. When considering the nature of these opportunities and the circumstances that brought them about, it becomes apparent that several key factors are at work. First among these is the miniscule size of the EOD community relative to the size of the traditional Navy. Because the expeditionary community makes up a relatively small portion of the Navy in terms of both manning and the number of mission sets to which it contributes, it is likely that the community's requirements are naturally assigned lower priority than those of the maritime force. The Navy must make choices regarding how it spends its resources to obtain the most satisfaction from its large, but nevertheless finite, resources. Consequently, it is plausible that an organization the size of the Navy would be unable to completely meet the needs of a minority of stakeholder organizations like EOD.

Another factor that contributes to the inefficiencies observed in the case of EOD is the diverse nature of the commands which make up NECC and their relative sizes. NECC is composed of 10 separate commands, with SEABEEs claiming more than half of NECC's personnel. At least on the West coast, this has resulted in the SEABEE culture dominating the NECC community and its requirements being given de facto higher priority.

Environment

Every expeditionary mission is different. While there are some similarities, the composition of the units deploying, the duration, and the environment where units are deployed can vary greatly. The following are some unique aspects of the operating environments that the expeditionary units face that make providing logistical support more challenging:

- Local vendors are used to provide as many supplies as possible.
- The dynamic, high-risk environment requires the latest technology to give the units that are deployed the best "edge" or competitive advantage possible.
- Information technology support is often unavailable during deployments due to the remote and/or austere environments.

Recommendations

The cause and effect analysis described earlier has led to the following set of recommendations for improving the logistical processes of EODESU:

- Information systems are highly inadequate and require multiple manual entry processes. Develop and introduce new information systems that will support expeditionary logistics. As an interim step, develop interfaces to enable single entry of data.
- Two important considerations to keep in mind before designing new information systems are (1) to first streamline the logistical processes and then design the information systems to fit the needs of that process, and (2) identify the data that will be needed to optimally manage the inventory and then design the information to capture those data elements.
- Currently, everything is on-the-job training with little knowledge capture or dissemination. Develop and deliver specific logistics training and education.
- When purchasing using government credit card (GCPC), the information is not tracked about which item is purchased, or how much or how often it is purchased. GCPC is a financial system and not a logistics system; it is used for tracking the amount of purchase but not what was purchased or its quantity.
- Given the large amount of money that passes through EODESU, having a full-time contracting official could possibly save money and time.
- There is a temptation to believe that because each expedition is unique and that the organization has always been able to “make it happen,” there is no need to improve processes from both efficiency and effectivity perspectives. However, there are always some commonalities between different expeditions and those commonalities should be identified and leveraged to achieve process improvements.

Expeditionary Logistics: Preliminary Concepts

Joint Publication 4.0 defines logistics as “planning and executing the movement and support of forces” (Joint Chiefs of Staff, 2013). Expeditionary logistics falls on the line between the operational and tactical levels. There are several definitions for expeditionary logistics available in various military instructions and publications. NECC adopted the expeditionary logistics definition stated in Navy Tactical Reference Publication 1–02, which defines expeditionary logistics as

> the science of planning and carrying out the movement and maintenance of an armed force organized to accomplish a specific objective in a foreign county. In its most comprehensive sense, those aspects of military operation that deal with design and development, acquisition storage, movement, distribution, maintenance, evacuation, and disposition of materiel; movement, evacuation, and hospitalization of personnel; acquisition or construction, maintenance, operation, and disposition of facilities; and acquisition or furnishing of services.

The levels of logistics correspond directly to the three levels of war: strategic, operational, and tactical. Strategic logistics focuses on organizing, training, and equipping the SOF forces, whereas operational logistics provides the link between tactical requirements to strategic capability in order to accomplish operational goal. They provide theater-wide logistical support, closely monitor in-theater shortfalls, communicate shortfalls to strategic sources, and continuously match tactical requirements with strategic recourses. Finally, tactical logistics primarily focuses on providing key services to support battles and engagements.

Two primary key areas of focus for ExLog are Sustainment and Combat Service Support. Sustainment provides forces the necessary equipment and services to maintain and/or prolong operations until successful mission completion. Effective sustainment allows combat commanders and expeditionary forces to have depth to seize, retain, exploit, and conduct decisive operations. Combat Service Support allows forward operating forces to have necessary supplies, equipment, transportation needs, and various services to support elements in theater at all levels of war.

Expeditionary logistics is challenged with the "tyranny of distance" since it often operates in areas far from Navy supply and distribution chains. Expeditionary logisticians often rely on host nations for support and make heavy use of local contracts, vendor support, and commercially available supplies.

ExLog: Functional Areas

Expeditionary Logistics is comprised of six functional areas: supply, maintenance, transportation, general engineering, medical, and other service (food, disbursing, postal, MWR, etc.). The main three components of logistics are supply, maintenance, and transportation.

- *Supply* functions as a materiel and financial management support that is similar to Supply Department afloat. The functions include ordering, procurement, receipt, stowage, and inventory control of repairable and consumables items.
- *Maintenance* functions as a team responsible for developing and performing all maintenance policies and procedures. In addition, they are also responsible for all equipment maintenance that preserves, repairs, and maintains reliability.
- *Transportation* takes care of movement of personnel and materiel from one point to another. They are well versed in worldwide ports of embarkation, debarkation, inter-theater, and intra-theater locations.
- *Expeditionary Engineering* is primarily a function of the Naval Construction Force, commonly referred to as "Seabees." Seabees can be deployed independently or can be imbedded into other expeditionary units. Seabees are capable of a wide range of construction services such as combat engineering, rapid runway repair, facility damage repair, combat engineering, bridge and road construction, and maintaining facilities ashore. In addition, they also provide responsive support in disaster recovery operations and perform civic action construction projects to improve relations with other nations.
- *Health services* include all medical, dental, and all health-related functions (combat and non-combat) to include: health maintenance, entomology, medical readiness of personnel, food service sanitation, treatment of casualties, and medical evacuation.
- *Other Logistic Services* function as a general area that includes services such as food, post, disbursing, exchange, billeting, legal, barber, laundry, and other administrative services and functions.

Comparison of Expeditionary and Commercial Logistics

Table 1 shows a comparison of expeditionary logistics and traditional commercial logistics along multiple dimensions. Expeditionary logistics often operates in foreign countries, in areas far from traditional Navy supply and distribution chains. Consequently, expeditionary logisticians often rely on host nations for support and heavily depend on local contracts, vendor support, and commercially available supplies.

Expeditionary logisticians support expeditionary situations that are substantially different and challenging as compared to those faced by the logisticians supporting traditional commercial operations. The stock keeping unit (SKU) variety-to-volume ratio—which describes the ratio of the number of different types of SKUs relative to the total volume of demand—is typically much higher in expeditionary operations. Meaning, the assortment of items is relatively high given the overall relatively low volume of logistical support demand. Table 1 provides a comparison of expeditionary logistics and commercial logistics along several dimensions.

Table 1. Comparison of Expeditionary vs. Commercial Logistics

Nature of Operation	Expeditionary Logistics	Commercial Logistics
Location	Foreign Country	Domestic and/or Foreign
Duration	Short Term	Long Term
Occurrence	Irregular	Routine
Demand	Variable	More predictable
SKU Variety-to-Volume Ratio	High	Low
Operational Tempo	Unpredictable	Steady
Level of Risk	High	Low
Desired Service Level	Very high due to low on-hand inventory levels	Medium to high due to the availability of local or regional distribution hubs
Distribution Dispersion	Low demand across many locations to serve few customers at each location	Use of large distribution centers or retail locations to serve many customers

References

Apte, U., & Kang, K. (2015). *Naval expeditionary logistics* [Project report]. Monterey, CA: Naval Postgraduate School, Office of Naval Research (GSBPP).

Joint Chiefs of Staff, Chairman. (2013). *Joint logistics*. Washington, DC: DoD.

Kundra, S., Brown, L., & Donaldson, C. (2014). Assessment of logistical support for expeditionary units [MBA professional report]. Monterey, CA: Naval Postgraduate School.

Reeves, D., & Brown, S. (2017). *Assessment of logistics effectiveness for expeditionary units* [MBA professional report]. Monterey, CA: Naval Postgraduate School.

Yoho, K., & Apte, U. (in press). Navy expeditionary logistics. *The Case Journal*.

Appendix: ELO/Gear Issue Process

Mobile unit submits requirement request using a DD Form 1149 (Requisition and Invoice/Shipping form). Supply Department (for PGI) or Materiel Department (for TOA) checks its WASP if the item is in stock.

(a) If the item is in stock, it is delivered to the unit.

(b) If the item is out of stock, check if it is a Navy NSN item.

(i) if Navy NSN item (30%):

Order through R-Supply
The order goes through Navy Supply System and funds are subtracted
When the item arrives, R-Supply is updated
The item is issued to the platoon and WASP is manually updated

(ii) If Non-Navy NSN (70%):

If cost > $3,000 or performance period > 90 days, send to Contracting; otherwise, open purchase:

Order through R-Supply
Funds are obligated using GCPC and paid to the vendor
When the item arrives, R-Supply is updated
The item is issued to the platoon and WASP is manually updated

Additional Papers

	A Systems Theory Based Examination of Failure in Acquisition System Reform Charles B. Keating, Old Dominion University Joseph M. Bradley, Patrona Corporation Polinpapilinho F. Katina, Old Dominion University Craig Arndt, Defense Acquisition University ***Industrial White Paper Briefing: Monterey Bay Regional Spaceport Lowest Priced Technically Acceptable COTS Launch System*** Kelly Keith Weigel, Trans Universal Energy, LLC ***In the Fullness of Time: Towards Realistic Acquisition Schedule Estimates*** Raymond Franck, Brig Gen, USAF (Ret.)—Naval Postgraduate School Gregory Hildebrandt, USAF (Ret.)—Naval Postgraduate School Charles Pickar, USA (Ret.)—Naval Postgraduate School Bernard Udis, University of Colorado at Boulder ***Searching Hidden Links: Inferring Undisclosed Subcontractors From Public Contract Records and Employment Data*** M. Eduard Tudorneau, University of Arkansas at Little Rock Keith Franklin, University of Arkansas at Little Rock Arnold Rego, University of Arkansas at Little Rock Ninging Wu, University of Arkansas at Little Rock Richard Wang, Massachusetts Institute of Technology ***A Method for Identification, Representation, and Assessment of Complex System Pathologies in Acquisition Programs*** Charles B. Keating, Old Dominion University Polinpapilinho F. Katina, Old Dominion University Keith F. Joiner, University of New South Wales Joseph M. Bradley, Patrona Corporation Raed M. Jaradat, Mississippi State University ***A Mathematical Framework to Apply Tradespace Exploration to the Design of Verification Strategies*** Alejandro Salado, Virginia Tech Farshad Farkhondehmaal, Virginia Tech ***A Review of Trusted Broker Architectures for Data Sharing*** John R. Talburt—University of Arkansas, Little Rock

Additional Papers Continued

	Towards Game Theoretic Models for Agile Acquisition Scott Rosen, The MITRE Corporation Kelly Horinek, The MITRE Corporation Alexander Odeh, The MITRE Corporation Les Servi, The MITRE Corporation Andreas Told, The MITRE Corporation ***Optimizing Contract Modifications Under One Universal Mod*** Thomas Graham, nGAP Stanley Sydor, nGAP Mason Beninger, nGAP Paige Glaze, nGAP ***Assessing Vulnerabilities in Model-Centric Acquisition Programs Using Cause-Effect Mapping*** Jack Reid, Massachusetts Institute of Technology Donna H. Rhodes, Massachusetts Institute of Technology

A Systems Theory Based Examination of Failure in Acquisition System Reform

Charles B. Keating[1]—is a Professor of Engineering Management and Systems Engineering and Director of the National Centers for System of Systems Engineering (NCSOSE) at Old Dominion University (ODU) in Norfolk, VA, USA. He received a BS in Engineering from the United States Military Academy (West Point), an MA in Management from Central Michigan University, and his PhD in Engineering Management from Old Dominion University. His current research focus is on complex system governance, system of systems engineering, and management cybernetics. [ckeating@odu.edu]

Joseph M. Bradley—received his PhD in Engineering Management and Systems Engineering at Old Dominion University (ODU) in Norfolk, VA, USA. He holds the degrees of Professional Engineer and Master of Science in Mechanical Engineering from Naval Postgraduate School and Bachelor of Engineering from The Cooper Union. He is currently Principal Engineer at Patrona Corporation, a small defense consulting company headquartered in the Washington, DC, and area president of his own small consulting firm working with clients in government and industry. Dr. Bradley's areas of research include complex system governance, systems theory, competency models, and performance measurement systems. His research has been published in the *Systems Engineering*, *Naval Engineers Journal*, and *International Journal of System of Systems Engineering*. [josephbradley@leading-change.org]

Polinpapilinho F. Katina—is a Postdoctoral Researcher for the National Centers for System of Systems Engineering (Norfolk, VA). He serves as an Adjunct Assistant Professor in the Department of Engineering Management and Systems Engineering at Old Dominion University (Norfolk, VA) and is an Adjunct Assistant Professor in the Department of Engineering and Technology at Embry-Riddle Aeronautical University—the Worldwide campus. He received his PhD in the Department of Engineering Management and Systems Engineering at Old Dominion University (Norfolk, VA). He received additional training from, among others, Politecnico di Milano (Milan, Italy). His profile includes more than 70 peer-reviewed papers in international journals, conferences, and books. Dr. Katina is a founding board member for the International Society for Systems Pathology (Claremont, CA). [pkatina@odu.edu]

Craig Arndt—received his DEng in Electrical Engineering from the University of Dayton, MA from the Navy War College, MS degrees in Human Factors Engineering and Systems Engineering from Wayne State University, and BS in Electrical Engineering from Ohio State University. Dr. Arndt is recognized as an international expert in biometric systems, human computer interface and human centered systems, image and signal processing, and artificial intelligence. He has served as a technical expert for the Army and Defense Science Boards, the National Science Foundation, the International Standards Organization, the IEEE, and other public and private technical organizations. [craig.arndt@dau.edu]

[1] Corresponding author

Abstract

The defense acquisition system has been the source of intense scrutiny and calls for reform for over four decades. This research is to examine the contributions of Systems Theory to enhance prospects related to acquisition reform. Systems Theory offers a set of principles, laws, and concepts that explain the behavior of complex systems. Although the acquisition system and constituent programs have been critiqued and examined from multiple perspectives, they have never been the subject of exploration from Systems Theory. Recent advances in Systems Theory have identified 83 different potential system pathologies that can result in degraded system performance or outright failure. System pathologies have been previously defined (Keating & Katina, 2012) as "a circumstance, condition, factor, or pattern that acts to limit system performance, or lessen system viability, such that the likelihood of a system achieving performance expectations is reduced." Following a brief introduction to Systems Theory, this paper reports on efforts to (1) briefly examine the current state of the defense acquisition system and programs, focused on successes, failures, major reform themes, and critical challenges for moving forward; (2) mapping of systems pathologies to provide a different "Systems Theory" based perspective of acquisition system reform as well as acquisition system development; and (3) suggest implications for acquisition system development based on contributions from Systems Theory. The paper concludes with future research directions for Systems Theory contributions to the acquisition field and reform efforts.

Introduction

The defense acquisition system has remained under continual scrutiny since its inception. Failures in acquisition have been as numerous as are the attempts to explain those disappointments (Berteau, Levy, Ben-Ari, & Moore, 2011; Francis, 2008, 2009; Rascona, Barkakati, & Solis, 2008). Unfortunately, problems in acquisition continue to exist, and arguably are increasing in frequency and severity. Arguably, the defense acquisition system falls short on the traditional essential attributes that are used to delineate a "system." These attributes, following decades of systems literature (Kramer & de Smit, 1977; Beer, 1978; Sykttner, 1996; Clemson, 1984; von Bertalanffy, 1968) include minimal characteristics of *boundary* (specifying what is included and excluded from the system), *environment* (all that exist external to the system boundary), *input* (matter, energy, resources, information crossing the boundary), *transformation* (processing of inputs to produce something of value), *outputs* (products of value consumed external to the system), and *feedback* (support for regulatory adjustment to make corrections necessary to maintain stability). The Defense Acquisition Management System has been referred to by the DoD 5000 as both a "framework" and an "event-based process." Processes and events, while they can be aspects of a system, fall short in the most fundamental characteristics for classification as a system.

Our point is not to criticize defense acquisition, or to challenge different formulations of defense acquisition as a "system." However, simply calling something a "system" does not make it a system, except in the very loose interpretations of the term. In fact, the mischaracterization may preclude discoveries and insights that might accrue from the more formal appreciation of, and accountability for, making attributions. Instead, our objective is to suggest that a more rigorous formulation and classification as a "system" may yield new insights into familiar unresolved acquisition system reform issues.

In previous work (Keating et al., 2017) related to acquisition system difficulties, there have been several "systemic" inconsistencies identified, coupled with the suggestion that a reformulation of issues from a stronger Systems Theory base might deepen understanding.

Among these "systemic formulations" were included (1) *Sprawling Complexity*—exponentially increasing complexity that exceeds the present capacity to sufficiently absorb to limit negative impacts, (2) *Process and Event Centric Orientation*—emphasis on the critical processes and milestones as the central focus for execution and development, (3) *Complication as a Response Strategy*—increasing the regulation and proliferation of controls to address increasing system complexity, (4) *Output versus Outcome Emphasis*—focus on the output based cost, schedule, and technical performance aspects of systems as primary versus the outcome based problem/need fulfillment aspects of systems, and (5) *Achievement of Control Through Excess Regulation*—emphasizing control of complexity by additional regulation by ad hoc and fragmented additions versus purposeful "systemic" design for control. While Systems Theory is not being offered as a panacea to this situation, nevertheless it does offer an alternative viewpoint from which the dialog might be shifted.

Systems Theory is a somewhat polarizing term, without a substantially agreed-upon definition. In fact, following the work of Adams et al. (2014), Table 1 depicts several definitions for Systems Theory.

Table 1. Definitions for Systems Theory

Definition	Source
The formal correspondence of general principles, irrespective of the kinds of relations or forces between the components, lead to the conception of a "General Systems Theory" as a new scientific doctrine, concerned with the principles which apply to systems in general.	Von Bertalanffy (1950)
General systems theory is the skeleton of science in the sense that it aims to provide a framework or structure of systems on which to hang flesh and blood of particular disciplines and particular subject matters in an orderly and coherent corpus of knowledge.	Boulding (1956)
A new way of looking at the world in which individual phenomena are viewed as interrelated rather than isolated, and complexity has become a subject of interest.	Klir (1972)
General Systems Theory and the Systems Approach grapple with the issue of "simplicity" and "complexity" by which the relationships among systems and subsystems are decided. The problems of "optimization" and "suboptimization" are central to explaining the fruitless efforts of systems designers who reach for the "summum bonum" while settling for a "second best."	van Gigch (1974)
Systems theory is a unified group of specific propositions which are brought together to aid in understanding systems, thereby invoking improved explanatory power and interpretation. It is precisely this group of propositions that enables thinking and action with respect to systems.	Adams et al. (2014); Whitney et al. (2015)

Interestingly, the extension of Systems Theory into the domain of system acquisition is noticeably absent in the literature. While not totally unexpected, it is somewhat surprising that a field so heavily steeped in the acquisition of complex systems has somehow not routinely incorporated the most fundamental aspects of Systems Theory into the field. Given the scarcity of Systems Theory in acquisition, we suggest that there are three major contributions that Systems Theory can make related to a new and novel perspective of acquisition reform (Figure 1). The incorporation of Systems Theory might provide new insights and contributions into the past "failures," "present" challenges, and "future" trajectory for acquisition reform.

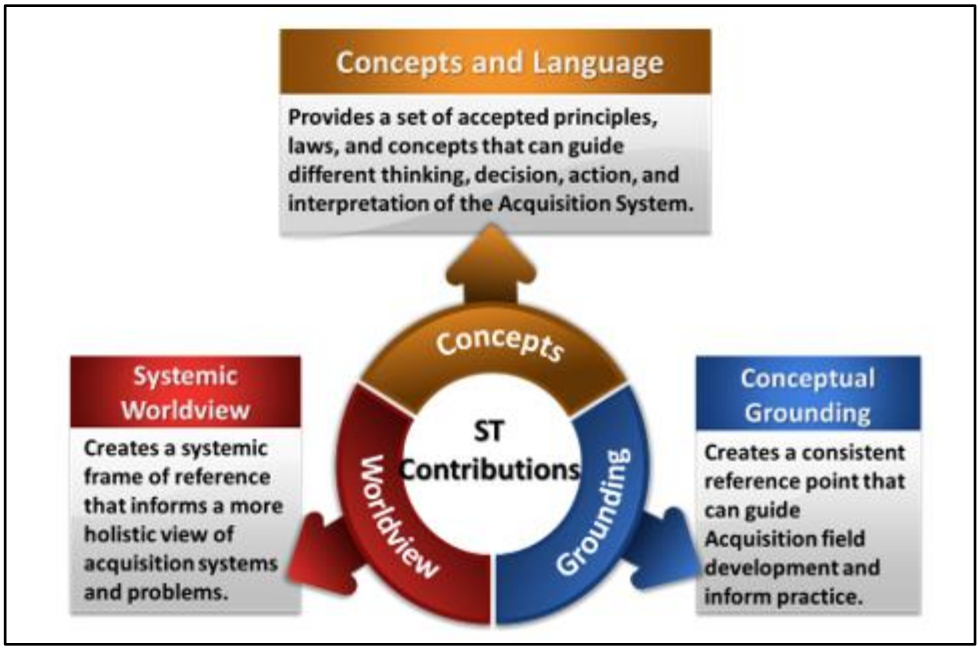


Figure 1. Contributions of Systems Theory for Acquisition

1. ***Concepts and language*** of Systems Theory can provide a basis to drive different thinking, decision, action, and interpretation related to understanding and explaining acquisition system difficulties. We think through language, and if we are to engage in a different orientation for potential breakthrough in acquisition reform, Systems Theory provides a conceptual foundation that has been largely absent from acquisition system development.
2. ***Conceptual grounding*** provided by Systems Theory offers acquisition system development and practice a theoretical grounding that appears to be absent in the field. The strong theoretical basis of Systems Theory can offer a rigorous theoretical grounding for the acquisition field and provide the basis for a stable and sustainable foundation. This can provide the acquisition field with a consistent reference point against which system development and reform can be anchored.
3. ***Systemic worldview*** provided by Systems Theory is consistent with the complex domain facing the acquisition system and practitioners. The systemic worldview is consistent with the complexity, ambiguity, contextually bound, and holistic nature of acquisition. This worldview can support thinking, decision, action, and interpretation that may provide potential new and novel insights to "move the equation" for acquisition reform.

This paper is focused on providing an alternative paradigm, Systems Theory, for viewing failure in acquisition system reform. This does not diminish the work, efforts, or results achieved by the individuals and entities engaged in trying to improve the acquisition system. On the contrary, our intention is to invite a dialog to further explore and understand the contributions that an alternative paradigm (Systems Theory) might provide to move the acquisition reform dialog in new and fruitful directions. To achieve our purpose, the remainder of the paper is organized around four primary objectives. First, we provide an overview of the state of acquisition reform, focused on highlighting several failure modes that delineate the system. In the section following, we elaborate a Systems Theory perspective through the introduction of pathologies (aberrations from healthy system functioning) as violations of underlying system propositions (concepts, laws, principles). After that, in the section titled Systems Theory Implications for Acquisition System Reform, we suggest implications of pathologies in relationship to acquisition system reform. The final section concludes the paper with implications for further research and development of Systems Theory as an alternative and insightful paradigm to better understand, and potentially shift the trajectory, related to failures in defense acquisition system reform.

State of Defense Acquisition System Reform

The state of the acquisition system is generally not considered to be strong. However, like many other topics in government, that assessment is not an entirely fair or straightforward answer to an extremely complicated question. The importance of the acquisition of weapon systems and other materials and supplies to equip the nation's armed forces cannot be overstated. If the Armed forces did not have the tools they need to fight, their existence would be threatened and so too would the existence of the nation itself. However, this acquisition is a function of government and not a function of industry, and therefore subject to the rules and regulations governing government. In his last report on the performance of the acquisition system, Under Secretary Kendall (2016) suggested that there were only four major steps to insuring success in acquisition. "(1) set reasonable requirements, (2) put professionals in charge, (3) give them the resources that they need, and (4) provide strong incentives for success." Kendall does, however, go on to say that the current system is much more complicated than this and, in many cases, does not allow basic good management to be the only factor in the development of systems.

These following sections focus on review of the current state of the acquisition system and provide some of the history of how the system was developed and the contracts that make the system operate in the ways that it does.

The Acquisition System

The acquisition system, as defined by the federal government, consists of many different parts and is not exactly the same in all parts of the government. The most complex version of the federal acquisition system is the Department of Defense (DoD) acquisition system. Since complexity is critical to issues with the acquisition system, we will concentrate on the defense acquisition system.

The defense acquisition system is actually three different systems that are linked together. These systems are (1) the acquisition system, which creates the systems and delivers them to the warfighter, (2) the requirements system (JSIDS), which generates the requirements from which the acquisition systems develops products, and (3) the Planning, Programing, Budgeting, and Execution (PPBE) process, which is the way the Department of Defense (DoD) asks for and gets the money it needs from Congress.

So, in reality there are three very complex processes that make up the DoD acquisition process. One of the most well-known attempts to document of complexity of the acquisition system is captured in the wall chart illustration of the DoD acquisition system (Figure 2), which is basically a flow chart of all three systems put in a single place. It does look complicated because the process is complicated. The processes level of complexity comes from several major sources. The first driver of complexity of the systems is the complexity of the programs (e.g., a Navy aircraft carrier is considered the most complex system ever designed). The second driver of complexity is the need to integrate defense systems into a very complex existing system with many interphases and relationships already in place. The next driver of complexity is the very harsh environment that defense systems must operate in; this drives complex and lengthy testing protocols. Also, complexity is driven do to all of the government rules and regulations that must be followed by the participants in the acquisition processes. Additionally, complexity is driven from the fact that many different stockholders, including many for profit companies, are trying to influence the processes in their favor. Finally, complexity is driven from a last major factor adding complexity to the system—the buildup of rules, processes, and reviews built into the system from each new leader and from generation after generation of Congress.

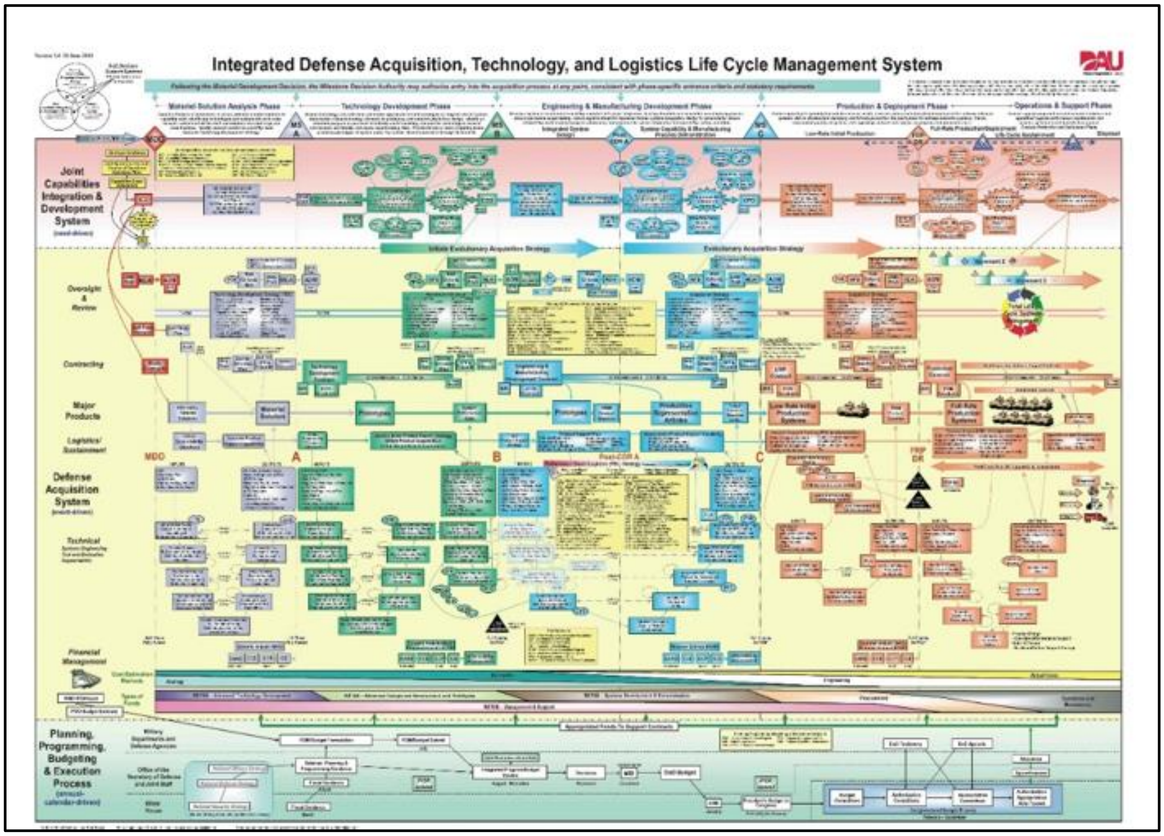


Figure 2. The Department of Defense Acquisition Chart (Wall Chart)

Success and Failure in the Acquisition System and Programs

Overall the defense acquisition system and the programs and products (plans, ships, weapons, etc.) that it produces can be evaluated as both a great success (delivering a wide range of very capable, very lethal systems) and also great failure (many programs overrunning cost and schedule, some to the point of cancellation). There have been many studies into the failures of the acquisition system. Some of the best documented failures include the following:

1. ***The Future Combat System (FCS)***—First introduced in 1999 by Army Chief of Staff Eric Shinseki, FCS was supposed to be a family of networked, manned, and unmanned vehicles and aircraft for the 21st century battlefield. With the Warfighter Information Network–Tactical (WIN-T) intended to support the FCS, it was supposed to be a wholesale re-envisioning of the ground force. However, the terror attacks of Sept. 11, 2001, short-circuited a 15-year operational pause that the military was hoping for to implement the

program. Spiral development and shifting requirements by the Army also resulted in costs ballooning by 25%. Finally, after $19,000,000,000 already spent and the program in the System Design and Demonstration phase, Secretary Gates cancelled the program in 2009.

2. ***The RAH66 Comanche***—22 years, $6,900,000,000 spent and zero helicopters. Originally conceived at the height of the Cold War, it was supposed to become the next generation of armed reconnaissance air support for the Army, replacing the Huey, Cobra, and Kiowa helicopters in the process. A textbook case in technology being superseded by current events, the Comanche also faced serious concerns over its ability to simply get off the ground when fully loaded. The program was cancelled in 2004 with two prototypes now on display.

Figure 3. RAH66 Comanche Helicopter

3. ***The XM2001 Crusader***—Intended to be the Army's next-generation mobile gun system, the Crusader (see Figure 4) was conceived in the early 1990s as a powerful new self-propelled howitzer (SPH). While it was designed to be lighter and faster than the existing M109A6 Paladin SPH, it was too similar to the existing, upgraded inventory. A system designed for a Cold War army, it was not widely supported by the Army Staff as it no longer aligned with the new operational concept. Ironically, many of the Crusader technologies were incorporated in the FCS family of XM1203 Non-Line of Sight (NLOS) cannons, which were subsequently cancelled as well.

Figure 4. XM2001 Crusader Mobile Gun System

4. ***CG(X)***—Known as the Next-Generation Cruiser in the early 1990s, it was part of the Navy's Surface Combatant for the 21st century program (see Figure 5). However, budget cuts resulted in the program being split up in 2001 with the destroyer variant being renamed the DD(X) and then the Zumwalt-class of destroyers. While the DD(X) is a case study in and of itself, the CG(X) actually looked like it might increase its allocation of ships before being abruptly cancelled in 2010. Deemed too similar in capability to the existing, upgraded Arleigh Burke-class of destroyers, the ship was never built, but not before spending more than $200,000,000 in development costs.

Figure 5. CG(X) Next Generation Cruiser

Equally, there are many success stories to be told in defense acquisition, including the following:

1. ***MC-12W Aircraft***—The U.S. Air Force needed more Intelligence, Surveillance, and Reconnaissance capability and so launched the Project Liberty program. The result was the low-cost MC-12W aircraft, which flew its first combat mission in June 2009, just eight months after receiving funds. It has since flown thousands of successful missions in Afghanistan and Iraq.
2. ***Harvest Hawk***—The U.S. Marine Corps needed a boost in close air support capabilities. In October 2010, just 18 months after announcing the program, the Harvest Hawk was in the fight. This inexpensive, reversible mod to a KC-130 not only puts steel on difficult targets, but also gets eyes on previously unseen locales.
3. ***Virginia-Class Submarine***—The U.S. Navy began the Virginia-class submarine program after terminating the unaffordable Seawolf program (see Figure 6). The USS *New Hampshire*, first of the Block II Virginias, came in eight months early and $54 million under budget, and that's on top of the $300 million cost savings which were already achieved on the Block II design.

Figure 6. Virginia-Class Submarine

Themes of Acquisition System Reform

oday's defense acquisition system is a product of decades of reform initiatives, legislation, reports, and government commissions. Major reform efforts began in earnest in the 1960s with Secretary of Defense Robert McNamara. His main reform efforts centralized control within the Office of the Secretary of Defense (OSD) and created the Planning, Programming, and Budgeting System for resource allocation. Throughout the latter half of the 20th century, each administration left its own mark on defense acquisition, focusing primarily on the acquisition process itself, as well as DoD management. However, many of the reforms recycled various schemes to shift decision-making authority from the services to the OSD, realign oversight and accountability responsibilities, and alter the process (adding and removing milestones, phases, and so forth). Despite these initiatives, cost and schedule growth continue.

The Pentagon has wrestled with reforming defense acquisition procedures for over 40 years and, during that period, over 120 defense acquisition reform actions and policies have been implemented. Of these, the 1986 Congress of the Goldwater-Nichols Defense Reorganization Bill has been one of the most wide-ranging and had the largest impact. This landmark legislation was intended to add both significant discipline and accountability to the defense acquisition process and focus the management and oversight of defense research development test and evaluation, which now consumes over $600 billion annually and continues to grow.

One of the major actions in the legislation was the establishment of the position of under secretary of defense for acquisition, or USD(A), to vest in one person the overall oversight responsibility of the defense research development test and evaluation process of the numerous systems in various stages of development and fielding. The hope was that this position would enable the defense secretary to have a single line of command, one office responsible for overseeing and streamlining the activities of the hundreds of large Acquisition Category I defense programs, such as ships and missile defense, all the way down to the smaller Acquisition Category IV programs, like small arms and body armor. But development times and costs went up, not down. A few years later, recognizing that injecting research and development breakthroughs was vital to retaining weapon superiority, the job was expanded to under secretary of defense for acquisition and technology, or USD(A&T), to assure that this cross-pollination was taking place efficiently. Again, development timelines and costs both increased The latest reorganization effort (splitting the function of the under secretary for acquisition, technology, and logistics into two), mandated by the FY17 NDAA (National Defense Authorization Act) will undo one of the major changes made by the Goldwater-Nichols Act of 1986, which codified the acquisition chain of command and was based largely on the recommendations of the 1985 Packard Commission. On the positive side, the creation of the new under secretary for research and engineering comes at a time when the DoD is working to regain its technological advantage, notably through its pursuit of a third offset strategy to renew and perhaps advance the competitive advantage of the United States and its military allies. This change should increase the emphasis on these efforts. On the negative side, this new organization may introduce an element of confusion and competition into the decision-making process by not having a single end-to-end process owner. Although acquisition reform continues, no reform processes or approaches have yet to overcome the challenges of complexities of the defense acquisition system.

Critical Issues and Challenges for Acquisition System Reform

Many issues and challenges remain for the acquisition system and for acquisition reform efforts. The next generation of acquisition reform will need to address many of the same issues as earlier reform efforts. However, given the many failures in parts acquisition reform, the answers and approaches need to be very different to meet the needs of an ever-more-quickly-changing world. The first priority of the next generation of acquisition reform will continue to be the need for speed. Whether the acquisition is a software development that takes two years and needs to take three months, or a new airplane that takes 15 years and needs to take four years to develop, we must find ways to make the acquisition system faster. The next major area that cannot be avoided is the need to somehow address the requirements process, to make it faster, more flexible, and more responsive to changes in technology and the future needs of the warfighters.

In order to make a faster and more flexible acquisition system work, we also need to find ways of delegating authority and accountability that empower key individuals to do their best work and deliver products in ways that will continue to be significant successes. Lastly, reform efforts will also need to address major congressional actions required to change many of the legal roadblocks that create problems in the defense acquisition system, including but not limited to, funding management and added levels of review and oversight.

In the following section, a path forward is forged through articulation of deep systemic issues (pathologies) affecting acquisition systems. While this is not posed as "the answer" to acquisition system woes, Systems Theory does offer a substantial departure from other attempts at modification of an unwieldy system.

Systems Theory Pathologies Perspective for Acquisition System Reform

For our present purposes, the nature of system pathologies in complex systems can be captured in the following critical points and their suggested relevance to acquisition practitioners and system development:

1. ***All systems are subject to the laws of systems***—Just as there are laws governing the nature of matter and energy (e.g., physics law of gravity), so too are our systems subject to laws (principles, laws, concepts defining the behavior, and performance of complex systems). These system laws are always there, always on, non-negotiable, non-biased, and explain system performance.
2. ***Violations of systems laws carry consequences***—Irrespective of noble intentions, ignorance, or willful disregard, violation of system laws carries real consequences for system performance. In the best case, violations degrade performance. In the worst case, violations can escalate to cause catastrophic consequences or even eventual system collapse.
3. ***Violations of systems laws generate associated pathologies***—Pathologies are circumstances, conditions, factors, or patterns that act to limit system performance, or lessen system viability, such that the likelihood of a system achieving performance expectations is reduced. When system performance fails to meet expectations, violations of systems laws are always in question.

In the examination of failures in the acquisition system, programs, and projects, violations of Systems Theory (manifest as system pathologies) should be considered as potential sources contributing to failures and dysfunctions. Following the systems pathology research (Katina, 2015a, 2015b, 2016a, 2016b, 2017; Katina & Keating, 2014, 2016) based in violations of systems propositions, the following summary table (Table 2) is provided. Three notes are necessary to guide interpretation of the table. First, we have referred to the principles, laws, and concepts simply under the banner of "propositions," following the nomenclature of Adams et al. (2014) so as not to overburden the presentation with the finer distinction between principle, law, and concept. In the end, they all inform our understanding and explanation of systems behavior/performance and their violation jeopardizes system performance. Second, we have presented the set of pathologies from a pragmatic perspective, attempting to remain free from a barrage of scholastic verbiage. While some depth will naturally be sacrificed in this delivery, our intent is to make the principles more approachable to meet our present objectives. A more thorough and "scholarly" deep accounting of the principles can be found in other composite works (Adams et al., 2014; Clemson, 1984; Hammond, 2002; Katina, 2015b, 2016b, 2017; Katina & Keating, 2016; Lespier et al., 2015; McDermott & Alejandro, 2017; Skyttner, 2005; Troncale, 1977;

Warfield, 1999; Whitney et al., 2015). Third, the principles are intended to provoke consideration related to the design, execution, and development of systems (e.g., acquisition). The role of the propositions is analogous to the use of "illities" (e.g., reliability, usability, affordability) in their system design role of informing design considerations, performance tradeoffs, and guiding development. Similarly, the systems propositions serve to inform complex system design, explain sources of performance variation, and support more enlightened inquiry to potentially drive system development from a different perspective.

Table 2 presents a set of Systems Theory derived propositions, a concise statement of their violation producing system errors, and speculation of applicability for Acquisition System Reform (ASR) and the Acquisition System (Acq Sys). The ASR and Acq Sys implications are speculative and anecdotal at best. However, lacking more rigorous explication, they broadly suggest that inclusion of Systems Theory (propositions) in Acquisition System Reform and Acquisition System development might better inform future acquisition system design, execution, development, and reform.

Table 2. Summary of Systems Theory Propositions (Principles, Laws, Concepts)

System Proposition	Concise Statement of Proposition Producing Errors (Pathologies)	Acquisition System (Acq Sys) Reform (ASR) Implication
Complementarity	A situation in which an entity ignores other perspectives/models that are not entirely compatible with the established-predominate perspectives of elements such as missions, goals and objectives. An entity in this case mistakenly assumes that there is only one "right" perspective.	ASR can benefit by inclusion of multiple "new" perspectives provided by Systems Theory.
Diminishing returns	Mistakenly assuming that continually increasing resources (e.g., number of staff) will have a corresponding increase in the productivity or performance of the system as a whole	Expecting more of the same approaches to ASR to be fruitful can be shortsighted.
Requisite hierarchy	There is insufficient regulatory capacity (levels of organization) to provide sufficient control of a system necessary to match that required by the environment.	Fragmented ASR system structure impacts efficient regulatory capacity.
Requisite knowledge	Sufficient knowledge is either not available, accessible, or actionable to provide sufficient regulatory capacity necessary to sustain consistency in system thinking, decision, action, and interpretation in response to environmental turbulence and internal system flux.	ASR is hindered by the knowledge system that appears somewhat incongruent to needs.
Requisite parsimony	System failure due to exceeding human capacity to simultaneously focus on multiple complex tasks. This number is limited to seven plus or minus two.	Acquisition System and workforce are stretched beyond capacity to respond.
Requisite saliency	System productivity is reduced due to having undifferentiated importance of system priorities—resulting in inconsistencies in priorities, decisions, actions, and interpretations.	Criticality in priorities for ASR do not appear to be congruent across entities.
Requisite variety	Regulatory capacity of the system fails to match that required to provide stability and sustain consistent performance in the midst of environmental turbulence and internal flux.	ASR environment complexity far exceeds regulatory response capacity.
Adaptation	Inability of internal structures of a system to change at a pace necessary to match that required in response to external disturbances to preserve system performance	ASR must address a system outpaced by the rate of external change.
Autonomy	Excessive limitations or lack of balance concerning the degree of freedom and independence of decision, action, and interpretation for constituents in a system	Increasing centralization of Acq Sys control/regulation diminishing local autonomy.
Balancing system	Inappropriate system balance in *Design* (ranging from self-	Acq Sys appears to be

System Proposition	Concise Statement of Proposition Producing Errors (Pathologies)	Acquisition System (Acq Sys) Reform (ASR) Implication
tensions	organizing to purposeful), *Change* (ranging from stable to unstable) and *Control* (ranging from autonomy to integration)	fragmented, unstable, and overly complex (regulated).
Basis of stability	Failure to provide sufficient resources or energy to move a system past a threshold to a new stable state, resulting in an inevitable return to the former stable state (maintain status quo)	ASR has not generated sufficient movement to significantly shift status quo.
Buffering	Lack of sufficient surplus resources, to provide for system stability beyond immediate needs, when confronted with unexpected increases in demand (threatening continued stability)	Exceeding cost, schedule, and performance targets questions sufficient buffer.
Circular causality	System failures due to nonlinearities that cannot be reduced to simplistic cause effect relationships, requiring consideration of multiple, and perhaps ill understood, causal relationships	ASR response appears to be piecemeal fixes largely self-organized and ill understood.
Consequent production	A system is only capable of producing what it produces, nothing more and nothing less—this does not necessarily match what was designed, intended, or desired.	Limited ASR focus on underlying system producing undesirable behavior.
Cybernetic stability	A system has an insufficient number of external connections necessary to provide stability and ease of adaptation to changing circumstances.	Ability of Acq Sys to adapt to volatile change appears inadequate.
Darkness in a situation	Knowledge/understanding of a system is always incomplete, fallible, and emergent over time with increasing experience gained through operation of a system.	ASR has been attempted with apparent limited and fallible system knowledge.
Dialectism	Inappropriate inquiry balance for detection and correction of error in a system between first order (staying within design) or second order (adjusting system design) learning emphasis	ASR modifications seem to skew to the first order versus second order inquiry and learning.
Emergence	Failure to compensate in system design or execution for occurrence of behaviors or performance in a system that could not be predicted in terms of timing, nature, or impact	Emergence in acquisition does not appear compatible with Acq Sys design.
Environmental-modification	Limitation in integrated design, strategies, and actions to deliberately and proactively attempt to influence the environment	Acq System appears to be reactive to env turbulence.
Equifinality	Failure to recognize that from different initial starting points, the same end state can be attained through different pathways and means—not just a singular path/design to achieve desirable states	Detail Acq event-process mapping appears to be rigid in the pathway to completion.
Equivocation	Inefficient communication channels not providing the intended signal (information/message) from one point (entity) to the next, resulting in lack of clarity, excessive noise, or misinterpretation	Anecdotal observation suggests that ASR lacks clarity of communication.
Eudemony	Overemphasis on a preferred set of affairs and motives (e.g., financial profitability) of a system above all other measures, sacrificing balance with other potentially meaningful measures	Acq Sys near exclusive focus on cost, schedule, and performance are limiting.
Events of low probability	Focus on events of a system without distinction as to their probability of occurrence, attempting to control for all scenarios and thus potentially jeopardizing fundamental system objectives	There is not sufficient knowledge to speculate on this proposition.
Feedback	Inadequacies in system scanning to identify fluctuations requiring adjustments to maintain system stability	Not apparent that ASR has been formulated as a system.
Flatness	Reduction of system stability by an inappropriate balance in the distribution of system control—generating an imbalance between administrative and productive functions	ASR does not appear to be under central control or development oversight.

System Proposition	Concise Statement of Proposition Producing Errors (Pathologies)	Acquisition System (Acq Sys) Reform (ASR) Implication
Frame of reference	The lack of consistent standards by which a system can be judged or existence of a common vantage point from which a system can be viewed	Acq Sys regulatory standards do not appear to present a common frame of reference.
Hierarchy	Lack of sufficient structure of a system (levels of organization) to provide sufficient regulatory capacity necessary to control a system to maintain stability	Acq Sys does not appear to have regulatory capacity sufficient for stability.
High-flux	The rate of arrival of correct resources in response to system failure is insufficient to provide continuing stability in response to a correctable perturbation.	There is not sufficient knowledge to speculate on this proposition.
Holism	Focus on individual system entities as the source of system performance, as opposed to performance stemming from interaction of those entities to produce what individually they cannot	The Acq Sys exists as a fragmented aggregate set of entities and standards.
Homeorhesis	System lacking mechanisms that provide ability to return it to a pre-set path or trajectory following an environmental disturbance	Other than generalities, ASR trajectory is not clear.
Homeostasis	System lacking ability to maintain essential variables, within limits necessary to maintain stability, in response to external disturbances	Arguably, the Acq Sys has not been in a stable state.
Internal elaboration	Excessive tendency of a system to increase interconnections, constraints, and controls (regulations) over time in ways that make them increasingly complicated and complex	ASR must deal with the sprawling complexity and complication of the system.
Iteration	Failure to move through repetition cycles in system development allowing quick error identification and increasingly deep understanding	ASR has not be explicitly developed or performed in an iterative fashion.
Least effort	Selection of high resistance (resources, constraints) paths to maintain system performance where less resistance paths could provide the same results with less expenditure of energy	ASR is engaging with high resistance for maintenance of the status quo as preferable.
Maximum power	Limitations in ability to increase intake capacity and transformation rate necessary to realize system productivity potential. Failure to keep up with demand.	There is not sufficient knowledge to speculate on this proposition.
Minimal critical specification	Introducing system constraints beyond those minimally necessary to maintain system performance—overconstraining system entities, wasting resources, and not improving performance	The Acq Sys appears to be overregulated to the detriment of performance.
Multifinality	Failure in realizing that from the same initial starting point radically different end states are possible—assuming approaches based on prior experiences will yield similar results is flawed.	There is not sufficient knowledge to speculate on this proposition.
Omnivory	Inability of a system's internal structure to be modified to accommodate a more diverse set of input resources to increase stability	There is not sufficient knowledge to speculate on this proposition.
Organizational closure	Incongruence in the essence of a system that provides coherence in system identity—providing consistency and unity in thinking, decision, actions, and interpretations for system related matters	The identity of the Acq Sys and ASR appear to lack clarity.
Over-specialization	An excessive degree of specialization such that a system lacks the ability to change and adapt to shifting circumstances and conditions	The Acq Sys appears slow to adapt to increasing change and rates of change.
Pareto	Expenditures of system resources to enhance productivity are not directed proportionally to those offering the greatest contribution for improvement.	There is not sufficient knowledge to speculate on this proposition.
Patchiness	Limited system design capacity to accommodate a diversity of resources from the environment, without needing to be	There is not sufficient knowledge to speculate on

System Proposition	Concise Statement of Proposition Producing Errors (Pathologies)	Acquisition System (Acq Sys) Reform (ASR) Implication
	structurally modified to accept different types of resources	this proposition.
Polystability	Failure to appreciate that stability of system entities does not imply stability can be directly translated to stability of the larger system	ASR should consider the constituent Acq systems.
Redundancy of potential command	Limitations in subsystem authority and independence to make decisions and take action on behalf of the system, limiting speed of response to identified opportunities, novelties, trends, and treats from the environment	Consolidation of decision authority limits autonomy and decision efficiency in the Acq Sys.
Redundancy of resources	Failure to provide redundant critical resources beyond those identified as necessary under ideal conditions—this optimal efficiency perspective assumes unforeseen circumstances will not occur.	There is not sufficient knowledge to speculate on this proposition.
Relaxation time	Introduction of too many simultaneous changes rendering a system incapable of processing or assimilating the changes and resulting in continual instability	ASR is being undertaken with multiple, and not necessarily integrated, efforts.
Resilience	Following a disturbance, lack of capability of a system to withstand the disturbance, either operating at a degraded level or outright failing to return to operation	The ability of the Acq Sys to withstand disturbances and function is speculative.
Robustness	Inability of a system to withstand a wide range of environmental disturbances without the necessity for system modifications	Constant flux in ASR suggests lack of robustness.
Safe environment	A system not acting to create a level of stability in the environment to reduce disturbances that might have a detrimental impact on system performance	The Acq Sys environment is not explicitly mapped, modeled, or understood.
Satisficing	Attempting to resolve issues by seeking the optimal (best) solution as opposed to a less resource-intensive solution that will work	There is not sufficient knowledge to speculate on this proposition.
Self-system	Failing to gain efficiencies by increasing autonomy of system entities to make decisions and initiate actions more their local level and require less energy (resources) to maintain	The Acq Sys does not appear to provide high levels of autonomy for decisions.
Separability	Failure to account for designs that permit such tight coupling of subsystems that small variations can spiral out of control to cause major negative consequences	The Acq Sys is tightly coupled making escalation of failures possible.
Steady state	Failure to account for overall system steady state being dependent on the continuing steady state of constituent subsystems—if a subsystem moves out of steady state, so too does the overall system.	There is not sufficient knowledge to speculate on this proposition.
Suboptimization	A focus on optimization of subsystems results in sacrifice of (optimal) performance of the larger system—all subsystems and the overall system cannot be simultaneously optimized.	ASR would benefit by representation as a system with defined subsystems.
Subsidiarity	Elevation of a local system issues/conflicts for resolution by a higher level (authority) system, when the resolution could be accomplished locally in harmony with higher level system objectives	ASR would benefit from close examination of decision authority level.
System context	Addressing a system independent of the context (unique circumstances, factors, trends, patterns) within which the system exists	The Acq Sys appears to be designed and regulated as a context free system.
First cybernetic control	System lacking ability to compare behavior/performance, or to make corresponding adjustments, based on continuous monitoring against a set standard	The narrow focus on cost, schedule, and performance is limiting in design for control.
Red Queen	System failure due to the inability to compete with other systems in the same environment—continually falling behind other systems by failing to make minimal improves to "just keep up"	There is not sufficient knowledge to speculate on this proposition.
Second cybernetic	Communications fail to provide regulatory capacity	ASR should consider

System Proposition	Concise Statement of Proposition Producing Errors (Pathologies)	Acquisition System (Acq Sys) Reform (ASR) Implication
control	necessary to address disturbances that impede system performance.	communications capacity.
Third cybernetic control	Attempting to make modifications (tinkering) to a system that is in control—a system cannot be brought into control if it has not first gone out of control.	ASR is attacked piecemeal with limited modifications for a problem system.
Transcendence	Failure to recognize that understanding might lie beyond rational, scientific, or determinate explanation—sometimes requiring explanation be taken on "faith" as belief without question	ASR lack of consideration of limitations in holistic knowledge is problematic.
Ultra-stability	Design sufficiency to fend off anticipated disturbances, but lacking the ability to fend off unknown disturbances without changing internal structures	The Acq Sys design ability to deal with emergent disturbances is questionable.
Undifferentiated coding	Failing to value knowledge or understanding that which cannot be attributed to direct observation of results and objective human sensing	ASR focus on "intangible" indicators would provide more holistic perspective.
Unity	Lack of an integrated system purpose or having an identity that establishes system uniqueness and serves to easily distinguish the system from other systems	Identity for ASR or Acq Sys could be more explicit, clear, and subject to development.
Viability	Failure to keep key system parameters in control and maintained within their set limits—questionable balance between autonomy and integration and between stability and adaptation	ASR might consider examination parameters and limits for Acq Sys viability.
Gödel's incompleteness	Operating on a system as though the frame of reference is consistent and complete—when in actuality it is not free from assumptions, infallible, or necessarily complete	ASR frame of reference does not appear as explicit.
Information redundancy	Insufficient reduction of probability for communication errors in a system due to a lack of "redundant" means used to transmit the communication	There is not sufficient knowledge to speculate on this proposition.
Morphogenesis	System failure to maintain stability following creation of a new and radically different structure	There is not sufficient knowledge to speculate on this proposition.
Morphostasis	Reduction of stability of a system by resisting change in favor of a preference for maintaining the existing status quo	The Acq Sys appears to have an emphasis on maintaining the status quo unthreatened.
Pareto optimality	Undertaking an activity to improve one aspect of a system with the mistaken belief that there will be no adverse effects on other aspects of the system	ASR appears as a well-intentioned set of disjointed activities tangentially related.
Purposive behaviorism	The purpose of the system is unguided and primarily based on intended, desired, or designed results as opposed to what the system produces.	Acq Sys purpose is not examined beyond stated intentions.
Recursiveness	Incorrectly assuming that a system exist independent and mutually exclusive of all other systems—in reality a system exists within a larger system and is comprised of (lower level) systems	There is not sufficient knowledge to speculate on this proposition.
Reification	Failure due to treatment of an abstract system (e.g., representation) as though it exists as a concrete reality	Much of the Acq Sys exists as incomplete representations.
Channel capacity	Inability of a communication channel to transmit different messages without being modified—design insufficiency to account for noise such that a message is not understood as intended	There is not sufficient knowledge to speculate on this proposition.
Genesis of structure	Failure to initiate and maintain forming structure through communications (flow of information among elements) necessary for continued system viability (existence)	The Acq Sys appears to be largely self-organizing directed by higher

System Proposition	Concise Statement of Proposition Producing Errors (Pathologies)	Acquisition System (Acq Sys) Reform (ASR) Implication
		authorities.
Synchronicity	Phenomena about a system appears to be meaningfully related but is ignored since its explanation is impossible in terms of cause-effect relationships and therefore not deemed meaningful.	ASR might focus on examination beyond simple cause-effect relationships.
Communication	Failure due to receiver(s) of information unable to receive information as intended by the sender—where the receiver does not understand the meaning and is not influenced as intended by the sender	There is not sufficient knowledge to speculate on this proposition.
Control	Inadequacy in the means necessary to provide regulatory capacity required to preserve identity of a system, permitting adaptation a maintenance of viability (continued existence)	Acq Sys regulatory capacity is in question as is clarity of system identity.
Dynamic equilibrium	Failure to maintain stability stemming from insufficient adjustment based on environmental shifts requiring system adjustments in response to maintain an equilibrium state	Acq Sys equilibrium appears questionable in response to environmental shifts.
Punctuated equilibrium	Failure to take into account that a system may experience long periods of stasis (relative calmness) that are interrupted by sudden bursts of change that were not expected and possibly catastrophic	ASR has gone through periods of fundamental change and periods of stasis.
Sociotechnicality	Failure due to misplacing preference, favoring either “technical” or “social” aspects of a system—when in actuality every complex system has both aspects and may shift their importance over time	ASR should focus on the social as well as technical aspects of the Acq Sys.
System boundary	Improperly establishing the demarcation between a system and its environment—without clear delineation of separation causing confusion as to what is to be included/excluded from the system	ASR should delineate the multiple system boundaries that denote the Acq Sys.
System environment	Lack of clarity for what lies outside the system and potential treatment of things outside of control/influence of the system as though they are within control boundaries of the system	The nature and articulation of the Acq Sys environment appear underdeveloped.

The presentation of this set of Systems Theory principles must be acknowledged for several considerations. First, this listing, although it is born from a wide breadth of existing literature of systems, cannot make the claim of being either absolute or complete. Second, the principles create the impetus for a different level of thinking, decision, action, and interpretation in creating conditions for improvement in fields (e.g., acquisition) struggling with increasingly complex systems and their problems. Third, the principles only create the conditions for different understanding of the acquisition system, and as such offer explanatory power as well as predictive power as to the future prospects for acquisition reform. We now shift our discussion to deeper examination of the implications of Systems Theory for acquisition system reform.

Systems Theory Implications for Acquisition System Reform

As we have seen from our present development, the acquisition literature is replete with calls for reform, improvement, and modification. However, a closer look at the acquisition literature suggests that the emphasis has been focused on the more tangible level of processes, tools, methods, and "new" structural ways of attempting to improve future prospects for meeting cost, schedule, and technical performance expectations. In a recent review of literature for acquisition system development, Keating et al. (2017) examined a distribution of the literature over an 11-year period for 151 journal articles across five major classification categories (Figure 7), including the following:

- ***Tools***—Implements used to support accomplishment of a specific task or purpose
- ***Methods***—Systematic approaches that are performed to achieve an objective
- ***Models***—Representations that capture attributes against which comparisons can be made
- ***Methodologies***—Generalized frameworks that guide applications for the field
- ***Conceptual***—Fundamental underlying philosophical, theoretical, and axiomatic foundations that serve as a basis for the field

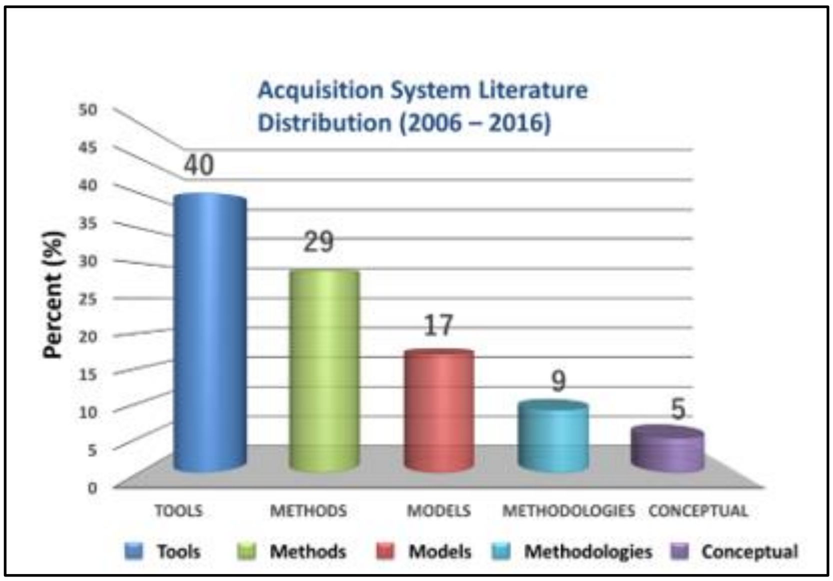


Figure 7. Literature Distribution for Defense Acquisition System

There are four primary conclusions based on this work: (1) There appears to be a heavy inclination toward the "practice" side of the acquisition system literature (Tools, Methods, Models) as fully 127 articles (84%) fit into these categories, (2) there was a noticeable absence of literature steeped in Systems Theory based domains, (3) a meager 14 articles (9%) address methodological aspects of defense acquisition, and (4) there was a noticeable scarcity of conceptual/theoretical articles (8 articles, 5%) where Systems Theory would be captured (although there were no Systems Theory based works).

In consideration of the present work in relationship to acquisition system development, five primary implications are offered:

1. *Acquisition system reform has proceeded without inclusion of Systems Theory*—this is not totally unexpected. Acquisition has developed as a practice based field. Notwithstanding the absence of System Theory, there is also a recognizable absence of consistent grounding in any theoretical basis. Therefore, the conclusion is offered that suggests an emphasis on a stronger theoretical linkage, which may include Systems Theory, might be beneficial for acquisition system reform.
2. *Systems Theory offers a different perspective and inquiry framework for examination of acquisition system reform*—Systems Theory places emphasis on understanding system design, execution, and development from the perspective of a well-grounded, mature, body of knowledge. Systems Theory provides a language, given as the set of propositions, which serves to explain the behavior/performance of complex systems while providing some predictive power.
3. *Acquisition system development breakthrough might be supported by focusing on the underdeveloped "conceptual" emphasis*—the scarcity of literature targeted to the conceptual (philosophical, theoretical, axiomatic) aspects of the acquisition system, suggests that this might be an area with substantial promise for enhancing acquisition system reform. As the preponderance of work has eluded this area, there might be significant breakthroughs to reform dilemmas.
4. *Focus on Systems Theory (propositions) violation might provide new and novel insights for acquisition system reform*—since the Systems Theory propositions have not been previously deployed in the development of the acquisition system, there is potential for new and insightful thinking. This might offer a shift in trajectory of acquisition system reform that has not yet been achieved.
5. *Acquisition Systems Theory*—a scan of comprehensive scholarly literature databases for "Acquisition Systems Theory" produced not a single article. This is consistent with finding that the conceptual (theoretical, philosophical, axiomatic) limited literature in the body of knowledge for the acquisition field. This suggests that there might be opportunity to "change the conversation" of acquisition system reform by the inclusion of theoretical development.

Systems Theory has broad-ranging implications for acquisition system reform. By any reasonable acknowledgement, acquisition system reform has met with difficulties. Nevertheless, as we have articulated, this has not prevented the success of multiple programs under the acquisition system. In looking for new and novel paths forward for the acquisition field, Systems Theory has been introduced as a body of knowledge with potential to elevate the acquisition field. We now turn to closing this work with an examination of conclusions and research directions.

Conclusions and Research Directions

In this paper, we have laid a foundation for the nature and role of Systems Theory in advancing present efforts related to acquisition system reform. Systems Theory has been presented as a potential contribution to better understand acquisition system reform. Additionally, Systems Theory might advance the acquisition field by introducing a new and novel language (to acquisition). This language can provide the basis for a different level of corresponding thinking, decision, action, and interpretation for acquisition system reform. In conclusion for this effort, three primary points are offered:

1. *Systems Theory as a Basis for Insights*—Systems Theory was presented as a set of propositions (laws, principles, concepts) that have been organized as an informing body of knowledge for the field. The set of propositions are applicable across systems, including the acquisition system. They serve to explain, and provide predictive power, for the behavior/performance of complex systems. As such, Systems Theory can provide explanatory analysis and insights to the acquisition system that have been elusive.
2. *Acquisition System Reform*—Systems Theory offers an enhanced perspective for acquisition system reform. It is interesting that the acquisition field has been relatively free from inclusions of Systems Theory. Even a rudimentary examination of Systems Theory provides a different perspective on acquisition. Perhaps acquisition system reform, continually being criticized for falling short of expectations, would benefit by the deeper examination from Systems Theory.
3. *Explaining Success and Failure*—The acquisition literature/programs are replete with both successes and failures. However, there has never been a thorough examination of the nature of success (and failure) from the perspective of Systems Theory. While Systems Theory has not been offered as a panacea to advance the acquisition system reform, it is portrayed as a new and novel approach to better understand critical issues in acquisition system development.
4. *Foundation for New Generation of Supporting Tools*—Bringing Systems Theory to life to support acquisition system reform requires movement beyond the conceptual (philosophical, theoretical, axiomatic) level. New Systems Theory based methods, tools, and techniques can be developed and tailored to the acquisition field to support practitioners faced with increasingly complex systems and problems.

Based on current explorations in Systems Theory application to acquisition system reform, several developmental avenues are suggested. While not offered as a complete set, these developmental directions will provide the foundation for a coherent (making sense for the acquisition field) and congruent (fit to address acquisition system reform) contribution to acquisition system development. These areas include *Systems Theory for Acquisition, Methods Development,* and *Applications* (Figure 8).

Systems Theory provides a knowledge base to explain the behavior/performance of complex systems. However, there is a noticeable absence of Systems Theory application to acquisition system reform. Although we do not know the reason for this absence, there are significant opportunities to expand acquisition system reform horizons through Systems Theory. While the essence of Systems Theory is not in question, there needs to be examination as to the direct application to acquisition field systems. We might certainly expect to make modifications in how the language is adjusted based on nuances of the acquisition field.

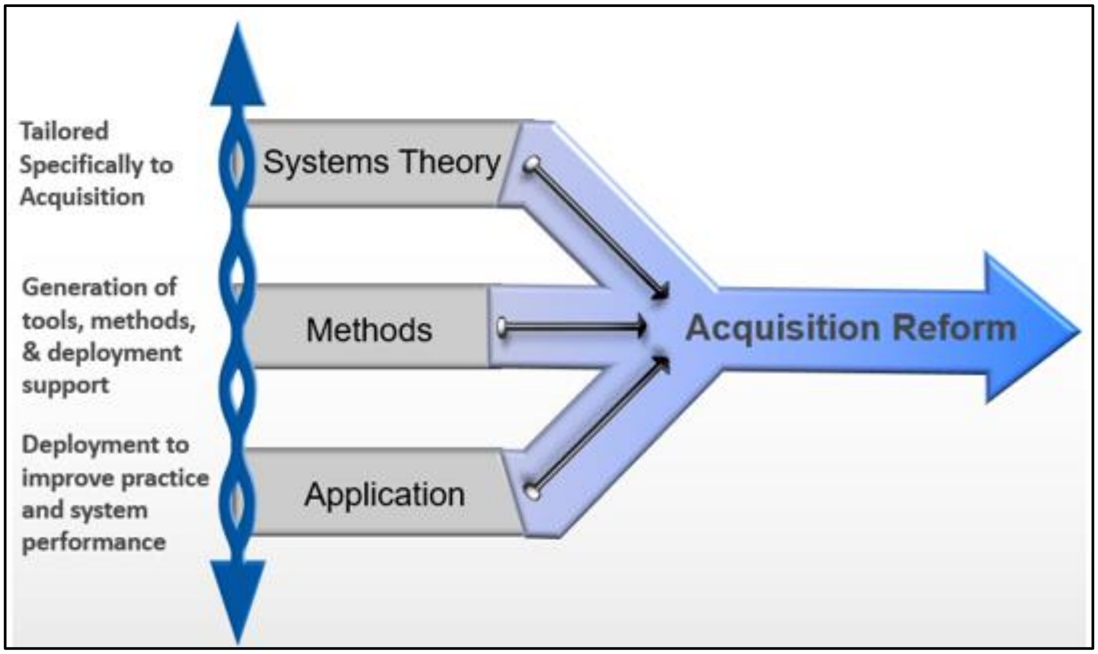


Figure 8. Simultaneous Development of System Theory, Methods, and Applications

Methods involves engineering of the science-based artifacts (tools, techniques, methods) to support enhanced capabilities that promote improved acquisition practice. Thus, *methods* finds its basis in Systems Theory and bridges the world of Systems Theory to the world of application through engineering of methods. Finally, *application* is focused on the deployment of methods-based capabilities to enhance acquisition practice. In the case of acquisition, this implies enabling practitioners with more sophisticated (Systems Theory based) methods to perform in their roles as acquisition professionals.

The path forward for application of Systems Theory for the acquisition field and practitioners is not without challenges. However, we have shown the promise that Systems Theory holds for acquisition system reform. There are no guarantees for the utility that will accrue for the application of Systems Theory to acquisition system reform. However, this exploration provides a level of confidence in knowing that Systems Theory offers a new, novel, and insightful perspective for engaging acquisition reform.

References

Adams, K. M., Hester, P. T., Bradley, J. M., Meyers, T. J., & Keating, C. B. (2014). Systems theory as the foundation for understanding systems. *Systems Engineering, 17*(1), 112–123. Retrieved from https://doi.org/10.1002/sys.21255

Beer, S. (1978). *Platform for change*. Chichester, UK: John Wiley.

Berteau, D., Levy, R., Ben-Ari, G., & Moore, C. (2011). *Defense industry access to capital markets: Wall Street and the Pentagon: An annotated brief.*

Boulding, K. E. (1956). General systems theory: The skeleton of science. *Management Science, 2*(3), 197–208. Retrieved from https://doi.org/10.1287/mnsc.2.3.197

Clemson, B. (1984). *Cybernetics: A new management tool*. Tunbridge Wells, Kent, UK: Abacus Press.

Francis, P. L. (2008). *Defense acquisitions: Zumwalt-class destroyer program emblematic of challenges facing Navy shipbuilding*. Washington, DC: GPO.

Francis, P. L. (2009). *Defense acquisitions: Cost to deliver Zumwalt-class destroyers likely to exceed budget* (1437909310). Washington, DC: GPO.

Hammond, D. (2002). Exploring the genealogy of systems thinking. *Systems Research and Behavioral Science, 19*(5), 429–439. Retrieved from https://doi.org/10.1002/sres.499

Lespier, L., Long, S., & Shoberg, T. (2015). A systems thinking approach to post-disaster restoration of maritime transportation systems. In *Proceedings of the 2015 IIE Annual Conference and Expo* (pp. 2262–2272).

Katina, P. F. (2015a). Emerging systems theory–based pathologies for governance of complex systems. *International Journal of System of Systems Engineering, 6*(1/2), 144–159. https://doi.org/10.1504/IJSSE.2015.068806

Katina, P. F. (2015b). *Systems theory-based construct for identifying metasystem pathologies for complex system governance* (Doctoral dissertation). Norfolk, VA: Old Dominion University. Retrieved from http://search.proquest.com.proxy.lib.odu.edu/docview/1717329758/abstract/29A520C8C0A744A2PQ/2

Katina, P. F. (2016a). Metasystem pathologies (M-Path) method: Phases and procedures. *Journal of Management Development, 35*(10), 1287–1301. Retrieved from https://doi.org/10.1108/JMD-02-2016-0024

Katina, P. F. (2016b). Systems theory as a foundation for discovery of pathologies for complex system problem formulation. In A. J. Masys (Ed.), *Applications of systems thinking and soft operations research in managing complexity* (pp. 227–267). Geneva, Switzerland: Springer International. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-21106-0_11

Katina, P. F., & Keating, C. B. (2014). Metasystem pathologies: Towards a systems-based construct for complex system deficiencies. In S. Long, E.-H. Ng, & C. Downing (Eds.), *Proceedings of the American Society for Engineering Management 2014 International Annual Conference*. Virginia Beach, VA: American Society for Engineering Management. Retrieved from http://search.proquest.com/openview/f93264cf060b406658d39c6916a22cc2/1?pq-origsite=gscholar

Katina, P. F., & Keating, C. B. (2016). Metasystem pathologies: Towards discovering of impediments to system performance. In H. Yang, Z. Kong, & M. D. Sarder (Eds.), *Proceedings of the 2016 Industrial and Systems Engineering Research Conference*. Anaheim, CA.

Keating, C. B., & Katina, P. F. (2012). Prevalence of pathologies in systems of systems. *International Journal of System of Systems Engineering, 3*(3–4), 243–267.

Keating C. B., Katina, P. F., Jaradat, R., Bradley, J. M., & Gheorghe, A. V. (2017). Acquisition system development: A complex system governance perspective. *INCOSE International Symposium, 27*(1), 811–825. Retrieved from https://doi.org/10.1002/j.2334-5837.2017.00395.x

Kendall, F. (2014, July 10). Testimony Before the House Committee on Armed Services.

Kendall, F. (2016, October 24). *Performance of the Defense Acquisition System: 2016 annual report.*

Klir, G. J. (Ed.). (1972). *Trends in general systems theory* (1st ed.). New York, NY: John Wiley & Sons.

Kramer, N. J. T. A., & de Smit, J. (1977). *Systems thinking: Concepts and notions*. New York, NY: Springer U.S. Retrieved from http://www.springer.com/gp/book/9789020705874

McDermott T., & Salado, A. (2017). Improving the systems thinking skills of the systems architect via aesthetic interpretation of art. *INCOSE International Symposium, 27*(1), 1340–1354. Retrieved from https://doi.org/10.1002/j.2334-5837.2017.00432.x

Rascona, P. M., Barkakati, N., & Solis, W. M. (2008). *DOD business transformation: Air Force's current approach increases risk that asset visibility goals and transformation priorities will not be achieved*. Washington, DC: GPO.

Skyttner, L. (2005). *General systems theory: Problems, perspectives, practice* (2nd ed.). Singapore: World Scientific.

Troncale, L. (1977). Linkage propositions between fifty principal systems concepts. In G. J. Klir (Ed.), *Applied general systems research: Recent development and trends* (pp. 29–52). New York, NY: Plenum Press.

van Gigch, J. P. (1974). *Applied general systems theory* (2nd ed.). New York, NY: Harper and Row.

von Bertalanffy, L. (1950). The theory of open systems in physics and biology. *Science, 111*(2872), 23–29.

von Bertalanffy, L. (1968). *General system theory: Foundations, developments, applications.* New York, NY: George Braziller.

Warfield, J. N. (1999). Twenty laws of complexity: Science applicable in organizations. *Systems Research and Behavioral Science, 16*(1), 3–40. Retrieved from https://doi.org/10.1002/(SICI)1099-1743(199901/02)16:1<3::AID-SRES241>3.0.CO;2-F

Whitney, K., Bradley, J. M., Baugh, D. E., & Chesterman, C. W. (2015). Systems theory as a foundation for governance of complex systems. *International Journal of System of Systems Engineering, 6*(1–2), 15–32. Retrieved from https://doi.org/10.1504/IJSSE.2015.068805

Industrial White Paper Briefing: Monterey Bay Regional Spaceport Lowest Priced Technically Acceptable COTS Launch System[1]

Kelly Keith Weigel—is a Space Commerce Entrepreneur; Co-Founder of Trans Universal Energy, LLC (Monterey Spaceport Operations); and Project Site Lead for the Monterey Bay Regional Spaceport. [weigel@transuniversalenergy.com]

Abstract

Proposed DoD Acquisition Study to focus on COTS Space Launch System for government and commercial markets. Level of interest is in the broad cost savings to the government and commercial customers.

Research Issue: The cost to launch vehicles and payloads into orbit and to station them in the upper atmosphere is prohibiting innovation and limiting the total number of systems that can be deployed. There is a need to study the life cycle of current and future COTS Launch Systems and their benefits to government and commercial procurement cycles.

Research Result: Lower costs via Lowest Price Technically Acceptable (LPTA) COTS launch solutions and systems are the desired research result. Smaller payloads and delivery vehicles with more capabilities will increase the amount of work being done while less money is being spent. Overall cost savings to the government activities and the civilian population they serve will be the outcome of the research result as well.

Center of Excellence Model to Protect Intellectual Property

During the compilation of this Acquisition Research Proposal, Intellectual Property Rights created parameters which restricted certain research materials from being published. A University of Florida suggestion is being deployed as a remedy solution. U. of Florida has suggested that a Center of Excellence (CoE) for Commercial Space Transportation (CST) chapter be assigned to Monterey. The CoE model would allow for the sharing and throughput of information as well as enable “unlimited distribution” of materials. The CoE acronym is likely to be CoE CST Monterey.

Proposed Consortium of Acquisition Research Partnerships:

- Academia: UCMBEST UC Santa Cruz, CSUMB, NPS, DLI Army Garrison Facilities
- Prime contractors: As specified in the MBRS Project Site SoCE
- Government: FAA AST CST, DoE ~ LLNL/Sandia Livermore, CA, DoD ~ DARPA

[1] Federal FY 2018 Q3; FY 2018 ~ FY 2038 Range Activity Projection

LPTA COTS Launch Areas of Endeavor:

- Range Work: Flight Field at the Marina, CA, Municipal Airport, OAR (KOAR)
- Sea and Air Launch activities in the Pacific Ocean 60 ~100 mi from OAR

Launch System Components: LTAVs, Legacy G.A. Approved Airframes, ATC UASoS Network, Other Experimental Horizontal Flight Frames, Hybrid Engines, Hybrid Space Vehicles, Energy Tech for Power Train

Proprietary: T.U.E., LLC (MSO) MBRS-MBSDC ~ Range Operations Opportunities

MBRS Project Site Relevant Information

- Worked for the Presidio of Monterey's DFMWR as IT/MIS-GISO, 2005 to 2012
- Became iGATE Client Company as "FOTDDO / T.U.E., LLC," 2012 to 2013
- Launched the MBRS Project Site in 2013 as Trans Universal Energy, LLC (2013 to Present)

Company Interest in Space Commerce Background

- Company Time Line—iGATE Client: March 1, 2012, to October 1, 2013
- MBRS Project Site Sole Focus—October 1, 2013, to the present day

Work with the DoE Incubator called the iGATE in Livermore, CA, from 2012 to 2013: "My long term interest in Astrophysics from the late 1970s was finally able to blossom in 2012. Elements and components for the advanced energy and space launch system were discovered at the iGATE. A clear path to commercialization was illuminated by the DoE and DOD GOTS examples for spin off technologies."

On March 1, 2012, our company signed up as an iGATE client, known as FOTDDO.biz. On January 14, 2013, our company incorporated as Trans Universal Energy, LLC. In March 2013, contact with the FAA AST CST was established which formed the basis of the current and ongoing reporting cycles. The specialized focus on the Flight System Operators license began in August 2013. From October 1, 2013, to the present day, our company has singularly focused on the MBRS Project Site. Guidance from the FAA AST CST continues to provide the milestones for commercial space launch activities, transportation operations, and related services. The Flight System Operators license will establish the required parameters for the MBRS "Spaceport" license. The MBRS Spaceport license is likely to be granted to the municipal airfield now known as OAR (KOAR) in the 2020s.

Acronyms Used in This Industrial White Paper Briefing

ATC	Air Traffic Controller
CoE	Center of Excellence
COTS	Commercial Off the Shelf
CSUMB	California State University Monterey Bay
DARPA	Defense Advanced Research Projects Agency
DFMWR	Directorate of Family Morale Welfare and Recreation
DLI	Defense Language Institute
DoD	Department of Defense
DoE	Department of Energy
FAA AST CST	Federal Aviation Administration Air and Space Transportation Commercial Space Transportation
FOTDDO	Fort Ord Technology Development and Deployment Office
G.A.	General Aviation
GOTS	Government Off The Shelf
LLNL	Lawrence Livermore National Laboratory
LPTA	Lowest Priced Technically Acceptable
LTAV	Lighter Than Air Vehicle
MBRS	Monterey Bay Regional Spaceport
MBSDC	Monterey Bay Spaceport Development Center
MSO	Monterey Spaceport Operations
NPS	Naval Postgraduate School
OAR (KOAR)	Call letters of the Marina Municipal Airport
SoCE	Scope of Contracting Environment
T.U.E., LLC	Trans Universal Energy, LLC
UASoS	Unmanned Aerial System of Systems
UCMBEST	University of California Monterey Bay Education Science and Technology Center

In the Fullness of Time: Towards Realistic Acquisition Schedule Estimates

Raymond Franck, Brig Gen, USAF (Ret.)—retired from the faculty of the Graduate School of Business & Public Policy (GSBPP), Naval Postgraduate School (NPS) in 2012. He retired from the Air Force in 2000 in the grade of Brigadier General. His active duty career included a number of operational tours and staff positions, and head of the Department of Economics and Geography, USAF Academy. His published work includes a number of journal articles and research reports in military innovation and defense acquisition management. [cfranck215@aol.com]

Gregory Hildebrandt, USAF (Ret.)—has had an Air Force career including assignments as an acquisition officer, Air Force Academy faculty member, and assignments at the Central Intelligence Agency and Office of the Secretary of Defense. Following his Air Force retirement, he has continued service with the RAND Corporation and NPS. His published work includes a number of journal articles in defense economics and RAND reports on acquisition issues. [ggh324@gmail.com]

Charles Pickar, USA, (Ret.)—is a member of the NPS faculty where he teaches project management, defense acquisition, and systems engineering. Before joining NPS, he led the Applied Systems Engineering Program Area at the Johns Hopkins University Applied Physics Laboratory. He is a retired Army officer with extensive experience in the U.S. defense industry, to include director and VP levels at Lockheed Martin, Northrop Grumman, and SAIC. He is the current Chair of the Systems Education Technical Committee of the IEEE Systems Council. His research and published work focuses on applying systems engineering and system dynamics analytical approaches to defense acquisition problems. [ckpickar@nps.edu]

Bernard Udis—is a Professor Emeritus of Economics at the University of Colorado at Boulder. He has also served at the U.S. Air Force Academy and the U.S. Arms Control & Disarmament Agency. His NATO research fellowship examined the costs and benefits of offsets in defense trade. A recognized authority on the economics of defense, his published work includes three books, plus a number of book chapters and journal articles. [bernard.udis@colorado.edu]

Abstract

This paper continues a research agenda started in 2016 with an aim of more realistic acquisition program scheduling estimates, especially for the development (SSD) phase. This, our third look at the scheduling problem, starts with a discussion of scheduling data, and how that data could be applied to help the DoD address this challenge. This section includes ideas on how to use acquisition data for the scheduling problem. Next, we present a case study that is the result of field interviews with senior DoD leaders. Finally, we present a discussion on using the system performance as a metric.

Introduction

Weapons system development projects are infamous for exceeding time and cost limitations. Often the reaction to this notoriety is changes at the policy level of acquisition. However, the problem may well lie somewhere else. This paper, like the two preceding papers in this series, suggests we may well be “lookin’ … in all the wrong places” (to paraphrase an old country song[1]) for the causes, because the causes may well lie inside the project and therefore not be readily addressed by policy changes.

While cost, performance, and schedule are critical variables in any acquisition program, Congress, the media, and policymakers generally focus on cost, with little attention devoted to the issues of schedule. Moreover, although the DoD has engaged in significant efforts to develop methods for realistic acquisition cost estimates, it has paid considerably less attention to schedules—their estimates and execution. To emphasize the challenge of schedules, Figure 1 provides a macro level view of the schedule problem. Over the past 20 years, Major Defense Acquisition Programs (MDAPs), as reported in Selected Acquisition Reports (SARs), averaged schedule overruns of more than 24 months. Schedule overruns occur for many reasons and this study examines some of those reasons.

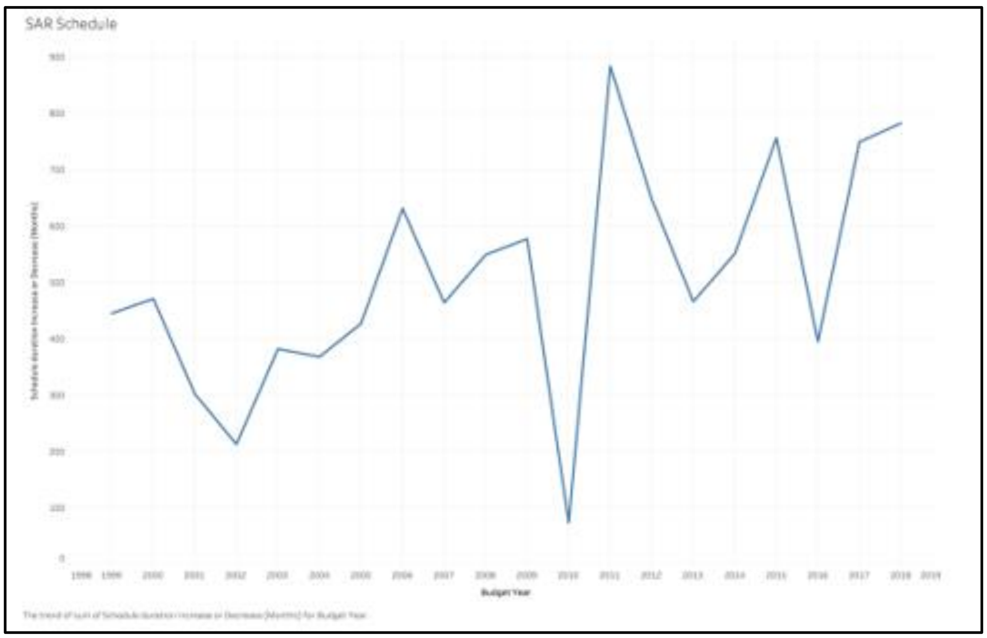


Figure 1. Sum of Schedule Overruns, 1998–2017 (Months)

We use a multi-faceted approach to examine weapons systems development scheduling to assess the current state and contributing causes of schedule estimating methodologies and suggest different ways to accomplish this difficult process. The overarching research question is as follows:

---

[1] From the words to a song written by Wanda Mallette, Bob Morrison, and Patti Ryan, and recorded by American country music singer Johnny Lee in June 1980.

> *What analytical techniques and approaches can be applied to schedule development/analysis to increase the efficiency and effectiveness of schedule estimating and execution?*

As long ago as 1988, Morris and Hough were critical of the practice of project management:

> Curiously, despite the enormous attention project management and analysis have received over the years, the track record of projects is fundamentally poor, particularly for the larger and more difficult ones. Overruns are common. Many projects appear as failures, particularly in the public view. Projects are often completed late or over budget, do not perform in the way expected, involve severe strain on participating institutions or are cancelled prior to their completion after the expenditure of considerable sums of money.

In fact, project management in general, and DoD project management in particular, has been dealing with these problems described by Morris and Hough for decades. We hope to inform these problems because, “when problems persist, practitioners and scholars are getting something wrong” (Christensen & Bartman, 2016).

This paper is the third in a series of investigations into alternatives to the way we do schedule estimation today and builds on the research agenda proposed by Franck et al. in 2016 and furthered in Franck et al. in 2017 (Franck, Hildebrandt, & Udis, 2016; Franck, Hildebrandt, Pickar, & Udis, 2017). We start with a discussion of scheduling data, and how that data could be applied to help the DoD address this challenge, and how system dynamics can inform. Next, we present a case study that is the result of field interviews with senior DoD leaders. Finally, we present a discussion on using parametric analysis.

The Dynamics of Project Management

The concept of time in project management can be divided into two steps: estimating task duration and building the schedule. Both processes require technical expertise and management savvy. First the technical process of estimating the duration of the project task must be determined. Once duration is established, the management process of project sequencing and scheduling must be defined.

Estimating Activity Duration

Surprisingly, little information is available in the literature on the “how” to estimate the elements of a schedule—the task duration. While the major defense contractors have formal in-company processes, little formal literature is available on the specifics of task estimation. Further, most available information on estimating task duration is found in project management textbooks, but even then, the specifics are scarce.

The PMBOK (Project Management Body of Knowledge) lists five methods for estimating project activity duration. These methods include (Project Management Institute [PMI], 2017):

- Expert Judgment
- Alternatives Analysis
- Published Estimating Data
- Project Management Software
- Bottoms-up Estimating

Expert judgment acknowledges that technical and engineering experts should be able to estimate the effort necessary to accomplish tasks and translate those estimates to duration. This assumes the chosen experts have significant experience in the execution of those tasks, and are therefore competent to judge time required (Hughes, 1996).

Alternatives analysis recognizes that activities or tasks can be accomplished in different ways—alternatives. These different ways include defining different techniques, differing levels of resources, and using different machines.

Published estimates are databanks that gather resources measures. These measures include hourly rates by skill level, acknowledged production rates for various development, and manufacturing activities. In most cases, this data is available internal to the organization. However, there are data companies that track and report this data. An example is the IEEE-USA Salary & Benefits Survey. This data is often available for different locations in the United States as well as worldwide.

Project management software is not really an estimation method. Instead, it provides a means to identify and organize information necessary for resource estimates.

Finally, an engineering or bottoms-up estimate is a comprehensive schedule (and cost) process that starts at the work package level and aggregates costs to build a complete estimate. Bottoms-up estimates are necessary when schedule activities cannot be accurately estimated using another technique. As the name implies, bottoms-up estimates start at a level of activity or task that can be confidently estimated. The activities are then rolled-up to the required level. These estimates are extremely work intensive but are also the most accurate.

Other recognized methods include parametric techniques. A parametric or top-down estimate builds an activity estimate for the development project from historical data comparing variables through a statistical relationship. All the methods listed are used to estimate the length of time each of the activities or Work Breakdown Structure tasks lists. "Simply stated, the duration of an activity is the scope of the work (quantity) divided by a measure of productivity" (Hendrickson, Martinelli, & Rehak, 1987, p. 278).

Thus, activity duration estimation establishes the actual time required to complete discrete tasks in an overall project, while project scheduling fixes the start and end dates, as well as execution approaches of the project. Once the overall schedule is established, management activities driven by either time and/or resource constraints will determine the actual execution of the project (Schwindt & Zimmerman, 2015). The analogy that comes to mind is that of an orchestra. The individual instruments (and of course, the musicians) are the discrete tasks of the project. The orchestra leader is the project manager, and the music score is the "plan" the orchestra leader uses to execute the "project." Building on this information, the next step in this effort is to identify schedule data that can be used to augment these estimating activities.

Schedule Data

While there is significant information available on DoD procurements, the overwhelming majority of that information is on cost. In order to effectively examine project schedules, we must be able to better understand those schedules. It is common knowledge that weapons system development projects overrun their schedules. However, we need to be able to determine what causes schedule overruns, as well as an actual measure of the development time.

Data for this research was obtained from the Defense Acquisition Management Information Retrieval (DAMIR) database, a repository for, *inter alia*, the DoD Selected

Acquisition Reports (SARs). The SAR is a summary of the acquisition data of selected Major Defense Acquisition Programs (MDAPs). Table 1 provides a list of delay factors, as well as maximum and minimum delays as reported in the SAR during the period 1997–2017.[2]

Table 1. Delay Factors, Maximum Delays, and Minimum Delays, 1997–2017

Delay Factor	# instances	Maximum Delay (months)	Minimum Delay (months)
Administrative changes to schedule including updates to APB, ADM changes, as well as changes resulting from Nunn-McCurdy processes and program restructuring	460	168	5
Technical	291	60	4
Testing delays	283	66	1
Delay in availability of key capabilities/facilities (launch vehicle/testing facilities/IOT&E units)	3	13	6
Budget/Funding Delays	52	43	1
Delays attributed to the Contractor	50		
Delays because of Rework	16	4	1
External events such as inflation, earthquakes, labor strikes, etc. (*Force Majeure*)	4	4	1
Delays due to Contracting/Contract Negotiation/Award delays	29	27	1
Actuals (updating previously reported dates to actual occurrence)	172	13	-39

These delay factors suggest program managers (PMs) should plan for the time necessary to deal with oversight, information reporting and both the time takes, as well as the impacts of decisions—internal and external to the program. As the GAO pointed out in a 2015 study, the program office overheads associated with administrative activities added, on average, two years to complete:

> Programs we surveyed spent on average over 2 years completing the steps necessary to document up to 49 information requirements for their most recent acquisition milestone. This includes the time for the program office to develop the documentation and for various stakeholders to review and approve the documentation.

[2] The data described are from an unpublished study by the author of the delay factors for DoD program 1997–2017. The study is an initial attempt at quantifying schedule delays in program execution with the intent of using those delays to better inform project planning.

Figure 2 provides a trend line and forecast of the delays identified. Using this data, the forecast total delay hours across all programs in 2019 would be 712 hours, and in 2020 that forecast would increase to 729 hours.

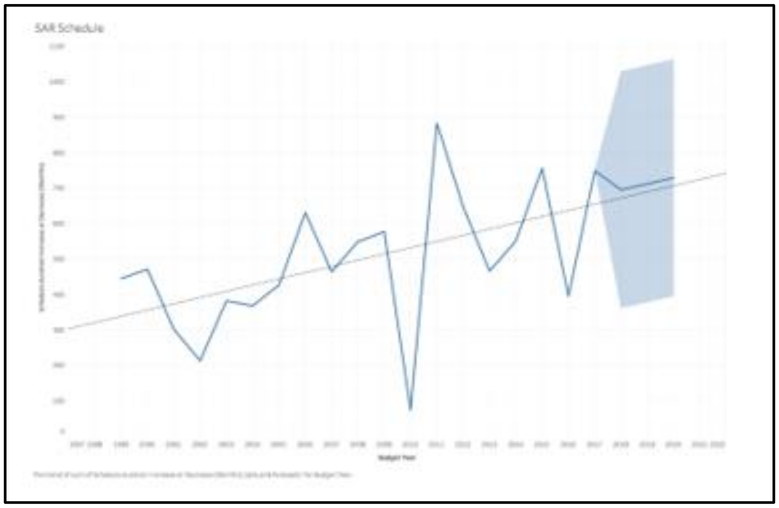


Figure 2. Trend Line Showing Forecasted Schedule Increases

Applying the Data

Our previous paper introduced the rework concept, shown in Figure 3. As noted, the CPM/PERT approach to scheduling precludes the use of data at the program schedule level. And, while some companies track task estimation data, that data is often proprietary and more focused on technical process estimation (Godlewski, Lee, & Cooper, 2012).

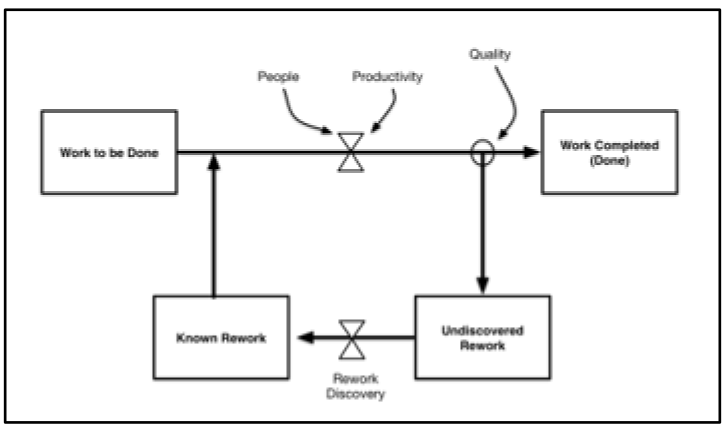


Figure 3. The Rework Cycle

The basic assumption that work proceeds as planned in the network from start to finish is naïve at best (Franck et al., 2017). System dynamics can account for the feedback that results from decisions made in the execution of a project. A project network using CPM/PERT techniques depends on each task being completed in the defined order established. While most PMs attempt to maintain that order, the reality of dynamics intervenes. That reality means that network analysis cannot capture the progress of a project (Williams et al., 1994).

A tool used in system dynamics to capture cause and effect is a causal map. The causal map becomes a tool used for the development of a model of the delay factors identified. Figure 4 is an initial causal map capturing some of the identified factors in weapons system program schedule delays. The factors shown are a subset of those identified for brevity in this paper.

Delay factors plus the effects of rework, decision wait time, tasks start delay, and other disruptions result in the PM (or PMO) recognizing a schedule problem (delay in the critical path). Invariably, the PM must take action to attempt to return the project to the equilibrium expressed as being on schedule. Thus, the PM could approve overtime, reschedule, or take some other mitigation. The pressure to get back on schedule is driven by many factors including cost considerations, pressure from the oversight organizations, and in weapons systems development, the necessity of delivering capability to the warfighter in the most efficient time. Regardless the reason, the PM "does something." The plus and minus signs indicate the effect of the actions taken.

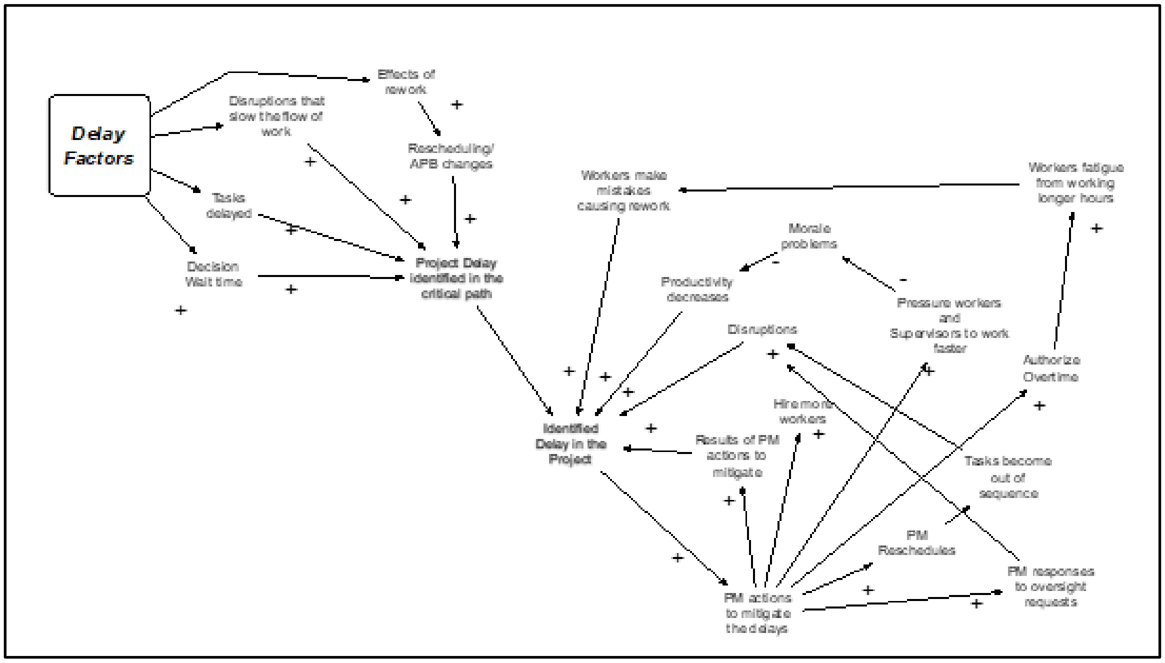


Figure 4. Delay Factors Triggers for Project Delays
(Howick, 2003)

A project is a dynamic system with feedback loops and, invariably, decisions taken to address one problem have an impact on, or create new problems. For example, approving overtime does initially address schedule issues as more work is being done in shorter periods. However, a recognized problem of overtime is fatigue. Fatigue causes workers to make mistakes, and those mistakes result in having to redo the work, thus perpetuating problems that were thought solved.

Similarly, hiring more workers causes more problems. Assuming the new workers have the requisite skills, they need to be trained/acclimated to the actual project situation. In the *Mythical Man Month*, Brooks (1995) explained how this concept works in software development. In reality, it is universal.

Finally, while many of the delay factors identified from the SAR analysis can be explained in Figure 4, others require further examination. One of the biggest challenges is the area of decisions, both internal and external. The internal decisions drive many of the actors discussed above. However, the PM must also deal with external decisions that can eventually impact the development.

Figure 5 is a notional graphic that represents a generic decision cycle in the context of the rework cycle. While the results of this data analysis included rework, the majority of the identified delay factors were decision focused. Those decision centric factors included represent this decision cycle. The notation is shown between the work to be done and work completed boxes because many of the decisions identified occur outside the project manager's purview. The exogenous factors identified cause either reactions to those factors, or force other internal decisions. While not normally a part of the rework cycle, we suggest that a formal appreciation of a decision cycle, and the time it takes for decisions to be made both internal as well as external to the program management cycle, must be considered.

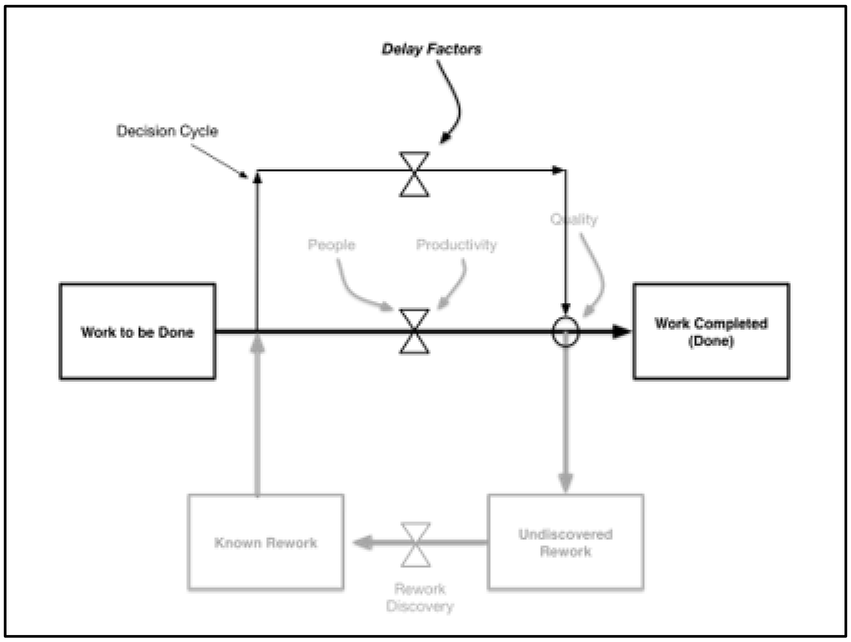


Figure 5. Notional Decision Cycle Added to Rework Cycle Diagram

Conclusion

This section continued the discussion on using system dynamics to better understand, plan, and execute defense acquisition programs. This section presented schedule information gleaned from Selected Acquisition Reports, and suggested a model to show how that information can be best understood in the context of the decisions necessary to model weapons system acquisition programs. To be clear, we are not advocating to replace the CPM/PERT methods used today. At best, system dynamics is an adjunct to those methods in use. Instead, we suggest that we should recognize the dynamics at play in any weapons system development, and once recognized, use the appropriate tools to better our execution.

No program manager sets out to overrun a schedule; "However, clients increasingly value not only cost and schedule control but cost and schedule certainty" (Godlewski et al., 2012, p. 18). Those clients for defense acquisition products seek certainty as well, both in cost and schedule. It is no secret that current methods for estimating and executing schedule are insufficient. In fact, certainty is one of the potential benefits of this examination of schedule factors. Project certainty starts in effective schedule planning by using the right tools.

F-35 Case Study[3]

Much has been written about the F-35 program, in many venues. Defense acquisition professionals know a lot about “what” has happened. “How” and “why” it has happened is less clear. Our last essay (Franck, Hildebrandt, Pickar & Udis, 2017) undertook an inquiry as to the “hows” and “whys” of this case. We asked how a program that traces its lineage to the Common Affordable Lightweight Fighter became the F-35—which is not very common (Bogdan, 2012), definitely not lightweight, of debatable affordability (see GAO, 2017; Capaccio, 2018), and arguably not a fighter (Airpower Australia, 2017).

The publicly-available literature was not terribly enlightening, although a few interesting clues were discernible. We closed with an intent “to learn more in future inquiries” (Franck et al., 2017, p. 420). Since then, the field interview method has brought new insights to many aspects of the F-35 program.

Given space limitations, we concentrate on some useful hypotheses we’ve gleaned—the assessment of which is for further inquiry. These hypotheses[4] concern program management, technology and engineering, and the lure of new technologies. Careful readers will note that they are not mutually exclusive and are interrelated in a number of ways.

Program Management

The management narrative can be organized as poor program structure from the start: an underequipped and over-burdened program office—all of which enabled bad decisions.

Program Structure

The program turned out to be well-designed to fail. Basically, Lockheed-Martin (LM, the prime contractor) had considerable discretion and control over a highly complex program with a vague set of requirements. Moreover, the incentive structure was not well designed (“poor” according to at least one authority). This produced a Principal-Agent problem (e.g., Kreps, 1990, Chapter 16) with the Principal (DoD) unable to fully monitor the agent’s (LM’s) behavior, or to incentivize good behavior. One result was a strained relationship between LM and DoD (“worst I’ve ever seen”; Bogdan, 2012).

The program strategy reflected a number of optimistic framing assumptions. These included joint programs saving money, plus new, but untried, methods expected to significantly reduce risk and time. This latter set included the assumed benefits of recent acquisition reforms and better simulation methods expected to reduce flight testing. All this led to an aggressive schedule—involving tight timelines with a high degree of concurrency accepted *a priori* (Blickstein et al., 2011, p. 37).

When these assumptions were not borne out, schedules stretched out and costs grew. The RAND Root Cause Analysis, for example, concluded “optimistic cost and schedule estimates” constituted a major cause of program difficulties (Blickstein et al., 2011, p. 37).

---

[3] We are greatly indebted to a highly-placed, well-informed DoD official for many of the insights that underpin this section of our paper. Chatham House Rule applies.

[4] Although readers will likely not agree with all the details, few, if any, will be surprised.

Program Office

The F-35's DoD management team was assigned a task that included serious complexities in both technical and management dimensions. Moreover, the management difficulties included coordination of 11 stakeholders (three U.S. and eight international) with varied operational needs while complying with the U.S. ITAR (International Trade in Arms) regime.

Additionally, cascading effects of program difficulties made the work even more complex. One example was weight growth early in the program (precipitated in part by entering development with a slender weight growth margin), which necessitated a larger engine, which in turn necessitated a major redesign of the fuselage to accommodate the larger engine (Blickstein et al., 2011, p. 53)—with one major result being cost growth and schedule delay. The acquisition strategy turned out to be something of a "house of cards."

Given its highly complex and demanding mission, the F-35 Program Office was woefully underequipped at crucial junctures. Requirements discipline in the formative period has been characterized as "weak" and unable to deal effectively with a number of changes internal to the program (e.g., tech insertions, revised development plans) and external (e.g., threat evolution). In addition, there were, at times, mismatches between Program Office needs and personnel skills aboard.

Some tools of program management were inadequate—particularly for schedules. From a program management perspective, schedule management tools proved hard to use, not well tied to resource use, insufficiently flexible to account for risk and program perturbations, and not supported with data from historical experience. As program difficulties arose, there was no credible means available to estimate schedule implications.

These are, of course, difficulties that afflict any defense acquisition program. However, new, complex, difficult, advanced systems like the F-35 suffer more. Another difficulty was rotating new program executive officers (PEOs) every few years. Accordingly, both the opportunity and incentive to reorient the program were in very short supply. This particular pattern was broken in 2012 with an indefinite-term PEO.

In addition, as problems continued, the Program Office was subject to a rather onerous oversight regime, with attendant political pressures and constraints. The one-year F-35B probation period is one example (Franck et al., 2012, esp. pp. 57–59).

Program Execution: Bad Decisions

The factors cited above facilitated bad decisions. The flawed assumptions that underpinned the acquisition strategy did not receive sufficient scrutiny (perhaps related to leadership tenure). In an atmosphere of pervasive optimism, relatively pessimistic assessments (such as the CAIG report in 2001) had little apparent effect on program management (Blickstein et al., 2011, p. 37). Requirements remained in some degree of flux well into the program life, with corresponding effects on program stability.

Heavy reliance on test data (e.g., reliance on simulations and test data from non-scale airframes) greatly delayed the test program when those presumptions proved inaccurate.

The F-35 Helmet Mounted Display (HMD) was a major technical advance—with great promise but high risk and no guarantee of success. However, a natural programmatic hedge, head-up display (HUD), was cancelled early in the program. This meant that lags in HMD development became a major threat to program success (Bogdan, 2012).

Program Office personnel clung closely to a commonality standard among the three models, with cost growth and delays associated with fixing one model's problems among all three models. (This seems to make sense if the F-35 is one unified program; less so, if there are three programs with commonalities.[5])

Technology and Engineering

The optimism that set the theme for the management strategy also pervaded the technology assumptions. There was a strong proclivity to underestimate the difficulties and risks. While, for example, there was a fair amount of experience with stealthy aircraft designs within the U.S. defense industrial base, the F-35 was nonetheless a major leap forward. As RAND's Root Cause Analysis noted, the basic technical requirements were very demanding. This is illustrated in Table 2. Given the high degree of commonality specified for the F-35, if one model needed to meet the design objectives in the table, all models need to achieve those objectives. It took considerable ingenuity to design an airplane whose morphology accommodated all these requirements (Blickstein et al., 2011, esp. p. 37).

Table 2. Required Features for F-35 Design

(Adapted from Blickstein et al., 2011, Table 4.6, p. 49)

	STEALTH	STOVL	SUPERSONIC
Engine Inlets	Small	Large	Specific shapes
Fuel Capacity	Internal only	Small	Large
Airframe Shape	Specific (radar signature)	Specific (weight distribution)	Specific (speed regime transitions)
Materials	Stealthy airframe skin	Light skin for vertical landing	Strong skin (speed regime transitions)

Accordingly, there was little margin for error or unexpected difficulties; one example was the 6% allowance for increased weight. That reserve was exceeded early in the program, which necessitated a major redesign exercise (Blickstein et al., 2011, pp. 47, 53).

Given the demanding nature of the original design and slender margins for error, there was nonetheless a definite willingness to push the technical envelope. Thus, for example, the Helmet Mounted Display (discussed above) was a major technical advance—with a natural hedge (HUD) discarded early.[6]

There was likewise a propensity to trust new and promising, but not fully validated, engineering methods. These included computer simulations substituting much of the testing normally accomplished in the air. The result was a test program generally behind and in a catch-up mode (e.g., see DOTE, 2016, esp. p. 31).

[5] LtGen. Bogdan (2012) eloquently stated the separate-programs perspective.

[6] Reasonable people can disagree as to whether this is a management issue, technical issue, or both.

The Attraction of New Technologies

Technology insertions occurred with some frequency during the F-35 development program. These included the Autonomic Logistics Information System (ALIS) and the Helmet-Mounted Display. ALIS seems to have been regarded as merely the logical extension of onboard aircraft diagnostics (Steidle, 1997, p. 9). However, more than a decade later, problems with ALIS were (rightly) viewed as an existential threat to the entire program (Bogdan, 2012).

Likewise, the evolution of the F-35 from an affordable, limited-capability companion for the F-22 (*inter alia*) to a "situational awareness machine" seems to be related to some major advances in sensor capabilities that the F-35 program adopted. (It's also true that the stakeholders were involved: "JAST … was … designed to have the smallest possible sensor suite and be dependent on external information sources … [But] most of the export countries did not have (those sources) in their inventory" and the F-35 became a battlefield information producer [Keijsper, 2007, p. 135]).

Such initiatives, taken in isolation, were undoubtedly viewed as sensible at the time. However, the cumulative effect of a series of sensible decisions can be a horrible end result.

The last word on the new technologies and platform performance issues might well come from General Deptula (2016):

> Current systems are largely expected to operate in a semi-autonomous fashion, with a basic level of collaborative engagement with other platforms. These shortcomings place pressure on individual assets to possess numerous internal capabilities. The complexity inherent to this approach drives lengthy development cycles, which in turn leads to requirement creep, time and cost overruns, and delays in capability. (pp. 6–7; emphasis added)

This looks like an indirect reference to the F-35 we're getting.

Some Questions for Further Investigation

1. Can an acquisition program schedule become self-stretching? A simplified version of this hypothesis goes something like this. System complexity entails a lengthy development program. Over time, various technical improvements present themselves—some of which are adopted. These technical insertions (even if done well) nonetheless add to system complexity or estimated program schedule (or both). This cycle is summarized in Figure 6.

While this influence diagram seems plausible, the strength of these connections and their total effects on program schedules are subjects for further inquiries.

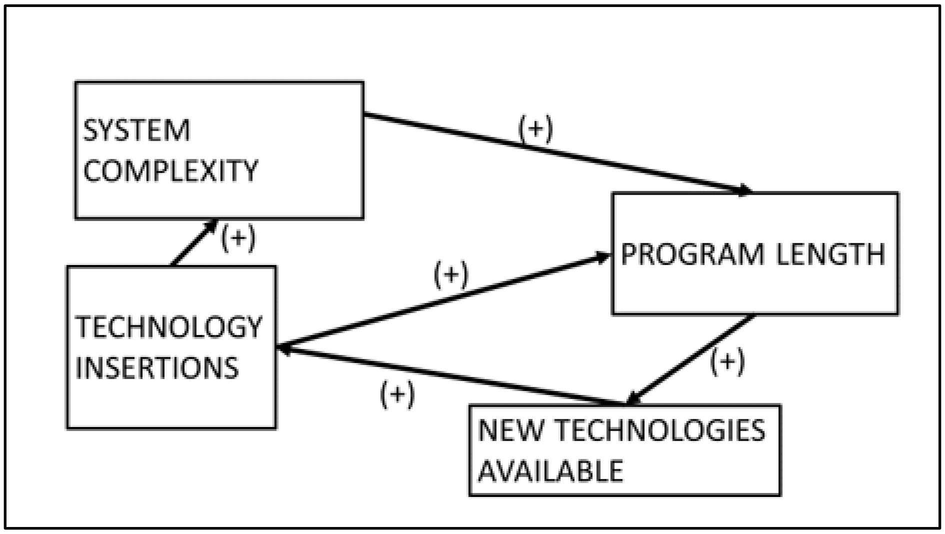


Figure 6. Self-Stretching Acquisition Program Schedules?

2. To what extent do weak schedule estimation and management tools affect program performance? There are excellent reasons to believe that scheduling estimates are sometimes not realistic. What schedule management tools do program managers and program offices lack? How can those gaps be addressed?

JCIDS Manual (CJCS, 2015, p. A-9) recommends tradeoffs among system cost, performance, and schedule. Program managers have reasonably good cost estimation tools, reasonably good indicators of system performance, but not good ways to estimate schedules—especially if the original program experiences requirements growth.

This question offers some scope for gap analysis—to be investigated through case studies and interviews with subject matter experts.

Measuring System Performance[7]

Updating combat system performance measures is important for at least two reasons. First, a better understanding of combat capability in the information age could significantly improve defense policy and planning. Second, a credible (preferably scalar) measure of combat capability could contribute much to schedule estimation—through better schedule-estimating relationships.

We've chosen to start with air-air combat systems. In a previous paper (Franck et al., 2017, esp. pp. 423–425), we explored a Lanchester aimed-fire model with various complications: stealth and command control, in addition to lethality and relative numbers. Results were interesting, but just a start.

Air combat in the near future will involve weapons committed from various types of platforms. Accordingly, we extend our previous model to include engagements "called in"

[7] This particular discussion has been abridged to fit the Proceedings' space limits. A more detailed version is available on request at cfranck215@aol.com.

from other platforms. One variant along this line is using non-stealthy aircraft (such as F-15s and B-52s) as weapons carriers whose targets are identified and assigned by fifth-generation aircraft—such as the F-35.

Accordingly, some aircraft ("scouts," S) will use their situational awareness capabilities to acquire hostile assets and then assign other aircraft ("weapons carriers," W) to engage them.[8] This assignment entails a useful networking capability. In this concept of operations, the shooter aircraft are primarily weapons carriers—and consumers of offboard sensors.

A Lanchester-Type Variant for Contemporary Air Combat[9]

Consider a stylized air battle scenario about one decade in the future:

> During the opening days, fighting focuses on the battle for air superiority as aircraft from both sides clash over contested territory. As the conflict continues, fifth-generation aircraft seek out, degrade, and destroy advanced SAMs in contested territory, creating a more moderate threat environment. This enables legacy aircraft to operate alongside their fifth-generation counterparts. (Harrigian & Marosko, 2016, pp. 7–8)

This suggests two major changes expected in the foreseeable future. The first is heterogeneous air combat forces: consisting of stealthy aircraft, plus a force of "legacy" combatants. The initially contested airspace contains stealthy fighters, with any older aircraft being quickly eliminated in that area (Barrett & Carpenter, 2017, esp. p. 5). However, those non-stealthy platforms can actively participate as weapons launchers whose fires are assigned by the fifth-generation aircraft acting as scouts—or other assets with command control capabilities.

Second, air combat will no longer be merely platform-on-platform engagements, but rather network-on-network, information-centered combat. One manifestation of this line of reasoning is the "kill web" concept, which features highly-networked forces with decentralized lethality and sensor capabilities, but most importantly decentralized decision making. Kill-web units take independent action, and are not "micro-managed" (Timperlake, 2017).

A related idea is the "combat cloud": "a model where information, data management, connectivity, and command and control (C2) are core mission priorities. The combat cloud treats every platform as a sensor, as well as an 'effector'" (Deptula, 2016, p. 1). In particular, operational decision making is spread throughout the network, with the "entire area of responsibility … functioning as a CAOC [Combined Air Operation Center]" (Deptula, 2016, p. 7).

---

[8] Given the F-35's limited internal weapons carriage, the role of finding hostiles and assigning others to engage is likely the primary role. We (Franck & Udis, 2016) have suggested "joint *scout* fighter" as a more descriptive name than "joint strike fighter."
[9] Taylor, Vol. II (1983, Section 6.13, pp. 318 ff.) provides a rigorous exposition of a starting point for our model.

The Data-To-Decision Problem

Kill webs and combat clouds are very promising. However, proficiency in network-based combat is a military advantage only to the extent it leads to better decisions than the enemy's (Gouré, 2018). Moreover, decentralized decision making is integral to the kill web and combat cloud concepts of operation. That makes achieving a reasonable degree of unity of effort a significant problem.

A simple example suffices to illustrate the point. Suppose there are two targets (A, B) of equal value, with associated (decentralized) decision makers (DA, DB). Suppose also there are two remotely-located weapons available for assignment (a, b), and that probability of kill varies with both weapon and target, as shown in Table 3.

Table 3. Simple Weapons Assignment Problem

WEAPONS	TARGETS	
	A	B
a	.9	.8
b	.7	.2

Clearly the optimal assignment is Weapon "a" to Target "B," and Weapon "b" to Target "A"—with 1.5 targets destroyed on average. However, structuring and solving such an assignment problem generally assumes a central authority with information that's both timely and sufficient.

In a *decentralized* decision mode, both Decision-Makers A & B (DA & DB) will note that Weapon "a" is better for his target. But there's only one Weapon "a." If DA happens to call in Weapon "a" first, then DB is stuck with Weapon "b," and targets destroyed declines to 1.1. (If DB calls in Weapon "a" first, then all is well, and 1.5 targets are destroyed on average.)

Timperlake (2017), in fact, proposes that the DA and DB simply ask which weapon is best for their target. That may or may not work out well. One ACC Commander, Gen Hawk Carlisle, posits (take everything from) "subsurface to on orbit," automatically piece it together, and "put it into the warfighter's hands in a way that … now they become the decision-makers" (as cited in Church, 2016).

Two comments: First, both perspectives assume a degree of situational awareness that goes beyond standard definitions such as "knowing real-time the current position, classification, condition and recent history of all items of military interest in both the physical and virtual battlespace" (Franck, 1995). Both Timperlake and Carlisle apparently assume that those decentralized decision makers also know target-weapons matchup characteristics well enough to make good choices among weapons available. This entails, *inter alia*, knowing plans (especially near future) of all relevant, friendly decision makers.

In short, the open literature indicates that translating shared situation awareness through a web of decentralized decision makers to produce a reasonable approximation of unity of effort is not yet completely understood, let alone solved. And the Air Force Air Superiority Flight Plan (U.S. Air Force Enterprise Capability Collaboration Team, 2016) apparently shares this opinion, recommending,

- a "data-to-decision campaign of experiments (to) examine how to fuse data from cloud-based sensor networks into decision quality information" (p. 7),
- "non-tradition concepts" for Battle Management Systems (p. 8), and

- development of new Command Control capabilities to provide “materiel and non-materiel solutions (that) should provide commanders in 2030 with the ability to synchronize forces across domains” (p. 8).

Assessing Air Combat Performance

Effective air combat forces are proficient in accomplishing the following tasks—which are generally accomplished sequentially, with accomplishment of all of them needed to ensure success:

- Cueing friendly forces of enemy activity (or early warning)—accomplished by assets with intelligence and reconnaissance capabilities
- Detecting, identifying, and tracking enemy forces—accomplished by surveillance systems
- Assigning forces to targets—command and control (C2) assets
- Engagement of targets—combat platforms and associated weapons
- Assessment of engagement results—surveillance systems

Timperlake (2017) essays a framework for contemporary combat capability called “payload utility.” This is the ability to acquire, engage, and destroy targets. What’s important for assessing near-future air combat is that the associated tasks are assigned to an entire network, with individual units calling on offboard resources within the network. For example, a target can be cued and detected by an early-warning sensor suite; identified and tracked by surveillance assets; be assigned to friendly forces by C2 assets; localized by a combat system—which engages the target through a weapon fired by yet another system; and then followed by an assessment of engagement results.

Simple models suggest that force sizes, weapons, stealth, and coordination are key variables in a credible measure of air combat capability. One look appears in Figure 7.

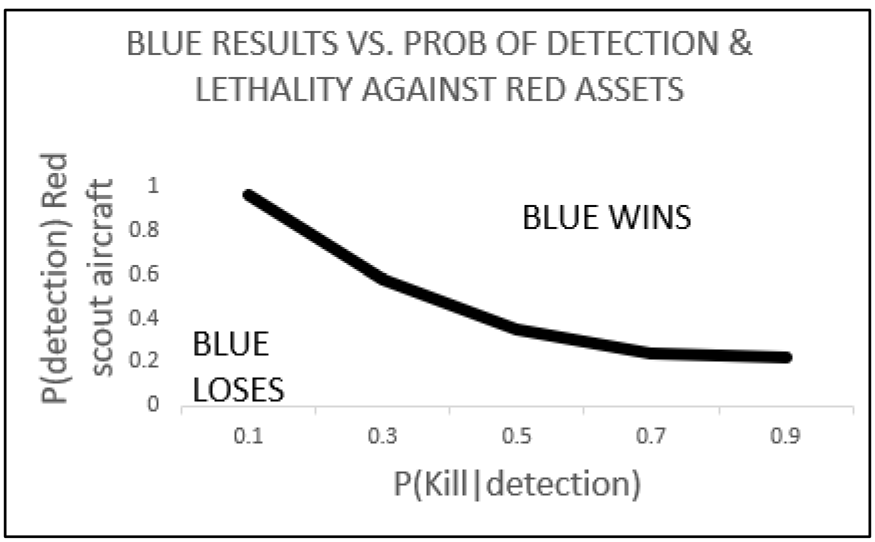


Figure 7. Combat Results VS. Detection and Lethality

The figure highlights important aspects of combat capability. First is the importance of “matchups” such as ISR versus stealth, and weapons versus targets’ countermeasures. Second is that (from a Blue perspective) it’s important to find Red targets (defeat Red stealth) and to have weapons that can defeat Red self-protection countermeasures. These aren’t terribly profound insights. However, they suggest we might be better off with fifth-gen weapons on fifth-gen airplanes than with fourth-gen weapons on sixth-gen airplanes. Finally, substitutions are indeed possible. For example, shortfalls in detection and tracking can be overcome with better weapons.

Next Steps

First, improve the model above to better account for the problem of coordinating a decentralized decision-making process. Among other things, we're hoping that the recommended Air Force studies and experiments along this line will include some unclassified results.

Second, try a more fine-grained approach to modeling future air combat. Agent-based simulation might be a useful method. This would, of course, be a major effort but with potentially major insights.

References

Abdel-Hamid, T. K. (1990). Investigating the cost/schedule trade-off in software development. *IEEE Software, 7*(1), 97–105. http://doi.org/10.1109/52.43055

Airpower Australia. (2016, March). Is the JSF really a fifth-generation fighter? Retrieved from http://www.ausairpower.net/jsf.html

Barrett, M., & Carpenter, M. (2017, July). *Survivability in the digital age: The imperative for stealth.* Arlington,VA: Mitchell Institute. Retrieved from http://docs.wixstatic.com/ugd/a2dd91_cd5494417b644d1fa7d7aacb9295324d.pdf

Blickstein, I. et al. (2011). *Root cause analyses of Nunn-McCurdy breaches* (Vol. I). Santa Monica, CA: RAND.

Bogdan, C. (2012, September 17). *Joint Strike Fighter requirements*. National Harbor, MD: AFA Air & Space Conference. Retrieved from http://www.af.mil/shared/media/document/AFD-120919-046.pdf

Brooks, F. P., Jr. (1995). *The mythical man-month: Essays on software engineering* (Anniversary Ed., 2/E). Addison-Wesley.

Capaccio, A. (2018, March 28). Air Force risks losing third of F-35s if upkeep costs aren't cut. Bloomberg News. Retrieved from https://www.bloomberg.com/news/articles/2018-03-28/air-force-risks-losing-third-of-f-35s-if-upkeep-costs-aren-t-cut

Christensen, C. M., & Bartman, T. (2016, Fall). The hard truth about business model innovation. *MIT Sloan Management Review*.

Church, A. (2016, February 26). Sensor to trigger in one challenging step. *Air Force Magazine Daily Report*. Retrieved from http://www.airforcemag.com/DRArchive/Pages/2016/February%202016/February%2026%202016/Sensor-to-Trigger-in-One-Challenging-Step.aspx

CJCS. (2015, January 23). *Joint Capabilities Integration and Development System (JCIDS)* (CJCSI 3170.01I). Retrieved from https://dap.dau.mil/policy/Documents/2015/CJCSI_3170_01I.pdf

Demeulemeester, E. L., & Herroelen, W. S. (2006). *Project scheduling*. Springer Science & Business Media.

Deptula, D. (2016, September). *Evolving technologies and warfare in the 21st century: Introducing the "combat cloud*" (Mitchell Institute Policy Papers, Vol. 4). Retrieved from http://docs.wixstatic.com/ugd/a2dd91_73faf7274e9c4e4ca605004dc6628a88.pdf

Director, Operational Test & Evaluation (DOTE). (2018). *FY17 programs, F-35 Joint Strike Fighter*. Retrieved from http://www.dote.osd.mil/pub/reports/FY2017/pdf/dod/2017f35jsf.pdf

Franck, R., Hildebrandt, G., Pickar, C., & Udis, B. (2017). Realistic acquisition schedule estimates: A follow-on inquiry. In *Proceedings of the 14th Annual Acquisition Research Symposium* (pp. 1–20). Monterey, CA: Acquisition Research Program.

Franck, R., Hildebrandt, G., & Udis, B. (2016). Toward realistic acquisition schedule estimates. In *Proceedings of the Seventh Annual Acquisition Research Symposium*. Monterey, CA: Acquisition Research Program.

Franck, R., & Udis, B. (2016, August 23). Why America shouldn't build sixth-generation manned fighters. *National Interest*. Retrieved from http://nationalinterest.org/blog/the-buzz/why-america-shouldnt-build-sixth-generation-manned-fighters-1744

GAO. (2015). *Acquisition reform: DOD should streamline its decision-making process for weapon systems to reduce inefficiencies* (GAO-15-192). Washington, DC: Author.

GAO. (2017, October). *DOD needs to address challenges affecting readiness and cost transparency* (GAO 18-75). Retrieved from https://www.gao.gov/assets/690/687982.pdf

Gouré, D. (2018, March 15). Getting the Navy's next generation network right. *Real Clear Defense.* Retrieved from https://www.realcleardefense.com/articles/2018/03/15/getting_the_navys_next_generation_network_right_113197.html

Godlewski, E., Lee, G., & Cooper, K. (2012). System dynamics transforms Fluor project and change management. *Interfaces, 42*(1), 17–32. http://doi.org/10.1287/inte.1110.0595

Harrigian, J., & Marosko, M. (2016, July). *Fifth generation air combat: Maintaining the Joint Force advantage* (Michell Forum No. 6). Mitchell Institute. Retrieved from http://docs.wixstatic.com/ugd/a2dd91_bd906e69631146079c4d082d0eda1d68.pdf

Hendrickson, C., Martinelli, D., & Rehak, D. (1987). Hierarchical rule-based activity duration estimation. *Journal of Construction Engineering and Management, 113*(2), 288–301. http://doi.org/10.1061/(ASCE)0733-9364(1987)113:2(288)

Herroelen, W. (2005). Project scheduling—Theory and practice. *Production and Operations Management.*

Howick, S. (2003). Using system dynamics to analyse disruption and delay in complex projects for litigation: Can the modelling purposes be met? *Journal of the Operational Research Society, 54*(3), 222–229. http://doi.org/10.1057/palgrave.jors.2601502

Hughes, R. T. (1996). Expert judgement as an estimating method. *Information and Software Technology, 38*(2), 67–75. http://doi.org/10.1016/0950-5849(95)01045-9

Jaafari, A. (1984). Criticism of CPM for project planning analysis. *Journal of Construction Engineering and Management, 110*(2), 222–233. http://doi.org/10.1061/(asce)0733-9364(1984)110:2(222)

Keijsper, G. (2007). *Joint Strike Fighter: Design and development of the international aircraft*. Pen & Sword.

Kreps, D. (1990). *A course in microeconomic theory*. Princeton, NJ: Princeton University Press.

Morris, P., & Hough, G. H. (1988). *The anatomy of major projects*. Hoboken, NJ; John Wiley & Sons.

PERT High Tech Projects. (2005). PERT High Tech Projects, 1–6.

Project Management Institute (PMI). (2017). *A guide to the project management body of knowledge (PMBOK® guide)—Sixth edition and Agile practice guide (English).* Newtown Square, PA: Author.

Schwindt, C., & Zimmerman, J. (2015). *Handbook on project management and scheduling* (Vol. 1). Springer.

Steidle, C. E. (1997, January–March). The Joint Strike Fighter program. *Johns Hopkins APL Technical Digest, 18*(1), 6–20. Retrieved from http://www.jhuapl.edu/techdigest/TD/td1801/steidle.pdf

Taylor, J. G. (1983). *Lanchester models of warfare* (Vol. II). Arlington, VA: Operations Research Society of America.

Timperlake, E. (2017, September 3). Shaping a way ahead to prepare for 21st century conflicts: Payload-utility capabilities and the kill web. Retrieved from Second Line of Defense website: http://www.sldinfo.com/shaping-a-way-ahead-to-prepare-for-21st-century-conflicts-payload-utility-capabilities-and-the-kill-web/

U.S. Air Force Enterprise Capability Collaboration Team. (2016, May). *Air superiority: 2030 Flight Plan.* Retrieved from http://www.airforcemag.com/DocumentFile/Documents/2016/Air%20Superiority%202030%20Flight%20Plan.pdf

Williams, T., Eden, C., Ackermann, F., & Tait, A. (1994). The vicious circles of parallelism. *International Journal of Project Management, 13*(3), 151–155.

Searching Hidden Links: Inferring Undisclosed Subcontractors From Public Contract Records and Employment Data

M. Eduard Tudoreanu—is Professor of Information Science at University of Arkansas Little Rock. Tudoreanu has expertise in human-computer interaction, information quality, advanced visualization of complex data, and virtual reality. He worked on visual data analysis and has extensive experience in software development and user interface design. Tudoreanu was the Founding Technical Director of the Emerging Analytics Center. He was the keynote speaker at ABSEL 2010 and served as a panelist for the National Science Foundation and Missouri EPSCoR. He earned his Doctor of Science degree in Computer Science in 2002 from Washington University in St. Louis. [metudoreanu@ualr.edu]

Keith Franklin—is a PhD candidate in Information Science at University of Arkansas Little Rock. He has over 20 years' experience in project management, IT, and Quality. He holds certifications in Project Manager (PMP), Information Security Manager (CISM), Risk and Information Systems Controls (CRISC), Quality Engineering (CQE), and Business Improvement (BIA). He has held management positions at Johnson & Johnson and SIMS Industries Medical, Pfizer (formerly Warner-Lambert division). He earned his Master of Business Administration from City University of Seattle and was honorably discharged from the U.S. Navy. [kxfranklin@ualr.edu]

Arnold Rego—is a PhD candidate in the Computer and Information Science program at the University of Arkansas Little Rock. He has expertise in data quality, data analysis, and data optimization. He earned his Bachelor of Science, Master of Business Administration, and Juris Doctorate degrees from the University of Arkansas Little Rock. [alrego@ualr.edu]

Ningning Wu—is Professor of Information Science at the University of Arkansas Little Rock. She received a Bachelor of Science and a Master of Science degree in Electrical Engineering from the University of Science and Technology of China and a PhD in Information Technology from George Mason University. Dr. Wu's research interests are data mining, network and information security, and information quality. She holds certificates of the IAIDQ Information Quality Certified Professional (IQCP) and the SANS GIAC Security Essentials Certified Professional. [nxwu@ualr.edu]

Richard Wang—is Director of the MIT Chief Data Officer and Information QualityProgram. He is also the Executive Director of the Institute for Chief Data Officers (iCDO) and Professor at the University of Arkansas Little Rock. From 2009 to 2011, Wang served as the Deputy Chief Data Officer and Chief Data Quality Officer of the U.S. Army. He received his PhD in information technology from the MIT Sloan School of Management in 1985. [rwang@mit.edu]

Abstract

Many prime contractors use subcontractors to meet the DoD requirements for Small Business Utilization, to incorporate specialized skills and to perform tasks that are not the prime contractors' core business. Some of these subcontractors are not explicitly made public for a variety of reasons, including security or competitive advantage. In such cases, because the involvement of these subcontractors in certain aspects of acquisition is not known, they can pose the risk of becoming a weak, stress point despite careful planning by the DoD. This paper examines whether such undisclosed contractors can be inferred through a combined analysis of both the published purchasing data (Federal Procurement Data Systems—Next Gen) and of additional pieces of information freely available on the web. In particular, this paper investigates employment data in an attempt to correlate changes in employment with negative modifications to contracts. It is possible that, when a contract is cut, the prime contractor may terminate a subcontract, resulting in layoffs of the subcontractor's employees. Employment numbers are published regularly and broken down to county and industry codes. This paper provides a preliminary analysis showing that drops

in employment occur when small contractors have relatively large reductions in DoD contracts.

Introduction

It is not uncommon for DoD contracts to involve multiple entities, and for some prime contractors to employ subcontractors to perform specific parts of the job. For a variety of reasons, including security, confidentiality, or competitive advantage, some of these subcontractors are not publicly disclosed. One large research question is whether the open society and transparent government of the USA provides the means to uncover some of these undisclosed subcontractors. A bounty of useful information is published by the local and federal government as well as private companies and individuals. While no single government source would likely allow such information to be determined, the combination of multiple data repositories might do just that, especially if the same undisclosed contractor is part of multiple contracts.

The work presented here is a preliminary exploration of whether a correlation exists between procurement data and employment information. The main contribution is examining whether reductions in contract amounts influence employment levels in certain industries and locations. While the longer term goal would be to find hidden contractors, the first step, and the topic of the paper, is to look at known contractors who suffered cuts and see if anything can be observed in (un)employment data around the time of the cut. The reasoning is that if employment data shows a reduction in most cases of contract reductions, then the chance of discovering undisclosed contractors would be high. Note that the discovery problem would likely not have an exact answer, but rather have a probabilistic one in the sense that certain locations might be found to have a high likelihood of being the home of an undisclosed contractor for a given type of industry, while others would have a negligible chance.

The preliminary data, as explored in this paper, reveals that reductions in contracts, which are large relative to the contractor's size, are correlated to a drop in employment in more than two thirds of the cases. The paper used two types of metrics, one linked to absolute employment levels for a county and industry type and the other, location quotient (2008), providing a relative measure of a county's employment in an industry sector relative to the nation. The second metric would be able to register a drop even if a particular location had an actual increase in employment, but that increase was not as large as the national trend. The location quotient showed a decline for 75% of the contract cuts studied.

The paper covers related work in the next section, followed by an explanation of the methodology used to obtain and process the data. Results, conclusions, and future work are the topics of the last two sections.

Related Work

One approach that used public acquisition databases combined with publicly available data is based on the Company ORganization and Firm name Unifier (CORFU) technique (Álvarez-Rodríguez et al., 2013). This is a stepwise (step-by-step) method that was used to reconcile corporate names in public contracts metadata. The research study used GeoNames REST Services for contextual information, and for filtering, the study used Google Refine to validate and reconcile the data. The corporate data was from multiple sources including databases from Forbes, Google Places, Google Maps, Foursquare, LinkedIn Companies, and Facebook. “This technique is applied to the ‘PublicSpending.net’ initiative to show how the unification of corporate names is the cornerstone to provide a visualization service that can serve policy-makers to detect and prevent upcoming necessities” (Llorens, Álvarez-Rodríguez, & Vafopoulos, 2015). In the study, there were over 40 million names that were extracted from public procurement datasets from Australia, the United States, the United Kingdom, and the CrocTail Project.

Another approach involved using nonprofits or commercial entities to create tools for finding and leveraging public data. Among the services that are provided were cross indexing with other data sources as well as data visualization tools. This project combined government data with information services from outside sources to provide context to the data. This technique is referred to as “mashups with other data sources” (Felten et al., 2009).

Search engines like Google are handicapped because they do not have access to agencies’ internal databases. Search engines have progressed, in spite of this limitation, and perform “a significantly better job at identify relevant information.” Even with this barrier, “private actors have repeatedly demonstrated that they are willing and able to build useful new tools and services on top of government data” (Felten et al., 2009) that provide better insight into the data.

Methodology

The data used in our analysis originated from Federal Procurement Data Systems—Next Gen (FPDS) and from the Bureau of Labor Statistics (BLS). The first source provides lists of federal contracts awarded during specified years, while the second publishes employment data. The data used in the research covered the period of 2013 through 2017.

Federal Procurement Data Systems (FPDS)

Our initial focus was to obtain information from FPDS on contracts for the Department of the Navy. The intention was to identify contracts with a funding reduction larger than at least a yearly salary of a qualified, skilled employee that was affecting small and medium-sized entities. From over 1.1 million entries pulled for the Navy, about 3,300 modifications to a contract were found that met the following criteria: i) the contract was reduced by more than $100,000, and ii) the vendor had between 10 and 200 employees. Both contract modification value and number of employees for the vendor are fields in FPDS.

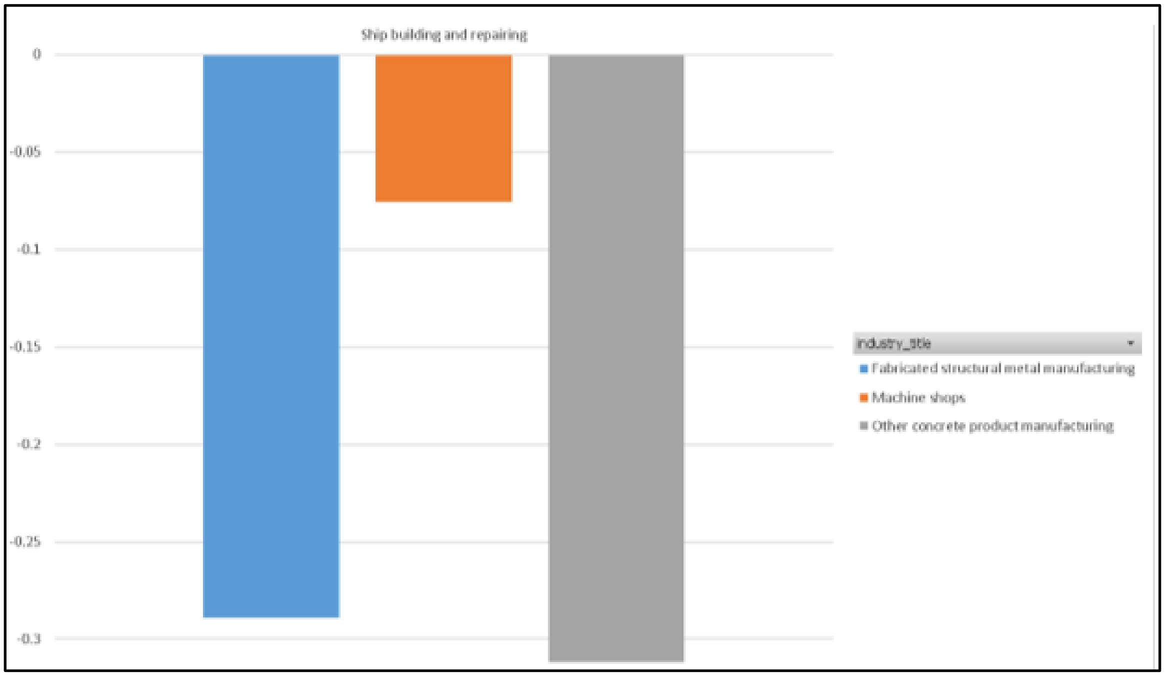
The correlation between FPDS and BLS data is in the process of being automated, thus for this work the analysis was performed manually. This limited our ability to process all contract modifications identified in the Navy data to a more manageable 33 instances of reductions in the amount of the contract, which amount to 1% of the Navy contract reductions for the studied period.

The selection focused on instances which had been the largest drop in contract amount per unit of annual revenue. We also restricted the industry to manufacturing (NAICS code beginning with 3) and IT services (NAICS code beginning with 581). Finally, to avoid some noise in the FPDS data, the selection only included entities whose revenue was larger than $100,000 (both NAICS code of the product and annual revenue of vendor are fields in FPDS).

Bureau of Labor Statistics (BLS) Employment Data

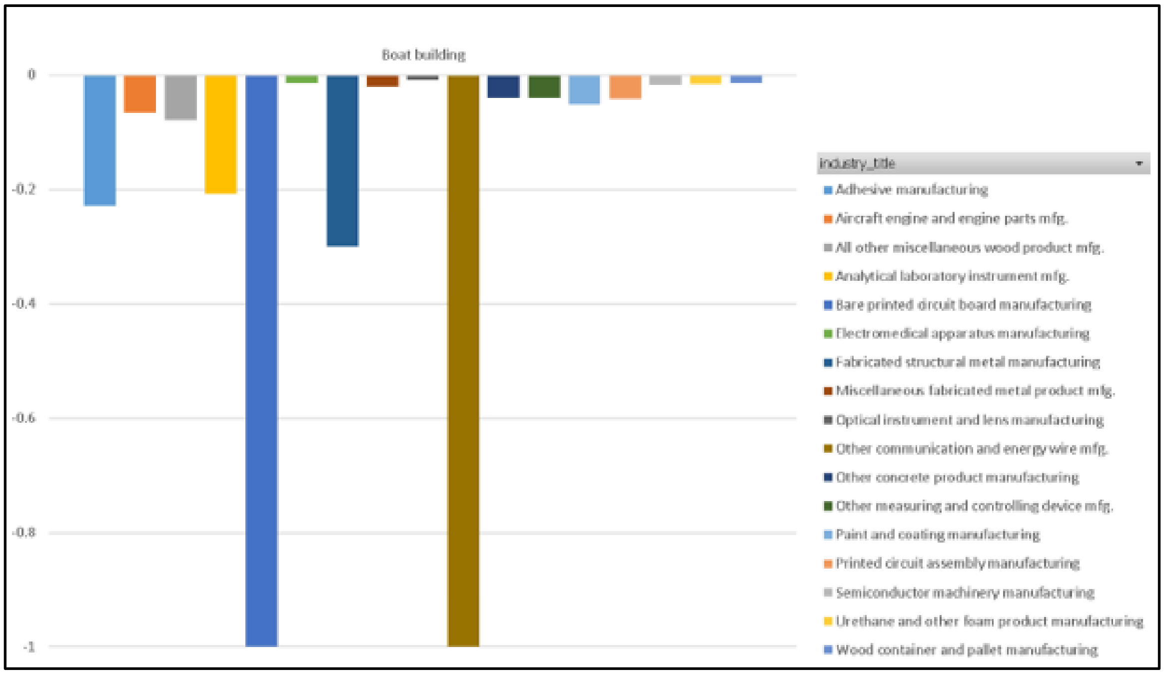
The employment data provided monthly absolute employment levels and monthly location quotients for each county and type of industry. The industry codes were also using NAICS (2017) to identify the type of employers. BLS uses the NAICS code slightly differently than FPDS in that BLS identifies the type of employer, while FPDS codes the type of product provided by an employer. It is possible that one employer may have products in different, yet related industry types.

Given that the single NAICS code existing in FPDS can be too rigid, a bundle of related industries were considered for any given location and date, as sampled in Figures 1 and 2. This approach is further supported by a secondary analysis we performed, which revealed that in about one third of counties (912 counties), a drop in employment in one industry was accompanied by a drop in related industries.



Note. The bars show relative change in employment from the quarter before to the next quarter.

Figure 1. Sample Industry-Type Bundle for Honolulu, HI, Third Quarter 2013



Note. The bars show relative change in employment from the quarter before to the next quarter.

Figure 2. Sample Industry Bundle for Essex, MA, Fourth Quarter 2014

Two metrics were used to estimate any potential changes in the employment situation: absolute employment level and location quotient (Bureau of Economic Analysis, 2008). The latter, location quotient (LQ), measures the local economy, county in our data, in relation to the rest of the nation. As such, it can capture cases where the absolute employment moves up or down in response to national trends unrelated to any local factors (local factors such as a cut in a Navy contract). A drop in LQ value is likely explained by a worsening of the local economy unrelated to national trends. Note that LQ can decrease even when employment increases, signifying that the increase is not taking place as quickly as in the rest of the USA.

Results

The 33 contracts studied in this paper are listed in Table 1 together with the two labor metrics of interest. This analysis took into account three months preceding and three months following the contract change to determine its impact. Averages were computed for both of the three month periods. The difference between those two averages, in percentage, is used for the rest of the analysis, both for absolute employment levels and location quotient.

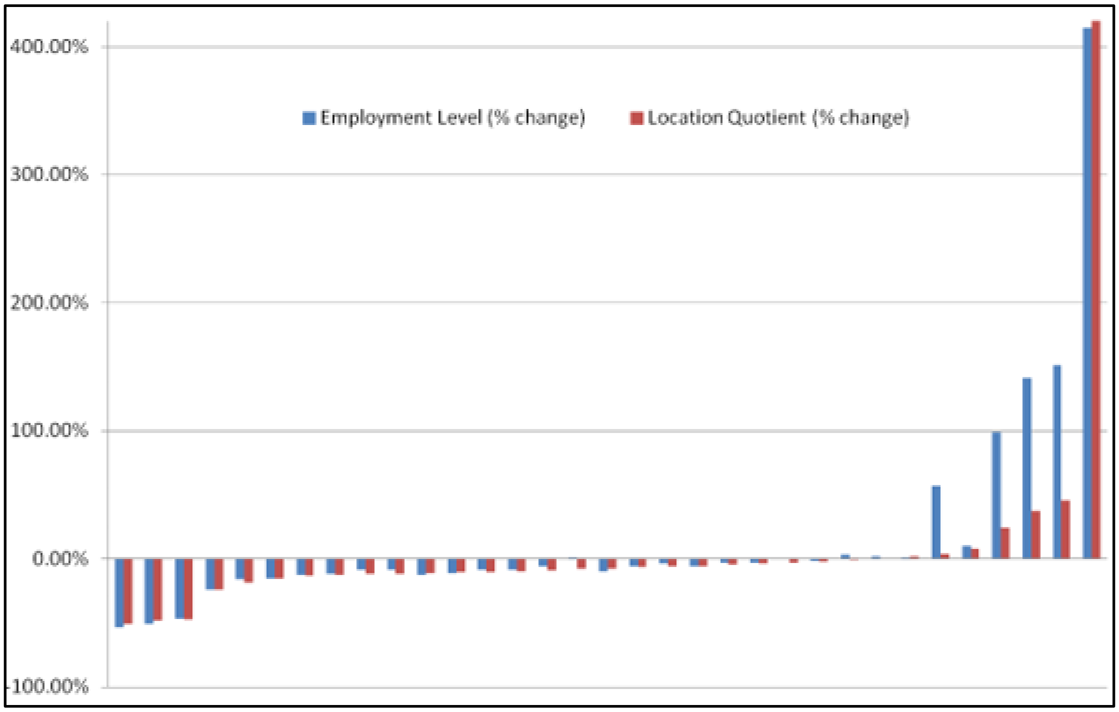
Table 1. List of 33 Instances of Contract Modifications

Name of Vendor / Cut Date	Percent change in employment level (the three months before the cut compared the three months after)	Percent change in employment location quotient (the three months before the cut compared the three months after)
AMC DEFENSE TECHNOLOGIES	**-2.25%**	**-2.89%**
9/10/2014	-2.69%	-3.47%
12/10/2014	-1.82%	-2.31%
ARNOLD DEFENSE AND ELECTRONICS	**-18.09%**	**-17.38%**
8/19/2015	-24.23%	-24.25%
7/15/2016	-12.63%	-11.27%
BARBER-NICHOLS INC.	**-15.52%**	**-15.16%**
9/28/2016	-15.52%	-15.16%
B-K MANUFACTURING CO.	**1.29%**	**1.72%**
6/3/2014	1.29%	1.72%
BOSTON SHIP REPAIR	**-53.26%**	**-51.10%**
12/5/2016	-53.26%	-51.10%
CD BIOSCIENCES INC	**57.38%**	**4.10%**
10/3/2014	57.38%	4.10%
CHAE&NAM UNIVERSE INC	**-5.67%**	**-5.75%**
8/23/2016	-5.67%	-5.75%
CORONET MACHINERY CORP.	**-12.10%**	**-12.68%**
9/25/2014	-12.10%	-12.68%
DUCWORKS	**10.05%**	**7.95%**
7/6/2016	10.05%	7.95%
EAG-LED	**-8.24%**	**-10.24%**
11/22/2016	-8.24%	-10.24%
FORM FIT AND FUNCTION	**-5.98%**	**-6.59%**
7/3/2013	-5.98%	-6.59%
G I INDUSTRIAL-MARINE	**414.68%**	**1173.90%**
9/16/2016	414.68%	1173.90%
G SYSTEMS LP	**1.03%**	**-8.07%**
9/23/2014	1.03%	-8.07%
H L TECHNOLOGIES	**2.08%**	**0.48%**
3/8/2013	2.08%	0.48%
LOGAN ENERGY CORP.	**-2.98%**	**-4.69%**
4/15/2015	-2.98%	-4.69%
MANTEC SERVICES INC.	**0.64%**	**-3.28%**
5/20/2013	0.64%	-3.28%
MOBILE TECHNICAL	**-46.62%**	**-47.53%**

SYSTEMS		
3/8/2016	-46.62%	-47.53%
OWL INTERNATIONAL INC.	**-8.38%**	**-12.04%**
8/8/2013	-8.38%	-12.04%
PACIFIC ENGINEERING INC.	**-16.25%**	**-18.94%**
4/21/2015	-16.25%	-18.94%
PELATRON	**-8.53%**	**-10.75%**
9/4/2013	-8.38%	-12.04%
5/6/2014	-8.67%	-9.57%
RIBCRAFT USA	**-11.98%**	**-11.88%**
11/5/2014	-11.43%	-10.61%
3/2/2017	-12.69%	-13.52%
SEA-LECT PLASTIC CORPORATION	**129.60%**	**35.31%**
11/20/2013	141.25%	37.46%
1/15/2015	98.89%	24.04%
1/14/2016	151.49%	45.60%
SUMMIT SOLUTIONS OF COLORADO	**-6.68%**	**-6.88%**
11/20/2014	-9.98%	-7.99%
3/26/2015	-3.39%	-5.77%
SWIFTSHIPS	**-51.04%**	**-47.73%**
8/30/2016	-51.04%	-47.73%
SYSTEMS APPLICATIONS&SOLUTIO	**3.39%**	**-1.04%**
5/12/2014	3.39%	-1.04%
WHITAKER TECHNOLOGIES	**-5.67%**	**-8.84%**
6/10/2014	-5.67%	-8.84%

Note. Aggregated data for each vendor as well as individual reductions by date are provided.

In a clear majority of the cases included in the study, a reduction in a contract is correlated with a drop in employment indicators, when comparing the average employment numbers the quarter (three months) before the reduction date with the quarter after. The absolute employment level was reduced in 22 out of 33 instances, or 66.6% of time. Similarly, the location quotient displays a drop in 25 out of 33 cases (75.7%). Figure 3 shows an overview of the two metrics. The data has a few outliers, especially on the positive side of the change, where the location quotient increased significantly in one instance.



Note. The x-axis lists all 33 cases of contract reduction and is ordered by location quotient. The right-most red bar is not shown in its entirety as it would reach up to almost 1,200%.

Figure 3. Percentage Change in Employment Level (Blue) and Location Quotient (Red) Before and After a Reduction in a Federal Contract

The study also examined the correlation between the magnitude of the reduction and the magnitude of the change in employment indicators, but in the limited cases (33) that were available, the correlation was relatively weak. Even when removing positive outliers (the right side of Figure 3), R^2 was about 0.3 for both employment level and quotient. For the revenue of a vendor to change in employment, the correlation was only marginally better ($R^2 \approx 0.4$).

Conclusion and Future Work

The paper presented a preliminary study that shows that large reductions in federal contracts are correlated in a majority of cases (66% or 75%, depending on the metric used) to drops in employment in a given region and industry. This finding shows that it is possible to determine the location of an undisclosed contractor by examining public employment data at the times when large contracts are reduced or simply reach the end of their period. Such undisclosed contractors are typically employed by larger government contractors to achieve confidentiality, security, or a competitive advantage. Depending on the situation, acquisition experts may need additional planning to protect such hidden contractors if security is desired, or may rely on data science to identify these contractors and avoid them becoming a weak link in the acquisition process.

The main contributions of the paper, besides the study itself, are the development of a framework for joining acquisition and employment data and the testing of the industry-type bundles for a given location and time frame. However, the main limitation of this study is the reliance on the quasi-manual joining of the acquisition (FPDS) and employment (BLS) data. As such, future work will focus on extending the database of DoD contracts and employment information and on automating the process of correlating industry types, locations, and dates. This will result in thousands more entries being analyzed, and the possibility of employing powerful statistical methods to better filter external factors (noise) from the analysis. Finally, in addition to studying reductions in contracts, we can analyze the effect that awarding new contracts or increasing the amount of existing contracts has on local job metrics.

References

Felten, E. W., Robinson, D., Yu, H., & Zeller, W. P. (2009). Government data and the invisible hand. *Yale Journal of Law and Technology*, 160.

Álvarez-Rodríguez, J. M ., de Pablos, P. O., Vafopoulos, M., & Labra-Gayo, J. E. (2013). Towards a stepwise method for unifying and reconciling corporate names in public contracts metadata: The CORFU technique. *7th Research Conference on Metadata and Semantics Research,* Thessaloniki, Greece. doi:10.1007/978-3-319-03437-9_31

Llorens, L., Álvarez-Rodríguez, J. M., & Vafopoulos, M. (2015). Enabling policy making processes by unifying and reconciling corporate names in public procurement data: The CORFU technique. *Computer Standards & Interfaces, 41*, 28–38.

Bureau of Economic Analysis. (2008). What are location quotients (LQs)? Retrieved from https://www.bea.gov/faq/index.cfm?faq_id=478

Office of Management and Budget (OMB). (2017*). North American Industry Classification System.* Retrieved from https://www.census.gov/eos/www/naics/2017NAICS/2017_NAICS_Manual.pdf

A Method for Identification, Representation, and Assessment of Complex System Pathologies in Acquisition Programs

Charles B. Keating[1]—is a Professor of Engineering Management and Systems Engineering and Director of the National Centers for System of Systems Engineering (NCSOSE) at Old Dominion University (ODU) in Norfolk, VA. He received a BS in Engineering from the United States Military Academy (West Point), an MA in Management from Central Michigan University, and a PhD in Engineering Management from ODU. His current research focus is on complex system governance, system of systems engineering, and management cybernetics. [ckeating@odu.edu]

Polinpapilinho F. Katina—is a Postdoctoral Researcher for the National Centers for System of Systems Engineering (Norfolk, VA). He serves as an Adjunct Assistant Professor in the Department of Engineering Management and Systems Engineering at Old Dominion University (Norfolk, VA) and is an Adjunct Assistant Professor in the Department of Engineering and Technology at Embry–Riddle Aeronautical University—the Worldwide campus. He received his PhD in the Department of Engineering Management and Systems Engineering at Old Dominion University. He received additional training from, among others, Politecnico di Milano (Milan, Italy). His profile includes more than 70 peer-reviewed papers in international journals, conferences, and books. He is a founding board member for the International Society for Systems Pathology (Claremont, CA). [pkatina@odu.edu]

Keith F. Joiner—received his PhD in Calculus Education, MS in Aerospace Systems Engineering through Loughborough University in the United Kingdom, and a Master of Management. He joined the Air Force in 1985 and became an aeronautical engineer, project manager, and teacher over a 30-year career before joining the University of New South Wales in 2015 as a senior lecturer in test and evaluation where he has since written 50 papers and articles linking research and practice. From 2010 to 2014, he was the Director-General of Test and Evaluation for the Australian Defence Force, where he was awarded a Conspicuous Service Cross. He is a Certified Practising Engineer and a Certified Practising Project Director. His fields of research include Aerospace Engineering, Six-Sigma Test Design, and Cybersecurity Testing and Education. [k.joiner@adfa.edu.au]

Joseph M. Bradley—received his PhD in Engineering Management and Systems Engineering at Old Dominion University (ODU) in Norfolk, VA. He holds the degrees of Professional Engineer and Master of Science in Mechanical Engineering from Naval Postgraduate School and Bachelor of Engineering from The Cooper Union. He is currently Principal Engineer at Patrona Corporation, a small defense consulting company headquartered in the Washington, DC, area, working with clients in government and industry. His areas of research include complex system governance, systems theory, competency models, and performance measurement systems. His research has been published in the *Systems Engineering*, *Naval Engineers Journal*, and *International Journal of System of Systems Engineering*. [josephbradley@leading-change.org]

Raed M. Jaradat—is an Assistant Professor of Industrial and Systems Engineering at Mississippi State University and a scholar in the U.S. Army. Dr. Jaradat received a PhD in Engineering Management and Systems Engineering from Old Dominion University in 2014. His main research interests include systems engineering and management systems, systems thinking and complex system exploration, system of systems, virtual reality and complex systems, systems simulation, risk, reliability, and vulnerability in critical infrastructures with applications to diverse fields ranging from the military to industry. His publications appeared in several ranking journals including the *IEEE Systems*

[1] Corresponding author

Journal and the *Computers & Industrial Engineering Journal*. His total awarded proposals exceed $2.9 million. [jaradat@ise.msstate.edu]

Abstract

Acquisition programs continue to struggle with increasing complexity. High degrees of emergence, interconnectedness, and uncertainty are the norm rather than exception. *The purpose of this research is to explore extension of ongoing research in complex system pathologies for acquisition programs*. Significant advances have been made in development of deeper understanding of the nature of pathologies (deviations from healthy system function) and their implications for performance of complex systems. Complex system pathologies represent “violations” of underlying system principles. These violations negatively affect system governance functions (control, oversight, accountability) resulting in degradation of system performance. Greater understanding of complex system pathologies offers insights to enhance complex system performance. This paper reports on the current state of development of a method to identify, represent, and assess systemic pathologies in complex systems. The method examined (M-Path Method) supports enhanced capabilities for pathology discovery, support for prioritization based on impact ranking, and provision of guidance for feasible strategic response across a spectrum of pathologies. Thus, the acquisition field and practitioners will benefit from results reporting on (1) acquisition field advancement through system science-based research into impediments to system performance, (2) providing a research-based method to improve acquisition program performance, and (3) reporting on successes and lessons learned from preliminary application of the method. The paper concludes with discussion of initial applications of the method, developmental areas, and guidance for acquisition practitioners.

Introduction

There seems to be a high level of agreement that acquisition of major systems continues to experience difficulties in meeting expectations under increasingly complex circumstances. There are plentiful accounts of “failures” of the acquisition system to produce on schedule, on cost, and on technical performance systems. This sentiment is echoed in the near constant criticisms from oversight bodies (e.g., Government Accountability Office) suggesting that there is much room for improvement in the acquisition field. There have been numerous attempts to explain the factors contributing to acquisition failures (Berteau et al., 2011; Francis, 2008, 2009; Rascona, Barkakati, & Solis, 2008; Smith et al., 2016). Unfortunately, without resolution, the demonstrable failures in acquisition programs and calls for reform persist. For example, Cilli et al. (2015) examined recent Government Accountability Office assessments of major acquisition programs, concluding that while attempts were being made to improve, the difficulties remain. Irrespective of a lack of reform success, efforts at Acquisition System reform continue (Bucci & Maine, 2013) and recognize the need to streamline the system and craft a more agile and flexible Acquisition System. Regardless of noble efforts and attempts to “improve” the acquisition system, the realistic conclusion persist that reforms have not had the desired impact. Instead, the continuing outward appearance of the acquisition system is that of a monolithic system. This system has not demonstrated that it is well suited for the complexity, speed, uncertainty, and ambiguity that exist in warfighting needs and environments characteristic of the 21st century.

There has been a continuing legion of reports, critiques, and calls for reform in the acquisition system. In fact, Fox (2012) was quick to point out that the calls for defense acquisition reform have been levied for decades and have continued to persist despite the continuing calls for change. There have been multiple corresponding investigations attempting to identify and explain the underlying causes contributing to unsuccessful acquisition efforts (Berteau et al., 2011; Francis, 2008, 2009; Rascona et al., 2008; Smith et al., 2016). Yet, the criticisms of the acquisition system persist and seem to be resilient to any of the remedies suggested to reform. The present state appears to be a system that, in the best case, appears to be severely debilitated. And, in the worst case is outright dysfunctional and "broken." Programs that can be offered as exemplars of successful acquisition endeavors seem to be a rarity. There is a short supply of successful exemplars of acquisition excellence, judged against usability, budget, and delivery schedule performance that meets or exceeds expectations. Successful acquisition endeavors are frequently studied in hopes that they will answer the riddle as to why acquisition programs so often fail and what might be done to improve the chances for success of future programs (Boudreau, 2007; O'Rourke, 2014). Presently, there is no satisfactory, or widely held consensus as to the path forward, much less the feasibility of successfully embracing that path. Looking to other countries for benchmarking and innovation in the hope that their smaller acquisition portfolios might provide a different vantage is an option. However extensive reviews like that of Joiner and Tutty (2018) comparing Australia's and the United States' defense acquisition systems, show that the U.S. initiatives provided at scale appear to deal better with complexity and that therefore the complexity problem for allies is worse.

There is not consensus on directions necessary for acquisition system development and reform. Characterizing reports critical of defense acquisition, Cilli et al. (2015) suggested, "In general, these reports call for early, robust, and rigorous assessments of a broader range of alternatives across a thorough set of stakeholder value criteria to include life-cycle costs, schedule, and performance" (p. 587). Given the present state of acquisition, this appears laudable, and possibly even to some extent infeasible. In a most recent publication, the Section 809 Panel (focused on making the DoD's acquisition system bold, simple, and efficient) January 2018 report provided recommendations such as marketplace framework, commercial buying, earned value management, and establishing of "offices" among other recommendations. The incorporation of these "recommendations" has yet to be established, much less their impact. Based on the present "disagreeable" state of the acquisition system, we must ponder the question, *"Why after over 40 years of acquisition reform do the critical performance issues not only persist in this field but seem to be worsening?"* In examination of a response, we suggest that, given the vexing nature of acquisition system problems, coupled with our inability to provide satisfactory reform, a different vantage point might provide insights. To this end, acquisition failure has not been rigorously examined from a systems theoretic perspective to formulate "systemic deficiencies" in the design, execution, and development of acquisition as a true "system."

There is much to be gained in deeper application of a systems theory perspective of acquisition to provide a different vantage point. To foster this perspective, the purpose of this paper is to *explore extension of ongoing research in complex system pathologies for acquisition programs.* There have been significant advances made in the development of pathologies (aberrations from normal or healthy system conditions) as a key to deeper understanding sources of failure for complex systems (e.g., acquisition). For this exploration we have focused on achievement of four primary objectives, including: (1) identify a perspective and method for identification of pathologies in a complex system, (2) suggest implications for pathology analysis for acquisition program development, (3) present a demonstration of results from pathology identification and assessment for acquisition programs, and (4) suggest implications for further development and deployment of pathologies as potential sources to advance acquisition system development. To achieve these objectives, the paper is organized to first suggest a systems explanation for current difficulties being experienced in the acquisition system. Second, as a deeper systems elaboration of systemic sources of failure in complex systems, we elaborate the concept of system pathologies. This elaboration anchors the notion of complex system failure sources to violations of underlying systems principles fundamental to the design, execution, and development of all complex systems. Third, a method for the discovery and characterization of systems pathologies is developed. A short demonstration of the implications of the method for deployment in the acquisition system is presented. The paper closes by suggesting a direction for further development and elaboration of pathologies for complex systems in general, and implications for the acquisition system in particular.

A Systems Perspective of Challenges in the Acquisition System

To begin a dialog on the systems formulation of acquisition difficulties, we posit six central themes (Figure 1), consistent with and extending several earlier works (Keating et al., 2017a, 2017b; Bradley, Katina, & Keating, 2016). These considerations provide a systemic frame of reference (views) for the modern landscape of defense acquisition. While these characteristics are endemic to modern systems in general, the particular emphasis of the Defense Acquisition System is intended to invite a different level of dialog, exploration, and "systemic" understanding.

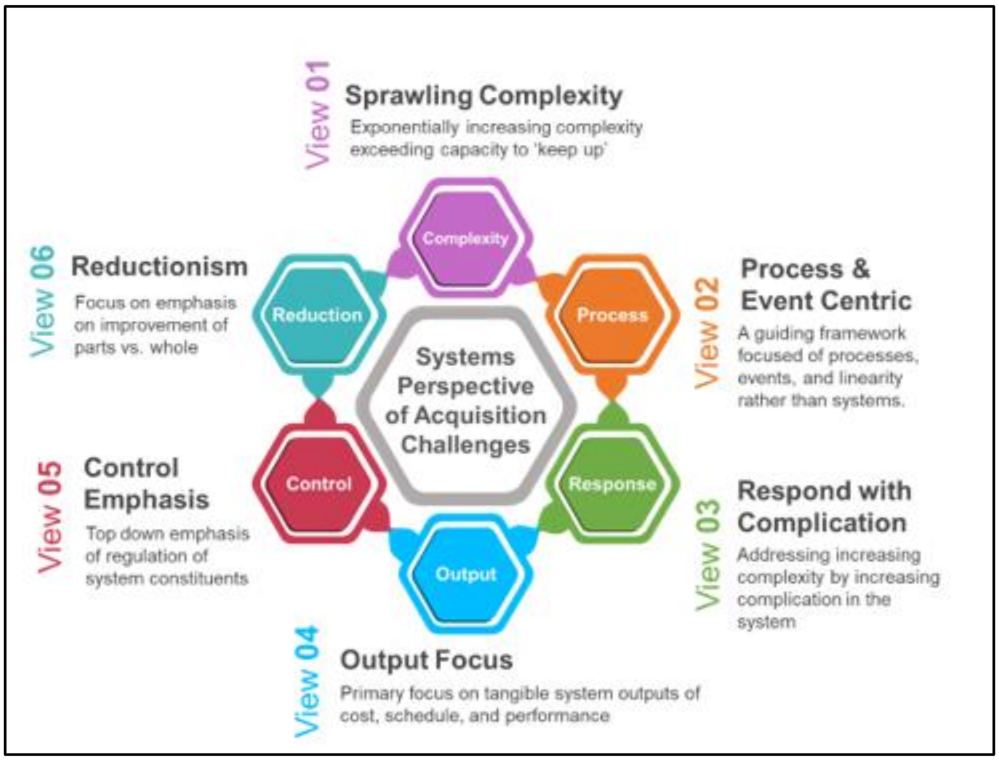


Figure 1. Six Systems Perspectives (Views) for the Defense Acquisition Field

1. *Sprawling Complexity Exceeds Capacity of the System to Absorb*. As the complexity of systems being acquired have increased exponentially, so too has the acquisition system which procures them. Such system elaborations are expected. However, lacking integrated, purposeful, and accountable management of increasing complexity, leaves the acquisition system behind and lacking capacity to "match" the complexity in the system(s) they are charged to acquire. For the Defense Acquisition System this suggests that the calls for reform, increased agility, boldness, simplicity, and efficiency, as

well as other such suggestions by numerous authors, are perhaps summed up in Kendall's (2014) congressional testimony stating, "Our system over time accumulated excessive levels of complex regulatory requirements that are imposed on our program managers and other acquisition professionals. … One thing I hope we can all agree on is the need to simplify and rationalize the bureaucratic burdens we place on our acquisition professionals" (p. 6). Similarly, Australia's First Principles Review (Peever et al., 2015) found, "Acquisition teams must comply with over 10,000 Defence Materiel Organisation specific policies and procedures which includes 35 policy and procedure artefacts totaling around 12,500 pages on procurement processes and controls" (p. 14). Despite these intricate processes, that review went on to find that acquisition had "great difficulty measuring and monitoring real performance" and that there was "a disconnect between customers and the purchaser as well as multiple and unnecessary handover points which increase complexity and risk."

2. *Process and Event Centric Focus*. The Defense Acquisition System is proclaimed to be a "system" without further qualification. However, facing the most rudimentary articulation for classification as a system (e.g., defined boundary, entities, transformation, outputs/outcomes, etc.), the Acquisition System falls short. In reality, the "Acquisition System" is a collection of elements for which the precise representation is not presented, operated, or evolved holistically as a system. The DoD 5000 reference to the Defense Acquisition Management System as both a "framework" as well as an "event-based process" supports this conclusion, as the "system" falls short of the notion of system from the most basic articulations of a system in the literature. The ramifications of a focus limited to process and event focus include missing development opportunities from a systems perspective (versus process or event perspective). The intent of pointing out this systems perspective is not to discredit defense acquisition. On the contrary, by inviting a more "systemic" perspective of the Acquisition System, the potential for future shifts in design, analysis, and development might become available.
3. *Response to Increasing Complexity Relegated to Increasing Complication.* The original intent of the Federal Acquisition Regulation (FAR) was quite straightforward, as it attempted to provide an efficient approach to the acquisition of material necessary to support government functions. However, since the introduction of the FAR, it has continued to elaborate in structure, volume, and become increasingly complicated (having many parts and pieces). New regulations, extended processes, and implementation of new controls have all acted to make the FAR much more complicated than the original document. As an example of this increasing complication, Federal Acquisition Regulation expanded from 1,953 pages at introduction in 1984 to 2,193 pages by 2014, with the DFAR supplement adding another 1,554 pages and each of the services initiating a host of their own "specialized" implementation guides, instructions, directives, and memorandums (Friar, 2015). This "explosion" has served to make the FAR more complicated, as deficiencies and calls for reform have continued, and arguably, escalated.
4. *Emphasis on an Output Versus Outcome Focus.* The Defense Acquisition System is clearly focused on achievement of outputs. Outputs are those tangible, verifiable, and objective elements that serve as products of a system and provide value consumed external to the system. Outcomes are the not

necessarily tangible effects of a system. These effects are more related to the fulfillment of purpose/need, not easily verifiable, subjective in nature, and primarily focused on meeting expectations for problem resolution. As such, outcomes exist as related, but removed in nature and scope, from outputs achieved from a system (Salado & Nilchiani, 2017). The "iron triangle" for acquisition programs has been, and remains, focused on the outputs of cost, schedule, and performance. It is hard to read a criticism of the current state of affairs for acquisition that is not targeted to one or more of these tangibly measurable (output) elements. However, we suggest that these indicators are "systemically" limited in their ability to capture the true indication of performance in acquisition. While these indicators (cost, schedule, performance) are necessary indicators of system performance, they alone do not provide sufficiency as a set of judgments of Acquisition System performance. For example, Cilli et al. (2015) point out the sunk costs of five programs between 2006 and 2011 in excess of $32 billion. While this "failure" is easily marked from the cost metric, it is disingenuous to capture the essence of program failures only on the cost dimension. Deficiencies that permitted those failures might be found beyond the cost, schedule, and technical performance triad. This invites consideration of a much wider view for examination of acquisition reform, beyond the simple cost, schedule, and performance triad.

5. *Global Control*. From a systems perspective, control is about providing constraint of a system only to the degree to which is necessary to assure continued performance (Keating et al. 2014). Excess constraint in a system (control) wastes resources and limits local autonomy (independence of decision, action, and interpretation). Thus, for acquisition, the less control invoked makes for a more cost-efficient system—since constraint is not free and escalates costs of a system. Acquisition programs aptly refer to these excessive controls with such terms as overregulation, bureaucracy, and excessive constraint without evidence of commensurate value added to the system. The near constant state of acquisition reform (Fox, 2012; Schwartz, 2014) supports the increasing elaboration of system controls in ways that do not necessarily enhance performance. This does not demean the improvements achieved, or those suggested (e.g., Panel 809 recommendations) in reform processes, but instead suggests that a different (systemic) viewpoint of control might shift the landscape for acquisition program design, execution, and development.
6. *Reductionism as a Driving Paradigm*. At the basic level reductionism is understanding systems by ever deepening "reduction" to more finite components whereby system performance is held in understanding component level behaviors. For acquisition, this frame of reference is evident in the development of the "acquisition system" as a fragmented assemblage of processes, procedures, regulations, and standards. It is not a large leap to surmise that the present acquisition system has emerged through a series of well-intentioned additions over time. One would be hard pressed to claim that the current acquisition system was either purpose built as a whole, or currently performs as a unity. In contrast to reductionism is holism, a central tenet of Systems Theory. Holism is focused the central notion that understanding of system performance or behavior is achieved by understanding the interactions among components, rather than the

components themselves. This shift in thinking paradigm lies at the center of a more holistic, versus reductionist, perspective of acquisition. Arguably, the acquisition system has been developed and evolved from a reductionist perspective. This has placed primary emphasis on the constituent element development (e.g., processes, laws, offices, procedures, regulations) as opposed to the interactions of those elements to understand system performance and drive reform.

This systems perspective for the Defense Acquisition System is intended to suggest that a different frame of reference might be helpful. Our intention is to invite a dialog to further exploration and understanding of the current system, while offering insights into issues in design, execution, and development of the system from an alternative frame of reference. For our purposes, the alternative frame of reference is focusing on understanding system difficulties through discovery of underlying pathologies (aberrations from "healthy" functioning of a system). To achieve our purpose, the remainder of the paper is organized around four primary objectives. First, in the following section, we provide a grounding perspective of "system pathology" in relationship to complex system dysfunctions. Second, we elaborate a method for discovery of system pathologies in complex systems. This method, the M-Path Method (Metasystem Pathologies Method), suggests that the source of system dysfunction might be found in the metasystem (higher level integrating functions) that ultimately produce system behavior/performance. Third, we examine a preliminary application of the explanation of acquisition difficulties based on the perspective provided by pathologies. Fourth, in the final section of the paper, we conclude with implications for further research, contributions, and application development of pathologies for a different set of insights to support Acquisition System development.

Pathologies as a Source of Dysfunction in Complex Systems

Certainly, understanding of system performance, including acquisition, involves discovery of conditions that might act to limit that (i.e., acquisition system) performance. Previous research related to systems theory and systems theory-based methodologies offers insights that provide explanation for aberrant conditions affecting performance (Keating & Katina, 2012). These aberrant conditions have been labeled as pathologies, defined as "a circumstance, condition, factor, or pattern that acts to limit system performance, or lessen system viability <existence>, such that the likelihood of a system achieving performance expectations is reduced" (Keating & Katina, 2012, p. 214). Pathologies have a rich development and have been anchored in Systems Theory (the set of laws and principles that govern behavior of all complex systems) and Management Cybernetics (the science of system structural organization).

For grounding our present exploration, we introduce two key points related to the nature and role of pathologies in complex systems—pathologies and their relationship to Systems Theory. First, pathologies have been extensively developed for application to the design, execution and development (governance) of complex systems (Keating & Katina, 2012; Katina, 2015). Complex System Governance (CSG) provides a set of "coordinates" to locate the existence of a pathology. This location is identified to nine different functions essential to continued viability of a complex system. For succinctness, Table 1, drawn from the work of Katina (2016) presents a summary of the nine essential metasystem functions of a complex system. The "metasystem" acts to provide governance (design, oversight, accountability) of a complex system (following Keating & Bradley, 2015; Keating et al., 2017) through the following:

- **Control:** constraints necessary to ensure consistent performance and future system trajectory.
- **Communications:** flow and processing of information necessary to support consistent decision, action, and interpretation throughout the system.
- **Coordination:** providing for effective interaction to prevent unnecessary instabilities within and in relationship to entities external to the system.
- **Integration:** maintaining system unity through common goals, designed accountability, and maintaining balance between system and constituent interests.

Table 1. Metasystem Functions for a Complex System

Metasystem Function	***Primary Role of the Function***
Metasystem five (M5): Policy and identity	To provide direction, oversight, accountability, and evolution of the system. Focus includes policy, mission, vision, strategic direction, performance, and accountability for the system such that (1) the system maintains viability, (2) identity is preserved, and (3) the system is effectively projected both internally and externally.
Metasystem Five Star (M5): System context*	To monitor the system context (i.e., the circumstances, factors, conditions, or patterns that enable and constrain the system).
Metasystem Five Prime (M5'): Strategic system monitoring	To monitor measures for strategic system performance and identify variance requiring metasystem level response. Particular emphasis is on variability that may impact future system viability. Maintains system context.
Metasystem Four (M4): System development	To provide for the analysis and interpretation of the implications and potential impacts of trends, patterns, and precipitating events in the environment. Develops future scenarios, design alternatives, and future focused planning to position the system for future viability.
Metasystem Four Star (M4): Learning and transformation*	To provide for identification and analysis of metasystem design errors (second order learning) and suggest design modifications and transformation planning for the system.
Metasystem Four Prime (M4'): Environmental scanning	To provide the design and execution of scanning for the system environment. Focus is on patterns, trends, threats, events, and opportunities for the system.
Metasystem Three (M3): System operations	To maintain operational performance control through the implementation of policy, resource allocation, and design for accountability.
Metasystem Three Star (M3): Operational performance*	To monitor measures for operational performance and identify variance in system performance requiring system level response. Particular emphasis is on variability and performance trends that may impact system viability.
Metasystem Two (M2): Information and communications	To enable system stability by designing and implementing architecture for information flow, coordination, transduction, and communications within and between the metasystem, the environment, and the systems being governed.

A second essential and fundamental grounding for development of pathologies is their linkage to Systems Theory–based laws/principles. For our present purposes, the nature of pathologies in complex systems can be captured in the following critical points and their suggested relevance to acquisition practitioners and system development:

1. **All systems are subject to the laws of systems.** Just as there are laws governing the nature of matter and energy (e.g., physics law of gravity), so too are our systems subject to laws. These system laws are always there, always on, non-negotiable, non-biased, and explain system performance. Acquisition practitioners must ask, “Do we understand systems laws and their impact on our system(s) design and performance?”
2. **All systems perform essential system functions that determine system performance.** These functions are performed by all systems, regardless of sector, size, or purpose. These functions define what must be achieved for maintaining viability of a system. Every system invokes a set of unique implementing mechanisms (means of achieving system functions) that

determine how system functions are accomplished. Mechanisms can be formal-informal, tacit-explicit, routine-sporadic, or limited-comprehensive in nature. These functions serve to produce system performance which is a function of previously discussed communication, control, integration, and coordination. Acquisition practitioners must ask, “Do we understand how our system performs essential system functions to produce performance and maintain viability?”

3. **Violations of systems laws/principles in design, execution, or development of a system carry consequences.** Irrespective of noble intentions, ignorance, or willful disregard, violation of system laws carries real consequences for system performance. In the best case, violations degrade performance. In the worst case, violations can escalate to cause catastrophic consequences or even eventual system collapse. Acquisition practitioners must ask, “Do we understand problematic system performance in terms of violations of fundamental system laws?”
4. **System performance can be enhanced through development of essential system functions.** When system performance fails to meet expectations, deficiencies in governance functions can offer novel insights into the deeper sources of failure. Performance issues can be traced to governance function issues as well as violations of underlying system laws. Thus, system development can proceed in a more informed and purposeful mode. Acquisition practitioners must ask, “How might the roots of problematic performance be found in deeper system issues and violations of system laws, suggesting different development directions?”

Given this brief introduction to pathologies in complex systems, following the recent work of Katina (2016) and earlier work of Keating and Katina (2012) a set of 53 pathologies have been develop in relationship to the metasystem functions provided earlier (Table 2). These pathologies are organized around the nine metasystem functions and serve to identify aberrations to normal (healthy) functioning of a complex system (e.g., acquisition).

Table 2. Pathologies Corresponding to Metasystem Functions

Metasystem Function	*Corresponding Set of Pathologies*
Metasystem five (M5): Policy and identity	M5.1. Identity of system is ambiguous and does not effectively generate consistency system decision, action, and interpretation.
	M5.2. System vision, purpose, mission, or values remain unarticulated, or articulated but not embedded in the execution of the system.
	M5.3. Balance between short-term operational focus and long-term strategic focus is unexplored.
	M5.4. Strategic focus lacks sufficient clarity to direct consistent system development.
	M5.5. System identity is not routinely assessed, maintained, or questioned for continuing ability to guide consistency in system decision and action.
	M5.6. External system projection is not effectively performed.
Metasystem Five Star (M5*): System context	M5*.1. Incompatible metasystem context constraining system performance.
	M5*.2. Lack of articulation and representation of metasystem context.
	M5*.3. Lack of consideration of context in metasystem decisions and actions.
Metasystem Five Prime (M5'): Strategic system monitoring	M5'.1. Lack of strategic system monitoring.
	M5'.2. Inadequate processing of strategic monitoring results.
	M5'.3. Lack of strategic system performance indicators.
Metasystem Four (M4): System development	M4.1. Lack of forums to foster system development and transformation.
	M4.2. Inadequate interpretation and processing of results of environmental scanning—non-existent, sporadic, limited.
	M4.3. Ineffective processing and dissemination of environmental scanning results.
	M4.4. Long-range strategic development is sacrificed for management of day-to-day operations—limited time devoted to strategic analysis.
	M4.5. Strategic planning/thinking focuses on operational level planning and improvement.
Metasystem Four Star (M4*): Learning and transformation	M4*.1. Limited learning achieved related to environmental shifts.
	M4*.2. Integrated strategic transformation not conducted, limited, or ineffective.
	M4*.3. Lack of design for system learning—informal, non-existent, or ineffective.
	M4*.4. Absence of system representative models—present and future.
Metasystem Four Prime (M4'): Environmental scanning	M4'.1. Lack of effective scanning mechanisms.
	M4'.2. Inappropriate targeting/undirected environmental scanning.
	M4'.3. Scanning frequency not appropriate for rate of environmental shifts.
	M4'.4. System lacks enough control over variety generated by the environment.
	M4'.5. Lack of current model of system environment.
Metasystem Three (M3): System operations	M3.1. Imbalance between autonomy of productive elements and integration of whole system.
	M3.2. Shifts in resources without corresponding shifts in accountability/shifts in accountability without corresponding shifts in resources.
	M3.3. Mismatch between resource and productivity expectations.
	M3.4. Lack of clarity for responsibility, expectations, and accountability for performance.
	M3.5. Operational planning frequently pre-empted by emergent crises.
	M3.6. Inappropriate balance between short-term operational versus long-term strategic focus.
	M3.7. Lack of clarity of operational direction for productive entities (i.e., subsystems).
	M3.8. Difficulty in managing integration of system productive entities (i.e., subsystems).
	M3.9. Slow to anticipate, identify, and respond to environmental shifts.
Metasystem Three Star (M3*): Operational performance	M3*.1. Limited accessibility to data necessary to monitor performance.
	M3*.2. System-level operational performance indicators are absent, limited, or ineffective.
	M3*.3. Absence of monitoring for system and subsystem level performance.
	M3*.4. Lack of analysis for performance variability or emergent deviations from expected performance levels—the meaning of deviations.
	M3*.5. Performance auditing is non-existent, limited in nature, or restricted mainly to troubleshooting emergent issues.

	M3*.6. Periodic examination of system performance largely unorganized and informal in nature.
	M3*.7. Limited system learning based on performance assessments.
Metasystem Two (M2): Information and communications	M2.1. Unresolved coordination issues within the system.
	M2.2. Excess redundancies in system resulting in inconsistency and inefficient utilization of resources—including information.
	M2.3. System integration issues stemming from excessive entity isolation or fragmentation.
	M2.4. System conflict stemming from unilateral decisions and actions.
	M2.5. Excessive level of emergent crises—associated with information transmission, communication, and coordination within the system.
	M2.6. Weak or ineffective communications systems among system entities (i.e., subsystems).
	M2.7. Lack of standardized methods (i.e., procedures, tools, and techniques) for routine system level activities.
	M2.8. Overutilization of standardized methods (i.e., procedures, tools, and techniques) where they should be customized.
	M2.9. Overly ad-hoc system coordination versus purposeful design.
	M2.10. Difficulty in accomplishing cross-system functions requiring integration or standardization.
	M2.11. Introduction of uncoordinated system changes resulting in excessive oscillation.

A Method for Discovery of Pathologies in Complex Systems

Katina (2016a) has developed a method for deploying pathologies in complex systems, entitled the M-Path Method (Metasystem Pathology Method). This method extends previous research related to problem formulation (Katina & Keating, 2014; Katina, 2015; 2016a; 2016b) and complex system governance pathologies (Katina & Keating, 2014; Keating & Katina, 2012). With respect to application of the M-Path Method, there are three qualifications necessary. First, the pathologies are of a generalized form. Therefore, their manifestation in "different" complex systems may be evidenced by surface level "symptomatic" conditions. The pathologies are rooted in the underlying dysfunctions of a system that produce "observable" surface symptomatic conditions. Thus, pathologies are not directly observable, but rather are inferred from observable/demonstrable conditions in a system. Second, pathologies have a degree of existence. They are not binary reducible, and thus have a "degree of existence," rather than a binary present/not present attribution. Third, pathologies represent "deficiencies" in the system design (structural organization of a system to achieve desired behavior/performance), execution (performance of the system design), or development (evolution of the design and design/execution interface). As such, pathologies produce real consequences related to system performance which can be measured across a range of possible impacts for a system. Given this essential grounding, we present the M-Path Method in five phases (Figure 2, adapted from Katina, 2016a).

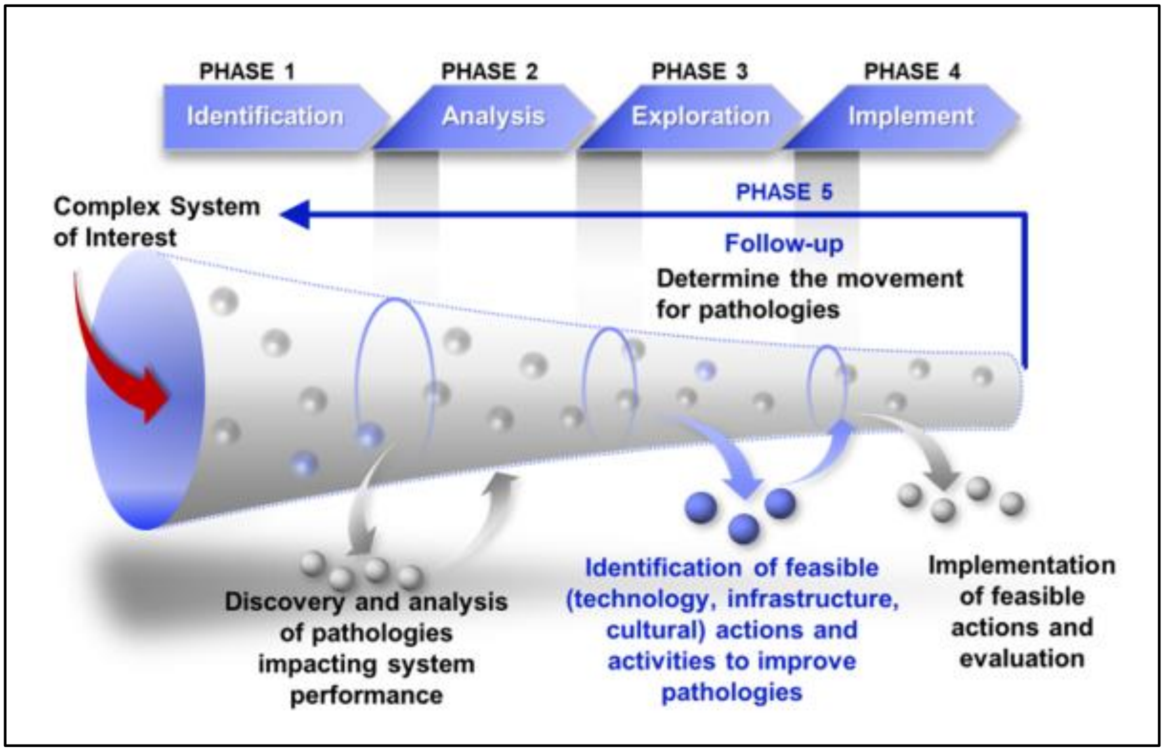


Figure 2. Phases of the M-Path Method for Pathologies

In essence, the M-Path Method is an approach that guides identification and assessment of specific pathologies that exist in a complex system. While not presented as complete or absolute, Table 2 was presented as the current state of pathology development for complex systems. This includes the 53 unique possible metasystem pathologies that can be assessed with respect to varying degrees of *existence, impact*, and *feasibility* to resolve. For conciseness we provide the following summary of the five phases of M-Path and their implications for pathology development.

Phase 1: Identification involves the identification and discovery of the degree to which the 53 pathologies exist for a given situation/system in focus. A determination is made with respect to the perceived existence and impact of pathologies. This determination is from the perspective of those practitioners who must design, operate, maintain, and develop the complex system.

This phase produces three essential pieces of information related to each of the pathologies: degree of existence, impact on system performance, and feasibility of addressing the pathology. *Degree of existence* is the level to which the pathology is deemed to be present—ranging from negligible to extreme. *Impact* is offers an assessment as to the degree to which the existence of the pathology is influential in affecting system performance. *Feasibility* captures the degree to which the pathology might be addressed (e.g., limited by technology, infrastructure, resources, safety, legality, authority, culture, etc.) with a reasonable confidence of success. Previous research has used a Web-based instrument (e.g., see Katina, 2016a) to capture this information. The data is represented in aggregate form (centroid of cluster for each pathology) and summary statistics (variance from cluster centroid). This provides input to a set of representations of the perspectives and their variability for the pathologies in a complex system of interest. Thus, support for further pathology analysis and exploration is supported by the process of "binning" pathologies based on levels of existence, their potential impact, and the feasibility of resolution.

Phase 2: Analysis examines the nature and implications of the unique landscape of pathologies for the system being examined. This phase is targeted to exploration of the specific implications each pathology holds for the particular system of interest. Pathologies do not have the same relevance, impact, or feasibility for resolution in a complex system.

The first phase only identifies the presence and impact of the 53 metasystem pathologies. The second phase involves an examination of nature and implications of the unique landscape of pathologies for the system interest. Driven by the kind of tools used in data collection of Phase I, the analyst collects and synthetizes the data into meaningful representations for further exploration of the pathologies. Invariably, this phase provides a deeper reflection on the pathologies identified from Phase 1. The following caveats apply to this phase:

- This analysis should *enumerate metasystem pathologies using measures of existence and impact,* and permit dialog concerning different notional evidence and support for the capture of the centroid of the pathology.
- The analysis should also indicate *variability in measures of degree and impact* as suggested by participants—in this case as taken from survey data. It is expected that each participant will not provide the same measure for the same pathology. This variability provides insights that might be further examined in Phase 3.
- The analysis should provide articulation of the *interpretation of feasibility* to address the pathologies identified by participants. This permits an explicit capture of the tacit nature of the pathology to be confronted and addressed, without a call for justification or loss of anonymity in the assessment.

Phase 3: Exploration guides an investigation into the meaning and system development implications for identified pathologies. An important preparatory aspect of this phase is the search for additional sources of data, anecdotes, and other supporting examples supportive of the classifications for the pathologies. This phase also maps existing initiatives and their expected potential contribution to identified pathologies. The result of this phase is a strategy and corresponding feasible actions designed to positively influence pathologies. Performance of this phase is critical to begin to make the pathology exploration actionable.

The results of Phase 2 are made available to system owners in preparation for further guided investigation into the meaning of the identified pathologies as well as their implications for system development. This phase involves a two-way dialogue between system owners and the analysts involving the general meaning of pathologies and exploration of the meaning in context for the system of interest. This dialogue is instrumental for articulating and/or voicing system of interest development implications in response to the discovered pathologies. Care must also be taken so as to provide support for the classification of pathologies along existence, impact, and feasibility. This might require gathering additional data, anecdotes, or supporting/refuting attributions for pathologies. It is during this phase that the existing initiatives (development activities underway in the organization) must be mapped against pathologies. This mapping enables discovery of strengths and weaknesses in system development in relationship to the existing pathologies. The results of this phase include a prioritized enumeration of pathologies based on feasibility. Feasibility is an indicator of the anticipated success should the pathologies be engaged for resolution. Ultimately, the output includes a set of strategies and corresponding actions designed to impact the identified pathologies.

Phase 4: Implement deploys selected responsive strategies and corresponding actions to address pathologies and provide for integrated system development. As with any system implementation, the response to pathologies will involve the potential for emergent conditions. Care must be taken to ensure that the implementation actions/strategies are monitored for their impact on the pathologies targeted for resolution.

The purpose of this phase is to ensure that selected responsive strategies are effectively deployed. Activities are based on what is decided in the previous phase. For example, this might include “develop and install a coordinated process for assessment of strategic monitoring” in response to metasystem pathology M5’.2. {Inadequate processing of strategic monitoring results} as identified in Table 2. Identifying this as an issue starts in Phase 1 by the initial identification and assessment of the suspected existence, impact and feasibility for resolution. Phase 2 continues to provide a more detailed examination and analysis of the pathology to identify sources and confirm existence of the pathology and veracity of attributions concerning the pathology. Phase 3 continues with a detailed examination to develop new initiatives while understanding and integrating current initiatives in relationship to the pathology. Phase 4 focuses on the planning for implementation of the responsive strategies and corresponding actions to address the pathology situation. An additional aspect of this phase is to set time expectations with respect to strategy/action deployments as well as definition of the expected contributions to address the pathology (reduction of existence, reduction of impact, or shifting feasibility to address).

Phase 5: Follow-up is focused on the examination of the impact for strategy and action execution in response to pathologies. While direct cause-effect is not possible, conclusions concerning the application of pathology responses should be examined for further implications.

This final phase is focused on examination of the effects of the strategies and actions undertaken to address pathologies. An established timeline, coupled with predetermined contribution expectations, can serve as a place-maker for a re-evaluation of the system by fulfilling two primary purposes: (1) to measure the effects of the strategies/actions as implemented in Phase 4 and (2) identification of new pathologies. Such efforts serve the role of continuous system development. This is essential since the system of interest is operating within a dynamic and most likely turbulent environment. Moreover, the deployment strategies might lose effectiveness over time, new pathologies might emerge, and new technologies might shift the landscape of pathologies along existence, impact, and feasibility dimensions. Therefore, navigating through the M-Path Method is truly continuous with each phase complementing and interrelated to the previous phases.

The application of pathologies for acquisition represents a new and novel perspective for understanding the nature of deficiencies in the acquisition system. A cursory look at the most recently recommended reforms (e.g., Panel 809 report), suggests that (1) while the reforms might be beneficial, they have been constructed and presented in independent actions, (2) the nature of targeting the recommendations does not suggest that the acquisition system is truly being addressed as an integrated system, and (3) without a direct linkage to systems based pathologies, several of the recommendations could be identified as addressing one or more pathologies.

Using System Pathologies: An Application Perspective for Acquisition

The CSG paradigm, and corresponding systems pathologies, has been written from first principles. This has been largely on a clean sheet, especially with respect to the current breakdown of disciplines in capability management in areas like defense acquisition. Those with experience working all aspects of such bureaucracies will know that they largely are beholden to four disciplines, and their corresponding systems, which are also societal memberships: project management, engineering, finance, and contractual law. It is possible, and often necessary, to be members of, and competent in, multiple disciplines. However, it is extremely difficult to be an effective change agent or foster a reform movement operating to shift a bureaucracy with all four disciplines simultaneously satisfied. In fact, the existence and propagation of pathologies in a system are just as likely to occur at the interfaces of these memberships, as to location within a specific membership. Bureaucracies like defense acquisition have largely operated on a project model to implement capability changes for the last 40 years, giving what appears superficially to be an ascendancy to project management, whereas engineering is responsible for developing, checking, certifying, verifying, validating, reviewing, gating and so forth—in short, the handbrake mechanisms. This portrayal explains why acquisition policies are so amorphous and unwieldly (Keating et al., 2017b), and acquisition reviews have called for common-sense changes (Peever et al., 2015; Patanakul et al., 2016; Fowler et al., 2017; Kendall, 2017; Keating et al., 2017b). CSG, and the associated M-Path Method is clearly a means to achieve simpler “due process” and thus governance without the need to revert to intuition. However, such approaches face a politically-charged battle involving the four disciplines.

Thus, the first challenge in implementing CSG and discovery of pathologies is one of recognition of the need for deeper exploration of familiar terrain from a vantage point that is not familiar. The issue for project management (PM) or systems engineering (SE) becomes if you begin with, “you do not understand,” then the bureaucracy will not understand CSG or the value of engaging M-Path for discovery of deeper systemic issues. The complex systems in CSG speaks of SE, yet governance in modern projects, while it adheres to SE gating, is primarily run by the P3O model of portfolio, program, and project management offices; that is, the PM discipline (PMI, 2003). Even then, while the P3O model usually is drawn as a hierarchy, the portfolio and program management offices are relative new-comers, ironically put in for better project governance, and thus projects have usually remained as pre-eminently powerful instantiations of funded futures for the chief executives of operations. This is often reflected in the portfolio and program officers being under-resourced and performing mostly administrative coordination. It is an easy step to understand the increasing complexity introduced by a model such as P3O, and the inherent pathologies stemming from a lack of purposeful design corresponding to the increased complexity. Also, the management hierarchies above acquisition projects have often remained in place with program and portfolio offices as add-ons—intensifying emergent “pathologies” where there is already a battle for control across projects. Arguably, the first P3O attempts at good governance of projects is still in relative infancy and being championed largely by the PM discipline, albeit with encouragement from the other disciplines. The inclusion of CSG, and the corresponding system pathologies perspective, could cynically be seen as a too-late version of P3O brought by the SE discipline. Thus, the risk-value-resource value proposition of engaging such an endeavor likely would be met with skepticism, particularly where there is a lack of understanding, desire, or intensity to move from a fragmented status quo. This status quo would be an ignorant view of the elegance and potential efficacy of the fresh thinking which cuts across both disciplines and offers a new and novel set of eyes to a fragmented, and admitted status quo delineated by existing and emergent problems—thus far resistant to resolution by existing paradigms, methodologies, or status quo thinking. However, the axiom that “culture eats strategy for breakfast every day” has to factor into how to get CSG and pathologies thinking into use. CSG pathology thinking offers nine new metasystem functions that are not beholden to either PM or SE, yet both disciplines would claim to perform these metasystem functions already. The key of course is to establish a structured dialogue on how efficiently and effectively they are done. This is particularly the case in the examination of status quo acquisition issues against the CSG metasystem and corresponding pathologies. Here the nine standardized dimensional metrics of CSG have a distinct advantage over current P3O governance measures, as they are directly linked to systemic deficiencies attributable to underlying violations of systems theory principles.

Another issue of resistance to CSG and pathologies incorporation is branding. For the terms “CSG paradigm” and “System Pathologies” to be embraced and sustained, there needs to be a simplistic power to attract and retain converts. Yet each word of CSG, metasystem, and pathology is, if anything, anathematic to simplistic power—complex, systems, governance, metasystem, pathology—are like five bullets capable of killing the best intentions and noble notions to improve systems. So, if CSG and the M-Path Method are to achieve implementation, especially in bureaucracies and not the utopia of a Silicon Valley start-up, they might need to be, unfortunately, deployed first from a Machiavellian perspective. Most importantly, CSG is unlikely to succeed if it does not leverage extant P3O management efforts towards good governance—the absence of already accepted approaches misses the opportunity to approach the unfamiliar from a familiar point. What is sacrificed in inappropriate assumptions might be made up in being able to continue the conversation. The remainder of this section outlines a suggested implementation approach aligning to these significant cultural barriers.

In pursuit of our goal, the next task is to target where within the P3O construct to fit the CSG paradigm and focus on pathologies. As CSG aims to govern systems for viability through the identification and reduction of pathologies (as threats to viability), CSG is more logically aligned with program management. Projects by definition are meant to be unique and to open and close, whereas acquisition programs are intended, designed, and focused to sustain operational capability. That both CSG and program management use biological adjectives like evolutionary acquisition, system viability, and pathologies indicates a healthy alignment.

If CSG is to constructively leverage the extant governance of program management, then the 53 pathologies developed in CSG around the nine metasystem dimensions are key. This is necessary to avoid introduction of CSG as an independent and mutually exclusive approach in competition to existing development methods. Currently, program management offices, where they seek improvement, usually do so for standardization and accreditation and through project management institutes. These institutes wield considerable influence and provide recognition predominantly within the PM domain of acquisition bureaucracies, in part equalizing established recognition within engineering disciplines. The PM institutes have been working on standards and metrics for P3O, with some proposals looking foundationally at aggregating assessment of realized benefits across a portfolio/program while others set generic standards. The most commonly used PM metrics focus on organizational maturity, developed from the four basic process improvement phases of standardize, measure, control, and continuously improve, but with program and portfolio governance added, as shown in Figure 3.

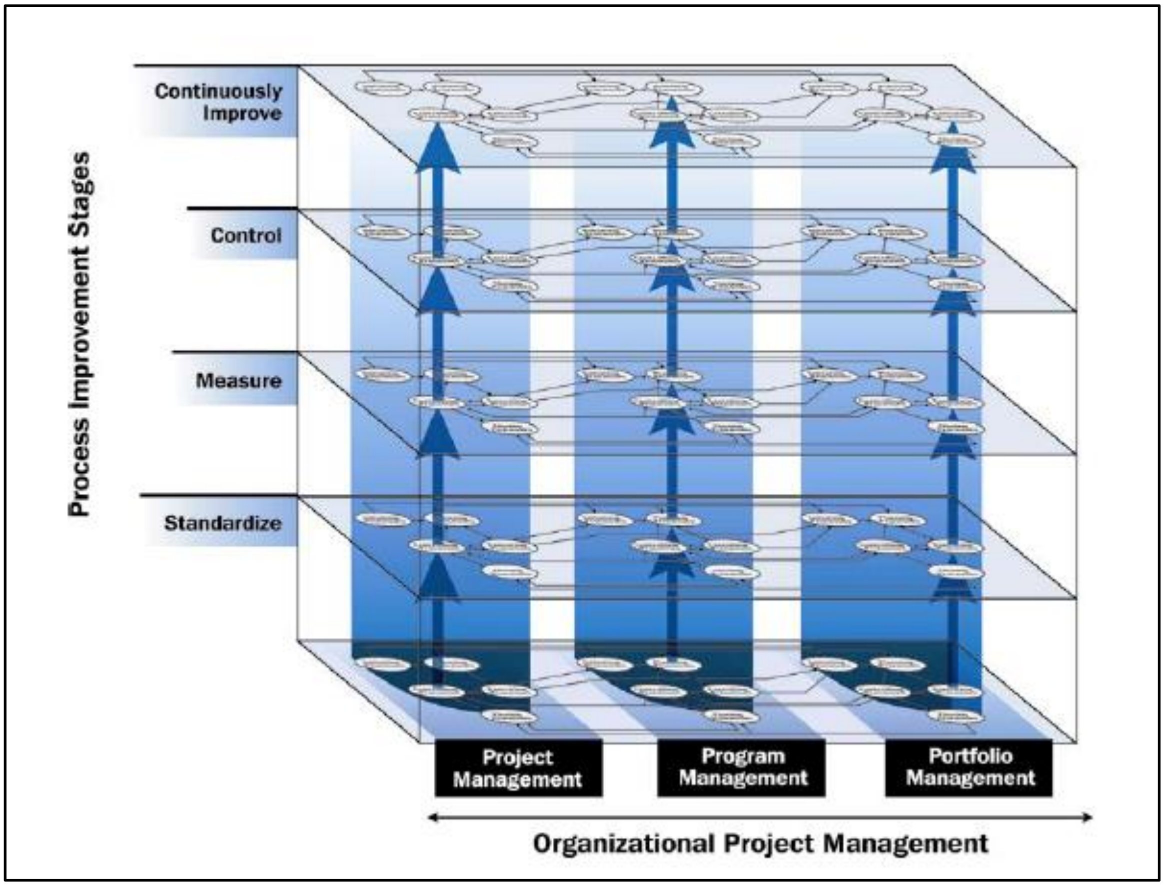


Figure 3. Common PM Metrics
(PMI, 2003, p. 28)

The maturity model metrics assess how the organization's processes compare to the model, usually as part of organizational accreditation for program or portfolio management. Scales are usually aggregated across project divisions and typically range up to five, where a three would be having defined project management processes and some measures of project achievement and feedback control, while a five would involve continuous and documented process improvement linked to better project achievements. Clearly, the CSG pathologies offer a substantially improved method to diagnosis program and portfolio governance, and if used correctly, ought to offer targeted process improvement, aligned with the first principles of CSG based in Systems Theory. Thus, P3O assessments offer a means for CSG pathologies to be adopted by accretion and improvement, analogous with how many manufacturing and some service industries use the six-sigma techniques without wholesale adoption and overt branding (Evans & Lindsay, 2014; Stamatis, 2016).

Adopting the CSG pathologies may be challenging for the PM disciplines alone, since as documented earlier, many governance elements remain relatively weak. A logical way to help achieve their use is to coopt the engineering domain, who, to be brutally honest, constructed most of the bureaucratic policies that beset acquisition as a check on projects delivering inadequate equipment. Much of the engineering hierarchy are skeptical at the vagueness of P3O project reviews and certainly of the traffic-lights and scoring of program office maturity. Putting the elegance and robustness of CSG pathologies into reviews of how program offices are structured and perform ought to gain engineering support. Thus, the PM discipline would lead the organizational reform, but using a robust model and CSG pathologies so the SE discipline can improve acquisition processes in a sound way. Put simply, CSG is not challenging the existence of PMOs, but is, rather, giving them better tools and techniques to refine the acquisition processes for more timely and effective governance.

A feature of CSG and the corresponding pathologies with potential to be a powerful change agent in PMO reviews, is the use of standardized assessment questionnaires. These questionnaires, taken by staff, can measure not only how effective the present state of governance is, but also how effective governance ought to be given the program objectives. This effectiveness can be examined in detail through the nine CSG meta-functional dimensions and their corresponding pathologies. In educational environments this is often referred to as the “actual” and “preferred” and is about the “fit” of an environment for an individual (Caplan & Harrison, 1993; Su, Murdock, & Rounds, 2015). Effort is therefore directed to where most P3O staff see governance is most deficient and it ought to lead to a satisficing level of necessary governance rather than constant pressure for utopia of an open five-point scale. Such a model is shown in the nine-point spider chart in Figure 4, where the actual web and preferred web show where there is the least fit and where effort in one dimension where governance exceeds assessed demand (e.g., capacity exceeding demand for the uncertainty dimension) can be traded off in another dimension where governance capacity falls below assessed demand (e.g., the complexity dimension; Jaradat, Keating, & Bradley, 2017). Such assessment of organizational governance and existence of pathologies depends on surveying the full spectrum of P3O; that is, both the governed and governing for 360-degree coverage of the system.

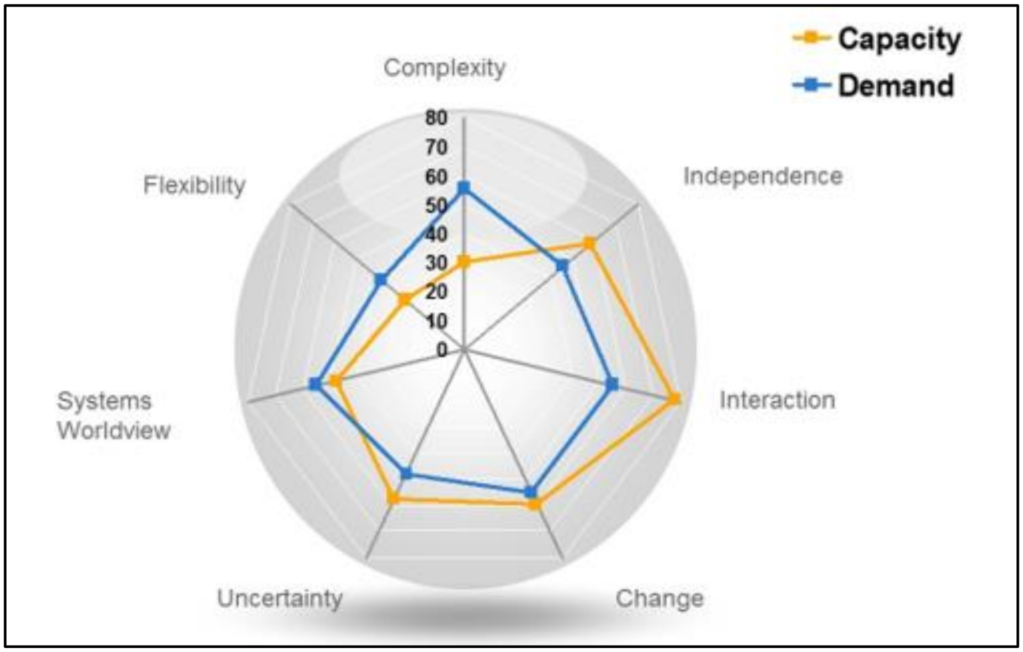


Figure 4. Complexity—Systems Capacity Versus Environmental Demand

Therefore, with that background, a general process of adopting the CSG paradigm and the constituent pathologies intent in industry, defense or other government acquisition department would be as follows:

1. With support from senior Systems Engineers, the portfolio of acquisition selects a program or series of program offices to undergo the CSG self-assessments of extant governance under the oversight of that program(s). The assessments are to compare the efficacy of CSG dimensions and pathologies against any existing program maturity improvement processes and metrics.
2. Each PMO doing entry-level CSG examines the preferred versus actual governance ratings of their P3O staffs to identify areas of expectation mismatch.
3. The associated pathologies, for areas of mismatch, are then workshopped with a representative sample of surveyed staff and affected process specialists to identify improvements which are then implemented.
4. After a suitable period of process embedding, the CSG self-assessments are repeated to examine if the mismatch(s) have been redressed and what, if any, further improvements are needed.

This approach has assumed a level of P3O management exists and has some form of self-assessment already focused on governance that can be leveraged. Where P3O management does not exist in name, it should exist in function and even if the self-assessment is ad hoc and not independent of general management, there exists a basis from which to acknowledge and begin. The M-Path Method is easily modifiable to fit to the approach proposed above.

Conclusions, Implications, and Directions of Pathologies for Acquisition

In conclusion, we examine two interrelated facets for further development of CSG and the corresponding M-Path Method for both practice and research for acquisition. The application of the current state of CSG and M-Path for the acquisition field holds promise in several areas.

Ultimately, CSG and the M-Path Method to discover deficiencies in the performance of CSG functions, offers significant contributions to help practitioners address some of the most vexing current, as well as future, system problems they must confront. CSG is not suggested as a panacea for all problems facing the acquisition system and programs. Instead, CSG is advocated as an emerging field with significant opportunity to provide value in the following areas (Figure 5):

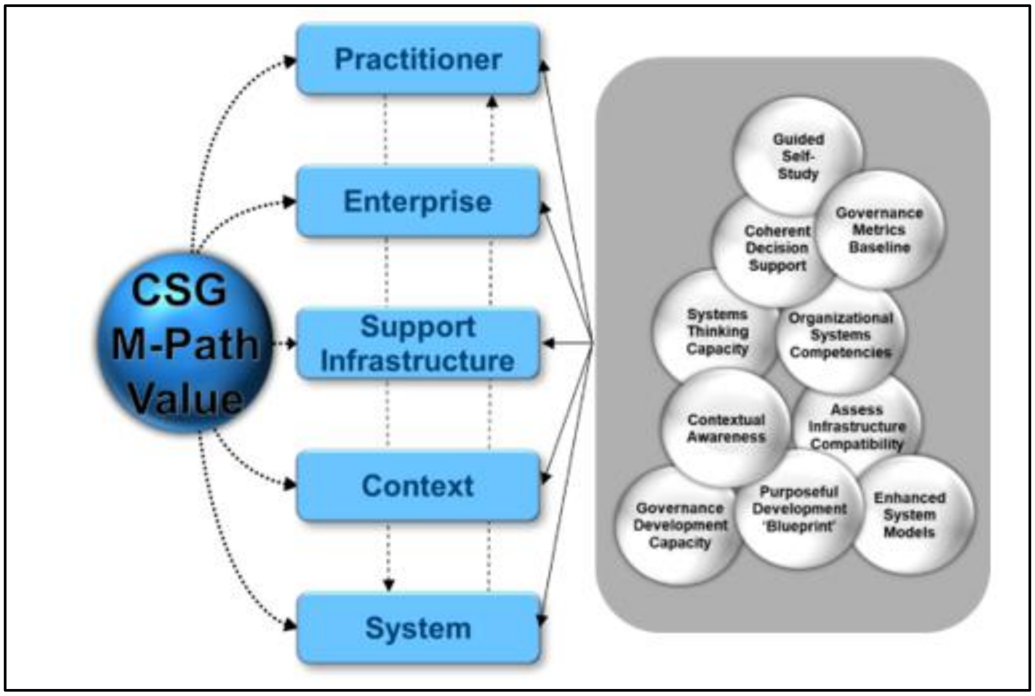


Figure 5. CSG M-Path Method Value Adding Across Multiple Levels

- **Rigorous Guided "Self-Study"** into CSG can provide significant insights into how the system (program, project, portfolio) actually functions. Although enterprises and their systems function routinely and successfully on a daily basis, as a matter of course practitioners are not particularly skilled, nor do they engage in deep reflection as to why, how, and what they do from a systems point of view. The gains to be made by reflective self-examination, from a systemic point of view, can reveal insights far beyond traditional methods of examination (e.g., Strategic Planning, SWOT analysis, Maturity Modeling, etc.). Thus, practitioners can examine a different level of analysis through "self-study" and experience insights in a "safe-to-fail" setting. Additionally, self-study might suggest the level of education/training that might be necessary for individuals and the organization to increase individual capacity and organizational competence necessary to engage in higher levels of systems thinking.
- **Coherent Decision Support** can be achieved by the big picture view of the governance landscape. This includes identification of highest leverage strategic impact areas and their interrelationship to the larger CSG performance gaps (through pathologies assessment). Thus, decisions for resource allocation can be better targeted. This allows steering away from activities that are simply intriguing without demonstrating the highest substantial benefit to the larger systemic governance concerns (e.g., pathologies high in existence, impact, and feasibility). In light of CSG development priorities, low contribution efforts can be eliminated, or resources shifted appropriately.
- **Governance Metrics Baseline** can be established to identify the present state of CSG functions as indicated by the pathologies. The set of unique indicators developed for a specific system of interest can provide a baseline that can be used to longitudinally establish the continuous progression of

governance improvement. In effect, the degree of improvement stemming from initiatives undertaken to improve CSG can be established. Therefore, the state and shifts in governance can be purposefully and actively planned, deployed, monitored, and adjusted as necessary.

- **Systems Thinking Capacity** of individual practitioners to engage in the level of systems thinking necessary to more effectively deal with the entire range of complex system problems can be enhanced through CSG application. These problems are a byproduct of modern acquisition enterprises and their systems. Effectiveness is achieved through development and propagation of CSG language, methods, and tools to assist practitioners in their efforts to design, analyze, execute, and evolve complex systems and their associated problems (Jaradat, 2015).
- **Organizational Systems Competencies** at the system (project, program, portfolio) level for dealing with complex systems and their derivative problems can be enhanced. This involves generation of knowledge, development of skills, and fostering abilities beyond the individual level to embrace problems holistically. For CSG, holism suggests competency development that expands beyond narrow technology centric infusions and the limiting cost-schedule-technical performance paradigm. Instead, enhanced competencies that span the entire range of sociotechnical considerations endemic to complex systems are an outcome from CSG engagement to identify, analyze, and address pathologies.
- **Enhanced Contextual Awareness** is a direct byproduct from the examination of system pathologies. Context exist as the circumstances, factors, conditions, or patterns that serve to enable or constrain performance of system functions. Thus, the wider consideration of system impediments provided by the M-Path Method can open the aperture of consideration of aspects for development of the system.
- **Assess Infrastructure Compatibility** necessary to support systems-based endeavors. This compatibility is necessary to formulate contextually consistent (feasible) approaches to problems, create conditions necessary for governance system stability, and produce coherent decisions, actions, and interpretations at the individual and organizational levels. The most exceptional system solutions, absent compatible supporting infrastructure, are destined to outright fail in the worst-case scenario and underachieve in the best-case scenario.
- **Governance Development Capacity** can be determined to help establish the feasibility of initiatives that can be undertaken with a higher probability of successful achievement. This does not minimize the degree of CSG discovered inadequacies that might exist in a system. However, it does take into account the current sophistication in system governance, the limiting/enabling context, and the individual systems thinking capacity that will influence what can be reasonably taken on with confidence of success. Minimally, consideration of feasibility for addressing M-Path Method generated issues can provide new insights into past successes/failures as well as cautions for impending future endeavors.
- **Enhanced System Models** generated through CSG M-Path Method deployment efforts can provide insights into the structural relationships, context, and systemic deficiencies that exist for governance of a system of

interest. These insights can accrue regardless of whether or not specific actions to address issues are initiated. The models can be constructed without system modification. Therefore, alternative decisions, actions, and interpretations can be selectively engaged based on consideration of insights and understanding generated through modeling efforts.

- **Purposeful Development Blueprint** development can provide focus for targeted advancement of the CSG functions. This accrues through resolution of priority M-Path derived issues in performance of system functions necessary to maintain system viability. While all viable (existing) systems perform the CSG functions and have pathologies, it is rare that they are purposefully articulated, examined, or addressed in a comprehensive fashion. Purposeful CSG development to resolve M-Path Method identified pathologies can produce a blueprint against which development can be achieved by design, rather than serendipity. This includes establishment of the set of "dashboard indicators" for CSG performance. These performance indicators exist beyond more traditional measures of system/organizational performance.

Ultimately, the CSG M-Path Method seeks to increase the probability of achieving desirable system performance (viability, growth, etc.) in the flux of a turbulent environment.

Further development of CSG pathologies for acquisition systems is focused on four critical challenges:

1. *Increasing Ease of Engagement*. CSG and corresponding pathologies are not easily understood, applied, or accessible for practitioners and the systems they manage. If CSG and pathologies are to become more mainstream there must be an increased emphasis on making the technologies (methods, tools, techniques, applications) accessible to practitioners. Accessibility must include the ease of engagement of pathologies, including (1) reduction of perceived risk for practitioners and systems subject to a thorough external analysis of design, execution, and development effectiveness, (2) efficiency in application resource demands, including time as well as cost, (3) emphasis on demonstrable value that can accrue from engagement in a CSG pathologies endeavor, and (4) potential linkage to ongoing or historical system development initiatives. Only by addressing these four areas can the probability of adoption of CSG pathologies be increased.
2. *Products-Insight-Action Triad*. Additional emphasis of CSG must focus on what tangible products (e.g., pathologies representation mappings) can serve as tangible artifacts of engagement in CSG activities. Absent tangible products, CSG is left at a conceptual level and is not likely to achieve full impact potential. Additionally, emphasis on the development of insights stemming from products will require processes that serve to guide exploration and interpretation of meanings generated from the CSG products. Finally, CSG pathologies insights must be made actionable to redesign, modify, and evolve governance for the system of interest. Irrespective of the intellectual grounding and products from CSG pathology applications, without corresponding actionable results, there is little possibility to make sustainable system improvements.
3. *System Ownership and CSG Accountability Acceptance*. Every system has owners who are ultimately responsible for the design of the system (functions), execution of the system (performance of functions to produce

results), and development of the system (to adapt to changing internal/external circumstances and address pathologies). Short of active engagement of “system owner” responsibilities and acceptance of accountability for CSG functions, there is little possibility to achieve the aims of CSG. Even though ownership might be agreed upon, the CSG responsibilities (e.g., pathology elimination) inherent in that ownership must be embraced. Without this mindset for CSG, the feasibility of effective engagement is minimal.

4. *Requisite Systems Capacity*. CSG pathologies are based in a strong underlying conceptual basis anchored in Systems Theory. Lacking a sufficient grounding, mindset, and acceptance of the “systemic” perspective, it is not likely that CSG pathologies will be effectively engaged or understood. Thus, care must to taken to ensure that an appropriate level of systems thinking capacity exist to adequately engage in CSG pathology related endeavors. Lacking the requisite systems capacity is likely to produce unsatisfactory results from CSG pathologies related efforts.

The acquisition field and practitioners are being called upon to deal with increasingly complex systems and the corresponding issues in their development. The emerging research in CSG and system pathology (e.g., see Troncale, 2013) pathologies can offer a new and novel way of framing acquisition issues across multiple levels, including projects, programs, and portfolios. CSG enables the design, execution, and evolution of critical metasystem functions necessary to maintain system viability. The identification of system pathologies provides a basis for a different level of thinking, corresponding decision, and alternative responsive actions. It is from this vantage point that we suggest engaging acquisition systems and their issues from a more systemic perspective creates possibilities for acquisition reform that, although desired, has yet to be realized.

References

Berteau, D., Levy, R., Ben-Ari, G., & Moore, C. (2011). *Defense industry access to capital markets: Wall Street and the Pentagon: An annotated brief.*

Boudreau, M. (2007). *Acoustic rapid COTS insertion: A case study in modular open systems approach for spiral development*. Paper presented at the IEEE International Conference on System of Systems Engineering, 2007.

Bradley, J. M., Katina, P. F., & Keating, C. B. (2016). Complex system governance for acquisition. In *Proceedings of the 13th Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Bucci, S. P., & Maine, E. (2013, December). *Thornberry initiative for effective defense acquisition reform appears promising* (Issue Brief #4106). Santa Monica, CA: RAND.

Caplan, R. D., & Harrison, R. V. (1993). Person-environment fit theory: Some history, recent developments, and future directions. *Journal of Social Issues, 49*(4), 253–275.

Cilli, M., Parnell, G. S., Cloutier, R., & Zigh, T. (2015). A systems engineering perspective on the revised defense acquisition system. *Systems Engineering, 18*(6), 584–603.

Evans, J. R., & Lindsay, W. M. (2014*). An introduction to Six Sigma and process improvement.* Boston, MA: Cengage Learning.

Fowler, S., Sweetman, C., Ravindran, S., Joiner, K. F., & Sitnikova, E. (2017). Developing cyber-security policies that penetrate Australian defence acquisitions. *Australian Defence Force Journal, 202,* 17.

Fox, J. R. (2012). *Defense acquisition reform, 1960–2009: An elusive goal* (Vol. 51). Washington, DC: GPO.

Francis, P. L. (2008). *Defense acquisitions: Zumwalt-class destroyer program emblematic of challenges facing Navy shipbuilding*. Washington, DC: GPO.

Francis, P. L. (2009). *Defense acquisitions: cost to deliver Zumwalt-class destroyers likely to exceed budget.* (1437909310). Washington, DC: GPO.

Friar, A. (2015). *Swamped by regulations: Perils of an ever-increasing burden.* Ft. Belvoir, VA: Defense Acquisition University.

Jaradat, R. (2015). Complex system governance requires systems thinking. International *Journal of System of Systems Engineering, 6*(1/2), 53–70.

Jaradat, R., Keating, C., & Bradley J. (2017). Individual capacity and organizational competency for systems thinking. *IEEE Systems Journal.* doi:10.1109/JSYST.2017.2652218

Joiner, K. F., & Tutty, M. G. (2018). A tale of two allied defence departments: New assurance initiatives for managing increasing system complexity, interconnectedness, and vulnerability. *Australian Journal of Multidisciplinary Engineering.* doi:10.1080/14488388.2018.1426407

Katina, P. F. (2015). Emerging systems theory–based pathologies for governance of complex systems. *International Journal of System of Systems Engineering, 6*(1/2), 144–159.

Katina, P. F. (2016a). Metasystem pathologies (M-Path) method: Phases and procedures. *Journal of Management Development, 35*(10), 1287–1301.

Katina, P. F. (2016b). Systems theory as a foundation for discovery of pathologies for complex system problem formulation. In A. J. Masys (Ed.), *Applications of systems thinking and soft operations research in managing complexity* (pp. 227–267). Geneva, Switzerland: Springer International.

Katina, P. F., & Keating, C. B. (2014). Metasystem pathologies: Towards a systems-based construct for complex system deficiencies. In S. Long, E.-H. Ng, & C. Downing (Eds.), *Proceedings of the American Society for Engineering Management 2014 International Annual Conference.* Virginia Beach, VA: American Society for Engineering Management.

Keating, C. B., & Bradley, J. M. (2015). Complex system governance reference model. *International Journal of System of Systems Engineering, 6*(1), 33–52.

Keating, C. B., & Katina, P. F. (2012). Prevalence of pathologies in systems of systems. *International Journal of System of Systems Engineering, 3*(3–4), 243–267.

Keating, C. B., Bradley, J. M., Katina, P. F., & Jaradat, R. M. (2017b). A systems theoretic-based framework to discover pathologies in acquisition system governance. In *Proceedings of the 14th Annual Acquisition Research Symposium* (pp. 352–376). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Keating, C. B., Katina, P. F., Jaradat, R. E., Bradley, J. M., & Gheorghe, A. V. (2017a). Acquisition system development: A complex system governance perspective. *INCOSE International Symposium, 27*(1), 811–825.

Kendall, F. (2014, July 10). Testimony before the House Committee on Armed Services.

Kendall, F. (2017). *Getting defense acquisition right*. Washington, DC: Under Secretary of Defense for Acquisition, Technology, and Logistics.

O'Rourke, R. (2014). *Case studies in DoD acquisition: Finding what works* (1437928390). Washington, DC: Congressional Record.

Patanakul, P., Kwak, Y. H., Zwikael, O., & Liu, M. (2016). What impacts the performance of large-scale government projects? *International Journal of Project Management, 34*(3), 452–466.

Peever, D., et al. (2015). *First principles review: Creating one defence*. Retrieved April 1, 2015, from http://www.defence.gov.au/publications/reviews/firstprinciples/

Project Management Institute (PMI). (2003). *Organizational project management maturity model (OPM3™).* Newtown Square, PA: Author.

Rascona, P. M., Barkakati, N., & Solis, W. M. (2008). *DOD business transformation Air Force's current approach increases risk that asset visibility goals and transformation priorities will not be achieved.* Washington, DC: GPO.

Salado, A., & Nilchiani, R. (2017). Understanding how problem formulation in systems engineering influences system affordability: A systems thinking investigation. *International Journal of Applied Systemic Studies, 7*(4), 227–256. doi:10.1504/IJASS.2017.089938

Schwartz, M. (2014). *Defense acquisitions: How DOD acquires weapon systems and recent efforts to reform the process*.

Smith, N. C., White, E. D., Ritschel, J. D., & Thal, A. E. (2016). Counteracting harmful incentives in DoD acquisition through test and evaluation and oversight*. ITEA Journal, 37,* 218–226.

Stamatis, D. H. (2016). *Six Sigma and beyond: The implementation process* (Vol. VII).

Su, R., Murdock, C. D., & Rounds, J. (2015). Person-environment fit. In *APA handbook of career intervention* (1; pp. 81–98).

Troncale, L. (2013). Systems processes and pathologies: Creating an integrated framework for systems science. *INCOSE International Symposium, 23*(1), 1330–1353. doi:10.1002/j.2334-5837.2013.tb03091.x

A Mathematical Framework to Apply Tradespace Exploration to the Design of Verification Strategies

Alejandro Salado—is an Assistant Professor with the Grado Department of Industrial and Systems Engineering at Virginia Tech. His research focuses on applying decision analysis to improve the practice of engineering, in particular in the areas of verification and validation. He is a recipient of the Fulbright International Science and Technology Award. Dr. Salado holds a BSc/MSc in electrical engineering (Polytechnic University of Valencia), an MSc in project management, and an MSc in electronics engineering (Polytechnic University of Catalonia), the SpaceTech MEng in space systems engineering (Delft University of Technology), and a PhD in systems engineering (Stevens Institute of Technology). [asalado@vt.edu]

Farshad Farkhondehmaal—is a PhD student of management systems engineering with the Grado Department of Industrial and Systems Engineering at Virginia Tech. He received his BSc in aerospace engineering from Sharif University of Technology in Iran. His research interests are systems engineering, model-based systems engineering, and systems dynamic. [farshad1@vt.edu]

Abstract

This paper is intended to disseminate some initial outcomes of the NPS Research Acquisition Program's "Tradespace Exploration for Better Verification Strategies" project. The research addresses the design of verification strategies in large-scale systems. Verification activities provide the evidence of contractual fulfillment. Thus, the importance of adequately defining verification activities in any acquisition program is unquestionable. Its significance extends beyond contracting though. The biggest portion of the development financial budget is spent in executing verification activities, and verification activities are the main vehicle in discovering knowledge about the system, which is key to reduce development risk. In current practice, the definition of verification strategies is driven by industry standards and subject matter expert assessment. This research addresses the main question of whether tradespace exploration can support the definition of more valuable verification strategies than current practice. We present in this paper a mathematical framework that enables the application of tradespace exploration to the design of verification strategies.

Introduction

Requirements lay at the core of system acquisition, given their contractual nature. Verification activities are executed to demonstrate fulfillment of those requirements. Hence, verification provides the evidence of contractual fulfillment. Actually, in several cases reaching an agreement about when a requirement is considered fulfilled is more important than agreeing on the requirement itself. Thus, their importance for acquisition in contracting is unquestionable.

Verification activities, which may take the form of a combination of analyses, inspections, and tests, consume a significant part, if not the biggest part, of the development costs of large-scale engineered systems (Engel, 2010). Verification occurs at various levels of a system's decomposition and at different times during its life cycle (Buede, 2009; Engel, 2010). Under a common master plan, low level verification activities are executed as risk mitigation activities, such as early identification of problems, or because some of them are not possible at higher levels of integration (Engel, 2010). Therefore, a verification strategy is defined as

> aiming at maximizing confidence on verification coverage, which facilitates convincing a customer that contractual obligations have been met; minimizing risk of undetected problems, which is important for a manufacturer's reputation, and to ensure customer satisfaction once the system is operational; and minimizing invested effort, which is related to manufacturer's profit. (Salado, 2015)

Essentially, verification activities are the vehicle by which contractors can collect evidence of contractual fulfillment in acquisition programs.

In current practice, the definition of verification strategies is driven by industry standards and subject matter expert assessment. Usually, the resulting strategy requires a higher cost than the initial budget allocated by the project. De-scoping activities are then performed, with qualitative evaluation of resulting risk, until agreement is reached by the engineering and project management teams. Such verification strategy is then agreed on with the customer, following similar dynamics. Sometimes in parallel, but often after agreement with the customer, the prime contractor tries then to impose its verification strategy to the lower level assemblies (developed by its subcontractors). This yields new negotiations and local trade-offs with each supplier. The same dynamics and approaches as described earlier are exhibited in these cases. Because the financial resources for such activities are usually committed at the early phases of a system's life cycle (INCOSE, 2011), succeeding in finding an optimal strategy is often limited by the amount of time and resources that are invested in its definition, which are often scarce.

Furthermore, current practice relies on non-normative methods that are based on subject matter expert assessments rather than on measurements, which questions the optimality of verification strategies currently defined in industry (Salado, 2015). This context leads to four major risks. First, there is a high uncertainty associated to the optimality of the selected verification strategy in terms of mitigated risk with respect to verification cost. Second, there is a lack of a quantitative risk associated to chosen verification strategy, which jeopardizes any mindful effort to execute informed trade-offs regarding execution of verification activities. Third, there is a high risk associated to the verification coverage of the selected verification strategy, which threatens the successful integration of components and the successful operation of the system. And fourth, there is a lack of alignment between stakeholder objectives and verification strategy, which leads to suboptimal decisions regarding the execution of verification activities.

Informed by the benefits of tradespace exploration in conceptual design (Ross & Hastings, 2005), the use of tradespace exploration was piloted in an actual industrial project to define a test strategy for a major satellite optical instrument (Salado, 2015). The results were positive, being able to identify a test strategy with the same level of value and lower risk to the customer with a 20% lower cost than using the industry benchmark (Salado, 2015). However, the work presented a number of limitations related to generality and normativity. This paper presents a modified framework to apply tradespace exploration to the design of verification strategies that overcomes those limitations.

Background

Tradespace Exploration

Traditional point design methods have been found to be ineffective in traditional concept design (Ross & Hastings, 2005). Such methods quickly anchor to a few solutions, limiting the perspective on potentially better solutions available in the larger solution space. As a response to such need, tradespace exploration techniques have been proposed (Ross & Hastings, 2005). They recognize that in multi-attribute decisions, a set of optimal solutions exists, as opposed to a single optimum solution. In this context, tradespace exploration consists in comprehensively populate the solution space with as many solutions as possible, identify its Pareto frontier or front (which is a set of solutions that provide maximum return for a given level of investment), and let the stakeholder choose a solution (Mattson & Messac, 2003; Ross & Hastings, 2005; Ross et al., 2004). Tradespace exploration has been proven to support design methods that are effective in resolving ambiguity and facilitating communication, understanding, and agreement between multiple stakeholders (Golkar & Crawley, 2014; Ross et al., 2004).

Verification in Large-Scale Engineered Systems

Verification of large-scale engineered systems may occur in every phase of their lifecycle (Engel, 2010), can take the form of a variety of methods (e.g., analysis, inspection, demonstration, test, or certification; Engel, 2010), and can take place at different integration levels (INCOSE, 2011). Designing a verification strategy consists of deciding which verification activity occurs at which point in time and on what integration level. For example, method selection may be driven by programmatic constraints imposed by customers and business goals, credibility of method validity by customers, and feasibility of the method (Engel, 2010; Larson, et al., 2009). Similarly, early verification, both in terms of assembly level and of lifecycle phase, may be desirable for mitigating the risk of failure or error (Engel, 2010; Firesmith, 2013), or because some system properties, attributes, or functionalities are not verifiable at higher levels of the assembly, or cannot be verified in some specific configurations (Firesmith, 2013). Respectively, late testing may also be desirable for mitigating the risks of damage during the integration and test campaign and of emergent behavior or properties of all constituting elements integrated together (Firesmith, 2013), or simply because some system properties, attributes, or functionalities can only be verified once a number of elements are operating together (Firesmith, 2013).

In addition, designing a verification activity is driven by finding the right balance between verification cost and the cost of failure corresponding to those ones not discovered by the verification strategy (Engel, 2010). Since the cost and time allocated to verification activities represents a significant amount of the whole system development cost and time, optimizing verification is important in the development of large-scale systems (Engel & Shachar, 2006). Using cost and time as target values, several optimization techniques have been proposed as underlying mathematical/numerical models to identify a preferred verification strategy: loss function optimization, weight optimization, goal optimization, and genetic algorithm optimization (Engel, 2010; Engel & Shachar, 2006). Despite the diversity of methods though, all of them output a single optimum solution, i.e., they are point design strategies. Hence, they present the same limitations as point-design methods employed in conceptual design, which have been identified in the previous section.

Tradespace Exploration Applied to the Verification Domain

The application of tradespace exploration to the domain of verification was piloted in an industrial project to design the test strategy for a satellite instrument (Salado, 2015). The approach provided positive results, enabling the project team to uncover a test strategy that was less risky at 20% lower cost than the solution that was initially defined by the expert team using conventional definition approaches (Salado, 2015). Figure 1 shows the process that was developed for applying tradespace exploration in that project. Essentially, the processes starts with a test campaign that contains all potential test activities as described in *Space Engineering—Testing* (ECSS, 2012), which is then parsed into its elemental test activities. Such activities are then characterized in terms of cost and value to the customer, together with some general rules that account for couplings between the various activities. Finally, combinations of the different activities are generated to populate the solution space and evaluated.

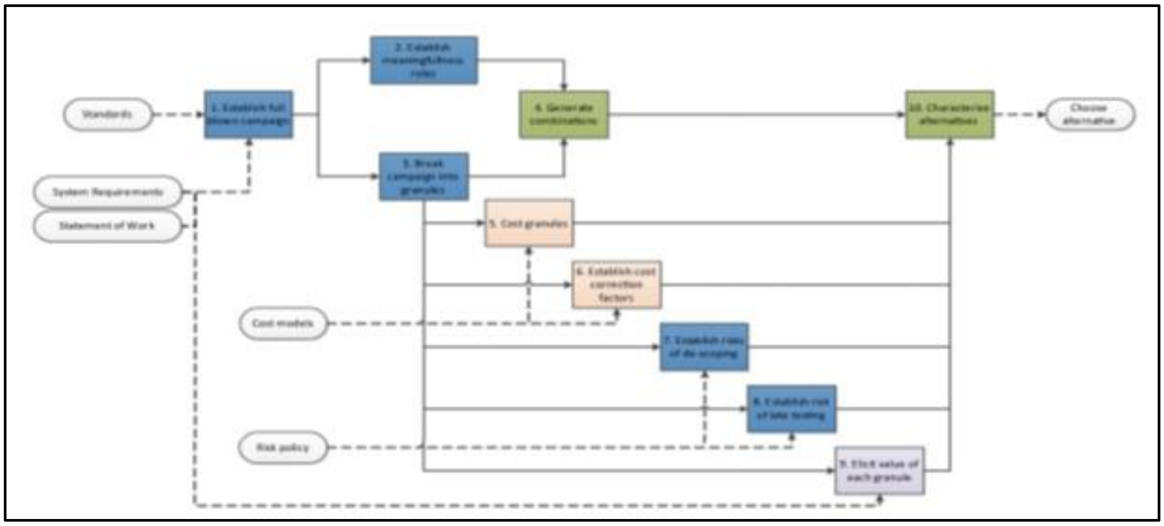


Figure 1. Tradespace Exploration Process Applied to the Design of a Test Strategy
(Salado, 2015)

While the application of such a process yielded positive results, the process had some limitations that disable it from being generalizable to other projects. Three limitations stand out. First, the process was defined only for test activities and not verification activities in general. This implied that each activity was associated to a particular system characteristic. As a result, the process did not cover cases in which various verification activities are employed to build up together the verification evidence for a single system characteristic. Second, the sequence in which the test activities were to be executed was fixed. That is, the solution space only contained alternatives created by selecting which verification activities would be performed, but only for a generic sequence. Therefore, a large portion of the solution space, containing different sequences of activities, was not explored. Third, valuation of verification strategies was qualitative and assumed a separable value function with respect to each verification activity. As we will discuss later, valuing verification strategies is not straightforward and demands a more sophisticated approach.

Salado's (2015) work was expanded to overcome some of its limitations. In particular, mathematical foundations of verification engineering were proposed to enable the generalization of the application of tradespace exploration to defining verification strategies (Salado, 2016). Of particular importance is the realization that the purpose of verification activities is to discover knowledge about the system of interest (Salado, 2016). Consequently, the value of a given verification activity is not absolute. Instead, it is a function of the previous knowledge about the system of interest. Hence, the value a verification activity depends, among others, on the verification activities that have been performed before it (Salado, 2016). This leads to two critical conclusions. First, sequence is a key driver of the value of verification strategies. Second, the value function for a verification activity may not easily be a separable function of its verification activities.

While the value of these dependency notions were showcased with a toy example, the mathematical foundations also present some limitations that disable it from facilitating automation in the population of the solution space, as well as on adequately valuing verification strategies. In particular, the mathematical framework did not capture sequence of activities, although it was recognized in the sample case, and valuation was done qualitatively, without identifying a rigorous mathematical framework to enable computations.

In this paper, we present a comprehensive framework that overcomes all limitations of previous work in the application of tradespace exploration to the design of verification activities.

A Tradespace Exploration Framework to Design Verification Strategies

Framework

We propose a framework that builds upon the two main activities of tradespace exploration: generation of solutions and positioning of solutions in the tradespace. The framework is depicted in Figure 2. The generation activity consists of creating as many solutions as possible leveraging a structural model. The location activity consists in evaluating the generated solutions with respect to a set of predefined criteria, which would then result in positioning every solution within the tradespace.

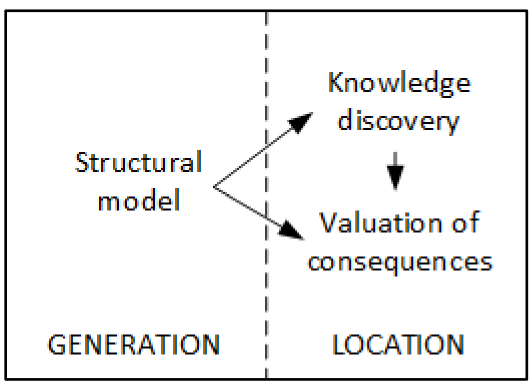


Figure 2. Proposed Framework to Apply Tradespace Exploration to the Design of Verification Strategies

The framework consist of three main elements, which are described in detail in the next sections. First, we make use of a mathematical model that describes the underlying structure of verification strategies. This model enables automating the generation of verification strategies through computational algorithms to populate the solution space of verification strategies. The model is built with set theory and graph theory. Second, we add machinery to the structural model of verification strategies to compute the knowledge they discovered. In other words, how verification strategies shape beliefs on the system containing or not containing errors as verification activities are executed. This machinery is built on Bayesian networks. Third, we valuate the consequences of executing a verification strategy. In particular, we provide expected value models to compute the cost associated with executing the verification strategy, as well as the expected cost to perform rework activities in case errors are actually found by a verification strategy.

A Mathematical Model to Generate Verification Strategies

In order to capture the dependencies between verification activities, we define a **verification strategy** S as a simple directed graph $S = (V, D)$, where V is a set of verification activities and D is a set of tuples of the form (a, b) such that $a, b \in V$. Then, V describes the verification activities that will be executed as part of S and D the sequence in which they will be executed. The solution space of verification strategies for a system z_0, denoted by $\sum(z_0, R)$, will therefore be given by all simple directed graphs that could be generated using all possible verification methods or procedures R on z_0. Mathematically, $\sum(z_0, R) = \{S = V, D): V = \Upsilon(z_0, R)\}$, where $\Upsilon(z_0, R)$ is the set of all potential verification activities that could be executed to provide information about z_0. This is given by

$$\Upsilon(z_0, R) = \bigcup_{i=0}^{n}\left(F(z_i) \cup \bigcup_{j=1}^{\#H_i} F(z_{i,j}) \right) \times R \text{,}$$

where

1. $z_1, \ldots, z_n$ are the systems that decompose z_0 in all of its constituent elements on which formal verification occurs. They are traditionally referred to as subsystems, components, or parts, among others.
2. $H_i = \{z_i, z_{i,1}, z_{i,2,} \ldots, z_{i,m}\}$ is the set of systems that are homomorphic images of system z_i, as defined in Wymore (1993). Note that a system is homomorphic to itself and hence it is included in the set. This set represents all models of system z_i that are used for verification. In practical terms, they can take the form of a mathematical model, a prototype, or the final product, for example.
3. $F(z) = \{p_1, p_2, \ldots, p_k\}$ is a parameterization of system z, where the definition of parameterization in Wymore (1993) is used. This parameterization is finite and represents the set of parameters of system z that need to be formally verified. For example, those parameters may represent the set of requirements that system z has to fulfill, and for which fulfillment needs to be proven through formal verification.
4. A **verification activity** v is defined as a tuple (p, r) , where $p \in F(z)$ and $r \in R$. A verification activity is therefore understood as the application of a verification procedure r to the discovery of knowledge about a system parameter p.

This mathematical framework overcomes the limitations of previous work. First, it recognizes the existence of various verification activities and the notion that different activities may be used simultaneously to verify a single system characteristic. Second, it incorporates the capability to distinguish verification strategies as a function of their sequences, not just their verification activities. Third, it does not impose any limitation on the valuation function in terms of separability. This will be shown later in the Valuing a Verification Strategy section.

A Bayesian Network to Capture Information Dependencies

Since the mathematical construct for describing a verification strategy presented in the previous section is a directed graph, it enables seamlessly embedding a Bayesian network to enable calculations related to beliefs or probabilities of errors existing in the system being developed and of the verification activities discovering those errors. We define now the Bayesian network machinery, as applied to model the knowledge discovery of a verification strategy. A detailed description of how to create the Bayesian network is given in Salado, Kannan, and Farkhondehmaal (2018). A summary follows.

Using the mathematical model presented in the previous section, consider a system z_0 built from components $z_{1,1}, \dots, z_{1,n}$ and a verification strategy $S = (V, D)$, where V is the set of verification activities $v_1, \dots, v_m$ and D is the set of tuples that capture information dependencies between the various verification activities $\{(v_i, v_j), \dots, (v_l, v_k)\}$, with $n, m \in$ and $v_i, v_j, v_l, v_k \in V$.

A Bayesian network that models such verification strategy can be constructed by combining three graphs. The first one contains directly the graph of the verification strategy, $S_1 = (V, D)$. The second graph the Bayesian networks contains captures the prior belief on the absence of errors in the various components that form the system and the system itself and the first verification activities executed on them. Mathematically, we can denote such graph as $I = (Z, A)$, where $Z = \{z_0, z_{1,1}, \dots, z_{1,n}\}$ and $A = \{v \in V : \exists z \in Z \text{ such that } P(v_{pass}|z) \neq P(v_{pass})\}$ and the outcomes of each verification activity v can only be v_{pass} and $\neg v_{pass}$. This graph captures the dependency between the prior knowledge about the components forming the system, including the system itself, and the first verification activities that are carried out in the verification strategy. Finally, the Bayesian network must contain the belief on the absence of error on the system z_0 as it relates to the belief on its components being absent of errors. Mathematically, we can denote such graph as $F = (Z, B)$, where $B = \{(z_i, z_j) : z_i, z_j \in V, P(z_i|z_j) \neq P(z_i)\}$. This graph captures the coupling between the different components forming the system, that is, how they inform the confidence on the proper functioning of the system.

In summary, the resulting Bayesian network is given by $BN = (V \cup Z, D \cup A \cup B)$.

Valuing a Verification Strategy

We have defined four value metrics for verification strategies:

1. The probability of the system exhibiting an error during operation, given that all verification activities were successful (note that this type of error relates to malfunctioning, not derived from reliability).
2. The minimum cost associated to the verification strategy, that is, the cost of the verification strategy assuming that no error is found during the execution of the whole verification strategy.
3. The maximum cost associated to the verification strategy, that is, the cost of the verification strategy assuming that errors are found and corrected as late as possible.
4. The expected cost associated to the verification strategy, which considers the possibilities of finding and correcting errors along the execution of the verification strategy.

The four metrics can be combined in a common tradespace, where cost and probability of the system exhibiting an error are on the axes, and the different ranges of cost are shown with bars. Now we define the four metrics mathematically.

Metric 1. The probability of the system exhibiting an error during operation, given that all verification activities were successful, is directly given by the Bayesian network described in the previous section. Hence, no further description is needed.

Metric 2. The minimum cost associated to a verification strategy is directly the investment necessary to execute the verification strategy. Simplistically, $c_{ex}(S) = \sum_{v \in S} c_{ex}(v)$, where $c_{ex}(v)$ is the cost of executing verification activity v.

Metric 3. The maximum cost associated to a verification strategy is given by the investment necessary to execute the verification strategy and the cost of fixing all possible errors, which are identified on the last verification activity where they could be identified (in terms of sequence of activities).

Metric 4. The expected total cost of a verification strategy is given by $E[c_{TOTAL}(S)] = E[c_{ex}(S)] + E[c_f(S)]$, where $E[c_f(S)]$ is the expected cost of fixing errors. We assume in this paper that an error is fixed as soon as it is discovered and that a fixed error does not reemerge once it has been fixed. Under these conditions, we define $E[c_f(S)] = \sum_{i=1}^{\infty} \sum_{j=1}^{\#V} P(e_{i,j}) \cdot P(d_{i,j} | e_{i,j}) \cdot c_f(e_{i,j})$, where $P(e_{i,j})$ is the probability that the system exhibits error e_i when verification activity v_j is executed, $P(d_{i,j} | e_{i,j})$ is the probability that verification activity v_j can discover error e_i (the discovery event is denoted by $d_{i,j}$), and $c_f(e_{i,j})$ is the cost of fixing the error e_i when discovered by activity v_j. An error e_i will be exhibited by a system during the event v_j if at least one of two conditions is met. The first one is met when the error emerges after completion of v_{j-1} and before completion of v_j. The second one is met when the error has emerged earlier, but has not been discovered by previous verification activities. Hence, $P(e_{i,1}) = P(e_i \text{ em } 1)$ and

$$P(e_{i,j}) = \sum_{k=1}^{j-1} \left[P(e_i \text{ em } k) \cdot \prod_{l=1}^{k} (1 - P(e_i \text{ em } l\text{-}1)) \cdot \prod_{m=k}^{j-1} \left(1 - P(d_{i,m} | e_{i,m})\right) \right] + P(e_i \text{ em } j) \cdot \prod_{k=1}^{j-1} (1 - P(e_i \text{ em } k)),$$

for $j \geq 2$, where $P(e_i \text{ em } j)$ is the probability that error e_i emerges after completion of v_{j-1} and before completion of v_j , and $P(e_i \text{ em } 0) = P(d_{i,0} | e_{i,0}) = 0$. The effect of the entire strategy is then incorporated by noting that the probability of an error being exhibited during a certain verification activity depends on its inherent nature of appearing at that point, as well as on the inability of the verification strategy to identify it earlier, if it emerged at an earlier point. It should be noted that these dependencies are defined by the Bayesian network presented in the previous section.

Challenges in Implementing the Framework to a Sample Case

Overview of the Problem

The application of the presented framework to a sample case is an ongoing effort. While the results will be presented at a different venue, we discuss in this section the challenges associated to operationalizing the mathematical framework presented in this paper.

The sample case in Salado et al. (2018) is used as a starting point. The system of interest is a simplified version of the Electric Power System (EPS) of the FireSat satellite (Wertz & Larson, 1999). A hierarchical breakdown of the system structure is depicted in Figure 3. The system model captures different levels of development maturity in the components that build the system (ECSS, 2009). Specifically, it is assumed that the EPS and PCDU need to be fully developed, the SA is based on an existing unit but needs some modifications, and the battery is recurring from a previous program.

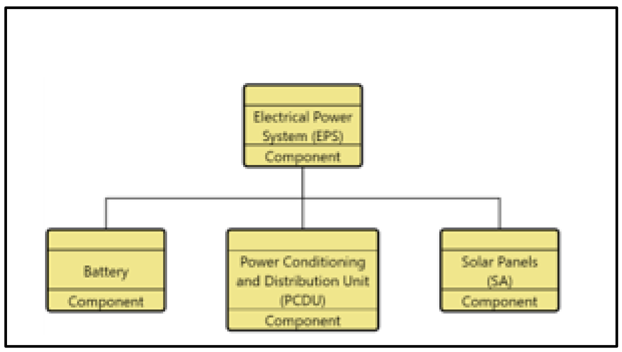


Figure 3. Simplified Firesat EPS Physical Hierarchy
(Salado et al., 2018)

For simplicity, we also assume that there is only one system characteristic that is verified and that verification can be achieved by analysis, test, or analysis and test on each building block in Figure 3.

Process

As discussed in the previous section, we start to populate the tradespace by applying the combinatorial effort to a fixed pattern of sequences of verification activities. This is done to limit the necessary computational effort, in particular in terms of eliciting conditional probabilities. We use as a base case the notional verification strategy defined in Salado et al. (2018), which is depicted in Figure 4. It reflects the order in which verification activities are executed (from top to bottom), as well as the information dependencies between them in the form of arrows. The verification strategy is defined as generic, without targeting any specific system characteristic.

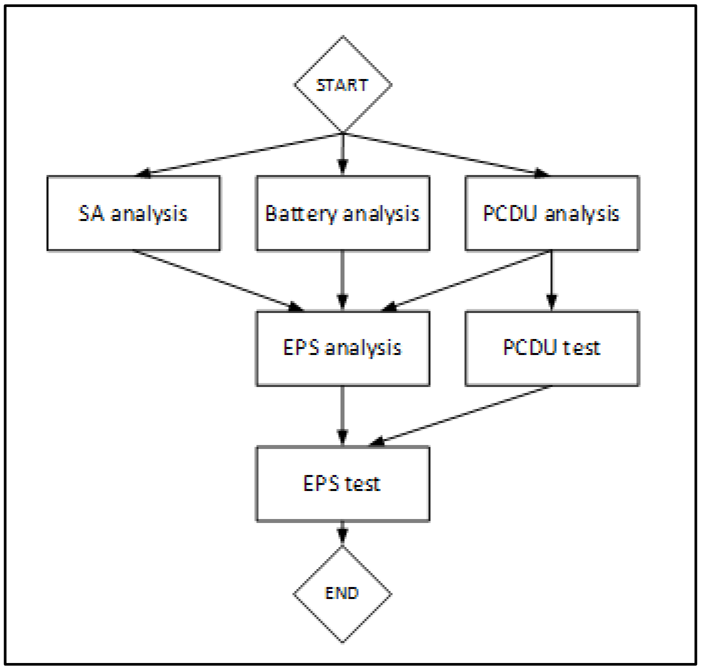


Figure 4. Base Case Verification Strategy
(Salado et al., 2018)

Therefore, the first tradespace is formed by all verification strategies that can be formed by applying or not applying each of the verification activities in Figure 4. Then, we incorporate a relative sequence change, leading to the modified base strategy and incorporate to the tradespace all verification strategies resulting from the combinations of applying or not applying each of the verification activities. Finally, we incorporate another relative sequence change, leading to a second modified base strategy and repeat the operation.

Valuation of verification strategies is also performed in two stages, in order to limit computational effort. First, each verification strategy in the tradespace is characterized by its minimum cost and the knowledge they discover. As previously described, minimum cost is defined as the execution cost of the strategy and knowledge discovery is defined as the probability of the system exhibiting once operational after all the verification activities in the sequence have been successful. Second, a subset of verification strategies that provide a similar knowledge discovery is selected and their expected and maximum costs, which depend on the actual finding of errors and exercise of repair costs, are calculated, forming a new tradespace.

Challenges and Way Forward

Operationalizing the presented mathematical framework into a computational code has presented some challenges. Two deserve particular attention.

The first one, inherent to tradespace exploration, is the exponential growth of the tradespace with verification activities and system characteristics. However, problem size becomes larger and more intricate due to the dependencies between the activities. In particular, it should be noted that given a set of n possible verification activities, there are 2^n sets of verification activities, and $4^{\binom{n}{2}}$ directed graphs for each one of the sets of verification activities. In addition to the common methods employed in tradespace exploration to reduce the size of the problem, identifying in advance the independence between knowledge generation helps in identifying equivalent sequences, thus reducing the size of the problem.

The second issue is related to the availability of conditional probability tables for each verification strategy. Conditional probability tables depend heavily on the specific sequences and reusing and adjusting them seems to be non-trivial. In a worst-case scenario, a dedicated set of conditional probability tables would need to be created for each verification strategy in the tradespace. Of course, this is infeasible. Furthermore, the nature of conditional probabilities in verification strategies makes it difficult to create a model that can be used to automate the generation of conditional probability tables. This problem can be overcome with a sufficiently large database of historical performance of verification strategies executed on systems similar to the system of interest. Given the lack of a publicly available database of such kind, we are currently developing a synthetic database of verification strategies for Earth observation satellites to support this study.

Conclusions

We have presented in this paper a framework to apply tradespace exploration to the design of verification activities. The framework is built on mathematical machinery that enable the automated generation of verification strategies, the computation of the knowledge they discover, and the valuation of the consequences of executing them.

The proposed frame.work overcomes the limitations of previous work. In particular, the proposed framework recognizes the existence of various verification activities and the notion that different activities may be used simultaneously to verify a single system characteristic. Moreover, it is able to capture the dependencies between verification activities, enabling distinguishing verification strategies as a function of their sequences, not just their verification activities. Furthermore, the proposed framework does not impose any limitation on the valuation function in terms of separability.

Finally, we have discussed the challenges that we are finding when operationalizing the mathematical framework to apply it to a sample case. The effort is ongoing and is planned to be completed within the timeframe of the NPS Research Acquisition Program's "Tradespace Exploration for Better Verification Strategies" project.

References

Buede, D. M. (2009). The engineering design of systems: *Models and methods*. Hoboken, NJ: Wiley.

ECSS. (2009). *Space engineering—Verification*. Noordwijk, The Netherlands: European Cooperation for Space Standardization.

ECSS. (2012). *Space engineering—Testing*. Noordwijk, The Netherlands: European Cooperation for Space Standardization.

Engel, A. (2010). *Verification, validation, and testing of engineered systems*. Hoboken, NJ: John Wiley & Sons.

Engel, A., & Shachar, S. (2006). Measuring and optimizing systems' quality costs and project duration. *Systems Engineering, 9*(3), 259–280. doi:10.1002/sys.20056

Firesmith, D. (2013). *Common testing pitfalls and ways to prevent and mitigate them: Descriptions, symptoms, consequences, causes, and recommendations*. Boston, MA: Addison Wesley.

Golkar, A., & Crawley, E. F. (2014). A framework for space systems architecture under stakeholder objectives ambiguity. *Systems Engineering, 17*(4), 479–502. doi:10.1111/sys.21286

INCOSE. (2011). *INCOSE Systems engineering handbook* (v.3.2.2). San Diego, CA: Author.

Larson, W. J., Kirkpatrick, D., Sellers, J. J., Thomas, D., & Verma, D. (2009). *Applied space systems engineering.* McGraw Hill

Mattson, C. A., & Messac, A. (2003). Concept selection using s-pareto prontiers. *AIAA Journal, 41*(6), 1190–1198.

Ross, A. M., & Hastings, D. E. (2005). 11.4.3 the tradespace exploration paradigm. *INCOSE International Symposium, 15*(1), 1706–1718. doi:10.1002/j.2334-5837.2005.tb00783.x

Ross, A. M., Hastings, D. E., Warmkessel, J. M., & Diller, N. P. (2004). Multi-attribute tradespace exploration as front end for effective space system design. *Journal of Spacecraft and Rockets, 41*(1), 20–28.

Salado, A. (2015). Defining better test strategies with tradespace exploration techniques and pareto fronts: Application in an industrial project. *Systems Engineering, 18*(6), 639–658. doi:10.1002/sys.21332

Salado, A. (2016). *Applying tradespace exploration to verification engineering: From practice to theory and back again*. Paper presented at the Conference on Systems Engineering Research (CSER), Huntsville, AL.

Salado, A., Kannan, H., & Farkhondehmaal, F. (2018)*. Capturing the information dependencies of verification activities with Bayesian networks*. Paper presented at the Conference on Systems Engineering Research (CSER), Charlottesville, VA.

Wertz, J. R., & Larson, W. J. (1999). *Space mission analysis and design*. Microcosm.

Wymore, A. W. (1993). *Model-based systems engineering*. Boca Raton, FL: CRC Press.

Acknowledgments

This material is based upon work supported by the Naval Postgraduate School Acquisition Research Program under Grant No.N00244-17-1-0013. The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

A Review of Trusted Broker Architectures for Data Sharing

John R. Talburt—University of Arkansas, Little Rock. [jrtalburt@ualr.edu]

Abstract

Sharing data across organizational boundaries must strike a balance between the competing data quality dimensions of access and security. Without access to data, it can't be used and, consequently, is of no value. At the same time, uncontrolled access to data, especially sensitive personal data, can result in dire legal and ethical consequences. This paper discusses the trade-offs between security and access for three styles of trusted broker architectures in hopes that this will provide guidance for organizations trying to implement data sharing systems.

Background

There are three important topics that provide the background for the discussion of data sharing architectures. These topics are

1. The difference between data sharing and data exchange
2. The need for persistent data integration to enable data sharing
3. The trusted broker as a vehicle for data sharing

Each of these topics is explored in the following sections of this paper.

Data Sharing Versus Data Exchange

Although data exchange is a form of data sharing, it carries a particular connotation with respect to control and time. Data exchanges in which one organization transfers data to another organization are typically one-to-one, episodic exchanges. Even though there may be a standing agreement between the two organizations, the actual exchanges are usually time-bound in the sense that the sourcing organization provides a snapshot of its data at the time of the transfer to the consuming organization with no further interaction or updating of the data by the sourcing organization. Often the data sharing agreement will require the consuming organization to stop using and delete the data within a specific time period. Even in the case that there is not a specific end date, the fact that the quality and utility of unmaintained data tends to decrease over time has essentially the same effect.

The transfer of data also implies a transfer of control. Once in the hands of the consuming organization, the data are processed and manipulated by the consuming organization without the direct participation of the sourcing organization. The sourcing organization only has indirect control vis-à-vis any requirements or constraints in a written "exchange agreement."

Furthermore, the benefits of using the exchanged data accrue primarily to the receiving organization. Even if there is a reciprocal transfer of data from the receiving organization back to the sourcing organization (true data exchange), the benefits derived from the aggregation of the data are not necessarily reciprocal. Each organization may receive some type of benefit from the exchange, but these are not necessarily the same shared benefit.

On the other hand, data sharing posits a system accommodating participation by several organizations all having joint control and continual access and all deriving mutual benefit from the use of the data. Although data sharing can be done for all types of entities (persons, places, and things), the data sharing systems discussed here are those designed to share data about people (person entities) and organizations (legal entities). In the master data management parlance, data about persons and organizations are referred to as "party" data. Examples of organizations benefiting from data sharing systems include healthcare (providers and payers sharing patient data), law enforcement (city, state, and national agencies sharing criminal and persons-of-interest data), and government (service agencies sharing data about their clients).

The primary mutual benefit derived from data sharing systems is an improvement in the data quality dimension of completeness. Whereas each organization holds a "piece of the puzzle," when the data are shared, all organizations have the complete picture. At the same time, better completeness leads to better data analysis and insights. For example, a physician is able to make a better diagnosis when he or she is able to view all previous tests and treatments rather than just those done within his or her clinic.

In order to gain a complete view of party information over time, the data sharing needs to enable persistent data integration based on shared, persistent identifiers of the parties (people and organizations). The next section discusses the concepts of transient and persistent data integration.

Transient Versus Persistent Data Integration

Data integration processes can be classified into two different categories based on the persistence of the integration. The two processes are illustrated in Figure 1.

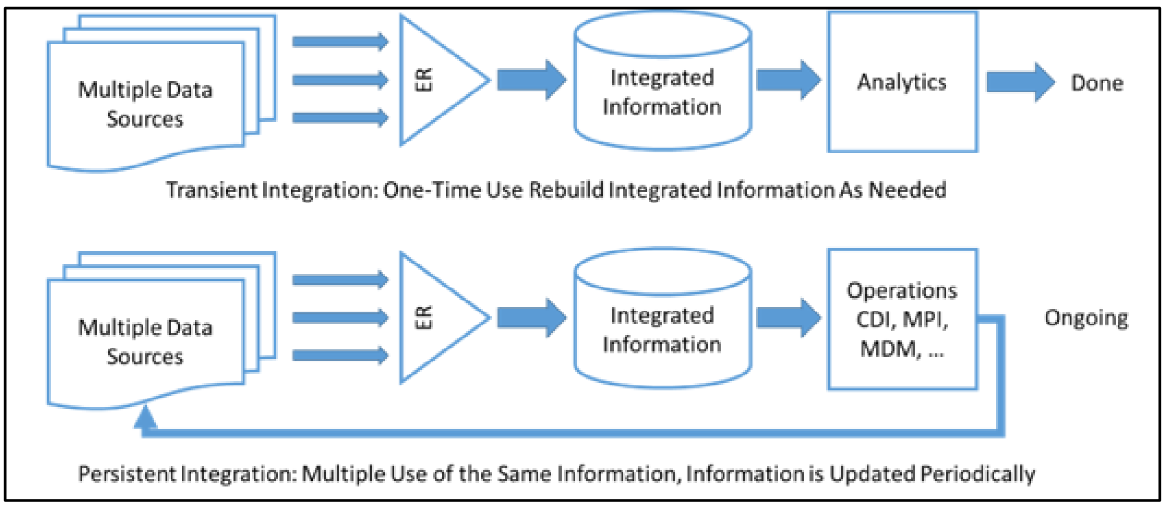


Figure 1. Transient Versus Persistent Data Integration

There is a difference between data integration processes used to support projects such as data analytics as opposed to data integration supporting ongoing operations such as MDM. Both processes rely on ER to link together the information related to the same entity of interest (e.g., the same patient).

In a transient data integration process, ER alone is sufficient. Once the analysis of the integrated information is complete, the process (project) is over. Transient data integration represents data integration on demand. Each time a new analysis or project is needed, the ER process is run again to re-integrate the information. Bringing together data on demand for analysis is sometimes referred to as "data wrangling," a quick and dirty data integration process not intended to be operationalized.

On the other hand, there are many other data operations for which transient data integration is not well-suited. In particular, MDM cannot be managed effectively using transient data integration. The two motivations for using persistent data integration instead of transient data integration are reuse of effort and changes in characteristic information over time.

Taking the same sources of information and re-integrating them again and again can expend a lot of time and effort, especially if the volume of information is large. For example, suppose you want to generate a report of total sales by customer on a weekly basis. With transient data integration, all of the customer sales transactions must be merged and the ER process run against the combined file. In a persistent data integration process, the groups of sales records from one week are saved, and only the new sales from the next week are incrementally added to the previous groups.

Another reason for persistent data integration is that the characteristic information describing real-world entities changes over time. Characteristic information represents the properties that, taken together, uniquely define a specific entity. Characteristics are also called “identity attributes.” For example, there may be many students in a college with the first name LINDA. Fewer students, but more than one, may have the last name HENDERSON. But only one student will have 1987-11-06 as her date of birth and live on ELM STREET. The values of these attributes taken together identify a particular student.

The problem is these identity attributes are not immutable. LINDA may move to a new address, or she may marry and begin to use her husband’s surname. These changes speak to the need to store and manage identity information in order to consistently identify the same entity over time.

Identity Management and Persistent Identifiers

An Entity Identity Information Management (EIIM) system is an additional layer of ER functionality, namely, the ability to assign the same identifier to references to the same entity over time (i.e., from process to process). EIIM is the collection and management of identity information with the goal of sustaining entity identity integrity over time (Zhou & Talburt, 2011). Entity identity integrity is the state where each entity managed by the system is assigned a unique identifier, and distinct entities are assigned distinct identifiers (Maydanchik, 2007). Entity identity integrity is a basic requirement for MDM systems.

EIIM comprises several processes. Each EIIM process uses aspects of ER and data structures representing the entity identities to accomplish specific goals. These work in concert to maintain the entity identity integrity of master data over time. EIIM is not limited to MDM. It can be applied to other types of systems and data as diverse as RDM systems, referent tracking systems (Chen et al., 2013), and social media (Mahata & Talburt, 2014).

Although ER is a necessary component of MDM, ER is not in itself sufficient to manage the life cycle of identity information. EIIM is an extension of ER in two dimensions, identity knowledge management and time. The knowledge management aspect of EIIM relates to creating, storing, and maintaining identity information. The knowledge structure representing a master data object (entity) is an entity identity structure (EIS).

An EIS is essential to assuring an entity under management in the MDM system is consistently labeled with an identifier from process to process. The EIS stores the identity information of the entity along with its identifier so both are available to future processes. Persistent entity identifiers are not inherently part of ER. At any given point in time, the only goal of an ER process is to correctly classify a set of entity references into clusters where all of the references in a given cluster are equivalent. The labels (identifiers) of these clusters serve as the identifier of the entity. Without storing and carrying forward identity information in the EIS, the identifiers assigned to entities and entity references may change in future processes.

Figure 1 illustrated the problem. It shows three references, References 1, 2, and 3. References 1 and 2 are equivalent references, and Reference 3 is to a different entity. References 1, 2, and 3 are in File 1 along with other references when the first ER process runs. The same References 1, 2, and 3 are also present in File 2, and the same ER process is run against File 2 at a later time. In both cases, if the ER process is accurate, it will classify References 1 and 2 as equivalent and place Reference 3 in a different cluster. The problem is that where there is no communication between the processes, even though the references are clustered properly, they will most likely be labelled differently (i.e., given different identifiers), as shown in Figure 2.

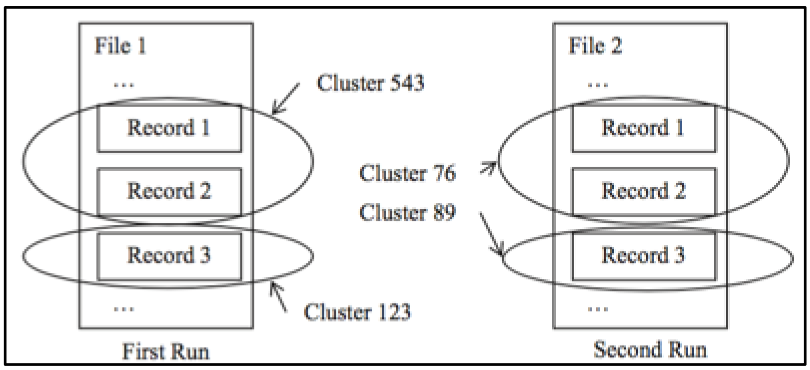


Figure 2. ER With Consistent Classification but Inconsistent Labeling

In the first run, the cluster containing References 1 and 2 is identified as Cluster 543, whereas in the second run, the same cluster is identified as Cluster 76. The purpose of the EIS is to provide the communication between processes so that a reference to an entity in one process is labeled by the same identifier to the same entity in later processes.

ER that only classifies records into clusters representing the same entity is sometimes called a "merge-purge" operation. In a merge-purge process, the objective is simply to eliminate duplicate entity references. In this context, the term "duplicate" does not mean that the references are exactly the same in content. Instead, it means the references are for the same entity. To avoid this confusion, the term "equivalent" is preferred (Talburt, 2011) to describe references to the same entity.

The term equivalence also avoids confusion arising from designation of "matching" references. References to the same entity do not necessarily have matching information. For example, two records for the same customer may have different names and different addresses. Conversely, matching references do not necessarily refer to the same entity. Two references can be quite similar when important identity information is missing. For example, references to John Doe, Jr., and to John Doe, Sr., may be classified as matching records if the suffix Jr. or Sr. has been omitted from one or both of the references.

Unfortunately, many authors use the term "matching" to mean two references are very similar and also to mean they reference the same entity. The result can often be confusing for the reader. Reference matching and reference equivalence are different concepts and should be described by different terms.

The ability to consistently assign a cluster the same identifier when an ER process is repeated at a later time requires an EIS to carry forward identity information from process to process. The storage and management of identity information in the EIS is what enables the persistence of entity identifiers. Persistent identifiers are the value add EIIM brings to ER.

Different EIIM systems implement different strategies for identity information management. Figure 3 shows one example from the OYSTER open source system (Talburt & Zhou, 2013; Zhou et al., 2010).


```xml
<root>
  <Metadata>
    <Modifications)
      <Modification ID=”1” OysterVersion=”3.2” Date=”2013-03-29 04.51.07” RunScript=”Run001” />
    </Modifications>
    <Attributes>
      <Attribute Name=”@RefID” Tag=”A” />
      <Attribute Name=”Phone” Tag=”B” />
      <Attribute Name=”FirstName” Tag=”C” />
      <Attribute Name=”StrNbr” Tag=”D” />
      <Attribute Name=”LastName” Tag=”E” />
      <Attribute Name=”SSN” Tag=”F” />
    </Attributes>
  </Metadata>
  <Identities>
    <Identity Identifier=”X9KTZ5GOQ5RVHOWV” CDate=”2012-03-29”>
      <References>
        <Reference>
          <Value>A^ListA.A953698|C^ANTONIOV|D^247H|E^CARDONA|F^196369947</Value>
          <Traces>
            <Trace> OID=”X9KTZ5GOQ5RVHOWV” RunID=”1” Rule=”@”/>
          </Traces>
        </Reference
        <Reference>
          <Value>A^ListA.A989582|C^ANTONIOV|D^5221|E^CARDONA|F^196369974</Value>
          <Traces>
            <Trace> OID=”X9KTZ5GOQ5RVHOWV” RunID=”1” Rule=”@”/>
          </Traces>
        </Reference
      </References>
    <Identity Identifier= ...
    ...
  </Identities>
</root>
```


Figure 3. Example of an EIS in XML Format Created by the OYSTER ER System

During processing, the OYSTER EIS exists as in-memory Java objects. However, at the end of processing, the EIS are serialized in an XML document. The XML structure encapsulates extensive metadata to reflect not only the entity identifier (OID value), but also many other aspects of the process such as the sources of the references, the label of the matching rule that linked a reference into the cluster, the identifier of the process (run) when the reference was brought into the cluster, and other information. The XML format also serves as a way to serialize the EIS objects so that they can be reloaded into memory at the start of a later ER process.

Components of Entity Information

To better understand the implementation of EBDI, it is helpful to think of entity information as having three components:

1. A unique and persistent entity identifier. This is a single value assigned as the unique identifier of a particular instance of the entity. It is used to label all information related to the entity. The intent is the identifier should not change (persist) over time. However, processing errors may mandate changing the identifier. An important EBDI goal is to continually improve the identification process and thereby minimize changes to entity identifiers. Best practice is to make the identifier entirely anonymous, such as using a Really Simple Syndication (RSS) hashing algorithm on some identity fields to create an identifier value. An example is the entity identifier "X9KTZ5GOQ5RVHOWV" shown in Figure 2. Encoding information into the identifier can increase the likelihood the value of the identifier may have to be changed to reflect changes in the encoded information.
2. Entity identity specific information. These are identity attributes, sometimes called characteristic information (ISO 8000), that, when taken together, distinguish one entity from another. Another important goal of EBDI is the reuse of identification effort. In the EBDI process, the identity attribute values of entity records are resolved to the entity identifier. The entity identifier is then used to label the record. Later processes can then simply use join or sort records by the entity identifier to aggregate entity information, relieving these processes of the need to apply their own (and possibly less accurate) matching logic.
3. Application specific information associated with an entity (multiple values).

The identity management components are shown in Figure 4.

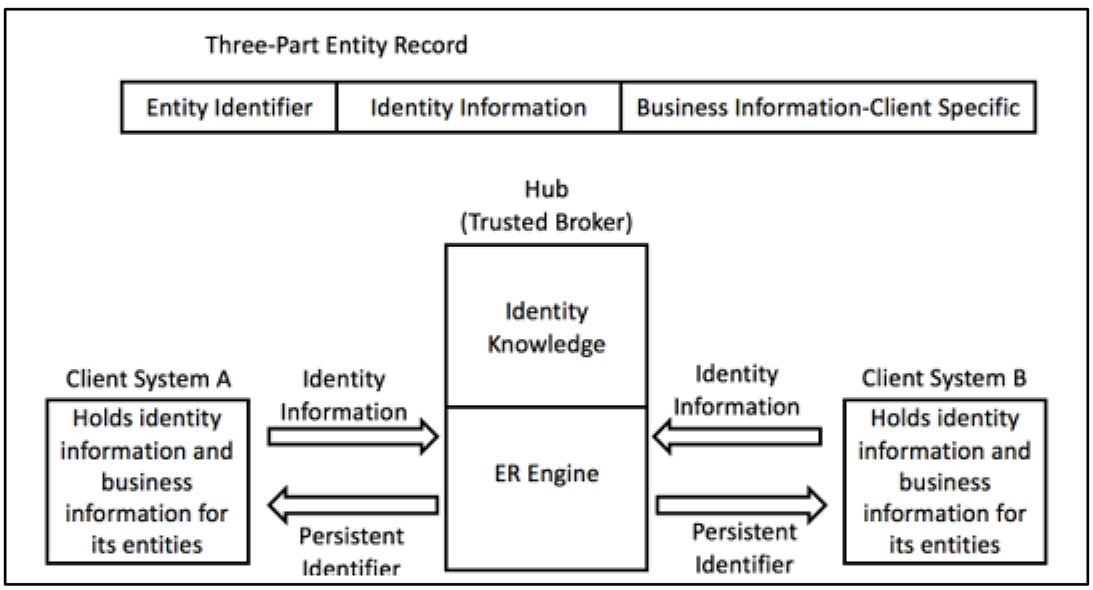


Figure 4. The Identity Management Component

Entity-Based Data Integration (EBDI)

Entity-Based Data Integration (EBDI) is the process of bringing together information about the same real-world entity. From an EBDI perspective, an entity is a set of real-world objects that have distinct identities, for example, customers of a business, patients of a hospital, students of a school system, products of a manufacturer, and locations of service. The fundamental issue with EBDI is that in the modern world, we describe these entities as data records in an information system. As we all know, the blessing and curse of storing information in a computer is how easy it is to copy and replicate information. As a result, there are often many different records in the information system all describing the same entity.

Entity-based data integration has a broad range of applications in areas such as law enforcement (Nelson & Talburt, 2008), education (Nelson & Talburt, 2011; Penning & Talburt, 2012), and healthcare (Christen, 2008; Lawley, 2010). For example, a hospital patient will have separate records for each hospital visit. In addition, each visit or encounter will generate many additional records such as a record for each laboratory test, each treatment, each drug administered, and so on. Aside from the records of medical treatment, there will also be billing records associated with each visit and treatment.

Consequently, EBDI is a two-step process (Talburt & Hashemi, 2008). When integrating entity information from multiple sources, the first step is to determine whether the information is for the same entity. Once it has been determined the information is for the same entity, the second step is to reconcile possibly conflicting or incomplete information associated with a particular entity coming from different sources (Holland & Talburt, 2008, 2010; Zhou, Kooshesh & Talburt, 2012). Master data management (MDM) systems play a critical role in successful EBDI by providing an EIIM process, enabling the MDM system to consistently identify and label references to the same entity.

Three Trusted Broker Architectures for EBDI

- External Reference Architecture
- Registry Architecture
- Transaction Hub
  (Berson & Dubov, 2011)

External Reference Architecture

In the external reference architecture (shown in Figure 5), the IKB is a large cross-reference table connecting equivalent references located in the various client systems. The EIS in the IKB are entirely virtual, only containing pointers to the references to a particular entity. None of the actual entity identity information for a particular entity is stored in the IKB.

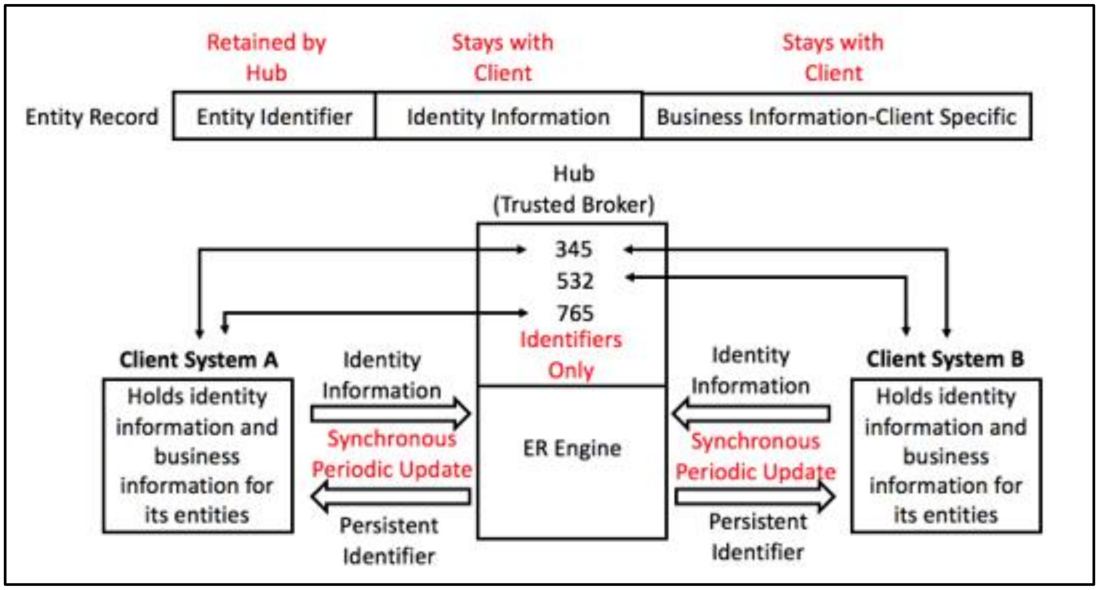


Figure 5. The External Reference Architecture

Both the identity attribute values and the application-specific attribute values of the source record reside in the client system, as shown in Figure 6. The advantage of external reference architecture is that changes to an entity identifier taking place in one system can be more easily propagated to all other client systems where the same entity is referenced.

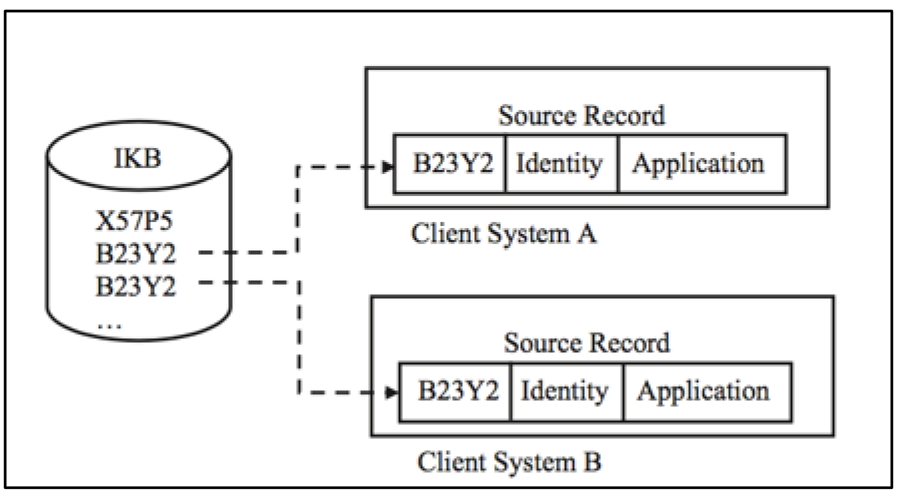


Figure 6. External Reference Schema

The external reference architecture works best when the governance policy allows for distributed authority to make master data changes in several different client systems. It does not work as well in systems where a large number of new source records must be ingested and identified on a regular basis. In systems implementing external reference architecture, the identity information needed for matching must be marshaled on demand from the client systems where it resides.

External Reference Architecture Pluses

- The hub (broker) does not retain identity information between updates, so there is less risk of identity leakage.
- Identifiers are anonymous.
- Lightweight IT Infrastructure, periodic bulk processing, no transactional processing, less staffing.
- Less complex ER processing is required.
- Detection of ER linkage errors is done by client personnel.
- When found, errors are reported to the hub, where they are corrected during periodic processing.

External Reference Architecture Minuses

- Linkage between client records is only as current as the last update.
- All clients must provide all of their identity information at the same time to the hub in order to update the cross-reference table.
- Because the hub does not hold identity information, inquiries for specific entities must be made through a client.
- Only after an entity is identified by a client can the hub can tell if other clients have records for the same entity.
- Analytics is less agile. Must first collect and combine information from each agency onto one platform before starting analysis.

Registry Architecture

A more common architecture for MDM applications ingesting and managing large volumes of input data is registry architecture. In registry architecture, each EIS in the IKB contains a collection of identity attribute values representing the entity under management. Each EIS has an identifier serving as the master identifier for the entity across all client systems. The amount of identity information retained in the EIS will vary from system to system and on the choice of EIS strategy as discussed earlier (e.g., survivor record, attribute-based, record-based, etc.).

In registry architecture, each reference is divided into two parts, as shown in Figure 7. The values for the identity attributes are kept in the IKB. The IKB must retain sufficient identity information so that when a new source record is introduced, the system can correctly determine if it references a previously registered entity already having established entity identifiers or if it references a new entity which must be registered in the IKB with a new entity identifier.

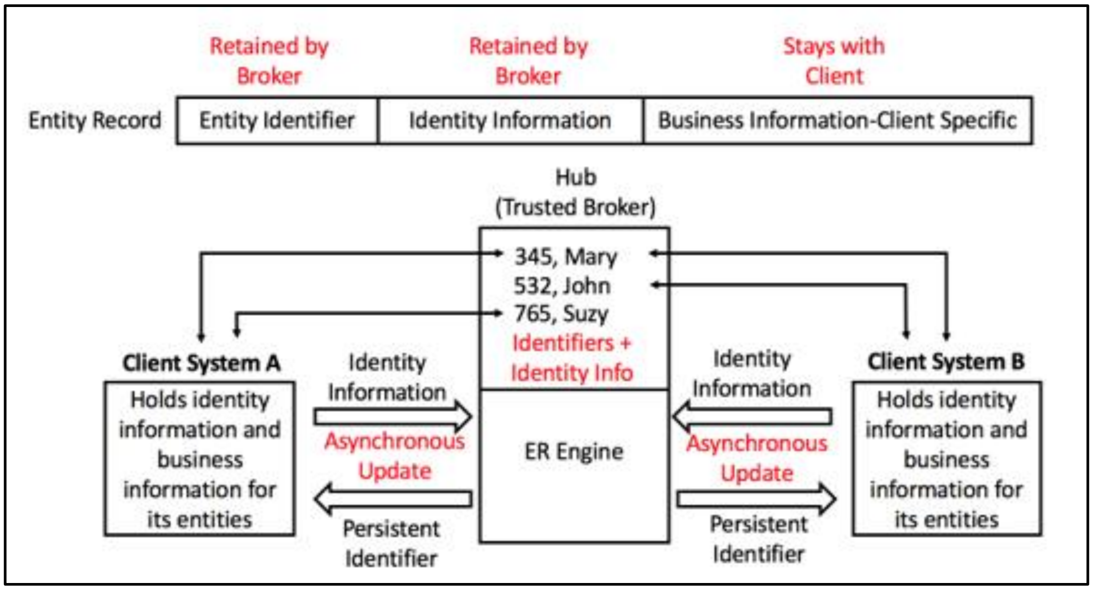


Figure 7. The Registry Architecture

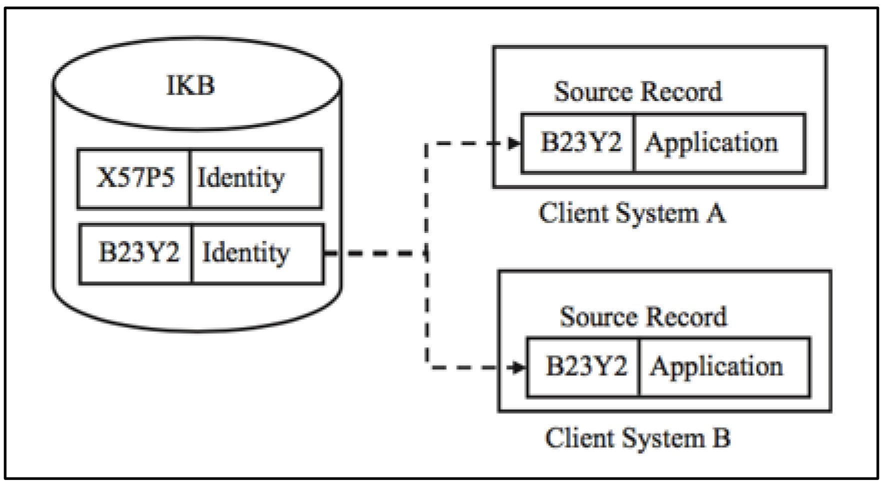


Figure 8. Registry Architecture Schema

A third possibility is if a new source record carries additional identity information providing evidence that two EIS initially thought to represent distinct entities are actually references to the same entity. For example, a source record having both a current and previous address for a customer connects an EIS created for that customer at the previous address and another EIS created for the same customer at the current address. As a result, one of the entity identifiers is retired, usually the most recently assigned, and the identity information in the two EIS are merged.

In registry architecture, the values for client-specific attributes are retained in the source records residing in the client systems. The two halves of the source record the identity values and the client-specific values are linked together by the entity identifier. The shared link identifier also does away with the need to store the identity values of the reference in the client system. The replacement of entity identity values with a unique identifier for the entity is called semantic encoding. Semantic encoding is one of the fundamental principles of the ISO 8000-110:2009 Standard for Master Data Quality.

In registry architecture, the IKB and the systems are loosely coupled. It is usually the responsibility of each client system to synchronize its entity identifiers with the identifiers in the registry through periodic batch or interactive inquiries to the registry. The registry architecture is typical for most CDI systems primarily providing an identity management service (i.e., appends link identifiers to source records on demand).

Registry architecture is sometimes used to provide anonymous entity resolution to several external clients in a trusted broker configuration (Talburt et al., 2005). Trusted broker architecture can be useful when each external client holds some information for entities also known to other clients but also manages information for some entities unique to the client organization. The clients want to collaborate and share information about common entities but do not want their exclusive entity information shared or exposed to other clients. This situation often arises in law enforcement, healthcare, and information sharing among government agencies.

The name comes from the fact that all of the clients must trust one neutral organization to host the hub of the registry. In addition, even though the hub internally maintains only one set of entity identifiers, it issues a different set of entity identifiers to each client. This means even though two different clients hold information about the same entity, the hub will give each client a different identifier for that same entity. The hub mediates the translation between client identifiers. In this way, the trusted broker also incorporates some features of the external reference architecture.

If Client A wants to know whether Client B is holding information about an entity of interest, Client A sends an inquiry to the hub organization. The hub organization can then translate the Client A identifier into its internal identifier and determine if Client B has information on the entity. The hub organization can also mediate policies or regulations on access. If, according to policy, Client A's inquiry is valid, then the hub can send the information from Client B to Client A using the entity identifier of Client A.

Registry Architecture: Pluses

- Clients can update information in the hub according to their own schedule and are not required to coordinate with other clients.
- Depending upon client update schedules, the linkage between clients can be more up-to-date than the external reference.
- Because the hub holds identity information, inquiries for a specific entity can be made at the hub and immediately find which clients hold information for that entity.
- Detection of linkage errors can be done by both hub and client personnel.

Registry Architecture: Trade-Offs

- The hub (broker) retains all identity information between updates, so there is a greater risk of identity leakage.
- Analytics on business information is still less agile.
- Must first collect business information from each agency onto one platform before starting analysis.
- Requires heavier IT infrastructure and staff to service ongoing updates and linkage corrections.
- Requires more complex ER functions including a system for publishing identifier changes (due to linking errors) to clients.

Transaction Hub: Manages Identity and Business Information

The transaction hub architecture is also a hybrid. It attempts to solve both the attribute partitioning problem and the synchronization problem at the same time. In this case, the hybridization is between the IKB and its client systems. In a transaction hub, the IKB stores the complete source record, both identity attributes, and application-specific attributes.

By incorporating the source records into the IKB, the transaction hub is simultaneously an MDM system and an application system. The transaction hub can be a good solution for situations where the system must process large volumes of new source references while at the same time servicing high volumes of inquiries for application-specific information because the application information is immediately at hand. There is no need to fetch the application information from a client system in order to service the inquiry. However, this is only feasible if only one or two applications are integrated with the hub; otherwise, the maintenance of the system becomes too complex to manage. Many financial systems incorporate the transaction hub architecture for MDM.

Transaction Hub Architecture

The transaction hub architecture is shown in Figure 9.

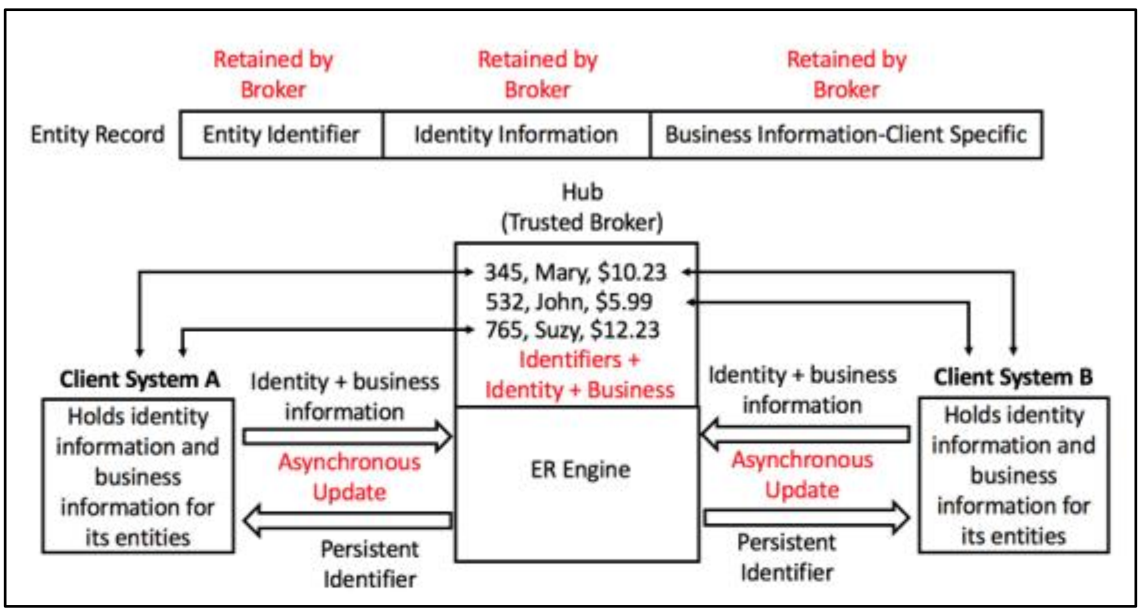


Figure 9. The Transaction Hub Architecture

Transaction Hub: Pluses

- The biggest advantage is that analytics is simpler than with registry or external reference. All of the information to analyze is already in one platform. There is no need to request and join information from clients.
- Clients can update information by their own schedule.
- Depending upon client participation, the linkage between clients can be more up-to-date than the external reference.

Transaction Hub: Trade-Offs

- The hub (broker) retains all identity and business information between updates, so there is a greater risk of not only identity leakage, but also leakage of business information.
- Requires more complex ER functions including a system for publishing identifier changes (due to linking errors) to clients.
- The ER system requires additional functionality to store and retrieve business information with different schemas.

Conclusion

The selection of an appropriate architecture to support data sharing depends primarily on finding an acceptable point of balance between data exposure risk and operational agility. Following the progression from data exchange agreements, external reference, and registry to full transaction hub leads to increasing aggregations of sensitive data in one system for longer periods of time. However, the same progression provides increasing ease of access and agility to support maintenance and data analytics. The real key to successful data sharing is a strong data governance program that implements and enforces compliance with the policies, assuring that the privacy and regulatory compliance of all the participating organizations are satisfied. Data sharing and MDM systems lacking strong data governance programs are doomed to failure.

References

Berson, A., & Dubov, L. (2011). *Master data management and data governance*. McGraw Hill.

Chen, C., Hanna, J., Talburt, J. R., Brochhausen, M., & Hogan, W. R. (2013, July). A demonstration of entity identity information management applied to demographic data in a referent tracking system. *International Conference on Biomedical Ontology (ICBO) 2013* (pp. 136–137). Montreal, Canada.

Christen, P. (2008, January). Febrl-A freely available record linkage system with a graphical user interface. In *Proceedings of the Australian Workshop on Health Data and Knowledge Management (HDKM): Conferences in Research and Practice in Information Technology (CRPIT*; vol. 80).

Holland, G., & Talburt, J. (2008). A framework for evaluating information source interactions. In C. Hu & D. Berleant (Eds.), *The 2008 Conference on Applied Research in Information Technology* (pp. 13–19). Conway, AR: University of Central Arkansas. Retrieved from http://research.acxiom.com/publications.html

Holland, G., & Talburt, J. (2010). An entity-based integration framework for modeling and evaluating data enhancement products. *Journal of Computing Sciences in Colleges, 24*(5), 65–73.

Lawley, E. (2010, March 29). Building a health data hub. *Nashville Post* (online version).

Mahata, D., & Talburt, J. R. (2014, August 13). A framework for collecting and managing entity identity information from social media. In *Proceedings of the 19th MIT International Conference on Information Quality* (pp. 216–233). Xi'an, China.

Maydanchik, A. (2007). *Data quality assessment*. Technics Publications.

Nelson, E., & Talburt, J. (2008). Improving the quality of law enforcement information through entity resolution. In C. Hu & D. Berleant (Eds.), *The 2008 Conference on Applied Research in Information Technology* (pp. 113–118). University of Central Arkansas, Conway, AR. Retrieved from http://research.acxiom.com/publications.html

Nelson, E., & Talburt, J. (2011, July). Entity resolution for longitudinal studies in education using OYSTER. In *Proceedings of the 2011 Information and Knowledge Engineering Conference (IKE 2011)* (pp. 286–290). Las Vegas, NV.

Penning, M., & Talburt, J. R. (2012, July). Information quality assessment and improvement of student information in the university environment. In *Proceedings of the 2012 International Conference on Information and Knowledge Engineering (IKE '12)* (pp. 351–357). Las Vegas, NV.

Talburt, J., & Hashemi, R. (2008). A formal framework for defining entity-based, data source integration. In H. Arabnia & R. Hashemi (Eds.), *2008 International Conference on Information and Knowledge Engineering* (pp. 394–398). Las Vegas, NV: CSREA Press.

Talburt, J., Morgan, C., Talley, T., & Archer, K. (2005). Using commercial data integration technologies to improve the quality of anonymous entity resolution in the public sector. In F. Naumann, M. Gertz, & S. Madnick (Eds.), *10th International Conference on Information Quality* (pp. 133–142). Cambridge, MA: MIT IQ.

Talburt, J., & Zhou, Y. (2013). A practical guide to entity resolution with OYSTER. In S. Sadiq (Ed.), *Handbook on research and practice in data quality* (pp. 235–270). Springer.

Talburt, J. R., 2011. *Entity resolution and information quality*. Morgan Kaufmann.

Zhou, Y., Kooshesh, A., & Talburt, J. (2012). Optimizing the accuracy of entity-based data integration of multiple data sources using genetic programming methods. *International Journal of Business Intelligence Research, 3*(1), 72–82.

Zhou, Y., & Talburt, J. R. (2011, November). Entity identity information management. *International Conference on Information Quality 2011.* Retrieved from http://iciq2011.unisa.edu.au/doc/ICIQ2011_Proceeding_Nov.zip

Zhou, Y., Talburt, J., Su, Y., & Yin, L. (2010, November). OYSTER: A tool for entity resolution in health information exchange. In *Proceedings of the Fifth International Conference on the Cooperation and Promotion of Information Resources in Science and Technology (COINFO '10)* (pp. 356–362). Beijing, China.

Towards Game Theoretic Models for Agile Acquisition

Scott Rosen—is the Chief Engineer for the Operations Research Department of the MITRE Corporation with more than 10 years of application experience. He holds a PhD and MS in Industrial Engineering and Operations research from The Pennsylvania State University and a BS in Industrial and Systems Engineering from Lehigh University. [srosen@mitre.org]

Kelly Horinek—is a Principal Analyst in the Center for Acquisition and Management Sciences Division of The MITRE Corporation in McLean, VA. She holds a BBA in Marketing from Texas A&M University, and is a Certified Professional Contracts Manager and a former Contracting Officer. [khorinek@mitre.org]

Alexander Odeh—is Lead Economist and Business Analyst in the Center for Acquisition and Management Sciences Division of The MITRE Corporation in Chantilly, VA. He holds a bachelor's degree in Economics from McGill University and a master's degree in Public Policy from George Washington University. [aodeh@mitre.org]

Les Servi—is a Group Leader in the Operations Research Department of The MITRE Corporation. He holds a PhD and MS in Engineering from Harvard University, and an ScM and ScB in Applied Mathematics from Brown University. He is a Fellow of INFORMS, a former editor of three journals, and a past member of two Defense Science Board studies. [lservi@mitre.org]

Andreas Tolk—is Technology Integrator for the Modeling, Simulation, Experimentation, and Analytics Division of The MITRE Corporation in Hampton, VA. He holds a PhD and MS in Computer Science from the University of the Federal Armed Forces, Munich, Germany. He is a Fellow of the Society for Modeling and Simulation and senior member of ACM and IEEE. [atolk@mitre.org]

Abstract

Game Theory has been applied to drive strategies for contract negotiations in the business world. This paper investigates the idea of applying game-theoretic utility models and strategies to provide a means to illuminate better contracting tradeoffs for the government. This additional insight is intended to provide strategies that move potential contractors into the government's preferred negotiation point and expedite the decision-making process in acquisition. The case studies presented in this paper focus on developing accurate utility functions that would enable such a game theory framework.

Introduction

Game theory has been a dominant research paradigm for studying conflict, bargaining, and negotiations for more than 50 years. It is widely applied throughout business to develop strategies that reflect priorities and tradeoffs. The government has an opportunity to leverage game theory in the federal acquisition system to improve outcomes and increase the agility of government acquisitions. As programs become more technical and complex, game theory can help decision makers identify strategies and leverage information to make data-driven decisions that reflect government priorities and tradeoffs. This paper explores the applicability of game theory to the federal acquisition process and provides a framework to help decision makers identify critical attributes and develop implementable negotiation strategies.

Industry plays a high-stakes game of survival as they act as a testbed for new technologies, and refine processes and strategies in their efforts to compete. Business, competitive, and technological pressures on industry necessarily drive a rapid pace of change and decision making. Industry has successfully applied game theory to develop business strategies that reflect for strategic decisions and negotiations in the business world. The government experiences similar decisions as competing missions, interests, and strategies actively shape acquisitions and budgets. Game theory provides proven insights and approaches that help both industry and the government develop baselines and strategies to create mutually beneficial solutions.

The federal acquisition process is governed by a system of clearly defined rules and regulations codified in the Federal Acquisition Regulation (FAR; 2017). The codification, publication, and adherence to a uniform acquisition system establishes a common understanding or common knowledge of the rules of engagement. Common knowledge, a central tenet of game theory, encourages industry to develop and execute rational business strategies that differentiate solutions and reflect tailored cost, schedule and performance tradeoffs. This creates the framework for achieving best-value through competing strategies and decisions. As a rulebook, the FAR ensures fairness and transparency in the acquisition process for all players, including both industry and the government. Structure and process of federal acquisitions seem to be well positioned to leverage game theory.

In game theory, successful negotiations require clear communication of the attributes. Federal source selections adhere to this principle by the mandatory disclosure of evaluation criteria as key discriminators or attributes. By advertising its source selection criteria and relative order of importance, the government signals its tradeoff considerations. Industry acts as players in the game by tailoring and offering solutions to the government to meet these considerations. One of the initial applications of game theoretic concepts is therefore helping the government identify and develop the key attributes or criteria as well as their importance relative to each other. To support such applications, game theory assumes rationality and that players attempt to maximize the outcome or their utility. Defining the utility for the government is one of the main challenges, as rarely all criteria can be measured in monetary values. Industry seeks to maximize their expected outcome or utility (e.g., profits, market share, etc.) by tailoring solutions that reflect the government's evaluation criteria. Similarly, the government maximizes its expected outcome or utility of industry solutions through its best-value tradeoff considerations. The government can leverage game theory to develop and execute negotiation strategies to improve decision making under uncertainty and contract performance.

Table 1 summarizes some of the similarities between game theory and the government source selection process and shows how the government inherently implements several key aspects of game theory.

Table 1. Similarities of Game Theory and Government Source Selection

Game Theory Principles	Government Source Selection
Players know the "rules of the game"	Federal acquisition process governed by well-defined rules and regulations codified in the Federal Acquisition Regulation (FAR)
Requires clear communication of attributes, priorities and outcomes	Mandatory disclosure of evaluation criteria that will be used to evaluate the proposal and their relative importance
Players are rational and seek to maximize their expected outcome or utility	Government maximizes expected outcome or utility through best-value tradeoff

Applying a game theory approach allows the government to objectively manage risk without compromising its cost, schedule, and performance tradeoffs. It provides insight into decision attributes and negotiating strategies that move industry into a preferred negotiation point while also considering the government's best value tradeoff constraints.

Quantitative Decision Support in Acquisition

The current federal acquisition system typically follows a structured and serial process outlined in the FAR to guide the government in navigating the complexities and cumbersome nature of the source selection process. This regimented system ensures both a fair and transparent acquisition and compliance with oversight stakeholders and the FAR. Unfortunately, these standard acquisition practices can result in the government spending too much time and money in negotiating and establishing contracts. Fortunately, FAR 1.102, Statement of Guiding Principles for the Federal Acquisition System, outlines an opportunity to introduce the agility and efficiencies of game theory by allowing strategies, practices, or procedures that are in the best interests of the government that are not specifically limited or prohibited by the FAR, Executive Order, or regulation. By enhancing negotiations and discussions, game theory complements the fairness and transparency provisions of the FAR. Game theory and utility theory offer an innovative and agile approach for driving bids closer to the desired attributes of most value, thereby saving time and money through improved negotiations and contract outcomes. FAR 1.102 grants government agencies tremendous acquisition flexibility, but the risk-averse nature of some agencies and acquisition professionals may limit their desire to leverage this flexibility. We are pursuing with this research a quantitative decision support framework that is intended to reduce risk and introduce acquisition agility and efficiencies for government and offerors. This game theory approach is consistent with FAR 15.305, Proposal Evaluation, as evaluations may be conducted using any rating method or combination of methods, including color or adjectival ratings, numerical weights, and ordinal rankings. Additionally, this approach is consistent with the April 1, 2016, Department of Defense (DoD) Source Selection Procedures (Section 2.3: Develop the Request for Proposals) where evaluation criteria may be quantitative, qualitative, or a combination of both. Although numerical or percentage weighting of the relative importance of evaluation criteria may not be used in DoD, assigning quantifiable or value tradeoffs in evaluating an offeror’s proposal is allowable and harmonious with the game theory approach.

Applicability of Game and Utility Theory for Acquisition Support

The application of game theory and utility theory can help facilitate the decision-making process in acquisitions. Utility functions provide a framework to translate player preferences into mathematical functions to which standard optimization techniques can be applied. Over a finite set of tradeoffs, there is a utility function that represents a rational preference ordering. This allows decision-makers the insight to tradeoff attributes and criteria based on their expected utility and affords decision makers a framework to make decisions among several alternatives that may result in several possible uncertain outcomes (e.g., what is the likelihood that Proposal X best matches the cost/schedule/performance tradeoff).

While government solicitations or Request for Proposal (RFP) identify the evaluation criteria and relative importance, the qualitative and subjective nature of evaluations may result in suboptimal tradeoff and selection analysis. By providing a mathematical framework, utility models have the capability to generate initial bids that reflect the government’s underlying preferences, help to compare and balance alternatives, and are transparent to everyone partaking in the bidding process. They have the potential of providing additional insight to objectively evaluate the government’s critical attributes and tradeoff considerations of the desired product. Moreover, through integrating the government’s utility function with cost constraints established by industry, a process for generating better initial bids can be established. One high level framework, originally suggested in Simon and Melese (2011), presents such a concept and is based on executing the following high level technical steps:

- Formulate the government's preferences that are modeled as utility functions parameterized by critical non-cost attributes, criteria or discriminators;
- Publicize government utility functions in RFP for industry to formulate and submit cost functions parameterized by the critical non-cost attributes, criteria or discriminators;
- Assess the uncertainty distribution or likelihood of success of various solutions; and
- Evaluate and optimize government objective function subject to industry provided cost functions and government utility functions to select preferred alternative subject to uncertainty distribution.

A wealth of research questions, however, present themselves when further designing and implementing this type of a framework in a real-world acquisition scenario, the first of which pertains to the type of utility model and the accurate and efficient calibration of the model for use as the foundation of such a framework. This paper focuses on utility models that can be used in this framework.

Survey of Game Theory Literature Relevant to Acquisition

Every scientific endeavor begins with a review of solutions and constraints that help to better define the research to be conducted. Therefore, in the early stages of the underlying research, our team conducted a literature survey. While we originally searched for applications of game theory in the context of procurement and related government activities, we realized quickly that we needed to address two additional topics as well.

As game theory and the optimization using game theoretic approaches relies heavily on the underlying value function, we extended our search to include utility theory with emphasis on utility functions that not only measure monetary values, but also express priorities and intangible preferences.

These led us to the third topic, multi-criteria decision making, as these methods are needed to support the utility-value functions that then can be used to optimize decisions and strategies using game theoretic approaches. The next three subsections will present a small subset of the literature evaluated by us with focus on the approaches we utilized in our study. This selection is neither complete nor exclusive.

Literature on Multi-Criteria Decision-Making

One of the most challenging tasks in acquisition is the selection of the criteria that have to be evaluated to reflect the preferences of the government. Wallenius et al. (2008) provide a good overview of the various methods that are currently in use, placing them into a historical perspective as well, as the same author group also conducted a similar review in 1992. Their overview is written from a management perspective. A more engineering leaning perspective is given among others by Parnell, Driscoll, and Henderson (2011).

Velasquez and Hester (2013) conducted a literature review and analysis of multi-criteria methods. They observe that outranking methods, which were prevalent in early approaches, were overtaken by value measurement approaches. Further they show that deficiencies can be overcome by combined approaches, although this requires a clear understanding of the advantages and disadvantages of the individual approaches, which are captured in a summarizing table in their conclusion section. Their paper assessed the more common methods of Multi-Criteria Decision-Making in order to benefit practitioners to choose a method for solving a specific problem, and they state clearly that this can only be the first step in selecting the right approach.

Agarwal et al. (2011) provide an alternative viewpoint on the selection of the best Multi-Criteria Decision-Making method by focusing on the proper evaluation and selection of suppliers, which is highly relevant in acquisition as well. An additional insight provided by them is the need to evaluate the suppliers based on the inputs of the strategic, functional and operational levels. They present that the

> implication of lean manufacturing and popularly used JIT approach has forced the researchers to shift the focus from the efficiency based model to quality based approach. The single criterion approach of the lowest cost supplier is no more accepted in this challenging and continuously changing environment. Thus, price or cost shifted down the line with respect to its importance in evaluating the suppliers, while the quality and delivery performance climbed up the hierarchy. (Agarwal et al., 2011, p. 808)

This insight is relevant for the government as well and needs to be addressed in the selection of the appropriate methods.

Both recent literature reviews show that there is no universal best solution, but that the selection of the best method is determined by the problem and may even require the use of problem-specific hybrid solution that require an in-depth knowledge of the problem as well as of the tools.

Literature on Utility Theory and Utility Functions

While the literature highlighted in the previous section focuses on the aggregation of multiple criteria in support of decision making, the references given in this section were evaluated regarding the definition of utility functions to reflect the preferences of the decision makers. Slantchev (2012) provides sound definitions of preferences and utilities to support decision making, including those to be made under uncertainty. As he is writing for political scientists, explanations and examples are easy to follow and do not require an in-depth education in gem-theoretic mathematical foundations.

If data is available that reflects preferences of earlier decision-making processes for either side of the negotiating partners, the methods and algorithms described by Afrait (1967) are still relevant. We assumed to be able to find more on the application of big data methods in support of the definition of utility functions, but it seems that this is still a topic of ongoing research and not predominant methods emerged so far.

An interesting variant for multi-issue closed negotiations addressing multi-time as well as multi-lateral negotiation strategies is described by Matsune and Fujita (2017), who developed not only the concept, but demonstrated it in an agent-based simulation environment. Theoretically, nothing speaks against applying these ideas for acquisition specific challenges as well, but we did not see any applications in this domain within our survey. What makes the application described in this paper so interesting is the ability to learn the opponents' utility information from observing their bidding choices within a strategy.

While the mathematics behind utility theory and utility functions is well understood, how to elicit the knowledge about their preferences from decision makers is still a challenge in itself. Our survey did not reveal any predominant strategy.

Literature on Game Theoretic Application for Government Solution

Obviously, every game theoretic insight can be applied to support government solutions better, but two of the evaluated papers deserve special attention, as they directly apply game theory to acquisition and government decision making.

Levenson (2014) provides an overview of the constraints of DoD procurement, showing why typical solutions from commercial markets are often not applicable and lead to undesired and unforeseen results. He describes the effects of fixed price and competitive price contracts and comes to the conclusion that

> only when one or more competitors offer innovations that truly reduce the costs of development and production does the government substantially benefit from competition over sole-source procurement without the adverse side effects of cost overruns. Distinguishing between true innovation and optimistic cost estimating, however, can pose a challenge for DoD acquisition officials. (Levenson, 2014, p. 437)

Blott et al. (2015) compiled a set of auction and game theory based recommendations for DoD acquisitions by synthesizing literature into specific military acquisition categories: procurement with unknown cost and no risk, items with known costs and existent but understood stochastic risk, and items with unknown costs and/or unknown stochastic risk. Some examples further evaluate if multiple competing vendors participate, and if the lot to be procured from several bidders, potentially at different stages of the project.

In summary, the literature survey provided sufficient examples of successful applications, but also the need for continuous research, in particular on how to elicit preferences and utilities from decision makers and apply these methods in a multiple criteria environment under the special constrains of acquisition.

Optimization With Game Theory

Before we go into the details of specific research conducted, the following section shall give an overview of the general concepts that will be addressed in the Selected Approaches section and the Case Studies and Results section.

The following optimization problem drives the application of the utility model. A decision maker has to choose from a set of solutions provided by vendors. The solution is defined by a set of weighted attributes. Furthermore, each vendor is involved with the mathematical optimization that is specific to their own individual cost constraints.

$$\text{Max } V(x) = \sum w_i v_i(x_i)$$

subject to: $\sum c_i x_i < B$

where: x_i = is the level for attribute i

w_i = the weight for attribute i

v_i = the single attribute value function for attribute i

c_i = the offeror cost function for attribute i

B = budget constraint for maximizing utility i

Solving this optimization formulation allows for vendors to generate bids reflecting their specific cost constraints. This yields initial bids that are more consistent with the government's preferences based on the levels of the key non-cost attributes of interest. The solutions to the above optimization formulation allow for stronger initial bids by the interested vendors. These solutions aren't necessarily final solutions or final bids but are an efficient means to getting the bidder close to what the government is looking for. This can more efficiently set up the next stage of proposal tweaking and negotiation on both sides. The rigor of this approach also allows for unambiguous documentation of the negotiation, selection, and provides means for repetition and further evaluation.

Moreover, RFP language can make it difficult for potential bidders to extract out what the most important attributes are for the government (e.g., when too many attributes are included and the evaluation criteria are unclear). Using this quantitative mathematical programming formulation instead allows for bidders to move directly towards those key attributes through an automated means.

To show the applicability of game-based approaches as discussed in these introductory sections, we selected three approaches to evaluate in more detail, which is the topic of the rest of this paper.

Selected Approaches

After conducting the literature survey, we applied three candidate approaches in our research. We selected them due to their perceived potential in being implemented in an acquisition procedure:

- Best/Worst Method (Rezaei, 2015),
- Multi Swing Method (Schmidt, 2017), and
- Functional Dependency Network Analysis (Garvey, Pinto, & Santos, 2014).

Beside their potential for application in an acquisition setting, all three approaches have calibration procedures that are not overly burdensome to the decision maker. They also complement each other. Testing of these methods, as discussed in the Case Studies and Results section, will further reveal the features of the acquisition scenarios where each approach does well. In order to conduct our research, we applied all three approaches to a small decision problem that involved just five attributes to get decision maker accustomed to the steps and procedures needed to be conducted. Furthermore, we applied the Best/Worst method to a larger, 20 attribute problem. These test cases are discussed in more detail in the Case Studies and Results section.

Initial collaborations and discussion with a government sponsor identified three best practices or considerations that impacted our utility function assessment procedure and resulted in the application of multiple assessment techniques.

The first is that the level of effort in developing the assessment procedure must be commensurate with the size, scope, and complexity of the acquisition. A day-long interview process to fit a model may be realistic for a highly complex multi-billion-dollar Acquisition Category (ACAT) 1 program but is not realistic for all acquisitions, and surely not for supporting a smaller research effort like ours. In contrast, a one-hour discussion may be sufficient for many complex acquisitions. Decision makers must balance competing objectives and may not have the luxury of time or resources to support a lengthy and involved process to support the development and calibration of assessment procedure. A long and drawn out initial assessment procedure may result in fatigue and complacency, which may lead to inconsistencies in preference articulation.

The second consideration is that assessment procedures must be adaptable so that they can be effectively applied to decision makers who are either more quantitative or qualitative in nature. However, our research showed that most acquisition professionals are comfortable with relative importance and prefer qualitative descriptions of their preferences. Introducing descriptive adjectives in place of numerical values, in many questions, can help alleviate this issue. Finally, there are many acquisition situations where there is a large attribute set that influence the decision. The size of this attribute set can be overwhelming for any decision maker. Therefore, preference modeling methods must be able to screen out attributes of minimal significance to isolate the critical non-cost attributes and the critical tradeoffs between those attributes. This supports an acquisition best practice of focusing on critical non-cost attributes to avoid diluting the importance of key discriminators.

Overview of Best/Worst Method and Extensions

The Best/Worst method originates from Rezaei (2015) and this research has extended the approach to work more smoothly for cases where there are a large number of attributes at hand and when the attributes are binary in nature (result in either a 0/1 or yes/no value). One of the Best/Worst method's features is its ability to perform calibration in a short series of questions. Moreover, these questions have the ability to be phrased to not be overly burdensome to the decision maker. From our observations, having simple and clear acquisition questions to identify key discriminators facilitates the acquisition and conforms to best practices.

Consistent with source selection practices, the procedure for the Best/Worst method starts with selecting the attributes or discriminators that effect the decision. Then feasible ranges are assigned for each of these attributes. The next step is the assignment of weights for each attribute reflecting the preferences and importance. This applies specifically to the attribute to identify key discriminators and does not apply numerical weights to proposals in the source evaluation process. This step begins with selecting the most important attribute as well as selecting the least important attribute. From there, comparisons are made to understand the relative importance of the most important attribute to each of the other attributes. In a similar manner, comparisons are then made to assess the relative importance of the least important attribute to each of the other attributes.

The question phrasing to the decision maker is the key to getting this approach to work effectively. The decision maker needs to be directly asked how much more important is the most important attribute for each of the other attributes individually. Mapping qualitative scales to numerical scales was shown to work well in our studies for preserving rank order. For instance, levels, such as, "just as important," "slightly more important," "more important," "significantly more important," and "extremely more important" were applied with good success while being mapped on a scale of 1–5.

The end goal of the Best/Worst assessment procedure is to obtain a preference function in the form: $V(x) = w_1 v_1(x_1) + w_2 v_2(x_2) + \dots \dots w_n v_n(x_n)$. The Best/Worst procedure primarily focuses on the weights. Suggestions in this paper for extending to the assessment of the single attribute utility functions $v_1(x_1)$ focus on fitting a function across sample points for each individual attribute. Sampling can be effective with just four points on the utility curve. When doing a qualitative mapping, those points can be referenced as the min, midpoint, target, and max. On a [0,1] scale those reference points were mapped to values of 0, 0.5, 0.75, and 1 respectively. The qualitative assessment questions can first focus on the target. Here the question is asked, "What is the value of this attribute that you would really want to have?" Then the level representing satisfactory for the attribute is assessed: "What level for this attribute is acceptable and would not hinder my use? It can be considered being like a minimum requirement that is not ideal but gets the job done." Then the maximum level for the attribute can be assessed: "What is the level for the most functionality that you could possible handle need—any more wouldn't make life any better." Finally, the minimum level for the attribute is assessed: "What is the maximum attribute level where there is zero utility or where you would have absolutely no use for this product if this attribute was at this level."

The Best/Worst method was extended to a large number of attributes (>20). In acquisitions we observed with our government sponsors, the number of attributes was typically quite large. The Government Acquisition Case Study With Large Number of Attributes section provides details on the application of the Best/Worst method extension to a government acquisition study. For large number of attributes, the procedure was updated in the following manner:

1. First do pairwise comparisons across adjacent pairs of attributes start at attribute #1 and then work down the attribute list.
2. Bin the attributes based on whether the attributes were more important than two attributes, one attributes, or no attributes. End up with three bins: prime, mid, low.
3. Reassess attributes in each bin to make sure they are in the right place.
    a. Ask for best and worst for each bin.
    b. Do pairwise comparison of best in mid and low bin with worst in the higher-level bin.
    c. Repeat 3A and 3B until no more changes are made.
4. Identify the attributes for inclusion into the Best/Worst method
    a. Take all attributes in prime bin.
    b. Take best and worst in mid bin.
    c. Take best and worst in low bin.
5. Best/Worst method is then implemented on attributes in the prime bin.
6. Best/worst method is then implemented on all other attributes kept above.
7. Ask the level of difference between the worst attribute in prime and the best in mid. This level of difference then becomes the difference level for the weights in prime bin and the weights in the remaining bins and the weights are then scaled accordingly.

After these assessment procedures are made the weights for the preference function can be solved through the optimization outlined in Rezaei (2015). The pairwise comparisons given at the beginning of the assessment procedure can also be used to solve for the weights more effectively as well as for validation of the results.

Multi-Swing Rollup Method

The Multi-Swing Rollup Method (MSRM) was developed by MITRE as a new aggregation method for multi-attribute decision problems (Schmidt, 2017). As discussed in the Literature on Multi-Criteria Decision-Making section, rolling up multiple values into one representative value is a general challenge, as already discussed in our literature survey. The MSRM is using a generalized addition tallying organization (GATO) approach. While the classical approach uses the weighted sum of the contributing decision factors, MSRM/GATO uses a non-linear combination in areas in which the simple addition leads to counterintuitive results.

MSRM starts with the definition of multi-swing tables to collect data and combine getting weights and utility functions in one user-driven process. These multi-swing tables are then multiplicatively rolled up and calibrated to fit to a percentage scale. The four steps of the methods are as follows:

1. selecting and quantifying the metrics for each contribution;
2. defining a scale for quantifying the overall score;
3. constructing the multi-swing tables for each contribution;
4. constructing and calibrating the rollup function.

Selecting and quantifying the metrics for each contribution starts with identifying the qualities the user is interested in. The result is a quality tree that identifies the contributions and the metrics used to quantify them. Examples are the resolution (metrics) for the display (contribution), or the battery capacity in minutes (metrics) that keeps the device functional (contribution). These examples will become clearer with the application presented in the Cell Phone Example section.

Defining a scale for quantifying the overall score of the attributes (not numerical scoring of proposals) ensures consistency when assessing the overall value increase or decrease when evaluating the individual contributions. MSRM recommends using a mapping of generally understood expressions to numerical values, such as ideal = 100%, very good = 90%, good = 70%, indifferent = 50%, poor = 30%, very poor = 10%, and not acceptable = 0%. The scale does not have to be symmetric. It is more important that it reflects the weighting priorities and preferences of the user.

Constructing the multi-swing tables is conducted for each contribution, starting with defining baseline with typical and acceptable values for each contribution. For each contribution, we define next a set of swing scores that can be better or worse than the baseline. For each contribution, a set of swing scores spanning all values that can occur in the selection process are collected and the swing rows constructed. If the value of a contribution is a show stopper, e.g., the battery life is too short to support operational use of the item, it is marked as such. In acquisition settings, every attribute that falls under a minimal value becomes a show stopper.

The baseline and all swing rows are then captured in one multi swing table. In this table, in each row only one of the values is changed in comparison with the baseline, so that a comparison with the baseline can be used to access an overall score using the expressions identifies in step 2 of the MSRM. While the baseline may be seen as good, a less screen resolution may decrease the value to indifferent, poor, or may even become a show stopper, while longer life span may result in a very good overall value.

Constructing and calibrating the rollup function uses the multi-swing tables as its foundation, As each row in the multi-swing table captures how the overall value changes when we swing one contribution at a time, a multiplicative roll-up approach can now be applied to compute how the value changes when several of such changes occur at the time. If, for example, the resolution decreases, resulting in a change value decrease of 20%, and the battery life gets shorter as well, decreasing the value by 10%, then the occurrence of both changes should result in a decrease of 28%. The idea is to multiply the individual effects to generate the combined effects.

While the approach naturally results in the elimination of all show stoppers (as the multiplicative approach results in a zero whenever one of the contributions is not acceptable), the positive results can multiply up to more than 100%, which can be addressed using rescaled proportion retention multipliers that ensure that no combination exceeds the 100% limits.

One of the remaining challenges is the combinatorial explosion with the increasing number of contributions. Our initial application shown in the Case Studies and Results section was limited to five attributes, but still required more than 45 minutes to build all multi-swing tables. On the positive side, the method allows the linear integration of new contributions after the initial set-up: a new attribute can be integrated without having to change the trade-offs between the already existing attributes.

Functional Dependency for Network Analysis

The last approach utilized in our research was originally developed for a systems engineering setting, but due to its general applicability, we decided to include it in our evaluation. The application of the Functional Dependency for Network Analysis (FDNA) methodology involves

1. data gathering,
2. preference inference,
3. quantifying accuracy, and
4. making predictions.

The data gathering step involves constructing an experimental design to capture data on the different attributes of the product in accordance to the decision maker's input. The preference inference step involves proposing specific preference models and using the gathered data to infer the defining parameters which are most consistent with the data. The quantifying accuracy step involves the application of cross-validation to assess the accuracy of the fitted preference model. Finally, the making prediction step entails converting a test case to the form selected in the first step and make predictions using the parameters inferred in the second step.

For data gathering with FDNA it is necessary to create a dictionary of qualitative descriptions of product attributes and an assigned numerical representation to each. In the acquisition setting, the dictionaries are highly reusable according to our experiences, although no study has been conducted to verify this observation. As a general practical matter, many spirals of potential dictionaries should be generated and tested to ensure that the definitions are neither to narrow or too broad so that the decision maker who is being models will be assigning a broad ranges of numerical preference scores to the anticipated set of optional designs. Then a set of optional designs of interest can be generated. Assuming the absence of *a priori* knowledge of the decision maker's preference, the designs are randomly generated.

Motivated by the work of Garvey and Pinto (2009) and Servi and Garvey (2017) two different preference models are included in the approach[1]:

$$f_s(P_s, \gamma) = \min_i [\ P_s^i + \beta_i] \tag{1}$$

or

$$f_s(P_s, \gamma) = \alpha_0 + \sum_i \alpha_i \ \ P_s^i \tag{2}$$

[1] If there were a larger amount of experimental data, it would have been more desirable to use the precise FDNA model documented in the references. Due to the limited size of the data, two different aspects of the FDNA were used for this analysis.

where P_s^i is the numerical level of preference of the ith characteristic of the sth experiment.

With the decision maker's evaluation of different attribute combinations, the values of β_i or α_i computed using the training data, which are rows comprising all attribute values and the resulting evaluation by the decision maker, it is possible to estimate the preferences of the decision maker.

For the case study discussed in the following section, the accuracy using (1) was found to be superior to that using (2), it is recommended that when predicting the comparative preferences of the decision maker to two alternatives, the prediction is made using only (1) and, in the case of a tie, (2) is used to break the tie.

Case Studies and Results

The model assessment procedures are applied and tested on two case studies to test their applicability in the acquisition setting. The first case study, described previously, is a cell phone purchasing example, easily understood by everyone, and was used for internal testing for all three selected approaches. The second case study, described in the section titled Government Acquisition Case Study With Large Number of Attributes, is a real-life acquisition scenario that added complexities not existing when methods like these are tested and presented in literature. We only discuss the Best/Worst method example exemplifying the challenges.

Cell Phone Example

A case study for buying a cell phone was first used to work through the question phrasings of each method in a simpler environment. The test subject or decision maker was given the scenario of purchasing a new smartphone, such as an iPhone or Samsung Galaxy. The decision maker was made known that there are dozens of alternatives to choose from. They are then made to envision that there are five main attributes that will affect their decision as to which smartphone to buy. Here are the five attributes that were told that effected their decision:

A. Weight [0–5 pounds]

B. Lifespan [0–10 years]

C. Screen resolution [0–4,000p]

D. Processing speed [0–10x]

E. Storage amount [0–500 Gigs]

Included above are the ranges of values that each attribute can take. The ranges are meant to exceed what is true in reality. For example, one cannot obviously have a cell phone of no weight and there are no phones in the market that weigh 5 pounds. After reading through the example and taking the role of the purchaser, a series of questions was conducted about their preferences in accordance to the assessment procedures for the three preference functions tested.

Application of Best/Worst Method

The application of the Best/Worst method began with the assessment of the weights for the five attributes of interest. The test subject was asked a series of trade-off questions and identified processing speed as the most important attribute and lifespan as the least important attribute. Through a series of questions relating the level of importance of each attribute with respect to processing speed the following vector was obtained: A_b = [5,4,1,1,2]. The numerical values in this vector specify how much more important processing speed was with respect to weight, lifespan, screen resolution, and storage amount. The scale for importance is on a range of 1–5. So, the first value of 5 represents processing speed being extremely more important than weight. As indicated by the 4 assigned to the second slot, processing speed is considered significantly more important than lifespan. The remaining values show that processing speed is equally important to screen resolution, to itself, and more important than storage amount. These are the same mappings to numerical values introduced in the Selected Approaches section.

After this the same questions regarding relative attribute importance were asked with respect to the attribute noted as the least important. This resulted in the following vector A_w = [1,1,4,5,3]. As shown in Rezaei (2015). These two vectors representing relative importance between each attribute and the best and worst attribute, respectively can be used to perform a least squares approximation to solve for the weights. The following numerical weights were obtained: [0.04, 0.12, 0.28, 0.39, 0.17].

The final step was to solve for the single attribute value functions pertaining to each attribute. Here the test subject was asked for each attribute to specify the minimum, midpoint, target, and maximum values for each of the five attributes and the question wordings introduced in the Overview of Best/Worst Method and Extensions section were applied. Table 2 presents the values obtained for the min, midpoint, target, and max for each attribute.

Table 2. Subject Responses to Min, Midpoint, Target, and Max Levels for Each Attribute

	Min	Midpoint	Target	Max
Weight	1 lb.	0.75 lb.	0.5 lb.	0.33 lb.
Lifespan	1 year	3 years	4 years	6 years
Screen resolution	400p	720p	1080p	2000p
Processing Speed	0.5x	1x	3x	5x
Storage Amount	64 gigs	128 gigs	250 gigs	500 gigs

The values in Table 2 are used to solve for the single attribute value functions $v_i(x_i)$ for all five attributes in this case study. For this case study, a simple second order polynomial was applied for fitting these single attribute value functions and the method of least squares was used for fitting. The weights for all five attributes can then be integrated into the single attribute value functions to obtain the following function for the preference model:

$$V(x) = w_1 v_1(x_1) + w_2 v_2(x_2) + w_3 v_3(x_3) + w_4 v_4(x_4) + w_5 v_5(x_5) \quad (3)$$

In order to test the accuracy of the Best/Worst method, a series of comparisons across six purchasing alternatives was performed. The following pairwise comparisons across all combinations of these purchasing options below were performed by the decision maker.

A. [2, 2, 2000, 0.75, 256]

B. [0.5, 5, 720, 2, 128]

C. [4, 1, 4000, 1, 64]

D. [1, 4, 720, 1, 256]

E. [2, 3, 1080, 2, 256]

F. [0.5, 3, 4,000, 4, 64]

The results of these pairwise comparisons can be applied to generate a ranking. The rankings obtained here are compared to rankings generated through the preference model sampled under these same alternatives. In addition, the proportion of the pairwise comparisons that are consistent between the decision maker and model was measured. There were 15 different combinations of pairwise combinations resulting from the six scenarios above. The preference model resulting from the calibration involving the Best/Worst method resulted in consistency amongst all 15 pairwise comparisons. That meant when the subject specified, for example, that alternative B was more preferable than alternative A, that the preference model outputted a larger value when inputting in the attribute levels for alternative B than when inputting the attribute levels for alternative A. As naturally follows, the rankings for all six alternatives were consistent as well. The case study demonstrated promise in the Best/Worst method to generate an accurate model in a short amount of time.

The entire assessment procedure was done in roughly 45 minutes for this scenario involving five attributes. The process can be supported by tools, and our research resulted in the definition of new tool support that is currently prototypically developed to support data collection, calibration, and presentation of the results. The acquisition professional doesn't have to provide all the details captured in this section, but should provide the comparisons and evaluate the rankings.

Application of MSRM Method

The Multi-Swing Rollup Method (MSRM) was applied in the same setting as the Best/Worst method, using the same experts to conduct the experiments. Using the same attributes as enumerated in Table 2, we defined one positive and one negative swing for each attribute, as shown in this Table 3.

Table 3. Attributes and Swing States (Green Variations Are Positive, Red Are Negative)

	Baseline	Variations	
(W) Weight	0.5	0.33	0.75
(LS) Lifespan	2	1	4
(SR) Screen resolution	1080	720	2000
(PS) Processing speed	4	2	5
(SA) Storage amount	256	128	500

Next, we defined the utility factor terms to be used to rate the comparisons between the baseline and the swings. In the discussion with the experts and decision makers, we ended up with a table that showed the semantic equivalencies between different families of terms describing comparisons, status descriptions, and grades, which the group was comfortable with (see Table 4).

Table 4. Utility Value Terms

Utility value

significantly better	ideal	A
much better	very good	A/B
better	good	B
little bit better	above average	B/C
average/mid-point	average	C
little bit worse	below average	C/D
worse	poor	D
much worse	very poor	E
showstopper	showstopper	F

Having five attributes with two swing states each does result in eleven entries. Using the utility terms, each entry was compared individually with the baseline to identify the overall change in utility by changing one attribute. Table 5 shows the individual utility contributions in percent that resulted from our discussions with the experts.

Table 5. Utility Value Changes in Percent Relative to the Baseline

	Weight	Lifespan	Screen resolution	Processing speed	Storage amount
negative	-6.7	-26.7	-26.7	-46.7	-20.0
positive	13.3	20.0	33.3	33.3	26.7

Using this information, the full multi-swing rollup table with all 243 entries can be created. The resulting table contains all possible combinations of multi-swings plus the baseline. The resulting overall utility is calculated by multiplying the individual changes. When ordering the table, the entry with all negatives obviously is the lowest, and the entry with all positives the highest, but all possible permutations in between are listed as well, showing the ranking of all alternatives, including the selected subset used in the Best/Worst method. As we derived the same ranking, this should at least consistency in the evaluation, no matter which of the first two approaches was used.

The assessment procedure conducted with our decision makers was shorter than for the Best/Worst method, but only because several of the results could be reused. In an internal comparison with in-house experts, the amount of time needed for the first two methods was approximately the same for the cell phone example.

Application of FDNA

In a previous section, we introduced the two different preference models that were motivated by the work of Garvey and Pinto (2009) and Servi and Garvey (2017). First, the term needs to be defined. The dictionary shown in Table 6 is comparable to the terms defined in Table 5 for the utility terms used in the MSRM.

Table 6. A Dictionary Assigning a Numerical Preference Level to the Preference Level of Qualitative Characteristics of iPhone

	(W) Weight	(LS) Lifespan	(SR) Screen resolution	(PS) Processing speed	(SA) Storage amount
0 - Crummy	heavy	1	good images	email, word OK	some added apps, best photos
1 - OK	not heavy	2	good for printing	very good non-games, slow video	apps and photos
2 - Good	light	3	great images, good enlarging	good for video	huge for apps and photos, some videos
3 - Great	ultra light	3+	very good for enlarging	everything great	virtually unlimited including videos

Next, we generated possible solutions for the five attributes important for the selection of the cell phone: weight, lifespan, screen resolution, processing speed, and storage amount. Table 7 shows the 27 generated cases, using the index numbers defined in the dictionary to specify the solution. The decision maker that graded the various solutions as captured in the column "evaluator."

Table 7. The Experimental Training Data

No.	Weight	Lifespan	Screen resolution	Processing speed	Storage amount	Evaluator
1	3	3	0	0	0	**0**
2	2	2	3	2	2	**3**
3	2	2	3	2	2	**3**
4	3	3	1	0	0	**0**
5	2	2	2	1	3	**2**
6	0	1	1	2	3	**1**
7	2	1	0	3	0	**0**
8	1	2	2	1	0	**0**
9	3	2	1	2	2	**2**
10	2	1	2	2	3	**3**
11	3	1	2	0	1	**0**
12	2	1	0	3	0	**0**
13	3	1	1	1	1	**1**
14	1	0	3	2	3	**2**
15	1	3	0	3	1	**1**
16	2	2	0	2	1	**1**
17	1	1	2	2	2	**2**
18	0	3	0	0	1	**0**
19	0	1	1	2	2	**1**
20	1	3	0	1	3	**1**
21	1	0	2	2	2	**2**
22	1	0	1	1	3	**1**
23	1	1	1	2	2	**2**
24	2	0	3	3	3	**3**
25	3	3	3	2	1	**2**
26	2	2	2	3	3	**3**
27	1	0	1	3	1	**1**

Given the data in the Cell Phone Example section, it is possible to exhaustively search for the integer values of β_i most consistent with the data in terms of the mean sum of squares error, $\alpha_0 = 1, \alpha_1 = 8, \alpha_2 = 6, \alpha_3 = 0, and\ \alpha_4 = 0$ as well as analytically solving for the values ofα_i most consistent with the data,($\beta_0 = -1.6204, \beta_1 = 0.2101, \beta_2 = 0.2219, \beta_3 = 0.4290, \beta_4 = .4375, \text{and } \beta_5 = 0.5724$). This leads to mean sum of square error of 0.19 when using equation (1) and a worse mean sum of square error of 0.44 when using equation (2).

For FDNA, however, the more precise approach to quantifying the error is using the method of cross-validation. Here, the values of β_i or α_i are computed using a random set of 8/9 of the data in Table 7 and then the accuracy of the prediction is computing using the 1/9 of the data not trained on. This was repeated numerous times. This lead to the conclusion that the mean sum of square when using equation (1) was 0.19 (and a standard deviation of 0.24) and when using equation (2) was 0.54 (with a standard deviation of 0.23). The conclusion, for this data, is that equation (1) leads to a superior model of this decision maker, which means that $f_s(P_s, \gamma) = \min_i[\ P_s^i + \beta_i]$ is the better model to capture preferences.

These examples given in the Cell Phone Example section exemplify the different application possibilities of the three approaches as well as their strengths and mutual support, reemphasizing the need for having a toolbox of different solutions in support of acquisition decisions.

Government Acquisition Case Study With Large Number of Attributes

An additional case study was also performed with a government sponsor involving an acquisition scenario consisting of 20 attributes to show the scalability of approaches. This is a more challenging case study in that the decision maker must go through a lengthy assessment procedure to make comparisons among a very large set of attributes. Another twist to this problem was that these attributes were binary in nature. Each attribute had an objective value and the government was only interested in if the attribute exceeded that value. So, each attribute has two levels (0, 1) to represent whether it met the objective or not. A new modification of the Best/Worst method was applied as presented in the Overview of Best/Worth Method and Extensions section to handle this scenario of having a large set of attributes. The types of attributes cannot be discussed in this paper, but the implementation of the procedure can be discussed. The first portion of the assessment procedure involved doing pairwise comparisons across the adjacent pairs of attributes starting with the first attribute. The resulting table of the results to these pairwise comparisons is shown to help further explain the approach (see Table 8).

Table 8. Pairwise Comparisons Across Adjacent Attributes

Attribute	1st Comparison	2nd Comparison
1	1 > 2	1 > 20
2	2 < 1	2 > 3
3	3 < 2	3 < 4
4	4 > 3	4 > 5
5	5 < 4	5 > 6
6	6 < 5	6 < 7
7	7 > 6	7 < 8
8	8 > 7	8 > 9
9	9 < 8	9 > 10
10	10 < 9	10 < 11
11	11 > 10	11 > 12
12	12 < 11	12 < 13
13	13 > 12	13 > 14
14	14 < 13	14 < 15
15	15 > 14	15 > 16
16	16 < 15	16 > 17
17	17 < 16	17 > 18
18	18 < 17	18 > 19
19	19 < 18	20 < 19
20	20 < 19	20 < 1

After this initial pairwise comparison is done, the attributes are binned based on whether they were more important than two attributes, one attributes, or no attributes. This results three bins, which are named prime, mid, low, respectively. The resulting bins are shown in Table 9 to further exemplify the approach.

Table 9. Binning of Attributes After Initial Pairwise Comparisons

Prime Bucket	Mid Bucket	Low Bucket
Attribute 1	Attribute 2	Attribute 3
Attribute 4	Attribute 11	Attribute 6
Attribute 8	Attribute 7	Attribute 10
Attribute 5	Attribute 9	Attribute 12
Attribute 13	Attribute 16	Attribute 14
Attribute 15	Attribute 17	Attribute 19
	Attribute 18	
	Attribute 20	

The next step is to reassess the attributes in each bin to make sure they are allocated properly. This is done through asking the test subject to first identify the most important and least important attribute in each bin. Then pairwise comparisons are done between the most important attribute in each bin and the least important attribute in each bin. After any reassignments are made, the test subject is then asked to identify again the most important and least important attribute in each bin. If there are any changes to these assignments, the process is repeated and the most important attribute in each bin is compared again with the least important attribute in each bin. The process repeats until no attributes can be exchanged between bins in this manner. The final binning of attributes for this case study is shown in Table 10.

Table 10. Final Binning of Attributes After Extra Validation Questions

Prime Bucket	Mid Bucket	Low Bucket
Attribute 1	Attribute 2	Attribute 18
Attribute 4	Attribute 11	Attribute 3
Attribute 16	Attribute 7	Attribute 6
Attribute 5	Attribute 9	Attribute 10
Attribute 13	Attribute 8	Attribute 12
Attribute 15	Attribute 17	Attribute 14
	Attribute 20	Attribute 19

At this point the Best/Worst method can be executed across a subset of these attributes. The first step in doing this is to identify the most important and least important attribute in each of the three bins. Note that due to the initial pairwise comparisons being made between adjacent attributes it is not necessary to compare every attribute with the most important attribute and every attribute with the least important attribute. In the prime bin, each attribute is compared with the most important and least important attribute. Then for the mid bin, the only the most important and least important attribute are compared with the least important attribute in the prime bin. Likewise, these attributes are compared with the most important attribute in the low bin. Then finally, the most important and least important attribute of the low bin are compared to the least important attribute in the mid bin. These measures of relative importance are again done on a scale of 1–5. After all of these comparisons are made, there is enough information to perform a least squares estimation to approximate the weights for all 20 attributes.

The nice feature about having attributes with a binary value is that it is not necessary to assess a single attribute utility function for each attribute. If the attribute meets the threshold then the utility is mapped to a value of 1 and if it does not meet the threshold it is mapped to a value of 0. Therefore, the weights can be used with 0, 1 terms for the attribute values directly to result in this equation for the preference model: $V = \sum_i w_i \, t_i$ where t_i = 1 if the value for attribute i meets its threshold level for the objective.

The most notable result of this new assessment procedure was that the interaction time with the test subject to train this preference model with 20 attributes was done in less than one hour. For this amount of time, it is expected that the decision maker can stay engaged for the duration of the assessment procedure and remain accurate and limit inconsistencies. Another promising feature of this method is that the initial pairwise comparisons can be held out from training of the model and used for validation. When reserving eight for validation, the model resulted in matching the decision maker in seven out of eight (87.5%) of the test cases.

As we did not conduct any further experiments on the degree to which these results are scalable or generally applicable in other domains as well, we do not want to oversell any results, but the observed trends do make sense within our experience. Due to time and budget constraints, we did not conduct similar experiments using the MSRM or FDNA methods.

Conclusions

This paper suggests the idea of applying game-theoretic models as a foundation for a quantitative decision-making framework in support of acquisition. Through successful calibration of utility functions, we suggest there is a strong potential to develop a framework that can more effectively illuminate strategies that move industry into the government's preferred negotiation point and expedite the decision-making process in acquisition.

The case studies presented in this paper focus on the potential for developing accurate utility functions that would enable such a game theory framework. The government's utility functions, representing their level of preference for attribute levels involved in a proposal, are the cornerstone for enabling such a decision support framework to be utilized effectively and accurately. In this paper, we examined three potential utility function calibration procedures from literature and adapt these procedures to examine the strengths and weaknesses of each approach. The three methods were applied in a case study involving the acquisition or purchasing of a cell phone. All three reveled benefits for different acquisition scenarios. The Best/Worst method showed robustness in handling a small or large number of attributes effectively. The MSRM method demonstrating the ability to capture sharp drops in utility in individual attributes. This is an important feature when some attributes present thresholds where the entire product becomes unusable by the government. The FDNA method showed the ability to work effectively when the decision maker is more qualitative in nature than quantitative.

The Best/Worst method was also extended to be more applicable to acquisition scenarios involving a large number of attributes (20 plus attributes). These situations are more common than the small number of attribute examples that are often provided in the toy examples in the literature. This paper provided an efficient procedure for screening and implementing the Best/Worst method when the attribute set is large. It was tested on a real-world government acquisition example and was shown to be able to calibrate an accurate preference function in under one hour of decision maker engagement time.

Future work involves integrating these utility function assessment procedures into a decision support framework that can enable potential bidders to maximize the fitted utility function with respect to their own specific cost functions, which are parameterized by the same attributes as the utility function. The sampling procedure of these cost functions along with the best optimization algorithm to apply is another area of research needing attention. The optimization algorithm must have the ability to generate solutions in near real-time in order for this decision support framework to be usable and effective.

References

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International Economic Review, 8*(1), 67–77.

Agarwal, P., Sahai, M., Mishra, V., Bag, M., & Singh, V. (2011). A review of multi-criteria decision making techniques for supplier evaluation and selection. *International Journal of Industrial Engineering Computations, 2*(4), 801–810.

Blott, J., Boardman, N., Cady, A., Elliott, J., Griffin, W., Mastronardi, N., & Quinn, P. (2015). *Auction and game theory based recommendations for DOD acquisitions* (No. USAFA-AM-14-189). Colorado Springs, CO: Air Force Academy.

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2018). Retrieved March 2018 from https://www.gpo.gov/fdsys/pkg/CFR-2017-title48-vol1/xml/CFR-2017-title48-vol1-sec1-101.xml

Garvey, P. R., & Pinto, C. A. (2009, June). Introduction to functional dependency network analysis. In *The MITRE Corporation and Old Dominion, Second International Symposium on Engineering Systems* (Vol. 5). MIT, Cambridge, Massachusetts.

Garvey, P. R., Pinto, C. A., & Santos, J. R. (2014). Modelling and measuring the operability of interdependent systems and systems of systems: Advances in methods and applications. *International Journal of System of Systems Engineering, 5*(1), 1–24.

Levenson, W. J. (2014). DoD acquisition—To compete or not compete: The placebo of competition. *Defense ARJ, 21*(1), 416–440.

Matsune, T., & Fujita, K. (2017, July). Weighting estimation methods for opponents' utility functions using boosting in multi-time negotiations. In *Proceedings of the IEEE International Conference on Agents (ICA)* (pp. 27–32). IEEE.

Parnell, G. S., Driscoll, P. J., & Henderson, D. L. (Eds.). (2011). *Decision making in systems engineering and management* (Vol. 81). Hoboken, NJ: John Wiley & Sons.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega, 53*, 49–57.

Schmidt, B. K. (2017). *Generalized Addition Tallying Operation (GATO).* Bedford, MA: The MITRE Corporation.

Servi, L. D., & Garvey, P. R. (2017). Deriving global criticality conditions from local dependencies using functional dependency network analysis (FDNA). *Systems Engineering, 20*(4), 297–306.

Simon, J., & Melese, F. (2011). A multiattribute sealed-bid procurement auction with multiple budgets for government vendor selection. *Decision Analysis, 8*(3), 170–179.

Slantchev, B. L. (2012). *Game theory: preferences and expected utility* [Technical report, Political Science Courses]. San Diego, CA: University of California.

Velasquez, M., & Hester, P. T. (2013). An analysis of multi-criteria decision making methods. *International Journal of Operations Research, 10*(2), 56–66.

Wallenius, J., Dyer, J. S., Fishburn, P. C., Steuer, R. E., Zionts, S., & Deb, K. (2008). Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Science, 54*(7), 1336–1349.

Disclaimer

This work was supported by the internal MITRE Innovation Program in support of Agile Connected Government. The views, opinions and/or findings contained in this report are those of The MITRE Corporation and should not be construed as an official government position, policy, or decision, unless designated by other documentation. This work has been approved for public release with unlimited distribution; Case Nr. 18-1141.

Optimizing Contract Modifications Under One Universal Mod

Thomas Graham—MBA, is currently the Chief Operating Officer of nGAP Incorporated. His background includes service in the U.S. Air Force working on research and innovation programs, with the objective of maintaining U.S. supremacy in the skies through programs such as IFAST. Graham is an expert regarding DoD automation related to Navy Ship Building Projects and has done extensive research on the subject for the last 30 years. In 2005, Graham, in collaboration with Senators McCain, Coburn, and Obama, facilitated the creation and acceptance of the law currently known as FITARA, a precursor to the Data Act passed in 2014. [tgraham@ngap.com]

Stanley Sydor, LT, USN (Ret.)—spent several years in accounting positions in both the public and private sectors after completing his U.S. Navy service as a Lieutenant. Enjoying application of financial concepts to market opportunities, he spent several years in the areas of product management, project marketing, and marketing, in positions such as Operations Controller and Director of Marketing for high-tech companies in the telecommunications sector. Sydor has spent the last 15 years in a variety of roles in the acquisition software industry. He holds a BS from Bryant University and an MBA from Pepperdine, and is a CPA. [ssydor@ngap.com]

Paige Glaze—received her PhD in economics from Texas A&M University in 2002 and since that time has been deeply involved in Navy Acquisition and Training Research Projects. For example, Dr. Glaze developed the methodology and planned to conduct a Post Implementation Reviews of NCTSS OOMA, an ACAT IAC acquisition programs to provide an automated management information system to Navy and Marine Corps aviation maintenance units. Dr. Glaze was also a key influencer in the preparation of CARD and Life Cycle Cost Estimate systems in support of Navy Milestone C. [paigeglaze@gmail.com]

Mason Beninger—CTFL, CTFL-AT, CMT, is a research analyst at nGAP Incorporated in the area of procurement policy, business processes, and web software system design. [mason@ngap.com]

Abstract

This paper looks into the process known as concurrent contract modification (CCM) and proposes a solution to automate it. While CCMs are not inherently disallowed, they do present contract logistical and administrative problems. The larger projects become, the more difficult it is to track, administer, and document proposed changes. Without an efficient means toward managing incoming changes, gaining any tangible and accurate reporting on project outcomes proves significantly challenging if not impossible.

What this paper proposes is a new approach to contract change management utilizing a software tool designed for ground-level operations that scale up to Contract and Program responsible stakeholders. Instead of relying on the output of contract writing systems, this system can be used to manage the execution of several related contracts under a single project with shared sources of funding. Focusing on automated infrastructure and a process for contract change management will allow for greater insight and accountability at program execution levels, in the case of the Navy, at the Regional Maintenance Center or Shipyard level.

Concurrent Modifications, The Act of Optimizing Multiple Contract Changes Under One Universal Modification in the Acquisition, Maintenance, and Program Management of Large Military Sea, Land, and Air Platforms

Introduction

This project explores whether current processes for contract modifications can be optimized to better respond to today's environment of rapidly changing requirements due to unforeseen events that occur during the building and maintenance of large military sea, land, and air platforms that affect delivery, cost, program management, and planning.

Critical variables that affect acquisition and maintenance efficiency include emergent technologies that impact outcomes and contribute directly to unforeseen cost overruns. The rapid pace of technological change creates challenges to the acquisition community due to requirements that may not always be fully known at the time of contract award. Currently, changes for any reason require multiple contract modifications, which over time have become unwieldy and inefficient to manage; moreover, the cost impacts of multiple contract modifications have become difficult to assess, therefore making auditability difficult if not impossible.

Using the Navy's shipbuilding environment as the point of departure for the analysis, this project develops a proof-of-concept alternative contracting system that allows concurrent contract modifications, whether executed in parallel or sequentially, to be prioritized according to tailorable rule sets in a manner that allows users to monitor, manage, change and report total contract award in real time. This proof-of-concept also aims to provide solutions to other complexities inherent in today's contracting environment, such as allowing for multiple contract types within and between Contracting Line Item Number (CLIN) structures and within a single contract award, and the management of multiple Technical Instructions, CLINs, and SubLine Item Numbers (SLINs).

The aim of the final contracting system is to create the required data relationships in a single system for the purpose of monitoring contract cost and technical scope in real-time, thereby increasing transparency and auditability.

Examples abound regarding the difficulties the DoD has in forecasting cost and managing changes that affect key elements in the building and maintenance of large platforms. This is especially true regarding seagoing platforms such as submarines, carriers, littoral combat ships, and destroyers. These examples include the following:

- Through the course of a decade, the Littoral Combat Ship's program went from an estimated cost of $220 million per ship to an average currently at $478 million apiece, with more changes afoot (GAO, 2016).
- The Navy's number one budget priority, the Columbia Class Submarine, has already projected cost overruns before the first platform is even built due to uncertainties regarding critical emergent technologies (GAO, 2016).
- DoD Contract Management, Weapons Acquisition, and Support Infrastructure Management are all represented on the GAO's High-Risk Ledger.
- Cost overruns are imprecisely estimated and continue to provide challenges to the DoD that significantly impact performance and outcomes, in particular: shipbuilding. Multibillion-dollar cost overruns are common and, in many cases, expected.
- Documenting these challenges has proven difficult, affecting auditability, transparency, and effectiveness. This impacts the nation's leading edge in maintaining global military superiority.

Providing an optimized automated process for concurrent contract modification that reports situation awareness in real time will significantly add to the goal of excellence in DoD acquisition.

The aim of this proposed approach to concurrent contract modification is a process, supported by an agile software tool, to coordinate serial changes in projects that involve one or multiple contracts to increase acquisition excellence through concurrent modification, or "real-time" situational awareness. To quote Socrates, "Knowledge of the right leads necessarily to right acts" (Gilje & Skirbekk, 2017). Therefore, the aim of this project is to give direct real-time access to the execution of program funds and activities via the reporting of contractual transactions at an elemental granular level "in real time." This will allow Acquisition and Program Management stakeholders to have access to global and granular information that is critical to effective real-time decision-making that affects cost, planning, and delivery outcomes. In short, the purpose of this project is to provide knowledge in the form of global, granular, structured contextual reporting on all acquisition program management parameters in real time as to ensure the right actions are taking place.

The federal government has struggled with providing a consistent award and budgetary data repository that includes truly accurate information. For example, the Digital Accountability and Transparency of 2014, or DATA ACT, requires the federal government to transform its spending information into open data (Data Coalition Organization, 2018). However, there have been speed bumps along the way to its implementation. For example, in a report released by the GAO in November 2017 regarding DATA Act compliance, the GAO found that the consistency of required data submittals to BETA.USAspending.gov was faulty at very high levels. The agency found that "approximately 94% of all records … differed sharply between budgetary and award records," making any real decision-making and analysis based on those records false at the outset. Regarding award sub data, the GAO reported that the actual award information was "inconsistent with agency sources for 62 to 72 percent of all awards" (GAO, 2017). Another GAO report found that personal services contracts from the DoD lacked accurate data, therefore, "proper management of personal services and other contracts contained inherent difficulties that impact performance, reporting, auditing, and closeout" (GAO, 2017). This anomaly makes the lessons learned process, the budgeting process, and other critical elements affecting optimal acquisition all but impossible.

While this project could involve the creation of an entirely new acquisition infrastructure, this approach is not designed to replace enterprise-wide software systems for contract writing and reporting. The method proposed is intended to reside at the Contract and Program Management level to provide Concurrent Change Management to bring full accountability to all program elements, including budgets, task order, technical/task instruction, and funds expenditure. It is intended to provide for seamless collaboration, all integrated into one enterprise that provides real-time visibility and reporting capability into all project activities in "real time." This project provides for accurate reporting to the penny, cross-referenced to one or all relevant acquisition activity. In other words, truly accurate information that is currently lacking.

What Are Concurrent Contract Modifications?

Concurrent contract modification (CCM) is the process of simultaneously processing multiple contract changes against numerous contract vehicles that affect a project or program—for example, the $800 million maintenance project (Harper, 2017) for an Ohio Class Submarine in San Diego, CA. One concurrent modification of $800 million could include, but is not limited to, a dearth of actions including additions, deletions, new work, payments, new funding, additional option exercise, delivery schedule extensions, stop orders, and terminations. CCM exists due to the operational need to adjust contracts to suit ever-changing requirements rapidly in venues such as the Navy shipbuilding environments. CCM has no precedent in the Federal Acquisition Regulations (FAR), but the process is a unique agency-specific interpretation of the FAR regarding contract modification. The agency using this process is the Navy. Further complicating the matter is the scale of interaction between government staff and contractors in support of ship maintenance projects. As of 2017, the operating costs for the U.S. fleet was $56 billion a year. However, testimony before the Subcommittee on Seapower of the U.S. Senate Armed Services Committee projected that total Navy operating costs would increase to an average of $102 billion per year through 2047 (Labs, 2017). Shipyards such as Puget Sound Naval Shipyard and Intermediate Maintenance Facility IMF in Bremerton, WA, executed nearly 2.3 million-man days of work and employed approximately 12,340 civilians, accounting for a large percentage of the current budget (Bradlet et al., 2017). Managing these budgets and reporting on them in a format acceptable to key stakeholders inside and outside the DoD, including Congress, has proven to be a challenge. It's normally accepted that the DoD, particularly the Navy, is "un-auditable" (Nader, 2014). One of the current priorities of the Navy is to be fully auditable by fiscal 2018. An optimized Concurrent Modification Process and automated infrastructure contribute significantly to this purpose.

A simple analogy follows for the purpose of illustrating the logistical challenges of a large project with many contractors. This is a fictional example but with relevance to any project manager. In this analogy, there are contractors working on a multimillion-dollar boat overhaul project. This boat's maintenance schedule calls for it being homeported no more than three months. The boat arrives, and work begins with the 10 contractors and their subcontractors. About three weeks into the overhaul, one of the contractor engineers find that the engine assembly's wiring is in a complete state of disrepair and requires an urgent fix. The maintenance schedule did not foresee the need for the wiring to be worked on. Therefore the project management staff has to decide whether to issue a new contract, issue a change order to one of the 10 contractors, or issue an instruction. The staff decides to issue a change order for one contractor who specializes in electrical engineering. The electronic engineering firm tasked with rewiring the engine block has to replace the wiring as fast as possible because it will delay other contractors. That forces other engineering contractors to stop their current work. That stoppage prevents the other contractors from continuing or completing their work and so on and so forth. The work change order then forces the project management staff to adjust the other contracts to reflect a new period of performance, delivery or start dates, etc., for some tasks. As the project staff begins making changes to contracts, it creates a ripple effect that eventually will throw off the project's timeline, including the budget and the ultimate release of the boat back out to sea.

Therefore, as a result of constantly modifying contracts, the project management staff develops a way to make changes in bulk against all contracts and at a later date officially modify the contracts affected. In this way, the staff has a running list of all changes made to the project's contracts and can promptly issue work orders, changes, or modifications to keep tasks on schedule. For small projects, this is potentially workable, but for larger projects, this creates opportunities for errors, miscalculations, and reporting mistakes. For instance, what happens if the project has to make a change to a change? If there is a list of changes not yet officially modified into the contracts, what is the actual value of the contracts currently? How much has been actually spent compared to the budget of the contracts? In essence, the aforementioned antidote is what concurrent modifications are—the struggle of keeping projects on time and on budget while at the same time conforming to regulations on contract administration.

Continuing with concurrent modification, the main issue in the discussion toward optimal practice is the management of hundreds of requests for contract changes from end-user stakeholders, program managers, on-the-ground engineers, specialists, and project managers against multiple contract vehicles with various periods of performance, contract types, cost types, cost ceilings, and various contractors with different reporting requirements. Additionally, each contract could have separate administrative staff that could be geographically spread across multiple jurisdictions. Without a viable concurrent modification engine, tracking changes, ensuring funds are available, and handling critical actions such as technical reviews become more and more unmanageable. As a result, without a Concurrent Modification Infrastructure, changes are aggregated and the responsible contracting officers then “modify” the appropriate contracts at a later date.

In the meantime, contracting officers can, under certain conditions, authorize work to commence with the expectation that the contract will be confirmed later (Naval Regional Maintenance Center, 2013). The only issue here is that in most cases, the contract is never confirmed. This makes closeouts, accurate reporting audits, and other critical elements of the process all but impossible. These challenges reoccur with every class of ship and are an ongoing problem for the Navy. Adding yet more confusion to this process is the effect of change orders on different contractors and their ability to deliver on time and within projected costs. Additionally, chains of changes on one contract can have a domino effect directly as well as indirectly, forcing other contracts to be changed that affect a project. Without effective Concurrent Modification Protocols, the result is a significant administrative and paperwork backlog to conform contracts, resulting in significantly less efficient reporting and often no reporting at all. It’s been said many times in Navy Pentagon Program Shipbuilding Offices that the choice is obvious: “Do we focus on building new ships or on closing out old platforms where there is no information available?” (Senate Armed Services Committee, 2015). We focus on building new ships. The fallacy here is that building new ships depends on lessons learned from building, maintaining, and closing out retired ships. If that data is consistently not available or outright lost, the problem is obvious. The Navy proposes alleviating this reality with Optimal Concurrent Modification functionality using new technology. This technology is at the forefront of this document.

Looking at the root issue of CCM from a pragmatic perspective, building and/or maintaining a ship presents a logistical problem regarding the program and contract administration. There exists no way to effectively track changes that everyone expects are bound to occur during the build or maintenance lifecycle. The lack of accurate and documented contract and program changes, especially those regarding award data, can be extremely detrimental. Take, for example, mandated systems such as the Federal Procurement Data System–Next Generation (FPDS–NG) and data integrity. A December 2016 Congressional Research Service Report on Defense Acquisitions Spending and Reporting warns that “decisionmakers should be cautious when using reported obligation data from FPDS to develop policy or draw conclusions. In some cases, the data itself may not be reliable” (Schwartz, 2016). While this paper will not make a judgment against the efficacy of FPDS–NG, the main theme of complaints regarding FPDS–NG is the lack of accuracy and missing information in system data sets. The Department of Commerce, Office of Inspector General (IG) in 2015 found that the department needs to improve the “process for entering accurate and reliable data into FPDS–NG and its controls to properly maintain and safeguard contract files entered into the system.” The IG found that undefinitized actions (UAs), contract actions issued as letters contracts, and other instruments used to meet an urgent requirement of an agency contained coding errors due to a lack of training. However, more distressing, the IG also found that actual “contract files and FPDS–NG data sheets were missing” (Office of Inspector General, 2015), rendering the information all but useless to the informed user.

In the DoD, the lack of an adequate process to track contract changes and their award dollar obligations has had disastrous effects on public and congressional relationships. These issues range from the Army's $6.5 trillion of "wrongful adjustments" in 2016, where the Army lacked receipts and invoices or simply made them up (Paltrow, 2016), to the Navy's massive procurement scandal involving the ongoing investigation within the Navy involving ship support contractor Glenn Defense Marine Asia (GDMA), a subsidiary of the Glenn Marine Group, "Fat Leonard" scandal (Paul, 2017).

The key takeaway from one review of these events is the need to improve methods for creating, managing, and generating award data that allows for a contextual understanding of the data sets. Understanding such sets will improve efficiency, create effective work solutions, and as an added benefit, catch waste fraud and abuse. An improved method for creating, managing, and generating award data will also enable easy, intuitive learning curves for end users on the ground, those that live and breathe the project. By giving end users easy to learn, easy to understand, and easy to use methods for direct involvement into the processing and reporting of contract changes, users in the field will acquire easy tools to do their work and contribute to efficiency and delivery outcomes. Rather than waiting for official audits and reports, the system proposed will give end users, privileged managers, and stakeholders the ability to recognize anomalies quickly and provide for prompt cost and time-saving response. For example, in a 2016 paper on procurement fraud in the DoD, the authors posit that "shifting the first line of defense against procurement fraud should be the procurement workforce managing the contracting process," not the contract auditors and fraud investigators. They added that "missing from the DoD's response to procurement fraud risk is a more strategic approach to fraud deterrence and detection that includes emphasizing procurement workforce training, contracting process capability, and internal control effectiveness" (Rendon & Rendon, 2016)

The challenge to solve is not only to automate the submission of changes, types of changes, contractual conditions, approvals, and notifications, but also to also track these actions in real time, with the end goal of producing elemental or granular data on each contract action. Information such as who made the request, who approved the changes, which account/ACRN was obligated, who is the contractor, where exactly is the place of performance, what was ordered, when invoices were paid and by whom, and lastly, why does this change need to happen and how does this change impact the outcome of the project.

Proposed Design Proof of Concept Research for Concurrent Modification Management and Alternative Contracting System

This section will explain system design, processes, and outcomes. As an overview, the proposed method addresses the following challenges in concurrent modification:

1. Multiple modifications executed in parallel with numerous accounting classification reference numbers (ACRNs) targeting various Subline Item Numbers (SLINs)
2. Various task or technical instructions
3. Multiple Contract Line Item Numbers (CLINs) dictating different contract reimbursement types, i.e., fixed fee vs. cost reimbursable vs. a combination of the two
4. Reconciling the previous items to determine total contract award in real time, for monitoring and managing CLIN ceilings
5. Warehousing large volumes of ancillary data electronically in an easily accessible format

This section is broken down into two parts: the design philosophy and business process explanation. From a computer software design and acquisition perspective, the wording and terminology are simplified. The particular case with concurrent modifications is, generally, a Navy-centric process, and terminology and policies may not or do not apply to other branches. This section attempts to encapsulate and generalize procurement process concerning software systems. The intent of the authors is to break down the government procurement process to the root elements and define capabilities for the design of a system.

Design Philosophy

The system is designed to address several challenges in the contract modification process.

1. Efficiently managing contract change requests with large groups of contracting specialists, requirement holders, managers, and contractors
2. Tracking approved changes and budgetary implications
3. Aggregating changes and applying legal modifications to groups of related contract vehicles
4. Creating data model linkages between budgetary accounts, contract level funds, and expenditures on the line item level
5. Enabling detailed expenditure reporting against contract modifications, in real time.

To summarize the process, a method of rapid change management and tracking called the Rapid Contract Change Management Model, or bicycle model, is introduced in Figure 1. The procedure outlines three areas of focus: the contracts represented as the seat, the left cycle representing change management process for requirements holders and managers. The gear or center represents the aggregation and reporting of priced and unpriced changes; the right sequence represents the legal contract modification process for contracting specialists. Lastly, reporting represent the handlebars that connect the contract.

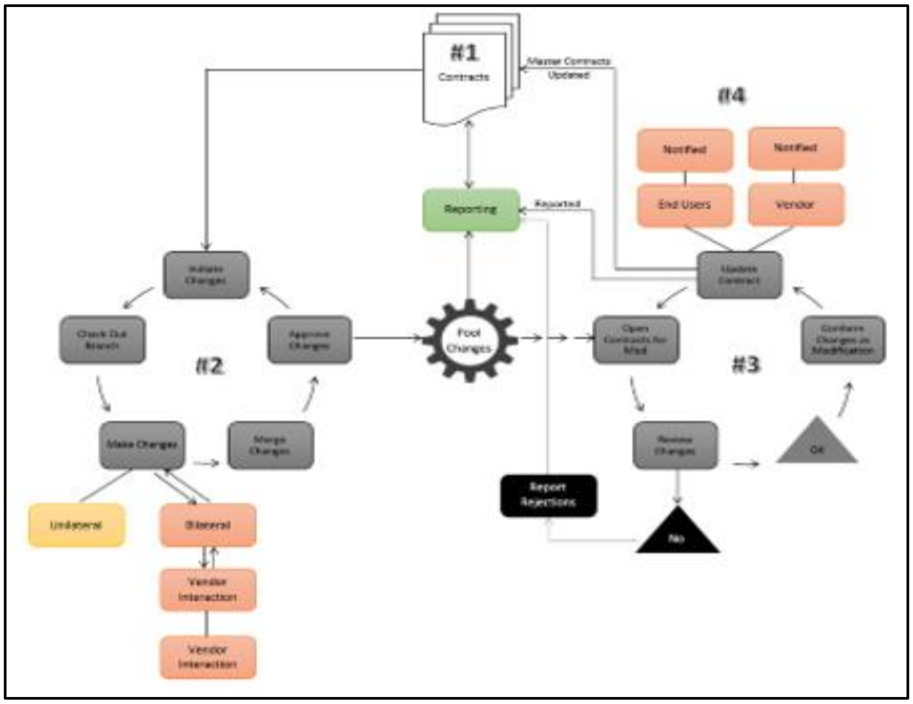


Figure 1. Bicycle Model

The model is designed to support three different stakeholder groups: requirements holders/managers, contracting officers, and specialists. The model is also designed to bind the various stakeholder groups together in an interactive environment that tracks overall contract changes and gives managers elemental level reporting capabilities necessary for effective decision-making. The model has four components.

Component 1: The Contract—Seat of the Change

The driver of change, the contract, is the platform for directing the entire change management process (see Figure 2). This presentation of the main contract file is designed for ease of use, ease of readability, and ease of learning. In the system, the primary determinants of change are new non-structural requirements and fund availability. The fund management functionality will be addressed later in the paper. The contract document itself is a pseudo-representation of the Uniform Contract Format (FAR, 2018) where critical performance is highlighted and focused. In this model's case, sections A, B, C, and G are primary, while the remaining sections are indirectly linked. For example, section F, deliveries or performance, and section E are connected to section B at the line item level. The purpose of this arrangement is to enable users to focus on the critical aspects of the contract's management and at the same time generate the required compliance data as the user works through the system. In other words, spend time managing procurements to achieve outcomes first rather than spending time filling paperwork for paperwork's sake. The result is a user experience that lets the system automate the mundane and free up critical attention to other areas of contract administration, all within the same infrastructure. The result is a focus on the optimal outcome while having the confidence that the system will manage compliance automatically.

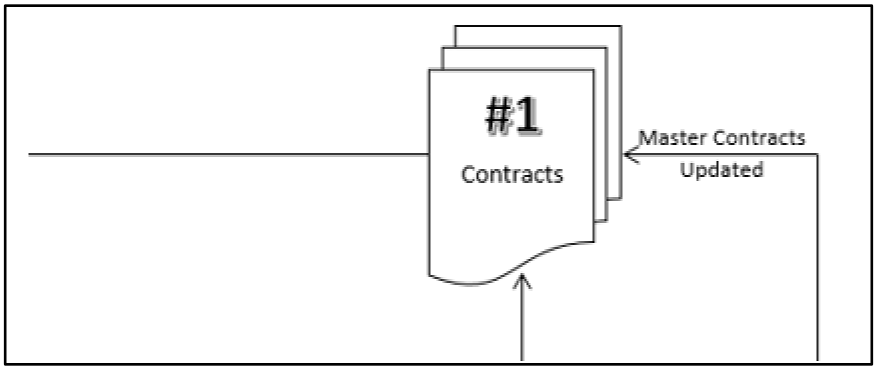


Figure 2. Contract Module of Bicycle Model

Line Item Level Management and Insight

A critical section of the contract and what the system brings to the fore is real time line item level management. From a data model perspective, the CLIN and SLIN relationships are what binds the contract together with stakeholder groups. Table 1 illustrates that data relationship.

Table 1. Data Relationship

ACRN	Number	Short Description	Cost Type	UOM	UNIT	RATE	Extend Price	Deliver/Performance Location	Contractor

Each column represents a relationship that builds the contract, binds parties together, and is subsequently affected when making contract changes. Any changes made to these data points result in a contract change that needs to be tracked, categorized, approved, consolidated, and legally modified. The overall goal of this approach is to make the contract file genuinely interactive, reportable, and friendly to end users.

Cost Type Management

In the system, cost type indicates how CLIN/SLIN pricing is determined and accounted for. Each line item has a cost type that requires the contractor(s), whether for new awards, modifications, or changes, to provide the relevant pricing information. The contracting user is presented with different template interfaces based on cost type. This allows contractors to price line items and at the same time allow government users to perform Independent Government Cost Estimates (IGCE) for each line item. The result is the ability to compare directly, line by line, the contractor's reported price and the government's estimates.

Contract Level Reporting Functionality

Key to this model is the idea of contextual, elemental, and relationally linked data. In other words, the data an authorized user can see is easy to read, understand, and comprehend. Rather than looking at aggregates, the system gives an elemental level at contract line level(s), fund expenditure level(s), vendor/contract level(s), task order/technical instruction level(s) and user level(s). To this effect, the data is organically generated as a result of utilizing the system rather than keying in data. In the model, every action during the change management process is documented, tracked, and reportable in required and ad hoc formats depending on user preference and privilege.

Component 2: The Change Management Cycle

The first construct in this model is the Master Contract File (MCF). The MCF is the simple data model representing component 1 addressed earlier. Any user attempting to create changes automatically generates a local Contract Branch visible to the user and those the user has chosen to collaborate with. The master and branch concept is the primary mechanism for organizing changes with a multitude of users against multitudes of contracts with multiple contractors.

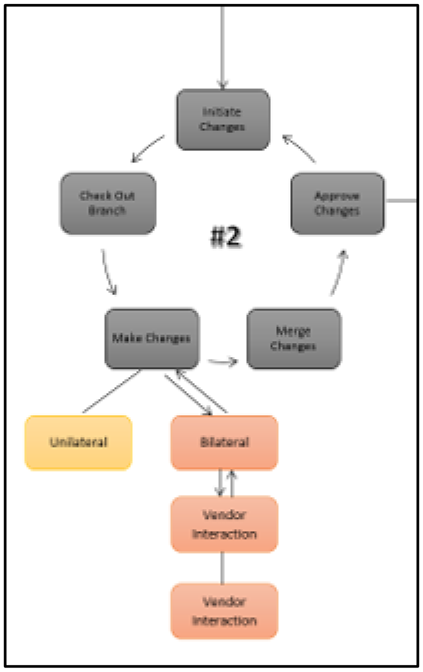


Figure 3. Change Management in Bicycle Model

This cycle involves the end-user assigned to enforcing contract performance and building change requests. This cycle has built-in workflow and is designed to receive changes continually. The sequence can handle multiple users making multiple applications for contracts.

1. A user would create a request for contract change (RCC). An RCC is a local copy of a contract, called a branch.
2. Depending on the scale of changes or collaboration, the user can invite other users to collaborate on an RCC.
3. Users then begin their work by making additions, subtractions, deletions, and other changes to line items, statement of works, fund management. They issue technical instructions/task orders and issue stop orders when applicable.
4. The system categorizes and analyzes the changes and determines each difference as either unilateral or bilateral.
5. Bilateral changes, depending on the nature of the modification, generate a need for contractor concurrency to conclude a supplemental agreement. In the event of a technical instruction or new work, a request for change is created and sent to the contractor to gather pricing and other information. The contractor sends back their response, and the data is applied to the requisite branch of the contract.

6. Once all changes are made, the initiating user then checks for duplicate or overlapping changes. Should there be overlapping or conflicting changes, the initiating user adjudicates the conflicts and merges the changes into one coherent document. Keep in mind that nothing is official until changes are approved.
7. The user then decides whether to continue allowing changes or stopping their acceptance. If the user leaves the branch open, more changes can be made and merged. If the user is ready, the changes are submitted to workflow.
8. The initiating user then sends the package of contract changes into a workflow process based on agency organizational rules and policies.
9. If the package of changes is approved, the package becomes a part of the concurrent modification. As mentioned previously, concurrent modification means one official action consisting of many changes rolled up into one universal modification. All concurrent changes can be rejected using this model.

In ending, the change cycle allows for continuous and controlled change management within a localized version of the contract. The deltas or changes are categorized and sorted by their FAR-defined types of modification. Finally, the cycle allows the user to continue making changes without a need for a pause in the process. As an aside, the next section continues the discussion on the format of the contract file.

Interactive Contract File Data Model

The goal is to transform the traditional electronic contract file, as specified by the UCF FAR 15.204-1, and treat the electronic contract file (ECF) as genuinely electronic. In other words, be a central repository for contract and award data that legally complies with the definition of an ECF that can be managed electronically. This includes the tracking of contract modifications, task instruction/orders, stop orders, funding allocations, administrative changes, protests, and close out—a "living contract" document so to speak. The contract file itself is the vehicle for direct management versus being a reference point on which to base managing the action.

The concept of a paperless contract file is not a new idea. Since 2000, the DoD has been implementing paperless contracting processes. To list a few cases, Standard Procurement System (SPS), Wide Area Work Flow, and many other systems geared toward support of a paperless environment (Sherman & Freeman, 2007).

The fundamental difference between a paperless contract file and a genuinely electronic one is the degree of interactivity between contracting stakeholders and resource owners. The second facet is the degree of use the data represents. To be clear, this is not an electronic filing system but a systematic automated method to manage changes and track them in real time—in other words, concurrent modification.

Pooling of Approved Changes

The act of pooling or aggregating approved changes serves as a controlled intersection for incoming contract changes. The point, represented as a gear, serves as the gatekeeper from changes or sub-modifications made from contract branches before they are released into the next cycle. Critical to this model is the reporting module sitting directly above the changes.

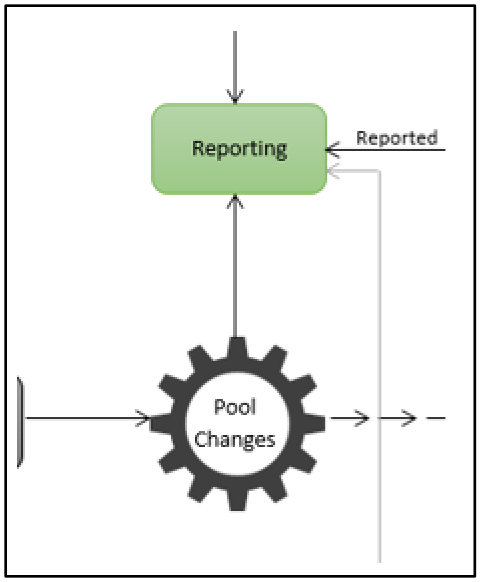


Figure 4. Change Pooling and Reporting in Bicycle Model

The reporting module is the central repository of change tracking—logging and analyzing contract deltas. Deltas, the real alphanumeric changes between the branch contract and master contract. The centrality of this reporting center allows managers and audit users to peer into the ongoing changes on any contract at any time. The reporting module will be revisited later in this paper.

A final note on this section revolves around compliance and adherence to the DATA ACT. As part of the user design, a more straightforward interface is used for the best user experience, but the back-end data model is structured with compliance in mind. The data model is represented twice—a simpler relational structure of the contract, the related changes, and the DATA ACT Compliant Extensible Markup Language (XML), and (extensible Business Reporting Language) (XBRL) based on Data Act Information Model Schema (DAIMS). The latter format option allows for seamless output to external systems for compliance while the former preserves a simpler user interface and experience.

Component 3: The Contract Conformance Cycle

The next cycle is the domain of the contracting officers, contract specialists, and resource managers (see Figure 5). The themes of this cycle are resource obligation, legal reviews, consultations, and notifications.

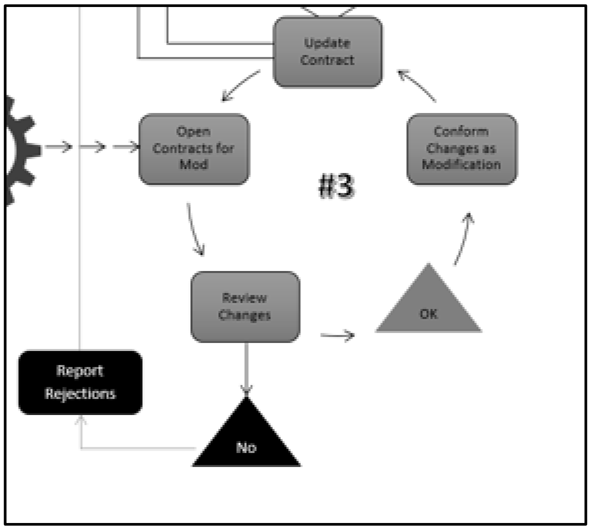


Figure 5. Contracting Conforming Process in Bicycle Model

The following section explains the process step by step.

1. A contracting officer (administrative or not) reviews contract changes incoming from the pool.
2. The contracting officer “opens” the affected contracts for legal modification and confirmation.
3. The contracting officer organizes what changes are going to be made in this modification cycle. While the contract is open for modification, transitions originating from the first cycle are allowed to pass through.
4. The contracting officer formally rejects or approves the collective batches of changes. Rejected changes are sent back to the originating user for adjustment and resubmission.
5. Once satisfied, the contracting officer “closes” the contract for modification. As a result, all incoming changes are held in the pool.
6. The contracting officer then “conforms” the changes into the master contract. The act of conforming does the following:
    a. Obligate/modify/remove funding from the affected CLINS/SLIN
    b. Textual changes in the document such as statements of work
    c. Confirms the task/technical instructions, task orders, work and work stop orders
    d. Changes the value of the CLINS/SLINS and therefore the value of the contracts
    e. Add/edit/remove contract provisions/clauses
    f. Add/remove CLINS/SLINS/ELINS
    g. Generate a “modification changelog”

    In essence, the conformation process is the application of the pooled changes into the master contract file.
7. The master contract file is updated.

The cycle is designed to handle not only changes against a single contract, but to manage changes against multiple contracts, all at the same time.

Component 4: Notifications and Reporting

The cycle now comes full circle as the originating users and contractors are notified (see Figure 6).

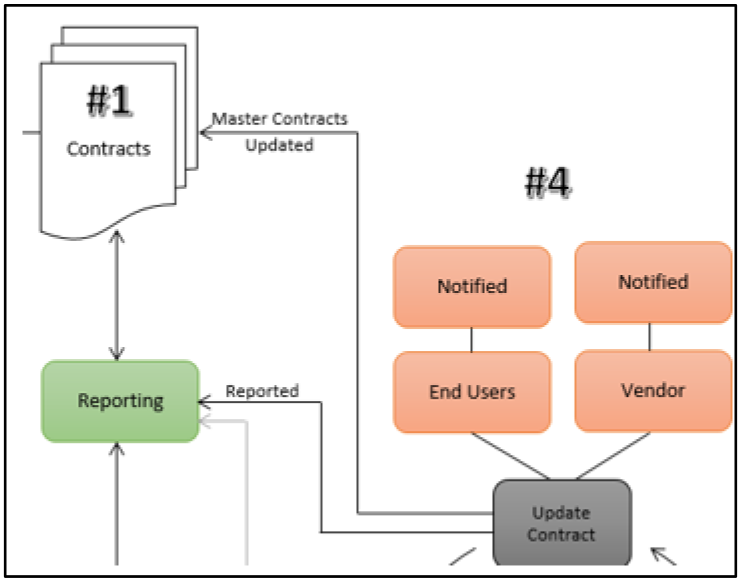


Figure 6. Notifications and Reporting in Bicycle Model

Contractors/vendor users receive several types of notifications depending on the nature of the change. In the event of a supplementary agreement, official contractor sign-offs are required by the contractor point of contact as part of the modifications process. If it's administrative and concerns the contractor, a summary of changes is sent that the contractor POC acknowledges.

End users—defined as on the ground engineers, project managers, contracting specialists, and contracting officers—are notified of the changes. Each end user who made a request receives a report of the summary of their differences that were modified and those that were not. The end user, looking at the master contract again, will see the updated contract and can initiate a new change.

Reporting Revisited

Approved changes have been conformed, and the result is exported to the reporting module. As mentioned previously, the reporting module looks at the two phases of the contract change process—the changes as they are made and the changes that were incorporated into the contract. The double entry of change allows auditors or managerial users critical insight into each contract as it changes in real time. These actions are available for review, desk audit, and official audit long after the contract is completed and closed out. Adopting this understanding Navy-wide would render Navy shipbuilding a fully auditable endeavor.

Business Logic

This section will show how each of the principled features is individually addressed. The format will be of a statement and direct answers.

1. Multiple modifications executed in parallel with numerous accounting classification reference numbers (ACRNs) targeting various Subline Item Numbers (SLINs)
    a. The change cycle, the left side, is responsible for managing all changes. Users would make their adjustments to the CLINS/SLINS and associated ACRNS as if they were editing the contract. The system analyzes and categorizes the changes and applies the relevant business logic. In the case of ACRN, where fund management comes into play, the act of adding/removing a fund cite, adding/lifting funding, or moving funds constitutes three separate changes.
    b. As explained in the previous section, the collection of changes goes through a vetting and approval process to remove duplicates and mediate conflicts. Once completed, all approved changes are pooled awaiting formal processing by the contracting officer(s) for the various contracts.
    c. Once the contract(s) are opened for conforming, the changes are sent through to the right side of the cycle.
    d. As soon as the modification has been grouped, approved, and funds appropriated, simultaneously, the package of changes are applied to the affected contracts.
2. Various Task or Technical Instructions (TI)
    a. Tis follow a similar process; however, should an instruction require pricing information, the vendor must respond. The response then must be reviewed and approved. Once approved, the task or technical instruction exists on the CLIN/SLIN structure or references a CLIN/SLIN structure.
    b. The procedure of generating Tis generates a series of indirect changes such as funding allocation.
    c. The group of changes related to a TI’s preparation is added to the more prominent catalog of changes made, which are then pooled, as mentioned earlier.
3. Multiple Contract Line Item Numbers (CLINs) dictating different contract reimbursement types, i.e., fixed fee vs. cost reimbursable vs. a combination of the two
    a. The CLIN/SLIN cost type determining reimbursements as specified in the master contract can be edited and adjusted by end users.
    b. As part of the analysis of the type of changes, the system prompts users to adjust the pricing/costs associated with the line if the type changes.
    c. These changes are made as are other bilateral edits to the change package that will be reconciled, approved, and finally submitted to the pool, ready for conforming.

   d. Funding requirements are taken into consideration when making changes in value to each cost type.
4. Reconciling the previous items to determine total contract award in real time, for monitoring and managing CLIN ceilings
   a. The system determines the total obligated value of the contract via the sum of all CLINs.
   b. The system also has a total budgeted contract value that is the sum of all the fund cites associated with the contract.
   c. The contracting officer can set ceilings on the global or contract level or via the CLIN level with configurable options to prevent new work until a review, or to alert users.
   d. While ceilings can be placed at the contract level, fund site and ACRN specific ceilings can be positioned to allow flexibility based on account or appropriation. For example, setting a threshold or time limit of money for 70% and a time limit for the first fiscal year of funds.
5. Warehousing large volumes of ancillary data electronically in an easily accessible format
   a. Reporting on these changes are mission critical. As explained earlier, the data structure allows the user to investigate several avenues of the model.
      i. Transaction layer, elemental view: a user can look at task/technical instruction level or task order level on each CLIN
      ii. Funding layer, macro view: a user can look at transactions against the funding instruments associated with each CLIN/SLIN
      iii. Change layer, contract view: a user can view all the transaction, changes, and activities on the contract
      iv. Project layer, program view: a user can see all deals and their transactions associated with a project. The following section will explain projects.
   b. Most importantly, the data presented is contextual, easy to read, easy to access, and easy to comprehend.

Large Program Management and Concurrent Modifications: The Bigger Picture

The bicycle model with a single contract to manage, while capable, is limiting in potential. When the context of contract management is adapted to a more extensive program view, the bicycle model provides the ability to manage large numbers of changes. A new construct is introduced called a project.

The traditional definition of a project is temporary in that is has a defined beginning and end in time, and therefore defined scope and resources (Project Management Institute, 2018). A project, as described in this paper’s context, is set as a data container for acquisition outcomes, funding, and managing users. A project can form a base unit to create the basis for the program. More importantly, the project is a container for program funding. A project is a central point for all procurement actions, from creating purchase requests, generating a request for contract change, managing contract tasks and line items, and tracking expenditures.

How does the system tie a program objective, program element line of account, contract, managing users, managing funding, and tracking changes? Projects are the answer. They exist to categorize, organize, and structure the data. The following graphical example is a simple, fictional naval maintenance program structure guide.

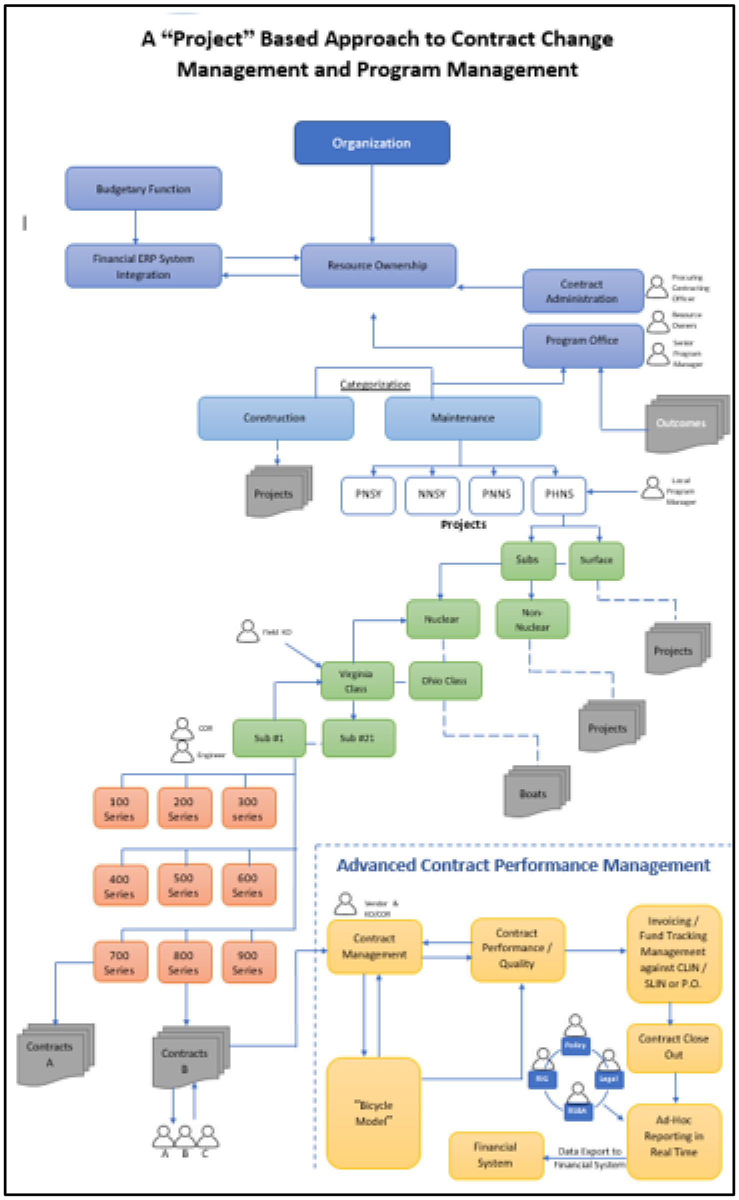


Figure 7. Project-Based Approach With the Bicycle Model Included

This example represents a complex organization with two functions: construction and maintenance. Each has funding appropriated with assigned program staff. Under the maintenance project is a child project for all activities at a particular shipyard. Under the shipyard project are projects for surface and subs, while further down is nuclear and non-nuclear. Below the nuclear class are the class of subs and finally under that, the actual sub itself. On that particular sub exists the various sections of the sub and at the root exists the various contracts supporting the serious projects. At last, the elemental level is the contract itself (see Figure 7).

By choosing the level and grouping elements, users have varying control on how many contracts can be modified at the same time, whether at the series level, the actual sub, or subclass level. The bicycle model itself is a small but important part of a larger program structure picture. This would be total visibility to support change management not only at the contract level but at the program level as well—in other words, CCM real time situational awareness.

Conclusion

This paper examines the process known as concurrent contract modification and proposes a solution to automate it. While CCMs are not inherently disallowed, they do present a significant contract logistical and administrative problem resulting in challenging outcomes. The larger projects become, the more difficult it is to track, administer, and document proposed changes and their impact on planning, cost, program management, and delivery. Without an efficient means toward managing incoming changes, gaining any tangible and accurate reporting on project outcomes proves significantly challenging, if not impossible.

This paper proposes a new approach to contract change management, utilizing a software tool designed for ground-level operations that scale up to contract and program responsible stakeholders. Instead of relying on the output of contract writing systems, this system should be used to manage the execution of the many related contracts under a single project with shared sources of funding. Focusing and bringing an automated infrastructure and a process for contract change management will allow for greater insight and accountability at program execution levels. In the case of the Navy, at the Regional Maintenance Center or shipyard level.

References

Bradlet, M., McMahon, M. E., Riposa, J., Kallimani, J. G., Bohman, A., Ramos, A., & Schendt, A. (2017). *A strategic assessment of the future of the U.S. Navy ship maintenance: Challenges and opportunities*. Santa Monica, CA: RAND.

Data Coalition Organization. (2018, March 3). *The data act*. Retrieved from https://www.datacoalition.org/issues/data-act/

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2018).

GAO. (2016). *Littoral combat ship: Congress faced with critical acquisition decisions*. Washington DC: Author.

GAO. (2017). *Columbia class submarine: Immature technologies present risks to achieving cost, schedule, and performance goals*. Washington, DC: Author.

GAO. (2017). *Improvements needed in how some agencies report personal service contracts*. Washington, DC: Author.

GAO. (2017). *OMB, Treasury, and agencies need to improve completeness and accuracy of spending data and disclose limitations*. Washington, DC: Author.

Gilje, N., & Skirbekk, G. (2017). *A history of western thought: From ancient Greece to the twentieth century.* Routledge.

Harper, J. (2017, March 22). Cost of new submarine could threaten Navy fleet expansion. *National Defense Magazine*. Retrieved from http://www.nationaldefensemagazine.org/articles/2017/3/22/cost-of-new-submarine-could-threaten-navy-fleet-expansion

Labs, E. J. (2017). *Testimony: Costs of building a 355-ship Navy.* Subcommittee on Seapower of the Senate Committee on Armed Services.

Nader, R. (2014, October 4). Giant Pentagon budget is unauditable year after year. Retrieved from https://www.huffingtonpost.com/ralph-nader/pentagon-budget_b_4046370.html

Naval Regional Maintenance Center. (2013). *Shipbuilding specialist desk guide*. Norfolk, VA: Naval Regional Maintenance Center.

Office of Inspector General. (2015). *Inaccurate reporting of undefinitized actions in the Federal Procurement Data System: Next generation.* Washington, DC: Department of Commerce.

Paltrow, S. J. (2016, August 19). U.S. Army fudged its accounts by trillions of dollars, auditor finds. Reuters.

Paul, P. (2017, November 11). What is Fat Leonard scandal? 440 Navy officials investigated for taking bribes. *International Business Times*.

Project Management Institute. (2018). *What is project management?* Retrieved from https://www.pmi.org/about/learn-about-pmi/what-is-project-management

Rendon, J. M., & Rendon, R. G. (2016). Procurement fraud in the U.S. Department of Defense: Implications for contracting processes and internal controls. *Managerial Auditing Journal*, 748–767.

Schwartz, M., Sargent, J. F., Jr., Nelson, G. M., & Coral, C. (2016). *Defense acquisitions: How and where DoD spends and reports its contracting dollars.* Washington, DC: Congressional Research Service.

Senate Armed Services Committee. (2015). *Senate Armed Services Committee holds hearing on the Navy posture in review of the proposed fiscal 2016 Defense Authorization.* Washington, DC: U.S. Senate.

Sherman, B. J., & Freeman, E. (2007). *Paperless policy: Digital filing system benefits to DoD contracting organizations*. Monterey,CA: Naval Postgraduate School.

Assessing Vulnerabilities in Model-Centric Acquisition Programs Using Cause-Effect Mapping

Jack Reid—is a graduate student with the Systems Engineering Advancement Research Initiative (SEAri) at the Massachusetts Institute of Technology. Reid is earning master's degrees in both Aeronautics & Astronautics and Technology & Policy. His research interests concern the design and management of complex sociotechnical systems, particularly with regard to the anticipation of emergent and cascading behavior. He received a BS in Mechanical Engineering and a BA in Philosophy from Texas A&M University and has experience with RAND Corporation and Sandia National Laboratories. [jackreid@mit.edu]

Donna H. Rhodes—is a principal research scientist at the Massachusetts Institute of Technology, and director of the Systems Engineering Advancement Research Initiative (SEAri). Dr. Rhodes conducts research on innovative approaches and methods for architecting complex systems and enterprises, designing for uncertain futures, and human-model interaction. Previously, she held senior management positions at IBM, Lockheed Martin, and Lucent. Dr. Rhodes is a Past President and Fellow of the International Council on Systems Engineering (INCOSE), and INCOSE Founders Award recipient. She received her PhD in Systems Science from T. J. Watson School of Engineering at Binghamton University. [rhodes@mit.edu]

Abstract

Acquisition programs increasingly use model-centric approaches, generating and using digital assets throughout the lifecycle. Model-centric practices have matured, yet in spite of sound practices there are uncertainties that may impact programs over time. The emergent uncertainties (policy change, budget cuts, disruptive technologies, threats, changing demographics, etc.) and related programmatic decisions (e.g., staff cuts, reduced training hours) may lead to cascading vulnerabilities within model-centric acquisition programs, potentially jeopardizing program success. This paper presents ongoing research that seeks to provide program managers with the means to identify model-centric program vulnerabilities and determine where interventions can most effectively be taken. Cause-Effect Mapping (CEM), a technique developed at MIT, is employed to examine cascading effects between emerging perturbations and terminal outcomes. Research begins with literature investigation and gathering results of past studies of relevance, including studies of model-centric environments and transformations from a traditional to model-centric engineering paradigm (sometimes referred to as the digital engineering paradigm), recent workshop findings, and related work on vulnerability assessment that may have implications for this work. The results are used to refine the CEM and analytic approach to develop a reference model for vulnerability assessment of model-centric programs. Usability of the resulting model is tested with selected research stakeholders.

Introduction

In a world where engineered systems are rapidly increasing in complexity, scale, and interoperability, there is an urgency to transform traditional practices. Digital transformation changes how systems are acquired and developed through the use of model-centric engineering practices and toolsets. While offering great benefit, new challenges arise from both technological and socio-cultural dimensions. This drives the need to examine and address vulnerabilities not only for products and systems, but also for the model-centric environments necessary for their acquisition and development. Ongoing research investigates the use of Cause-Effect Mapping (CEM) as a mechanism for better enabling program managers and system engineers to anticipate and respond to programmatic vulnerabilities as related to model-centric environments. A Reference CEM for model-centric enterprises is generated based on gathered research findings and used for discussion with subject matter experts. Information on uncertainties and leading indicators is collected. Analysis is performed to consider the cascading vulnerabilities and potential intervention options.

Motivation

Acquisition program management is grounded in management science and a sound set of practices evolved over decades; however, new challenges arise as acquisition becomes increasingly model-centric. Increasing availability and use of model-centric approaches and enabling technologies is transforming engineering from documentation-centric to model-centric. While good practices have emerged to support the shift to model-centric program acquisition, such programs experience perturbations over their lifecycles that introduce new vulnerabilities that may lead to cascading failures. For instance, perturbations may be caused by policy change (leading to IP disagreements), economic factors (leading to training cuts), or disruptive technology (leading to outdated infrastructure). Early detection and intervention of vulnerabilities can mitigate disruptions and failures. The research seeks to contribute to the vulnerability assessment state of practice for acquisition programs, both public and private, that increasingly depend on digital assets and model-centric environments.

Background

The following subsections describe model-centric engineering; vulnerability, hazard and risk analysis; cause-effect mapping; and programmatic vulnerabilities.

Model-Centric Engineering

Acquisition program management is grounded in management science and a sound set of practices evolved over decades; however, new challenges arise as acquisition becomes increasingly model-centric. Baldwin and Lucero (2016) state, “The DoD sees value in adopting digital engineering design and model-centric practices, enabling a shift from the linear, document centric acquisition and engineering process toward a dynamic digital, model-centric ecosystem. ”

The systems engineering field is going through a period of significant transformation (Peterson, 2017). Advances in computing, digital workflows, and multidomain-multiscale models are leading to new concepts and approaches for model-centric engineering (Piaszczyk, 2011; Reid & Rhodes, 2016; Glaessgen & Stargel, 2012; Puckek et al., 2017; West & Pyster, 2017).

Model-Centric Engineering (MCE) has been defined as “an overarching digital engineering approach that integrates different model types with simulations, surrogates, systems and components at different levels of abstraction and fidelity across disciplines throughout the lifecycle" (Blackburn et al., 2017). MCE involves using integrated models across disciplines, subsystems, lifecycle stages, and analyst groups. It uses models as the "source of truth" to reduce document handoff and allow for more continuous evaluation. This reduces communication time and rework in response to requirement changes. While many engineering organizations are applying various aspects of MCE (Glaessgen & Stargel, 2012; Kellner, 2015; Lockheed Martin, 2015), implementation is not without its difficulties. Enhanced infrastructure and new leadership capabilities are needed. Increased connectivity means the danger of improper access is heightened. Even with sound MCE practices in use, there are still many challenges that remain. Efforts are also ongoing to identify inconsistent policies in an organization using model-based tools (e.g., Krishnan, Virani, & Gasoto, 2017; Virani & Rust, 2016).

Most discussions of MCE focus on engineering practices and methods to overcome implementation difficulties. In any system, however, engineering is only a piece of the problem. Numerous human factors, business concerns, and organizational issues exist. The design and development of a system exists itself inside a sociotechnical system. Program managers and system engineers must learn how to identify and address programmatic vulnerabilities that pose threats to schedule and budget. Current program managers have significant experience with modern engineering processes. They can use this experience to identify and mitigate such vulnerabilities. Minimal experience exists specific to MCE and model-centric environments, however. This fact, coupled with the increased integration of models, means that emergent uncertainties (policy change, budget cuts, disruptive technologies, threats, changing demographics, etc.) and related programmatic decisions (e.g., staff cuts, reduced training hours) may lead to cascading vulnerabilities within MCE programs, potentially jeopardizing program success. New tools are needed to enable program managers to readily identify model-centric program vulnerabilities and determine where interventions can most effectively be taken.

The Defense Acquisition Glossary defines a *program* as "a directed, funded effort that provides a new, improved, or continuing materiel, weapon or information system, or service capability in response to an approved need." In this paper, this definition of *program* will be used. A *project*, on the other hand is acknowledged as being synonymous with program in general usage, but more specifically defined as "a planned undertaking having a finite beginning and ending, involving definition, development, production, and logistics support (LS) of a major weapon or weapon support system or systems. A project may be the whole or a part of a program" (Defense Acquisition University, n.d.).

Vulnerability, Risk, and Hazard Analysis

Numerous methods for analyzing vulnerabilities, risks, and hazards exist. These three interrelated terms have different definitions depending on the field and on the method of analysis. In this paper, a *hazard* refers to a system or environmental state that has the potential to disrupt the system. Examples include the existence of an iceberg at sea and tired operators. Hazards may not result in system failure, partly depending on the design of the system.

A *vulnerability* is the means by which the hazard might disrupt the system, thus it is through the vulnerability that the system is susceptible to the hazard. Vulnerabilities are best expressed as the causal series of events connecting a hazard to system failure. This is a generalization of common, field-specific usages of the term. MITRE's Common Vulnerabilities and Exposures (CVE) database defines a vulnerability as "a weakness in the computational logic (e.g., code) found in software and some hardware components (e.g., firmware) that, when exploited, results in a negative impact to confidentiality, integrity, OR availability" (The MITRE Corporation, 2015). In this definition, the same components can be seen: some structural means or “weakness” that can result in system disruption or “negative impact” if a hazard is present or the vulnerability is “exploited.” For example, the infamous Spectre security vulnerability is described by CVE as “systems with microprocessors utilizing speculative execution and branch prediction may allow unauthorized disclosure of information to an attacker with local user access via a side-channel analysis” (The MITRE Corporation, 2017). This is a neat summary of the hazard (an attacker), the means (side-channel analysis using speculative execution and branch prediction), and the disruption (unauthorized disclosure of information).

Risk is a measure of the probability of a system disruption and the consequences of that disruption. It is commonly expressed with just a statement of those two components (e.g., 1.25 deaths per 100 million vehicle miles). Risk can also be expressed as a multiplication of likelihood and consequence and can include other components such as detectability.

Common means of analysis include Fault-Tree Analysis (FTA); Failure Modes, Effects, and Criticality Analysis (FMECA, though sometimes reduced to FMEA); Systems Theoretic Process Analysis (STPA); and Event Tree Analysis (ETA).

FTA is a deductive, top-down analysis method where a failure mode is identified and all the possible causes of that event are laid out in sequences until the exogenous hazards are reached. Logic gates are used to connect the various hazards and intermediary events. An example FTA may be seen in Figure 1. Probabilities may be assigned to each hazard and thus a cumulative probability of the failure calculated. FTA is thus quite proficient in investigating the cause of failures afterwards, but is limited in its ability to identify all possible hazards. Additionally, it is somewhat limited by its arbitrary stopping point (i.e., where one chooses to define an event as an exogenous hazard).

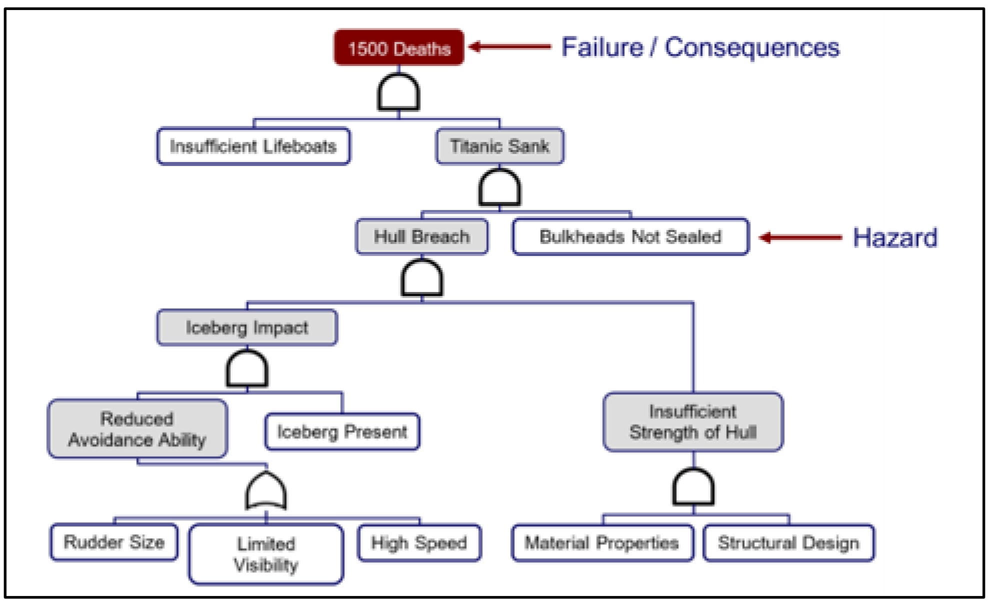


Figure 1. Simplified Fault-Tree Analysis of the Sinking of the Titanic

ETA is essentially an inverted FTA. Instead of starting from a failure and working backwards to a hazard, a hazard is selected and logic gates are used to assess potential consequences. This method is useful for predicting potential failures rather than determining the cause of an existing failure. It suffers from some of the same limitation as FTA. Additionally, it fails to examine the consequences of multiple concurrent hazards.

FMECA is an inductive method, similar to ETA, that seeks to tabulate all possible failures and then assess their severity, probability, detectability, and criticality. It excels at thoroughness but suffers from an inability to easily access multiple failures simultaneously. Additionally, its tabular format can be difficult to read. An example FMECA can be seen in Figure 2.

Function	Dispense amount of cash requested by customer				
Potential Failure Mode	Does not dispense cash			Dispenses too much cash	
Potential Effect(s) of Failure	Customer dissatisfied	Incorrect entry to demand deposit system	Discrepancy in cash balancing	Bank loses money	Discrepancy in cash balancing
Severity Rating	8			6	
Potential Cause(s) of Failure	Out of Cash	Machine jams	Power failure during transaction	Bills stuck together	Denominations in wrong trays
Occurrence Rating	5	3	2	2	3
Current Process Controls	Internal low-cash alert	Internal jam alert	None	Loading procedure (riffle ends of stack)	Two-person visual inspection
Detectability Rating	5	10	10	7	4
Risk Priority Number	200	240	160	84	72
Criticality	40	24	16	12	18

Figure 2. Portion of an FMECA
(Tague, 2004)

STPA takes a rather different approach and conceptualizes systems as control loops, as can be seen in Figure 3. The goal of STPA is to avoid focusing on exhaustively tabulating all vulnerabilities and attempting to quantitatively calculate probabilities. These can be difficult to do accurately for a system of any significant size. Instead STPA attempts to ensure that appropriate monitors and controls are in place for each component of the system (including its operators) so that *any* hazard is detected and addressed before it can cause a failure. In this way, it seeks to eliminate vulnerabilities while relying primarily on a qualitative, rather than a quantitative, assessment.

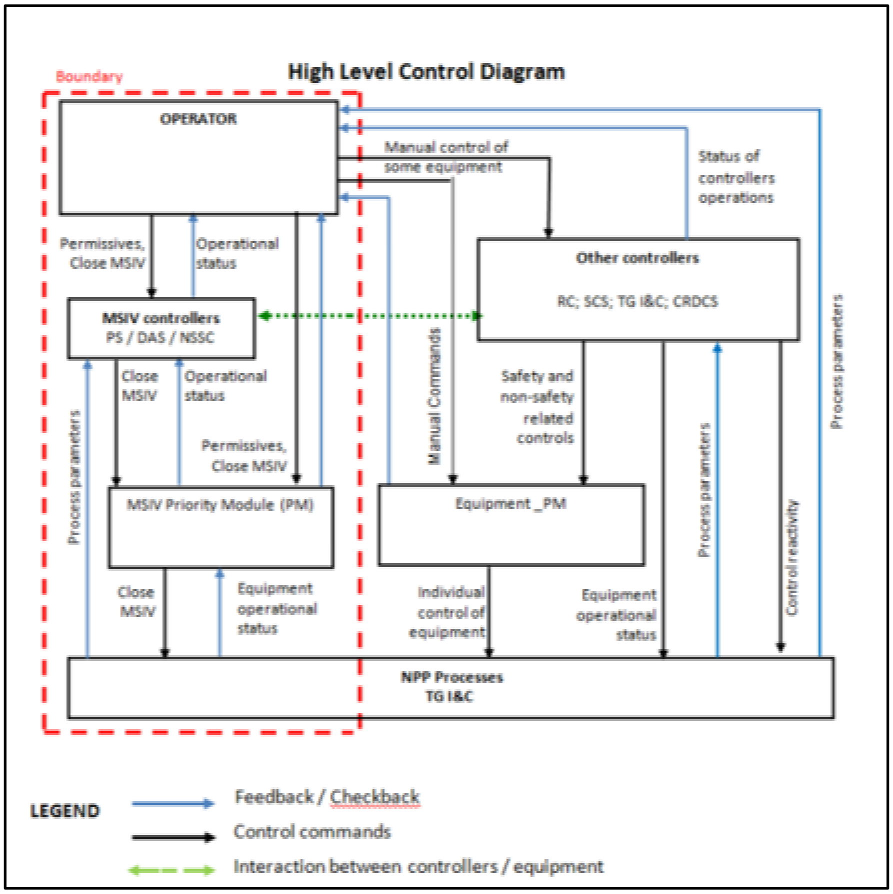


Figure 3. Example STPA Diagram
(Leveson, 2013)

Most vulnerability analysis methods fail to directly grapple with the problem of blame (though STPA does). Humans, engineers and program managers included, have a tendency to assign blame for a failure to someone or something other than themselves. FTA, ETA, and FMECA can enable this by allowing for an arbitrary “stopping point” (i.e., where the previous step in the causal chain is deemed the initiating hazard). In the Titanic FTA presented in Figure 1, for instance, why did we stop deconstructing the causes there? Were the designers of the rudder actually at fault? Or were the engineering standards poorly written? Were the owners of the boat at fault for installing too few lifeboats or should the government set a minimum required number of lifeboats? By adjusting the bounds of the analysis, it is easy to place blame on whomever the analyst desires.

STPA avoids this by (a) not assigning a specific “cause” of a failure and (b) by having every part of the system responsible for monitoring the other parts. Despite this, as the creators of STPA themselves acknowledge, the method has been criticized for its lack of a neat, one-page explanation of the causes of an accident (Leveson, 2013).

Cause-Effect Mapping

Cause-Effect Mapping (CEM) captures some of the benefits of STPA while still presenting distinct cause-effect paths. CEM has previously been applied to a case study of a Maritime Security System of Systems (Mekdeci et al., 2012) and to a supply chain case (Rovito & Rhodes, 2016). It consists of a mapping of causal chains that connect an exogenous hazard to a system degradation or failure, termed a *terminal event.* Each chain represents a vulnerability, sometimes called a *vulnerability chain* in order to emphasize that vulnerabilities are not discrete events. Terminal events are broadly defined and include any form of value loss. An example CEM (that lacks intervention points) can be seen in Figure 4. Similar to FTA, CEM is easily read in either direction, but it also allows for the simultaneous consideration of multiple failures and multiple hazards.

The hazards are external to the perspective of the defined user, and are thus sometimes called *external triggers.* An *intermediary event* is any unintended state change of a system's form or operations which could jeopardize value delivery of the program.

A CEM is not created for a system, but for a specific class of decision-maker. The hazards (referred to as "spontaneous events"; Figure 2) are exogenous from the point of view of the decision-maker that the CEM was made for. In this way, CEM avoids the aforementioned "blaming someone else" problem by making *all* hazards exogenous. The decision-maker only has control over the intermediary events. While she may not be at fault for any of the vulnerabilities, it is still her responsibility to address them.

CEM is fundamentally a qualitative analysis method, though it can be readily adapted into a quantitative form, by adding probabilities of transition to each intermediary. CEM provides immediate insight into which parts of the system warrant more detailed modeling. For instance, it may be useful to determine the likely time required for a specific vulnerability to complete. CEM enables classification of different vulnerability chains (by terminal event, by triggering event or type of triggering event, or by intermediary event). Additionally, it allows immediate identification of potential intervention points at intermediary events where multiple vulnerability chains intersect.

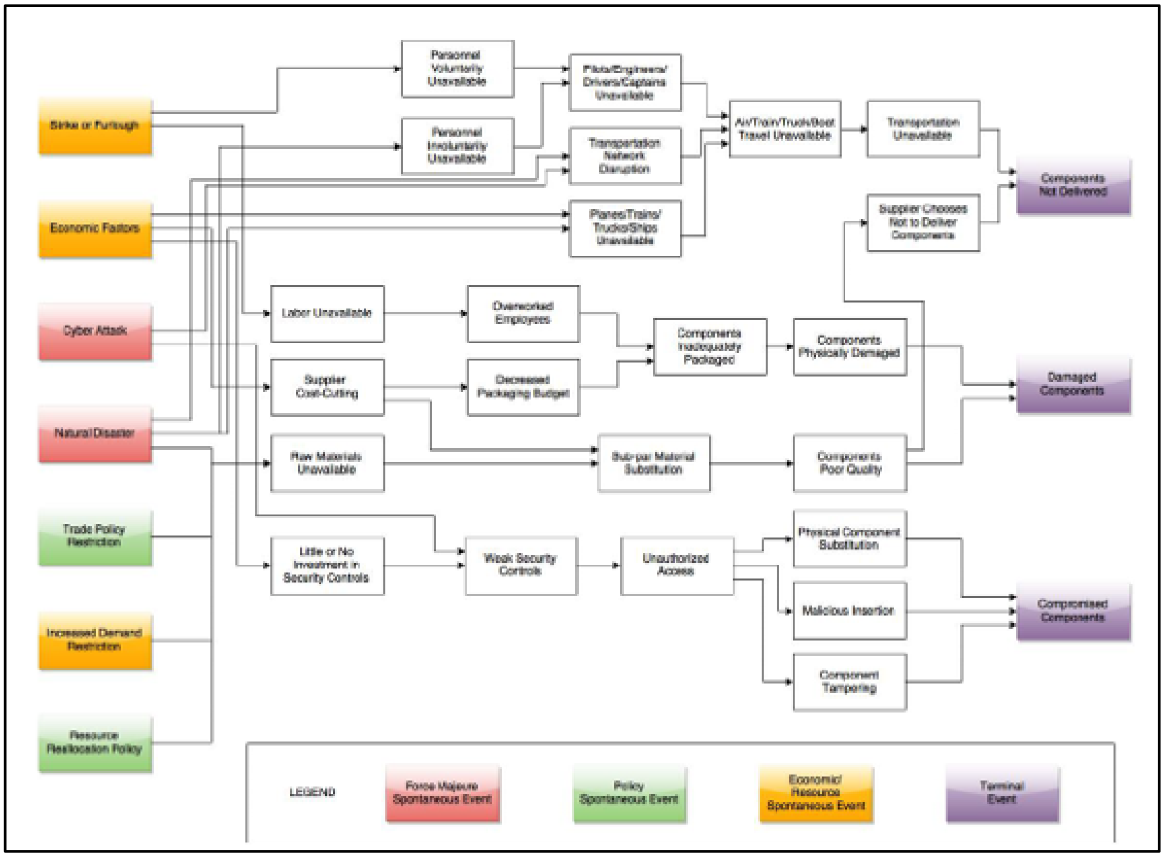


Figure 4. Example CEM of a Supply Chain
(Rovito & Rhodes, 2016)

Programmatic Vulnerabilities

Programmatic vulnerabilities differ from technological system vulnerabilities in a number of ways. Programmatic vulnerabilities tend to be more people-oriented, involving politics, economics, incentives, social interactions, and the like. They tend to be much less thoroughly studied, assessed, and understood, both in academia and in practice. Over the course of this study, a series of interviews was conducted with system engineers and program managers from a variety of fields, including defense, aerospace, manufacturing, and semiconductors. These interviews sought to provide insight into the following questions, in the context of a model-centric enterprise:

1. To what extent are program managers aware of programmatic vulnerabilities?
2. How do program managers conceptualize these vulnerabilities?
3. How do program managers respond to these vulnerabilities?
4. What vulnerabilities are present in MCE programs?
5. What cybersecurity vulnerabilities does MCE pose?

The first three questions provided some useful information regarding the status quo. Across all these industries, several facts were clear. First, many, if not most, programmatic vulnerabilities appear to be triggered by exogenous hazards beyond the control of the program manager. Some, such as poor scope or inadequate budget, are sometimes present before the first program manager joins a program. In general, program managers are at least aware of the potential for these hazards and in some cases can even see them coming. There was some variance in responses to these hazards, however. Some program managers attempt to do things like preemptive padding of a schedule using a multiplicative factor based on experience. Others used their own spreadsheets and tools for estimating the “real” schedule or cost (that is, the schedule or cost that would result from a potential hazard becoming real). Little to no formal risk or vulnerability assessment would take place, however, and responses tended to be reactive rather than proactive, contributing to the “program management via crisis management” paradigm. In general, the perception by interviewees was that program managers rely heavily on expertise, rather than on formal education. While some, such as the INCOSE PM-SE Integration Working Group’s Strategic Initiative, seek to directly address the exogenous hazards and others seek to provide risk registries that contain programmatic risks (Hall, 2018), there is a real need for easy to use tools for program managers to improve their ability to assess programmatic vulnerabilities and respond to them. The fourth question was intended to supplement and corroborate a Reference CEM, as discussed in the following section.

CEM for Model-Centric Programs

In this research, an objective was to develop a high-level cause-effect map for model-centric programs to serve as a reference for use by program managers. The intent is for this CEM to serve both as a standalone resource for such program managers, as well as a basis for organizations to construct their own, program-specific CEM with added detail. Additionally, this research sought to document the general steps to create and use CEM in general, as well as to conduct some initial usability testing of the usefulness of the Reference CEM.

Generating the CEM

Generating a CEM can be done in different ways and to different levels of granularity, depending on the need of the stakeholder. This process can be done with groups, such as project teams, as well as individually. The general process is as follows:

1. The stakeholder herself lists potential hazards posed to the program.
2. She then traces the consequences of each of these hazards through the intermediary events to the final terminal events.
3. The process is then done in reverse: She looks at the terminal events, adds in any that are still missing, and works backwards on how they might come about.
4. She then examines the causal connections between each intermediary event to see if there are any additional connections not previously noticed.
5. Finally, she consults lessons learned databases, case studies, and other experts to generate additional hazards, intermediary events, causal connections, and interventions, as well as to verify existing ones.

The Reference CEM shown in Figure 5 was generated through a combination of methods. At this time, there is little literature on programmatic vulnerabilities posed by MCE. Most negative case studies, that is, those that depict failures (Software Engineering Institute, 2007), and lessons learned databases (NASA Office of the Chief Engineer, n.d.), are from prior to the rise of MCE and thus deal with general vulnerabilities. Existing case studies that directly deal with MCE tend to be more positive, likely due to the rising popularity of the paradigm (Conigliaro, Kerzhner, & Paredis, 2009; Maley & Long, 2005; Martz & Neu, 2008). As a result of these, extrapolations from extant vulnerabilities had to be made, along with hypothetical inversions of the positive instances of MCE. Additional vulnerabilities were contributed by group brainstorming during the in-class activity discussed further in the section titled Usability Testing. All of these were supplemented and confirmed by the same interviews discussed previously.

The higher level of detail in the upper portion of the CEM, which includes issues such as training, misunderstood model assumptions, and level of trust in the models, represents the increased degree of concern that interviewees had about these issues. In general, the domain of aligning culture and expertise with well-designed and well-documented toolsets was of high priority. Failure to accomplish this had led to significant problems in past projects, but is viewed as a surmountable difficulty moving forward.

Using the CEM

Vulnerability analysis methods are most commonly applied either to the design or the operation of an engineered system. This is usually done to improve its design or investigate a failure. However, these methods can also be applied to the program itself. Instead of hazards such as “relief valve failure” and “solar flare” instead we have “hiring freeze” and “unexpected technological hurtle.” It can be difficult to assess likelihoods for such hazards, but even qualitative analysis can be useful. Similarly, terminal events are not “nuclear meltdown” or “loss of communications” but instead “schedule delay” or “failure during verification/validation.”

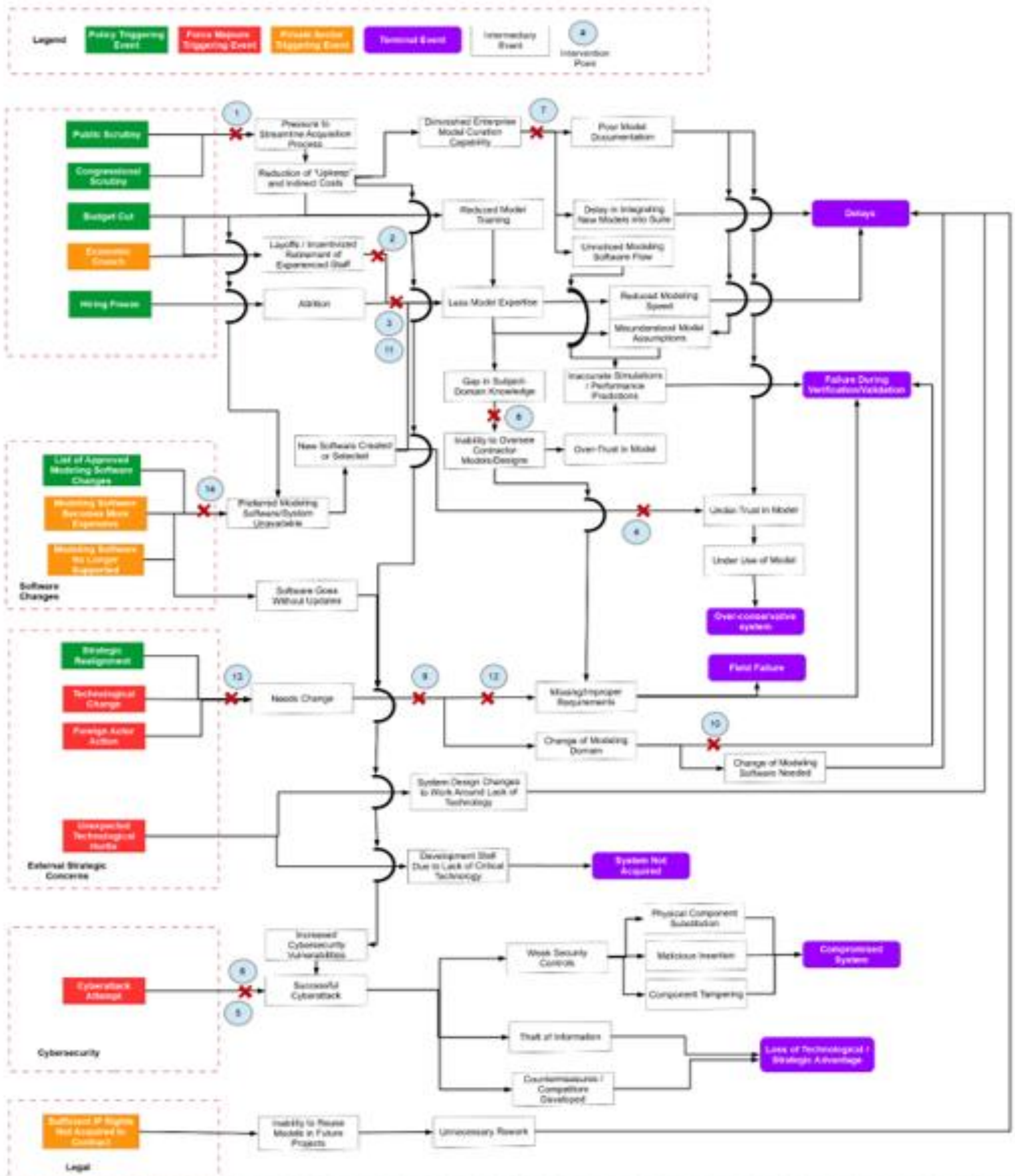
CEM in particular can be used to assess vulnerabilities in multiple ways and by different individuals. Four uses are described as follows:

- (A) By a Program Manager: Assessing potential future vulnerabilities and planning possible interventions
- (B) By a Program Manager: Determining specific vulnerabilities to address in response to the presence of a specific hazard
- (C) By the Program Organization: Changing program processes to mitigate or eliminate vulnerabilities
- (D) By Researchers: Organizing and classifying vulnerabilities into various categories or types

All of these start with the creation of a CEM for the organization’s standard program process or for a particular program. Once this is completed, additional steps can be taken, including

1. Identifying notable intermediary events and potential intervention points
2. onducting more detailed modeling of specific vulnerability chains
3. Classifying vulnerability chains to enable future study and potential mitigation.

While a program manager would be well-served by the creation of a CEM specific to their own program, there is some benefit in using a Reference CEM for model-centric programs in general. Such a Reference CEM can be seen in Figure 5.



Intervention Points

1) Initiate Internal Assessment and a PR Strategy
2) Initiate various non-monetary benefits (such as allowing pets at work or a 9/80 schedule) to encourage employees to stay
3) Seek to share resources and employees with other programs
4) Hire employees with prior experience with the new software
5) Actively conduct security training and monitor compliance
6) Isolate critical data
7) Create documentation and curation processes within the program
8) Institute handover periods to benefit from contractor expertise
9) Reevaluate training regime and needed fields of expertise
10) Increase the amount of testing conducted
11) Increase use of contractors/consultants (including previous employees) to maintain expertise level
12) Reevaluate requirements with the client and other stakeholders
13) Design for modularity to minimize impact on system
14) Negotiate with client to see if they are able to pay for the software

Figure 5. Reference CEM for Model-Centric Vulnerabilities (Preliminary)

Use (A) is most relevant for novice program managers or program managers using MCE for the first time. A senior program manager or team of program managers creates a CEM for their organization's program process. This CEM can then be provided to the novice for study and reference. The program manager can then learn what can go wrong and how to intervene. In this case, the CEM could be tied to a Lesson's Learned database, such as NASA's Lessons Learned Information System (NASA Office of the Chief Engineer, n.d.). This enables concrete examples and consequences to be linked to each vulnerability. One of the important factors here is that the CEM does not just present potential interventions, but it also places them in the appropriate part of the causal sequence. This enables the program manager to not only know how to intervene, but at what point.

Use (B) is relevant to all program managers, regardless of level of experience. Once a hazard manifests, the program manager examines the CEM to assess potential consequences and options. He can then respond quickly to head off any cascading effects. This may require additional analysis of a specific vulnerability chain or an individual intermediary event. System dynamics is a method particularly useful for this due to the preexisting models of many organizational phenomena (Rouwette & Ghaffarzadegan, 2013). For instance, *Attrition, Reduced Model Training*, and *Less Model Expertise* can be modeled by adapting the rookie fraction model shown in Figure 6 into the more MCE-relevant model shown in Figure 7. In this model, it is apparent that a hiring freeze (which would set the "Growth Rate" variable to zero) has no immediate impact, as rookies will continue to develop into experienced employees and model expertise will continue to accumulate. Overtime, however, the dearth of new rookies will result in fewer experienced employees, increasing the error rate. These kinds of long-term, indirect impacts are likely to become more common with increased use of MCE.

Use (C) is the traditional use of vulnerability assessment methods: to improve a design or investigate a failure. The program organization can change policies or create infrastructure to either mitigate or wholly eliminate certain vulnerability chains. For example, if the organization elects to only use modeling software produced in-house, the three hazards in the "Software Changes" grouping of Figure 5 are no longer relevant. Such a change could be costly though or even introduce new vulnerabilities, so careful analysis is necessary. In this use, the CEM is a visual representation of a risk registry, tabulating all possible hazards to the program and mitigation choices made (Hall, 2018).

In Use (D), CEM is used to organize and classify vulnerability chains. Two obvious classifiers are terminal events and hazards. Which is used to organize a CEM depends on whether the user wants to examine the causal chains forwards or backwards. Beyond this, however, more complicated classifiers are possible. As can be seen in Figure 5, external triggers that result in similar vulnerability chains are grouped together. By "similar," we mean that these vulnerability chains either involve many of the same intermediary events or that they involve the same part of the program. For instance, most of the intermediary events involving model curation and trust are located close to one another in the center-top of the figure. Once these groupings have been identified, they can be considered together, such as the "Belt-tightening" grouping, and common means of intervention considered.

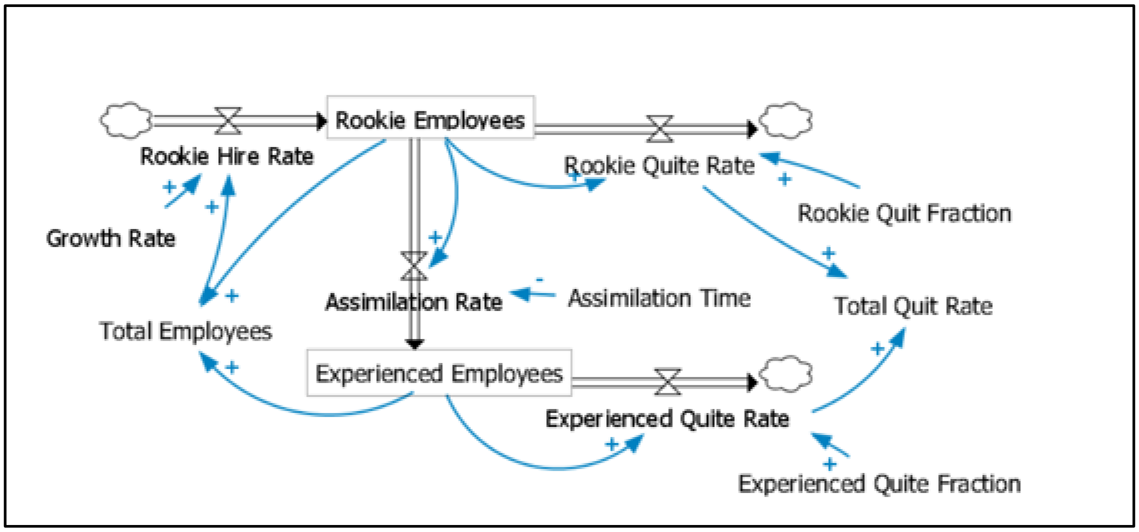


Figure 6. System Dynamics Model of Employee Training Rate
(Adapted from Sterman, 2000)

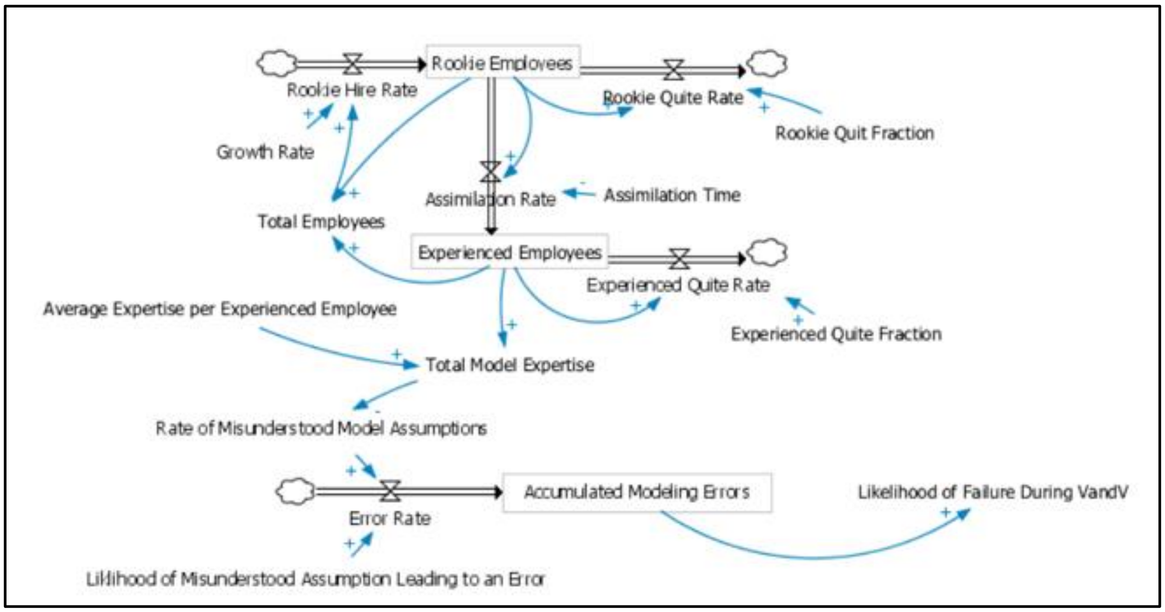


Figure 7. System Dynamics Model of Accumulated Modeling Errors

Usability Testing

While several potential use cases were proposed in the previous section, due to the scale, duration, cost, and uniqueness of major MCE programs, it is difficult to systematically test the utility of either the Reference CEM or of using CEM in general. Some simple usability testing was explored in the interviews with experts. Usability was also considered through analyzing the results of a scenario-based exercise on vulnerability analysis that had been generated in a graduate class activity involving techniques for investigating enterprises.

The classroom activity involved approximately 40 students from various backgrounds, most having prior experience in either industry or the military as systems engineers and/or program managers. The students were randomly divided into six groups of 5 to 7 students each, and each group was provided with the same "context" for the activity, as follows:

> You are a project manager for a vehicle manufacturer. Your current project is designing a lightweight troop transport vehicle for the U.S. military. It has a variety of high-tech components, including encrypted radio and satellite communication systems, an explosives detector, and night vision cameras. The design and testing process will take multiple years. Your company considers this a major project in terms of the resources put into it, the revenue received for it, and the potential for future military contracts. The military, as part of the contract, specified that the design and production process should predominately rely on models (sometimes called model-centric engineering) rather than written specifications.

Each group was provided with one or two selected external triggers or hazards to respond to. They were asked to discuss and record the potential negative impacts these hazards may have on the engineering environment and how they might act to mitigate these consequences. The hazards provided were as follows:

1. An unrelated military project (at another company) to design a next-generation missile defense system has ended up in the national news. That system has gone significantly over budget, has been repeatedly delayed, and still looks like it is a long way off from being completed. Public accusations of mismanagement and waste are being made, including frivolous travel and lavish company events. Congress and the Department of Defense are now carefully scrutinizing all other major projects for potential mismanagement or waste, including your project.
2. After recent elections, there is significant political pressure on Congress to reduce federal spending. As a result of this, they are making significant cuts to many agencies and programs, including the military. The decrease in government spending is likely to impact your company's other projects and may impact yours as well.
3. Government intelligence officials inform you that your company, and perhaps even your project, is likely to be the target of cybersecurity attacks based out of another nation.
4. This is your first project of this type (your prior experience was in designing civilian vehicles). You now have to choose whether your project will use the set of modeling software that you are accustomed to from the civilian projects or use another set that is more commonly used on military projects, but that you are unfamiliar with.

5. A recent economic downturn, coupled with a government that is cutting back on its military spending, has resulted in your company declaring a hiring freeze for an indeterminate amount of time.
6. Another project at your company is threatening to miss its deadlines. In order to get it back in line, its project manager is requesting that personnel from other projects, such as yours, be reallocated to hers.
7. The design requirements from the military that you are currently working with were put together with the idea of using this vehicle in a currently ongoing conflict. However, U.S. involvement in this conflict is winding down and the military is currently unsure where this vehicle will be used in the future. The context (and thus the requirements) may change during the design process. Furthermore, your models were created with the current context in mind.
8. A recent economic downturn, coupled with a government that is cutting back on its military spending, has resulted in your company providing incentives for early retirement to the more experienced, higher paid employees. You have no intention of retiring early yourself, but some of those working on your project might accept the offer.
9. In order to minimize rework and redundancy, your company has recently started an initiative pushing for increased reuse of components, designs, and models from one project to another. Your previous project also involved a night vision camera, but in a very different application context.
10. A particular piece of simulation software that your company has used on similar projects in the past is licensed from another company. The license contract is up for renewal soon and the price might go up significantly. You are uncertain if your company's executives will approve the license renewal.

After a period of 20 minutes, the students were taught about causal chains and use of the CEM reference model as a technique for investigating enterprise vulnerabilities. Each group was instructed to re-write the previously identified vulnerabilities and interventions as causal chains and map them on the provided CEM (Figure 8), as well as coming up with new ones. After another period of 20 minutes, their results were collected, a group debrief was given, and students shared general feedback on the class activity. [Note: The CEM presented in Figure 8 is similar to that in Figure 5 but not identical, as knowledge gained in interviews and usability testing has since been used to further develop the CEM.]

Several useful pieces of information could be garnered through analyzing the documented results of the class activity that had been conducted. In the out-briefing material, the participants expressed unanimous agreement that using CEM and conceptualizing programmatic vulnerabilities as causal chains was helpful, though the perceived degree of usefulness varied from “slightly” to “extremely.” Additionally, team out-briefs reported on four primary forms of how the CEM helped in their assigned scenario in the class exercise:

1. Identifying high priority intervention points: (70%)
2. Identifying new vulnerabilities: (55%)
3. Understanding the causal path / Reframing the concept of vulnerabilities: (45%)
4. Understanding interrelationships between vulnerabilities: (40%)

The relative importance of this first point was corroborated by the group outputs that were generated. It was clear that in this instance, when provided with a Reference CEM, the groups tended to focus on identifying where and how to intervene in the vulnerability chains. During the first round (without the CEM), most of the groups had presented their vulnerabilities and interventions as unordered lists of short phrases, typically unpaired (i.e., vulnerabilities were not matched with interventions). These short phrases were typically isolated events such as “team feeling more cautious” and “reputation damages.” The ultimate outcomes of these were assumed rather than explicitly stated. Once the CEM was introduced in the second round, the matching of interventions to vulnerabilities appeared to become much more clear, and most groups also identified additional interventions.

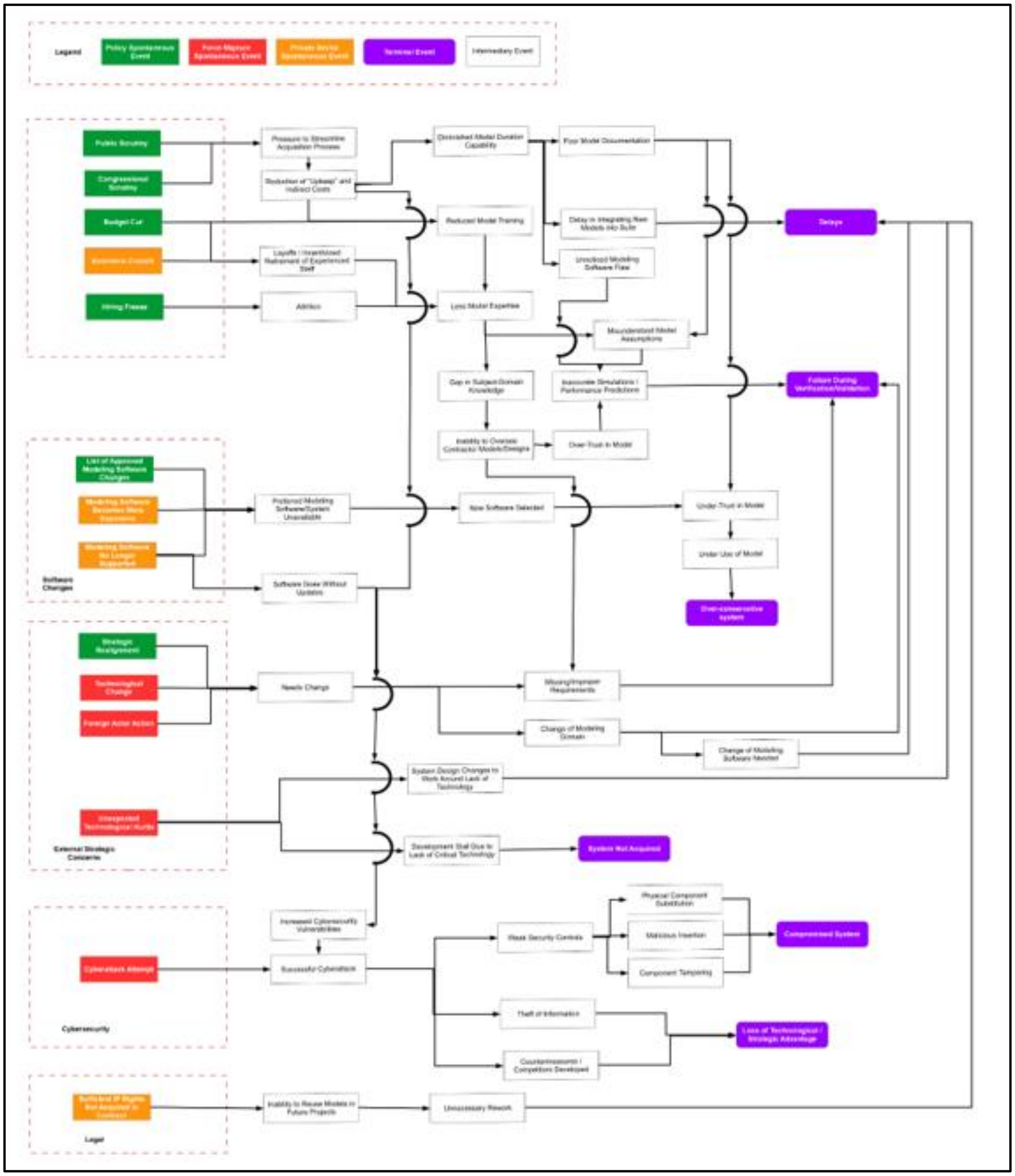


Figure 8. Reference CEM Used as a Basis for Usability Testing

Future Directions

Digital engineering is transforming the systems acquisition process (Zimmerman, Gilbert, & Salvatore, 2017), including the model-centric techniques and toolsets. Enterprises face new challenges in this transformation, including potential for new vulnerabilities within model-centric enterprises. While vulnerability analysis of products and systems is performed, examining vulnerabilities within an enterprise is less common. Vulnerability analysis of the enterprise becomes increasingly urgent given increasing complexity and interconnectivity in model-centric environments used to make system decisions. The interim outcomes of this research, including expert interview results, show the potential benefit of cause-effect mapping techniques and availability of a reference map for model-centric program vulnerability analysis. Additional expert interviews are planned with an expanded set of stakeholders.

Insights into usability were also gained through analyzing a data set that had been generated in a classroom setting. Accordingly, the results should be viewed as purely exploratory, but there appears to be good justification for future research to include conducting a controlled human-subjects research experiment (similar to the scenario-based exercise used in a classroom setting). Additionally, an important future research activity is to evaluate the Reference CEM on a pilot project in a real world model-centric engineering program. Further development of the Reference CEM is planned for next phase research, including more extensive investigation of cybersecurity vulnerabilities resulting from model-centric practices and infrastructure.

Conclusion

Acquisition programs increasingly use model-centric approaches, generating and using digital assets throughout the lifecycle. Model-centric practices have matured, yet in spite of sound practices, there are uncertainties that may impact programs over time. The emergent uncertainties (policy change, budget cuts, disruptive technologies, threats, changing demographics, etc.) and related programmatic decisions (e.g., staff cuts, reduced training hours) may lead to cascading vulnerabilities within model-centric acquisition programs, potentially jeopardizing program success. Ongoing research has led to a preliminary CEM Reference Model that aims to provide program managers with a means to assess, prioritize, and mitigate model-centric vulnerabilities. Usability testing of the reference model has shown positive benefits for practical use in assessing vulnerabilities of model-centric programs. Anticipated results are empirically-grounded vulnerabilities of model-centric programs and a cause-effect mapping reference model for identifying vulnerabilities and interventions.

References

Baldwin, K. J., & Lucero, S. D. (2016). Defense system complexity: Engineering challenges and opportunities. *The ITEA Journal of Test and Evaluation, 37*(1), 10–16.

Blackburn, M., Verma, D., Dillon-Merrill, R., Blake, R., Bone, M., Chell, B., … Evangelista, E. (2017). *Transforming systems engineering through model-centric engineering.* Hoboken, NJ: Systems Engineering Research Center. Retrieved from http://www.sercuarc.org/wp-content/uploads/2014/05/A013_SERC-RT-168_Technical-Report-SERC-2017-TR-110.pdf

Conigliaro, R. A., Kerzhner, A. A., & Paredis, C. J. J. (2009). Model-based optimization of a hydraulic backhoe using multi-attribute utility theory. *SAE International Journal of Materials and Manufacturing, 2*(1), 298–309. Retrieved from http://www.sae.org

Defense Acquisition University. (n.d.). *Glossary of defense acquisition acronyms and terms.* Retrieved from https://dap.dau.mil/glossary/Pages/Default.aspx

Glaessgen, E. H., & Stargel, D. (2012). *The digital twin paradigm for future NASA and U.S. Air Force vehicles*. Paper for the 53rd Structures, Structural Dynamics, and Materials Conference (pp. 1–14). Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20120008178.pdf

Hall, D. (2018). Risk identification challenge. INCOSE 2018 International Workshop, Jacksonville, FL.

Kellner, T. (2015). Wind in the cloud? How the digital wind farm will make wind power 20 percent more efficient. GE Reports. Retrieved from http://www.gereports.com/post/119300678660/wind-in-the-cloud-how-the-digital-wind-farm-will/

Krishnan, R., Virani, S., & Gasoto, R. (2017). Discovering toxic policies using MBSE constructs. In A. M. Madni & B. Boehm (Eds.), *Conference on Systems Engineering Research*. Redondo Beach, CA.

Leveson, N. (2013). *An STPA primer*. Cambridge, MA.

Lockheed Martin. (2015). Digital tapestry. Retrieved from http://www.lockheedmartin.com/us/what-we-do/emerging/advanced-manufacturing/digital-tapestry.html

Maley, J., & Long, J. (2005). *A natural approach to DoDAF*. Blacksburg, VA.

Martz, M., & Neu, W. L. (2008). Multi-objective optimization of an autonomous underwater vehicle. *Oceans* (Vols. 1–4), 1042–1050\r2248.

Mekdeci, B., Ross, A. M., Rhodes, D. H., & Hastings, D. E. (2012). A taxonomy of perturbations: Determining the ways that systems lose value. In *Proceedings of the 2012 IEEE International Systems Conference*, *Proceedings* (pp. 507–512). Vancouver, British Columbia: IEEE. https://doi.org/10.1109/SysCon.2012.6189487

The MITRE Corporation. (2015). Terminology. Retrieved February 20, 2018, from https://cve.mitre.org/about/terminology.html

The MITRE Corporation. (2017). CVE-2017-5753. Retrieved February 20, 2018, from https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2017-5753

NASA Office of the Chief Engineer. (n.d.). NASA public lessons learned system. Retrieved July 13, 2017, from https://llis.nasa.gov/

Peterson, T. A. (2017). INCOSE transformation strategic objective. In *INCOSE Webinar 106. INCOSE*.

Piaszczyk, C. (2011). Model based systems engineering with Department of Defense architectural framework. *Systems Engineering, 14*(3), 305–326. https://doi.org/10.1002/sys.20180

Rouwette, E., & Ghaffarzadegan, N. (2013). The system dynamics case repository project. *System Dynamics Review, 29*(1). https://doi.org/10.1002/sdr.1491

Rovito, S. M., & Rhodes, D. H. (2016). Enabling better supply chain decisions through a generic model utilizing cause-effect mapping. In *Proceedings of the 2016 Annual IEEE Sytems Conference.* Orlando, FL: IEEE.

Software Engineering Institute. (2007). Acquisition archetypes: Firefighting. Pittsburgh, PA.

Sterman, J. D. (2000). Coflows and aging chains. In *Business dynamics: Systems thinking and modeling for a complex world* (pp. 469–512). Boston, MA: Irwin McGraw-Hill.

Tague, N. (2004). Failure mode effects analysis (FMEA). Retrieved January 17, 2017, from http://asq.org/learn-about-quality/process-analysis-tools/overview/fmea.html

Virani, S., & Rust, T. (2016). Using model based systems engineering in policy development : A thought paper. In *Conference on Systems Engineering Research.* Huntsville, AL.

Zimmerman, P., Gilbert, T., & Salvatore, F. (2017). Digital engineering transformation across the Department of Defense. The *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*. https://doi.org/10.1177/1548512917747050

Acknowledgments

This material is based upon work by the Naval Postgraduate School Acquisition Research Programs under Grant No. N00244-17-1-0011.

THIS PAGE INTENTIONALLY LEFT BLANK

Acquisition Research Program
Graduate School of Business & Public Policy
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net