SYM-AM-25-314

Excerpt from the Proceedings of the Twenty-Second Annual Acquisition Research Symposium and Innovation Summit

Wednesday, May 7, 2025 Sessions
Volume I

Knowledge Based Metrics for Test and Design

Published: May 5, 2025

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

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School

Knowledge Based Metrics for Test and Design

Dr. Craig Arndt—has extensive experience as a senior executive and technology leader in research, education, engineering and defense, homeland security and intelligence technologies, with extensive experience as an innovative leader in industry, academia, and government. He is often called upon as an industry expert to chair major program reviews, lead new programs, and research efforts. Dr. Arndt currently serves as a principal research engineer on the research faculty of the Georgia Tech Research Institute (GTRI) in the System Engineering Research division of the Electronic Systems Lab. Prior to coming to GTRI, Dr. Arndt served as Professor and Chair of the engineering department at the Defense Acquisition University and as technical director of the Homeland security FFRDC at the MITRE Corporation. Dr. Arndt holds engineering degrees in electrical engineering, systems engineering, and human factors engineering and a Master of Arts in strategic studies from the U.S. Naval War College.

Dr. Valerie Sitterle—is a Principal Research Engineer and Chief Scientist for the Systems Engineering Research Division in the Electronic Systems Laboratory at the Georgia Tech Research Institute (GTRI), where her technical work synthesizes systems science, complex systems, and defense analysis. She is an expert with over 20 years of experience in engineering science, integrating engineering, natural, and physical sciences leading to the design and analysis of systems. Two primary areas of her work revolve around the design and assessment of defense systems alongside concepts of operation in theater environments, and the synthesis of architectures with decision methods, processes, and tools to meet the next generation of systems engineering challenges. She earned a BME and MS in mechanical engineering from Auburn University, an MS in engineering science from the University of Florida, and a PhD in mechanical engineering from the Georgia Institute of Technology.

Jeremy Doerr—has over 15 years of experience in software and systems engineering across several domains and is currently an MBSE researcher at Georgia Tech Research Institute (GTRI). He has a BS in software engineering and a master's in systems engineering from Georgia Tech. At GTRI, Jeremy's work ranges from basic research into technical language semantics to applied research, guiding DoD sponsors to adopt MBSE and integrated digital engineering approaches. Prior to GTRI, he successfully certified several flight deck systems using new and novel touchscreen technology while working for Gulfstream on the G500 and G600 platforms. Before Gulfstream, Jeremy worked in logistics for Walgreens, performing database and automation work in operations, as well as conducting timing optimizations studies and Lean Six Sigma projects on the industrial engineering side.

Abstract

The task of developing the best military equipment in the world has long fallen on the U.S. Department of Defense and the military industrial base that supports them. The United States made the decision years ago (and have succeeded in going to war with the best equipment since the second half of WWII (1943)) that their military would have the best equipment in world. As the 21st century continues to unfold, this commitment is becoming ever more difficult and more costly, and hard to execute in a timely manner.

Over the past few years, the leadership of the DoD acquisition community have listed the acceleration of development testing and fielding systems as their top priority. To try to make this happen, the DoD is implementing digital transformation. Another major part of accelerating the acquisition process has been a movement to integrate the design and test functions of the acquisition process. This includes moving test earlier in the development process.

When looking at the test and acquisition process, it is important to understand what the goal of test is in the development process. Traditionally, the goal of test has been to validate that a design will meet specific requirements created for the system. This traditional goal, however, is becoming less relevant, and the role of test as part of the development process is consuming much more test resources. So, what is the goal of test? If the role of test is to help ensure that we are developing the best product for our customer, then we might think of test's role being to increase knowledge about the future performance of a system still in design while there is time to improve the design. At a practical level this means two things. First, that testing should be designed specifically to support decision-making; the development of the Integrated Decision Support key (IDSK) was

intended to support this goal. Second, that we need to integrate all activities that provide additional knowledge about the future performance of the system together in meaningful ways to support decision-making.

In order to integrate and measure the amount of knowledge needed to make specific decisions (about things like requirements, risk, system design, and test resource allocation), we need to be able to measure the amount of knowledge needed for decisions and the amount of knowledge that we expect to generate in a given activity (including design, test, or history).

In this paper, we will demonstrate the development of a mathematical based knowledge metric and how it can be applied to specific DoD acquisition and test decision-making. The paper will document the development of the decision add and use it in practical programmatic decisions.

Introduction

Digital Transformation

In June 2018, the Department of Defense established its expectations for digital transformation in The DoD Digital Engineering Strategy. The strategy outlines five goals aimed at establishing a digital engineering environment for more rapid and effective development and fielding of weapon systems. The goals include the use of models to inform decision-making, establishing an infrastructure to enable the digital engineering methods, and transforming the workforce to adopt digital engineering methods over the acquisition life cycle.

Figure 1 was developed by the DoD to help communicate the different elements of the transformation effort. The development and use of standardized models is critical to the success of the transformation and the resulting advantages of digital engineering to the operations of all aspects of the department.

The DoD followed up this strategy with the release of formal guidance via DoD Instruction 5000.97, which ensures that the director of Operational Test and Evaluation (DOT&E) will utilize digital engineering methods to achieve their test objectives for operational assessment and Live Fire Testing. Also in 2023, DOT&E released their DOT&E Strategy Implementation Plan (I Plan), which includes objectives and key actions to develop digital, or model based Test and Evaluation Master Plans (TEMP) and Integrated Decision Support Keys (IDSK). As recently as December 2024, the department has released an update to DoD Instruction 5000.89 and five DoD manuals further refining the description and use of digital methods for the entire DoD test community (DoDM 5000.96, DoDM 5000.99, DoDM 5000.100, DoDM 5000.101, and DoDM 5000.102).

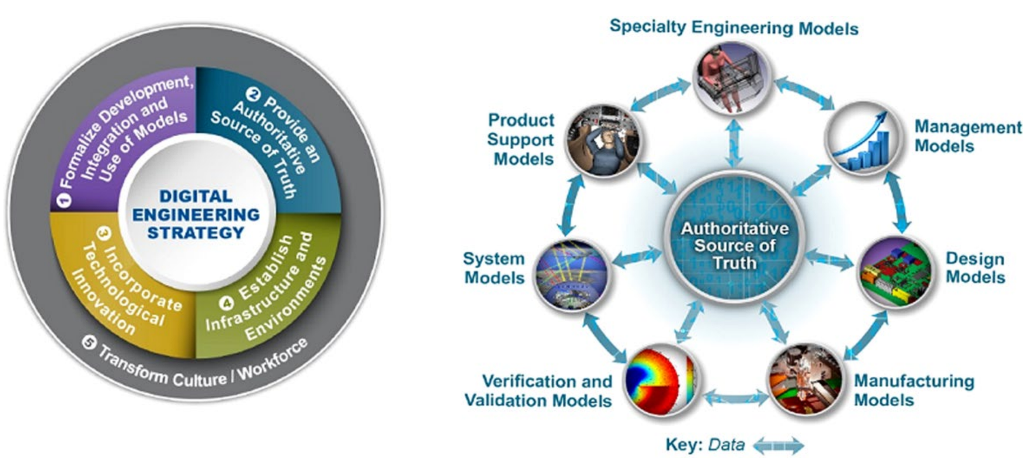

Figure 1 DoD Digital Transformation

Industry Trends / Knowledge Engineering

Knowledge engineering is a field that has grown and evolved significantly over the past few years and is now in many industries. Knowledge engineering is used to manage both knowledge in systems and development processes, and also knowledge created in manufacturing and use of systems to improve the performance and quality of a wide range of products. There are many applications of knowledge engineering in DoD applications. Some of these applications include the management and development of AI systems, and the management of data in operations, and combat systems. In industry and in the DoD, the need for and the management of knowledge is a becoming a critical concern in the development and use of systems.

Background

The Test and Evaluation Master Plan (TEMP) is one of the core artifacts in the DoD acquisition process (DoD 5000.01). However, the TEMP, and the test process as a whole, need to be understood as part of the larger research, engineering, development and acquisition process. Figure 2 is an illustration of the larger process needed to understand test and its relationship to requirements and the use of the system (mission data).

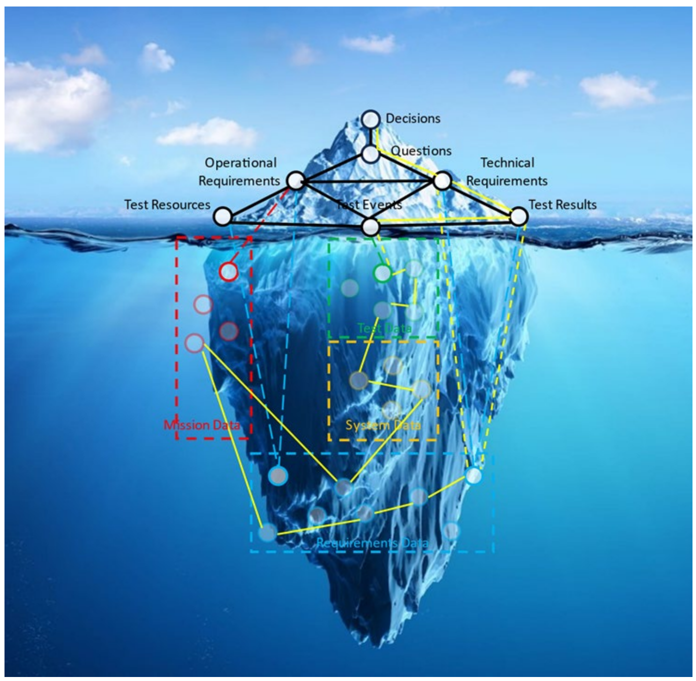

Figure 2 Development Iceberg

Prioritization of Speed in Defense Engineering

When we look at the priorities that the DoD has for process improvement to meet the significant challenges of the great power competition, it is clear that at the top of the list is the acceleration of the acquisition and development process. “As Undersecretary of Defense for Acquisition and Sustainment Dr. Bill LaPlante has emphasized, with some capabilities the department needs to be able to field several software revisions a day” (Shaffer & Whitley, 2024).

Defining the Requirement

Acquisition reform has taken many forms over the years. The defense acquisition system has been developed with many specific goals and provisions built into it. At this point, the acquisition system has a good deal of flexibility due to the adaptive acquisition pathways. However, in the ever-changing world of the 21st century acquisition, there is a priority for the rapid development of high-performance systems. High performance systems, however, create significant risk. Moving forward, this creates a growing need to better recognize high risk in technical development, and to accelerate the development of system when possible.

Historical Issues

In order to better understand the history of acquisition risk from the perspective of, we conducted a detailed requirements analysis. The primary sources for this analysis were Government Accounting Office reports and scholarly papers on the acquisition process. Congress and DoD leadership have long used the same source to formulate acquisition policy. Table 1 summarizes issues in nine major areas of defense acquisition.

Table 1

Area	Issues	Knowledge Gap	Risk	Source
1. Capability Need	Business case development	Is there sufficient detail in the business case allowing for clearly defined requirements?	Insufficiently developed business cases leads	GAO-23-106059, GAO-21-511T
1. Capability Need	Key stakeholders' project and technology knowledge to make appropriate decisions	Does the key stakeholders, have sufficient knowledge, training	Insufficient experience, training	Defense ARJ, October 2012, Vol 19 No, 4422-443, GAO-20-439, GAO-23-106059, GAO-24-106831
Alt. Multiple areas	Incorrect inflation assumptions	Does the AoA include current approved inflation assumptions?	Not incorporating approved inflation assumptions leades to cost over/underestimation.	GAO-24-106831
2. Decisions	Programs outside acquisition pathways	Are there any programs within the Service/Department which will impact this capability?	Limited oversight of non-AAF pathway projects	GAO-24-106831
2. Decisions	Production Decisions out of sync with testing	Has the program conducted prototype testing prior to making a production decision?	Testing the prototype after making production decisions	GAO-24-106831
4. Acquisition Strategy	Acquisition pathway flexibility	Are the requirements to switch between acquisition pathways acknowledged and deliberately planned for?	Allowing contracts which plan to use multiple acquisition pathways without a deliberate plan to address known pathway	GAO-24-106831
4. Acquisition Strategy	Official cost estimates as programs transition between pathways	Are the program's official costs developed and published prior to transitioning to a new pathway?	Insufficient cost development limits informed investment decision-making by perpetuating the sunk-cost fallacy.	GAO-24-106831
5. Requirements	Cyber-security/cyber-physical interconnectivity	Are the cyber requirements for the capability full developed?	Not identifying all cyber requirements leaves the capability vulnerable to non-kinetic/EW attacks	GAO-24-106831, DoDI 5000.90

5. Requirements	All or nothing approach to requirements development	Is the program using or facilitating the use of iterative requirements development?	A monolithic approach to requirements development limits adaptability as technology matures.	GAO-24-106831
6. Source Selection	Single contract for total program	Does the source selection include modular contracting terms?	Single, large-scale contracts limit incremental capability development	GAO-24-106831
6. Source Selection	Supplier/Defense Industrial base disruptions	How stable are the logistics pipelines for suppliers and the Defense Industrial Base?	Logistical disruptions increase production timelines and overall costs.	GAO-21-511T, GAO-23-106059, GAO-24-106831
6. Source Selection	Developed a software factory	Does the Service/Department have a Software Factory to serve as Software SMEs throughout a project's lifespan?	Software factories are designed to speed up software development and acquisition by providing consistent user feedback, secure DevSecOps	GAO-23-106059, GAO-24-106831
6. Source Selection	Diminishing manufacturing sources	Does the program rely on at-risk parts?	Reduced supplier options	GAO-20-439, GAO-24-106831
6. Source Selection	Cost-reimbursement contracts on major development items	Does the program use a cost-reimbursement contract or some other type like a fixed-price incentive contract?	Cost-reimbursement contracts introduce substantial funding risks	GAO-18-238sp
7. Design	Technology maturation/readiness level	How mature is the technology for the various components?	Overestimating technology maturity leads to extended timelines and increased costs	GAO-21-511T, GAO-23-106059, GAO-24-106831
7. Design	Key decision points are not clearly defined nor are the requirements for those decisions addressed	Is there sufficient detail to the information requirements for the project to transition between acquisition phases?	Not adequately addressing decision point information requirements allows for vulnerabilities and deficiencies	Defense ARJ, October 2012, Vol 19 No, 4422-443, GAO-20-439
7. Design	Inconsistent cost data	How consistent is the data reporting?	Inconsistent cost reporting data limits oversight and potential increases risks	GAO-24-106831
7. Design	Limited use of digital engineering	Does the program use digital engineering methods to increase efficiency throughout the project's lifespan	Relying solely on static models to measure impacts of design changes limits efficiency	GAO-24-106831
8. Testing	Testing procedures for cyber/cyber-physical vulnerabilities	Does the testing plan include early and frequent testing for potential vulnerabilities?	By not testing all known or potential cyber threats, the capability becomes	GAO-24-106831, DoDI 5000.90
8. Testing	Software testing limited to "testers," not "users"	Has the program provided incremental deliveries to users for feedback?	By limiting software testing to a team of "testers" as opposed to end-state "users," programs can experience substantial risks to costs	GAO-20-439

9. Fielding	Cyber deficiencies not corrected	Were there any late-stage cyber vulnerabilities not addressed due to funding or available timeline?	Not addressing all cyber vulnerabilities prior to fielding increases risk	DoD CyberSecurity Test and Evaluation Handbook, DoDI 5000.90
9. Fielding	Production lines not achieving statistical process control	Has the production process achieved statistical process control prior to full rate production?	Achieving process control prior to full rate production limits delays in fielding and downstream user reliability concerns	GAO-20-439
9. Fielding	Accepting serious deficiencies identified during testing	Is the program accepting equipment with serious deficiencies identified during testing as defined by the respective Service?	Accepting equipment for fielding with uncorrected serious deficiencies increases risk	GAO-18-238sp
8. Testing	Production representative prototype not tested in its intended environment	Has the program completed operational environment testing with a production level prototype?		GAO 23-106059

This requirement analysis reinforces our understanding of future needs of the acquisition community.

Role of Test in the Department of Defense

Traditional test and evaluation in the DoD has been focused on two aspects of test process. Specifically, operational test, which looks at whether new system can perform functions it was intended to perform in its intended environment, and development test, which is design to add in the development of the final system design. Adding to the traditional functions of test, we need to look at test as a means to forecast the future performance of a system in the field and to assess the health and risk of the development process.

Availability of Knowledge

As we look at of knowledge of the performance of the new system of interest, the function of testing takes on a different perspective. As we have discussed, test has traditionally been about the verification of requirements. In the current digital world of system development, we can also look at test as a knowledge source contributing to our ability to forecast future performance of the system. In order to improve the ability of decision-makers to evaluate risk in system development, it is critical that we use all the available knowledge about the system as it is developed to make these decisions.

In addition to traditional developmental and operational test data as a source for knowledge about the performance of a system, a great develop of knowledge can be gain be 1. Data about legacy systems that use the same or similar subsystems, 2. design of the system itself, and 3. Modeling and simulation of the system. For the purposes of this work therefore are using five specific classes of knowledge sources for future performance of the system under development. 1. Legacy systems data, 2. Design data, 3. Modeling and Simulation, 4. Developmental test data, and 5 operational test data.

Specific Needs of Decision-Makers

As we demonstrated in our analysis of the GAO reports, the acquisition community has a number of different needs including technology development. Historical, it is critical in development to accurately asses the amount of risk there is in the baseline plan for development test and fielding of the system. By more effectively using all of the available data from all available knowledge sources, we can better asses the dynamic risk in a specific program and program development approach. Note that the test program is part of the program development plan (baseline). When we look at the specific needs of decision-makers, it is instructive to look at critical use cases. In the case of acquisition leaders, there are two important use cases that stand out.

<u>Use Case 1:</u> High Risk, development programs: In the past 25 years, there have been a number of high-tech defense programs that have, significantly overrun costs and schedules, including the Future Combat Systems/DDG-1000/Railgun programs. Post program analysis of these and other programs have determined that the technical risk on these programs were much higher that program managers were aware of early in the program life cycles, even though the knowledge of these risks did exist. Use case 1 is there for the need to better assemble knowledge about program risk early in a program to better understand shortfall in knowledge about the system that would indicate high development risk.

<u>Use Case 2:</u> In the opposite case when the DoD is developing a new system with lower technical risk, it is important to also have a clear view of the technical risk profile of the system throughout its development. On programs with lower technical risk profiles, there is also a good deal of knowledge about the system design and future performance of the system that is known. By capturing more of this knowledge, program decision-makers can structure programs to accelerate schedules based on reasonable risk profiles, given that knowledge about the system allows them to reduce the cost and time of gathering additional (and redundant) knowledge.

These specific use cases inform the functional development, design, and implementation of a metric and metric reporting system within the digital program, and Model Based Systems Engineering (MBSE) methods that are currently being integrated into Defense program management and engineering.

Development of Digital Models in the Test Process

The digital models that are being developed as a part of digital test engineering are a key part of implementing knowledge-based decision making in the DoD. Specifically, the Integrated Decision Support Key (IDSK) was developed to link decisions to specific Knowledge sources and tests. The Model Based TEMP in kind was developed to link requirements, design, and test planning though digital modeling methods. This work extends these models, by the addition or other knowledge sources and the development of specific metrics analysis.

Model Based IDSK

To realize the IDSK's potential to positively impact acquisition outcomes and program decisions, the concept of a MB-IDSK developed using model-based systems engineering will address a majority of the shortcomings of the traditional IDSK and provide great benefits to decision-makers and all stakeholders across the acquisition and T&E enterprise. These benefits include (i) its ability to support the T&E-as-a-Continuum (Collins & Senechal, 2023) framework by integrating the IDSK into a program's digital engineering ecosystem, (ii) an MB-IDSK would provide mapping of decisions to development (i.e., acquisition) risk, test risk, and test resource models, thereby allowing for more sophisticated analysis including probabilities of success analysis, (iii) an MB-IDSK will expand the ability to link different aspects of the system design, capabilities, and testing to critical program decisions.

In Anyanhun and Arndt (2024), a MB-IDSK reference architecture (MB-IDSK RA) was proposed and developed to support digital transformation efforts of DOT&E. The motivation behind defining a MB-IDSK RA was based on the premise that an architecture should reflect the organization of the owning enterprise (CAS, 2022). Specifically, the MB-IDSK RA represents an essential tool to facilitate communication and alignment efforts of current and future IDSK architectures. Figure 3 depicts the IDSK architecture strategy as adapted from the DoD Comprehensive Architecture Strategy.

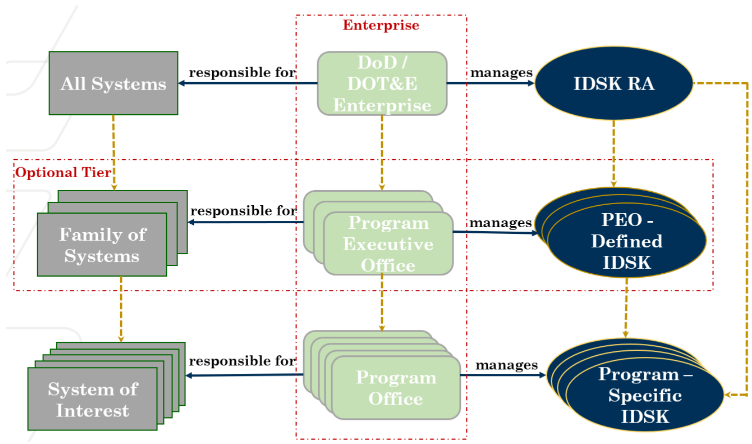

Figure 3. IDSK RA Architecture Strategy. Adapted from Figure 1 of the DoD CAS (CAS, 2022).

Equipping DoD acquisition programs with overarching guidance on how to leverage digital engineering for decision support is critical to achieving the enterprise-wide business and mission objectives of providing weapon systems at the speed of need and relevancy. As reported in CAS (2022) and Muller (2007), a Reference Architecture provides a method for focusing all architecture and design decisions.

Model Based TEMP

The MB-TEMP reference architecture captures the essence of the test planning and decision support domain relative to the needs of program offices, DOT&E, DTE&A, T&E practitioners and decision-makers. For the purpose of this article, abstractions and simplification concepts have been utilized in relation to how some diagram views appear and how they are presented in this work to enhance legibility. More importantly, the architectural strategy employed in the development of the TEMP RA results in a digital engineering artifact (tool) that, when instantiated, will seamlessly integrate into the digital engineering ecosystem of a program. Figure 4 is an example of the complete set of views that together make up the MB-TEMP RA description.

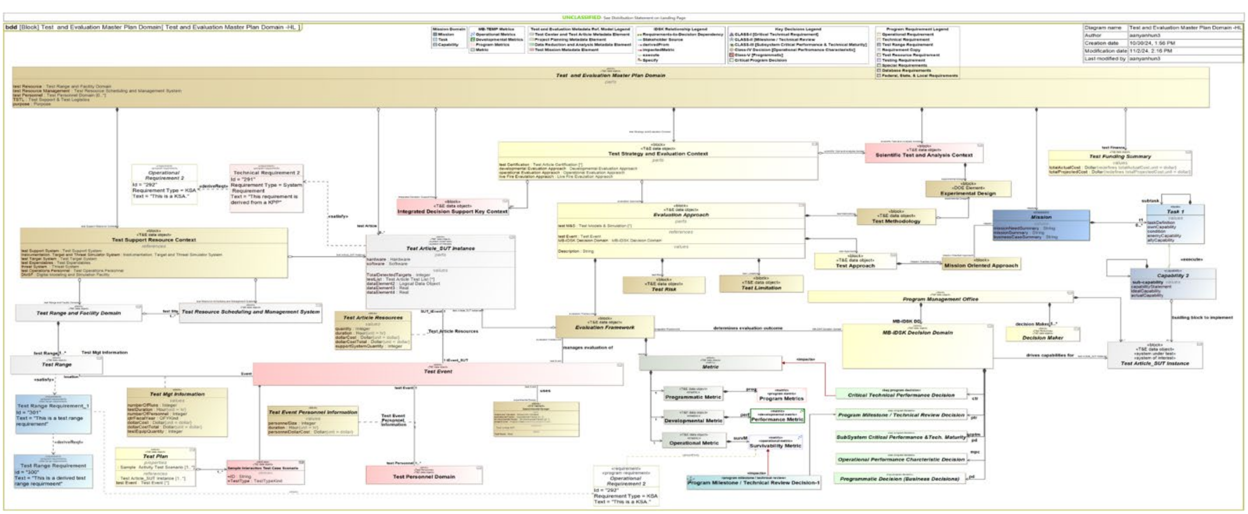

Figure 4. The TEMP Domain view of the MB-TEMP RA

The TEMP Domain view of the MP-TEMP RA provides crucial insights into the top-level composition of the TEMP domain. The RA view links together elements defined within the TEMP model and elements already defined in digital models that exist within a program's digital engineering ecosystem. These digital models include a program office model, requirements model, system model, SUT model, and test range models.

Development / Theory

There are two primary questions related to test and evaluation (T&E) relevant to effective and efficient fielding of warfighting capabilities: 1) Can we proceed to the next stage of development, and do we have an adequate understanding of the level of risk? 2) Do we have enough redundancy and reliability in our knowledge sources to accelerate the development and/or the test?

With respect to the first, the risk, we need to understand the requirements being levied, from request for proposal (RFP) to source selection, and be able to ascertain if we are proceeding at a higher risk than we realize. For the second, we need to determine if we know enough that we can truncate testing activities such that we do not undertake more testing than we need to do. For example, if we have enough redundancy and quality of data across relevant legacy performance data, design data, and modeling and simulation (M&S) data, can we reduce or otherwise compress the amount of data we need from developmental test activities? The development of a high-level characterization method to address risk and the maturation of knowledge for T&E activities is detailed in the following sections to answer these questions.

Knowledge Source Characterization Categories

There are five primary sources of data and information regarding performance of a capability under development that come together to produce knowledge about how that system will perform across the intended concepts of employment:

1. Legacy data and performance data from prior components or similar systems (K_1)
2. Design data (K_2)
3. Modeling and simulation data (K_3)
4. Developmental test data (K_4)
5. Operational test data (K_5)

These sources are typically sequential in time and in degree of "reliability."

Knowledge Based Metrics

Development of Knowledge for a Given Source

Every test decision should be linked to a specific knowledge source; this concept extends such that every key performance parameter (KPP) should also be tied to a knowledge source or sources and a test program. This concept forms the basis for how we produce an abstract representation of knowledge. While the ensuing approach is quantitative in nature for ease of propagation and understanding, there is a strong degree of qualitative expertise and assessment that underlies the numbers. The intent is that this approach will improve as use and experience mature its implementation.

Our approach develops a knowledge source characterization for a given data element that contributes to a specific K_j for a system under development and test, that is:

- $K_j(S)$ – total knowledge obtainable about a system (S) from a given source (j, 1 through 5 above).
- $K_j(s_i)$ – knowledge obtainable about a specific data element, here corresponding to specific part of the system or a subsystem relevant to the new system under development and test.

Each $K_j(s_i)$ is approximated as having three contributing dimensions: i) quality or fidelity of data from the given source, ii) similarity of data from the given source to the specified system or subsystem under current development and evaluation, and iii) completeness of data from the given source with respect to the KPPs being assessed for the specified system or subsystem under current development and evaluation.

All $K_j(s_i, t)$ will exist on the same range of [0,1], effectively representing a normalized level of representation. Similarly, each of the three contributing dimensions $K_j(s_{i\_quality})$, $K_j(s_{i\_similarity})$, and $K_j(s_{i\_completeness})$ are defined on [0,1]. In the absence of grounded data or experience to suggest otherwise, we used a simple minimum to bring dimensions of quality, similarity, and completeness together at a single point in time. The rationale for this is the maximum knowledge attainable for the specific measure or subsystem in question with respect to the relevant KPP should not exceed its minimum value across these dimensions. For example, if if $K_j(s_{i\_quality}, t) = 0.8$, $K_j(s_{i\_similarity}, t) = 0.8$, and $K_j(s_{i\_completeness}, t) = 0.1$, then $K_j(s_i, t) = 0.1$.

Next, we considered how $K_j(s_i, t)$ would come together to produce $K_j(S, t)$. A geometric mean is well suited to describing proportional and varying growth and is appropriate when the data in question may be sustainably different in either its properties (i.e., what it represents) or across its range. Intuitively, it represents the average position of the "center of mass" of a system of particles if each particle had the same weight. In this problem, with data, it represents the centroid of the finite collection of values for $K_j(s_i, t)$ across s_i = 1, 2, 3, … n. Table 2 provides an example for these steps for four subsystems. We assume that we start with zero knowledge, treating the accumulation of knowledge from legacy information as the starting point in our process, and that legacy data varies over time only due to discovery and effective interpretation of that discovery. Knowledge values can remain constant over multiple time points if no new information is gleaned from one step to the next.

Table 2. Example Calculation of Subsystem and then System Knowledge Accumulation for a Given Source Type over Time

	t0	t1	t2	t3	t4	t5
Kj(si_quality)	0	0.3	0.3	0.3	0.4	0.4
Kj(si_similarity)	0	0.2	0.25	0.25	0.25	0.4
Kj(si_completeness)	0	0.3	0.4	0.45	0.45	0.5
K1(s1, t)	0	0.2	0.25	0.25	0.25	0.4
Kj(si_quality)	0	0.1	0.1	0.1	0.1	0.1
Kj(si_similarity)	0	0.3	0.3	0.4	0.4	0.4
Kj(si_completeness)	0	0.1	0.2	0.2	0.25	0.25
K1(s2, t)	0	0.1	0.1	0.1	0.1	0.1
Kj(si_quality)	0	0.8	0.8	0.8	0.8	0.8
Kj(si_similarity)	0	0.2	0.25	0.25	0.25	0.4
Kj(si_completeness)	0	0.33	0.35	0.375	0.45	0.5
K1(s3, t)	0	0.2	0.25	0.25	0.25	0.4
Kj(si_quality)	0	0.6	0.6	0.6	0.7	0.7
Kj(si_similarity)	0	0.9	0.9	0.9	0.9	0.9
Kj(si_completeness)	0	0.8	0.8	0.8	0.85	0.85
K1(s4, t)	0	0.6	0.6	0.6	0.7	0.7
K1(S, t)	***0.00***	***0.221***	***0.247***	***0.247***	***0.257***	***0.325***

The subsequent steps considered integrative loss and the maximum attainable knowledge possible from a given source (i.e., legacy or developmental test). For the former, all of the point information gained through the development and test process will not perfectly combine into knowledge without loss. This loss will exist because perfect integration across disparate dimensions is exceedingly difficult in practice, and all of the dimensions, their attributes, and higher-order relational effects may not be visible. To capture this effect, a parametric equation was selected from the development in (McDermott et al., 2019) and adapted in meaning for application to the knowledge problem in this paper as follows:

$$x\_self\ (t) = x\_self\ (t-1) - sgn(\Delta) * \eta * |\Delta|$$

where:

x_self (t)	Knowledge at the current time t
x_self (t-1)	Knowledge at the prior time (t-1)
Δ	The difference in the current node state and the target state: (x_self (t-1) - x_target(t))
sgn(Δ):	The sign, or signum, function of Δ.
η	Eta. The shaping parameter that governs how the knowledge is matured (i.e., suddenly vs exponentially). $0 \le \eta \le 1$
|Δ|	The absolute value of Δ (also expressible as (Δ*sgn(Δ)).

The target, *x_target(t)* in the *Δ* above, is defined by the parents of the knowledge at time t, *x_self(t)*. Specifically, for the knowledge metric problem described here, the parents of each K(S, t) will correspond to its $K_j(s_i, t)$. Here, the sign of Δ will be negative only if some discovery invalidates or calls prior information into question.

The shaping parameter η controls how quickly *x_self(t)* approaches *x_target(t).* The closer η is to 1, the faster *x_self(t)* approaches *x_target(t).* The shaping parameter can be altered based on experience, the nature of the knowledge source, and/or the nature of the system under development and test. Moreover, there are multiple ways to define *x_target(t)* as a function of the parent contributions; a deterministic maximum or minimum, or any function of the parents are potential approaches. In this work, *x_target(t)* is defined as the geometric mean of the parent contributions as described previously. To illustrate the equation's behavior, Figure 2 illustrates the notional knowledge increase, *x_self(t)*, for a constant target *x_target(t)* = 0.5 across varying values of η.

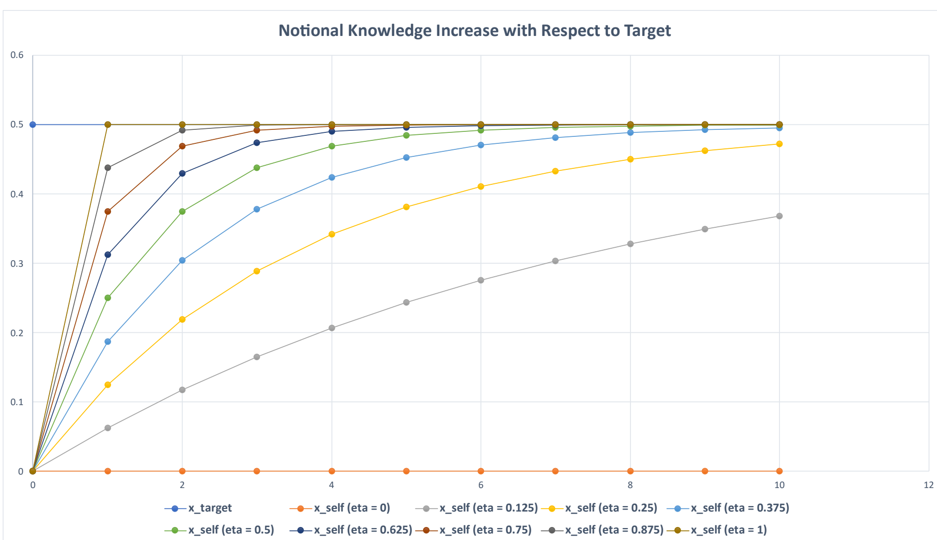

Figure 5. Illustration of Parametric Equation Behavior for Various Values of the Shaping Parameter

As per the description earlier in this section, each $K_j(S, t)$ will be defined by the geometric mean of its contributing $K_j(s_i, t)$. The parametric equation will be used to define the level of knowledge remaining after taking integrative loss into effect.

Specifically, $K_j(S_{\_parametric}, t)$ represents the actual knowledge level attained from $K_j(S, t)$ due to less than perfect knowledge integration for the whole system. Note that if the shaping parameter η is set to 1, then $K_j(S_{\_parametric}, t)$ will equal $K_j(S, t)$ (i.e., no integrative loss is considered).

Further, the maximum attainable knowledge possible from a given source needs to be defined based on the maximum amount of knowledge the development and/or test team anticipates is possible to gain from a given knowledge source (i.e., legacy data). Again, this value, $K_j(S_{\_max})$, will be defined by the team as a value in the range of [0,1], treating it as the decimal equivalent of a percent. $K_j(S_{\_max})$ is considered constant over time. $K_j(S, t)$ and $K_j(S_{\_parametric}, t)$ will be scaled against this maximum value, for example, $K_{j\_scaled}(S, t) = K_j(S_{\_max})* K_j(S, t)$. Continuing the example from 2, 3 illustrates these parts of the process.

Table 3. Continued Example Showing Integrative Loss and Scaling with Maximum Attainable Value for a Given Knowledge Source

Eta	**0.7**	(Constant over time)				
K1(S_max)	**0.2**	(Constant over time)				
	t0	**t1**	**t2**	**t3**	**t4**	**t5**
K1(S, t)	0.000	0.221	0.247	0.247	0.257	0.325
K1(S_parametric, t)	0.000	0.155	0.220	0.239	0.252	0.303
K1_scaled(S, t)	0.000	0.044	0.049	0.049	0.051	0.065
K1_scaled(S_parametric, t)	0.000	0.031	0.044	0.048	0.050	0.061

Development of the Knowledge Accumulation, Gaps, and Integration of Knowledge Sources over Time

We need to capture (i) how much total knowledge about a new system under development and test is being accumulated over time, and (ii) how far behind or ahead the process of knowledge accumulation is compared to its anticipated levels.

Total knowledge accumulated over time is represented through $K_j(S_{\_parametric}, t)$, the actual knowledge level attained from $K_j(S, t)$ due to less than perfect knowledge integration for the whole system. More specifically, this is represented by $K_{j\_scaled}(S_{\_parametric}, t)$, which places the value on the same comparative basis as the maximum, needed, and planned knowledge levels for the given source.

How far behind or ahead the knowledge development process is from what is needed is defined by evaluating the total knowledge accumulated against the amount of knowledge needed. Similar to $K_j(S_{\_max})$, the maximum level of knowledge deemed attainable for a given source, the amount of knowledge needed is also determined by the development and/or test team. These values may, however, increase in time. For example, $K_{j\_needed}(S, t)$ may be defined to increase linearly in time, reaching a maximum value equal to $K_j(S_{\_max})$. The knowledge delta, or gap, may be evaluated as $K_{j\_needed}(S, t) - K_{j\_actual}(S, t)$. Roughly, the actual knowledge corresponds to what the team has while the knowledge delta represents what the team lacks. Figure 6 illustrates these concepts as a continuation of the previous example for a single knowledge source.

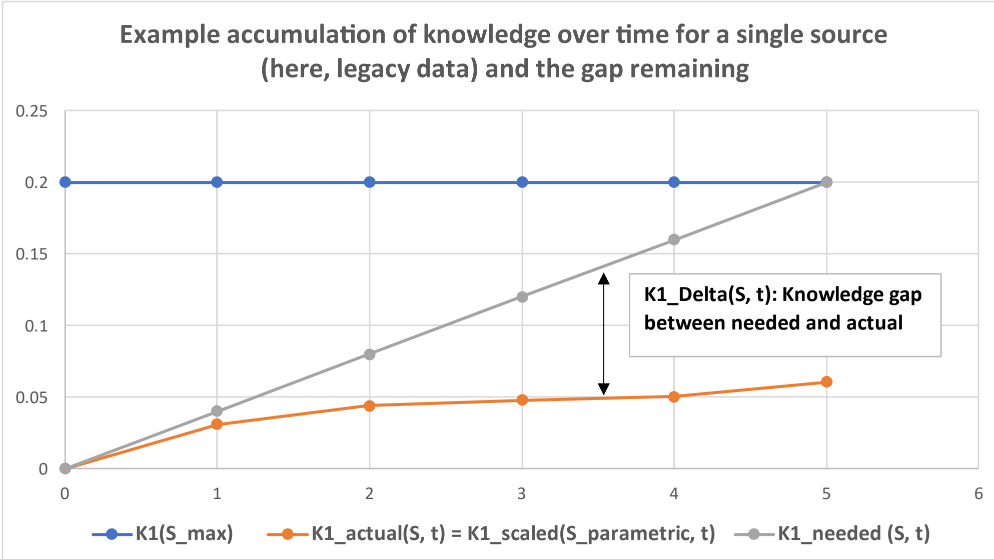

Figure 6. Example Knowledge Accumulation and Remaining Gap over Time

Figure 7 then shows how these concepts extend to combine knowledge sources over time. In the figure, the left graph shows an example where knowledge is not being gained sufficiently in the legacy data activities in comparison to what is expected. The right graph shows the potential effect of recognizing this early and shifting the collection of design data and M&S data earlier in time.

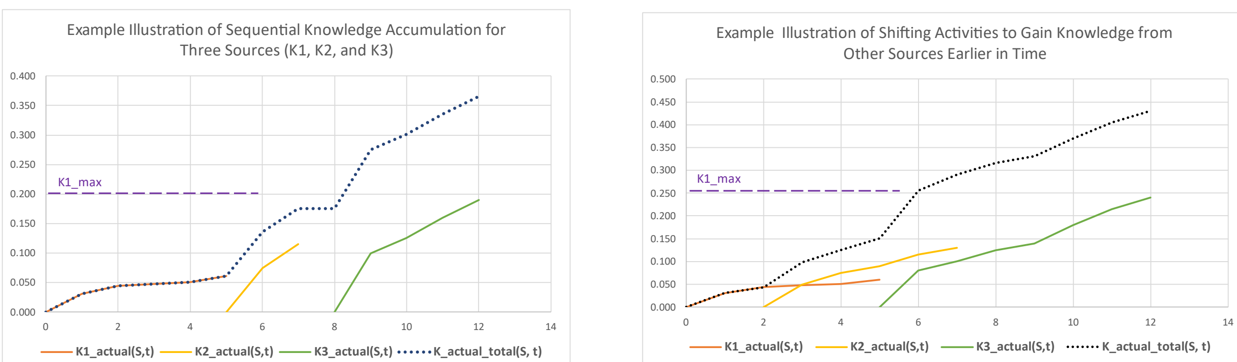

Figure 7. Example Knowledge Accumulation over Time for Three Knowledge Sources

Using the same knowledge accumulation profiles shown in Figure 7, Figure 8 illustrates how recognizing a knowledge deficit early and pushing other activities earlier can reduce the overall knowledge gap with which the development and test teams are proceeding.

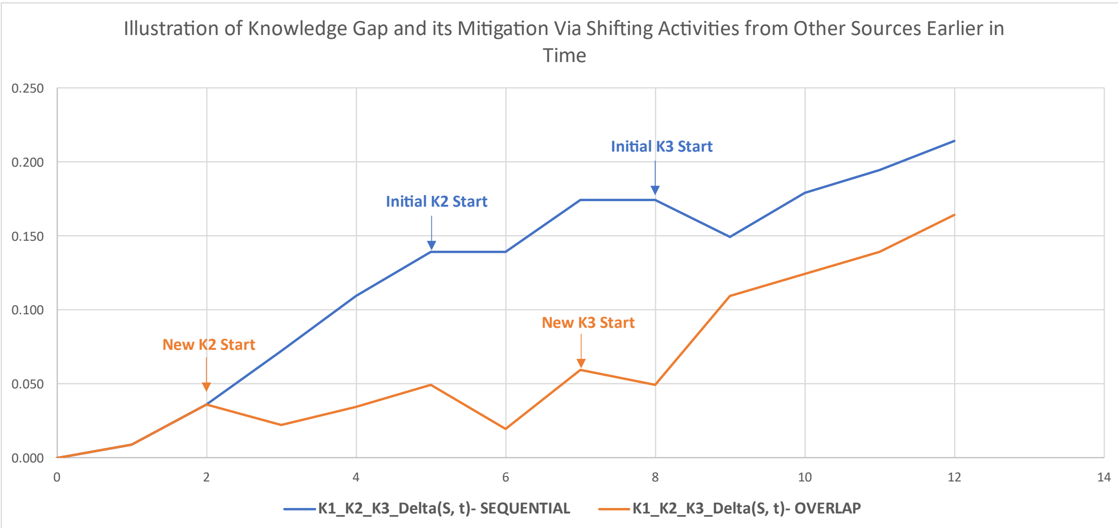

Figure 8. Illustration of Knowledge Gap Reduction via Shifting Activities Earlier in Time to Mitigate a Source Deficit

Discussion of Approach and Limitations

The math and associated methods for knowledge accumulation are simple and easily modifiable at any point in the procedure to better reflect reality as ascertained from experience. A significant challenge, however, is the quality of the values that start this knowledge metric approach. Specifically, if the evaluations of the dimensions $K_j(s_{i\_quality})$, $K_j(s_{i\_similarity})$, and $K_j(s_{i\_completeness})$ that come together to define each $K_j(s_i)$ are wild guesses, then that lack of grounded approximation will carry through the rest of the assessment. Again, each $K_j(s_i)$ represents the knowledge obtainable about a specific data element, here corresponding to specific part of the system or a subsystem relevant to the new system under development and test. This goes hand-in-hand with the quality of KPPs and the mapping of data creation activities to best evaluate the system's performance with respect to the KPPs.

Implementation and Use

In order for this process to and metric to be useful it needs to be implemented in both the digital thread of program and in a decision support tool for decision-makers. The metric will be integrated into the digital thread of the program in order to ensure that the data is automation updated as the design of the system changes, and as we gain additional knowledge sources.

Decision Support and Decision-Maker Displays

At a practical level this means the development and implementation of a set of Dashboards for program managers and other decision-making stakeholders. Figure 9 shows such a dashboard. What is displayed is a graphical representation of the current level of knowledge about the systems KPP (critical performance parameters) at specific key decisions points in the program, in this case before RFP release. On the display we see the current level of knowledge versus the expected amount knowledge and the resulting risk or opportunity. Additional dashboards will display the specific knowledge sources, and opinions to mitigate risk or exploit opportunity (accelerate schedule).

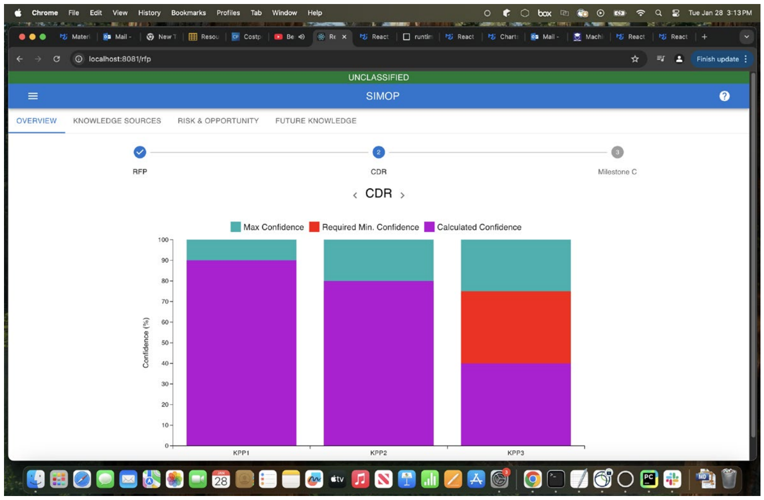

Figure 9 Program Risk by KPP by Time.

Knowledge Source Management/Test Program Trade Studies

As we talked about in the decision support section, the results of identifying knowledge shortfalls and surpluses of knowledge at specific point in time managing the development of a program will create the need and opportunity for adjustments in the baseline of the program. In the case of high risk due to knowledge shortfall, we would need to add design and development resources, and/or tests to increase knowledge. In the case of excess knowledge, the baseline for test and other knowledge sources can be modified to reduce current and future redundant sources of knowledge. However, in complex systems the relationships between the planning and execution of these knowledge sources is also complex. To facilitate better management, we as a part of this program crating standard models and characterizations of knowledge sources, so that we can manage the different knowing sources together. This allows us look at all of the knowledge sources, and make decisions about reductions or additional to knowledge sources the way you playing the video game “Tetris” (Figure 10).

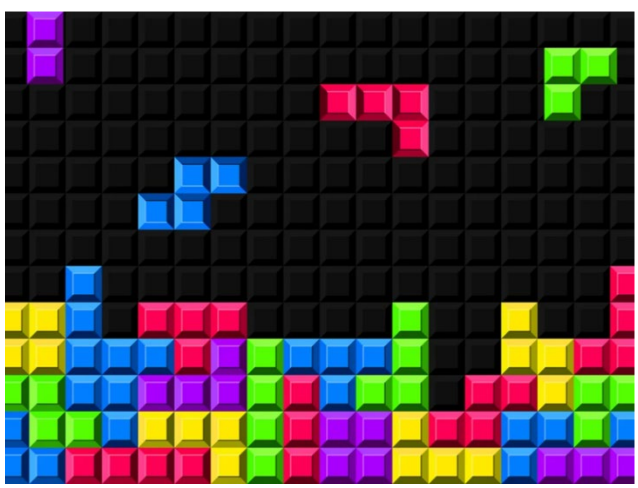

Figure 10 Test Program Dynamic Management

Knowledge Management Stack

We can better see how the knowledge management system is integrated with the rest of the models used in managing systems under the MBSE process. Figure 11 shows the bottom to top elements of the system. At the top we see the dashboards, and the knowledge source metrics that support the dashboards. Below that are the system level models that the knowledge metrics interface with, and below that are the system requirements and requirements models.

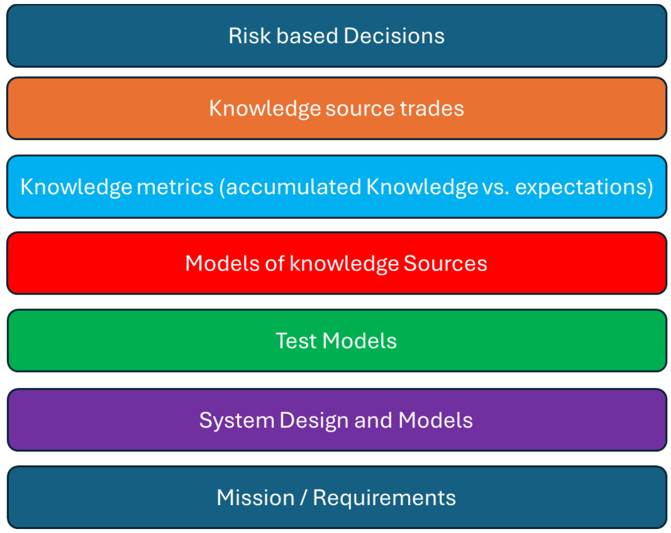

Figure 11 Knowledge Management Stack

Modeling

Central to the development of the knowledge-based metric is the modeling of the different knowledge sources. The models of the knowledge sources capture several key aspects of the knowledge source including 1. The knowledge source class, 2 the Knowledge source characteristics, and 3. The interfaces and links to the knowledge source. The classes of knowledge sources are 1. Legacy system performance, 2. System Design, 3. Modeling and simulation, 4. Developmental Test, and 5. Operational test.

Figure 12 is a SysML view of the knowledge source model that will exist for all knowledge sources. Embedded in the model for each knowledge source will be the following characteristics: 1. The knowledge source class, 2. The requirements that this knowledge source is linked to, 3. The design sub-systems related to this knowledge source, 4. The similarities of the knowledge source to the true system being designed, 5. The performance profile coverage for the specific knowledge source (including operating environment), 6. The reliability of the knowledge source, 7. The fidelity of the knowledge source, 8 the schedule associated with the knowledge source, 9. The cost associated with the knowledge source, and 10. The required inputs and predecessor event(s) to execute the knowledge source.

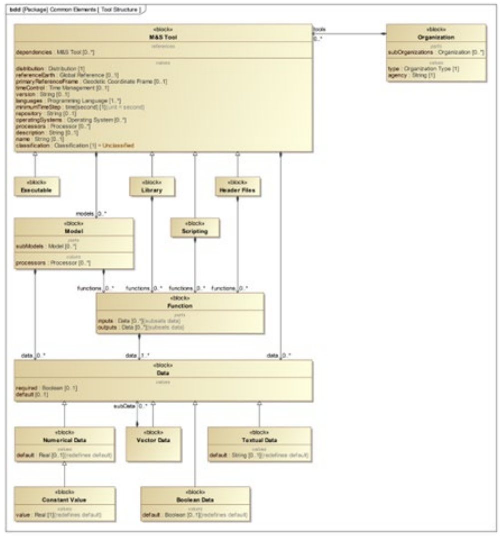

Figure 12 Knowledge Source Model

Once the different knowledge source are documented and modeled we will be able to see where there are gaps and redundancies in the knowledge we need to make decisions about the performance and development of the system of interest.

Integration with the Digital Thread

The models of the knowledge sources can then be integrated with the other models of the system including the requirements and test model described earlier in this paper. In addition, knowledge metrics can contribute significantly to informed risk management and decision-making within the program life cycle. In this way the metrics are both integrated into the digital thread of the system and are available as needed for decision-makers.

Conclusion and Recommendations

As we have seen in this paper there is an opportunity to take better advantage of available knowledge is critical, to our ability to inform and prioritize the two use cases that we highlighted earlier in this paper, first the ability to make better determinations about very high-risk technical projects, second to develop realistic approaches to accelerate development and fielding of systems when we have the knowledge needed to do so. For far too long the acquisition system's best risk management processes have been embedded in the minds and histories of its senior most program management and engineering team member.

Knowledge Source Management

The management of a program's test plan is a requirement for all programs and is captured in the Test and Evaluation Master Plan; likewise test program and test cases are developed and managed by Test and Evaluation working group and by the independent test authority. This covers operational test, and some modeling and simulation, and some developmental test. However, there is very little effort to capture the knowledge of system performance in legacy performance data and in design outside of the design process. As a result, much of the knowledge that is available about system is not captured, curated, and managed in a consistent and portable way to make it available when and where it is needed. This has led in the past to many poor assumptions and decisions being made for lack of visibility into knowledge that already exists.

Using Knowledge Metrics in Portfolio Management

The knowledge-based metrics once standardized have the potential to be used outside of individual program office. Like other digital processes in the MBSE umbrella a significant part of the metrics value is realized at a higher level than the individual program element. Because the knowledge sources and the metrics are math based, and created in model form, they can be aggregated and shared across programs to help manage portfolios of program in several ways. At the top level the metrics can be used to evaluate the status and risk of different technical development and testing efforts withing a portfolio of programs and make strategic decisions about where to spread or concentrate risk, and resources.

Using the Knowledge Metrics in Mission Engineering

Mission engineering is also an area where we can use the knowledge-based metrics. Mission engineering is the synchronization, management, and coordination of concepts, activities, technologies, requirements, programs, and budget plans to guide key decisions focused on the end-to-end mission.[1] The knowledge-based metrics will provide significant and important, information for mission-based decision making.

Using Knowledge Metrics in Service Level Budget Management

The use of this class of metrics may also have significant applications to budget management and capability versus budget trade space analysis. The knowledge-based metrics can be used to help provide data for trade studies.

Knowledge Based Interactions with Industry

Finally in order to make any of our advanced risk management methods work, we (the DoD) are going to need to work much closer with the vendor base to get better insights into the maturity of their designs, and other important knowledge sources that they rely on (legacy system data for sub-systems, design, and developmental test data) and that they often do not share. From a technical standpoint, the shift to MBSE gives the tools we need to capture and use this information, but the historical business relationships (contracting) that we have with vendors do not incentivize them to share this information. Both the technical and the business relationship with our vendors need to change significantly.

References

Anyanhun, A., & Arndt, C. (2024, September). *The use of AI and systems engineering processes in and for testing of AI-based systems*. SERC AI4SE/SE4AI Workshop,

Army Aviation and Missile Command. (2022). *Comprehensive architecture strategy (CAS)*. https://apps.dtic.mil/sti/pdfs/AD1185001.pdf

Arndt, C., Anyanhun, A., & Werner, J. S. (2023). *Shifting left: Opportunities to reduce defense acquisition cycle time by fully integrating test and evaluation in model based systems engineering*. Acquisition Research Program.

Beers, S. M. (2023). *Integrated Decision Support Key (IDSK) for capability evaluation-informed MDO decision-making*. ITEA MDO Workshop, July 2022.

Beers, S. M., Hutchison, S., & Mosser-Kerner, D. (2013). Developmental test evaluation framework: Focus on evaluation and analysis for acquisition success. *Phalanx*, *46*(3), 36–39.

[1] Mission Engineering Guild, November 2020. Office of the Deputy Director for Engineering Office of the Under Secretary of Defense for Research and Engineering.

Collins, C., & Beers, S. M. (2021). *Mission engineering, capability evaluation & digital engineering informing DoD technology, prototyping and acquisition decision*. NDIA S&ME conference, December 2021.

Department of Defense. (2020). *DoD Instruction 5000.89: Test and evaluation.*

DOT&E. (2022). *Office of the Director, Operational Test and Evaluation Strategy Update - Strategic Pillars, Final DOTE 2022 Strategy Update*. https://www.dote.osd.mil/Portals/97/pub/reports/FINAL%20DOTE%202022%20Strategy%20Update%2020220613.pdf?ver=KFakGPPKqYiEBEq3UqY9lA%3d%3d

Executive Services Directorate. (2020). DoD Instruction 5000.89. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500089p.PDF

IEEE/ISO/IEC 15288-2015. (n.d.). *IEEE SA-IEEE/ISO/IEC 15288-2015*. https://standards.ieee.org/ieee/15288/5673/

Gregory, J., & Salado, A. (2024). An ontology-based digital test and evaluation master plan (dTEMP) compliant with DoD policy. *Systems Engineering.*

Gregory, J., & Salado, A. (2024). dTEMP: From digitizing to modeling the test and evaluation master plan. *Naval Engineers Journal*, *136*(1), 134–146.

Gregory, J., & Salado, A. (2024). Towards a systems engineering ontology stack. *INCOSE International Symposium*, Dublin, Ireland.

ISO/IEC/IEEE. (2022). *Systems and software engineering—Architecture description* (document SO/IEC/IEEE 42010:2022(en)).

Muller, G., & Hole, Eirik. (2007). *Reference architectures; Why, what and how* (White paper). Architecture Forum Meeting, Embedded Systems Institute and Stevens Institute of Technology, Hoboken, NJ, March 12–13, 2007.

OpenCAESAR. (2023). *Ontological Modeling Language 2.0.0* [Online]. https://www.opencaesar.io/oml/

Shaffer, A., & Whitley, J. (2024, September 20). Accelerating defense innovation requires changes to the acquisition approach. *National Defense Magazine*.

Test and Evaluation Metadata Reference Model. (2011). *Range Commanders Council. Director, Operational Test and Evaluation (DOT&E) Test and Evaluation Master Plan (TEMP) Guidebook Version 3.1.*

Werner, J. S., & Arndt, C. (2023). *Development of digital engineering artifacts in support of MBSE based test planning, execution, and acquisition decision making*. Acquisition Research Program.

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net