Knowledge based Metrics for Test and Design

May 7th 2025

Dr. Craig Arndt, Dr. Valerie Sitterle, Mr. Jeremy Doerr
Georgia Tech Research Institute

Agenda

- Requirements
- Challenges
- Approach
  - Math
  - Model Based Data Sources
  - Standardized Knowledge source Modeling
  - Decision support
- Value Proposition
- Conclusion and Recommendations

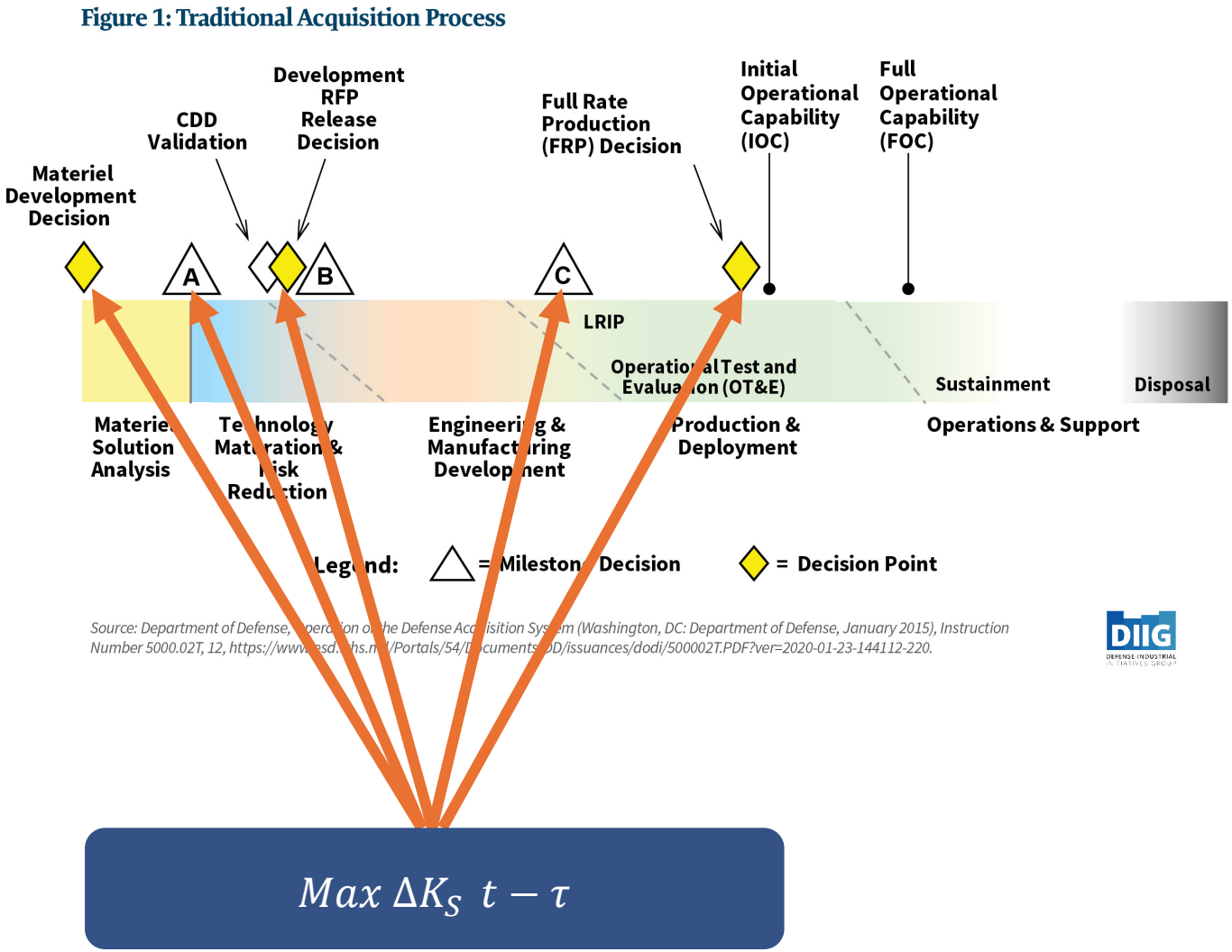
Decision Knowledge Requirements

- Along the acquisition lifecycle many decisions need to be made under uncertainty
- Increases (maximize) the amount of knowledge of a system K_S at decision points

$$Max\ \Delta K_s\ t - \tau$$

- Create earlier decision points based on meeting knowledge thresholds for the decisions at an earlier date

$$Threshold = K_S\ t - \tau$$

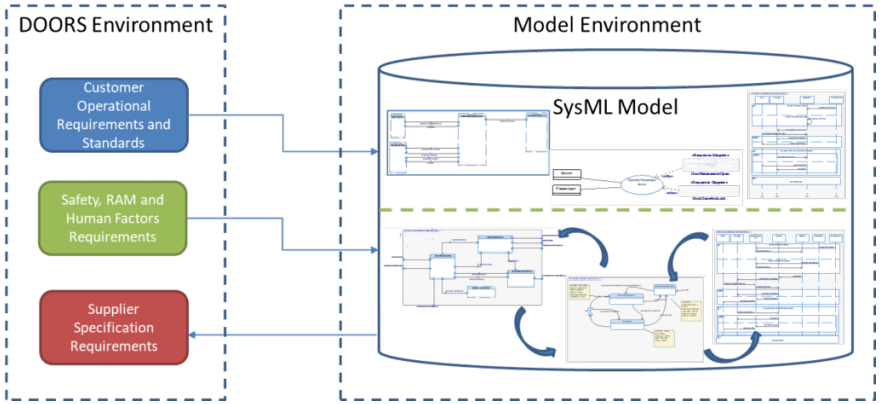


Model-Based Systems Engineering

- Challenge: MBSE allows for the movement of data across system boundaries. To support AI testing we need to also move data across time.

- Opportunity: By using MBSE as the framework we can insert technology including AI based predictive analysis to predict training sets and performance of future embedded AI systems.

- Need: In order to facilitate the ability to move data across both tasks and different parts of the lifecycle we need to integrate models across threat modeling, development and test.



Metrics

- Challenge: Current testing metrics are based on designing test around validation of performance to a specific requirement.
- Opportunity: To change the way we look at test from verifying performance to predicting future performance of the system.
- Need: Develop knowledge-based metrics that support an understanding of our level of knowledge of the system.

Requirements Development

- Baseline requirements, developed based on documented challenges to the acquisition system
- Different use cased developed and evaluated.
- Need for metrics that can be used by decision makers to evaluate
  - Risk
  - Opportunity for program acceleration

1. Capability Need	Business case development	Is there sufficient detail in the business case allowing for clearly defined requirements?	Insufficiently developed business cases leads to increased timelines and costs due to overestimating current knowledge and accumulating risks throughout the phases.	GAO-23-106059, GAO-21-511T
1. Capability Need	Key stakeholders' project and technology knowledge to make appropriate decisions	Does the key stakeholders, including the PMs, Financial Managers, Contracting Officers, and Decision authorities have sufficient knowledge, training, or experience to support making appropriate decisions?	Insufficient experience, training, or other related knowledge limits effective decision making and early risk identification which leads to risk accumulating across phases/milestones.	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 leads 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 impacts funds available for current JROC approved projects likely leading to an underestimation of costs & overestimation of capabilities in order to maintain momentum.	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 requires concurrent phases increasing the risk of operational ineffectiveness, higher costs and time-intensive design changes.	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 deficiencies violates the spirit of the AAF and does not adequately capture lifecycle costs or timelines.	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. This leads to long-term financial risk for programs by forcing quantity tradeoff or program sunset decisions before fielding.	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 once fielded.	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. It increases schedule and cost risks as user needs shift.	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 often leading to extended timelines and increased costs.	GAO-24-106831

Development

Metrics (Knowledge, Expectation, Risk/Opportunity)

For specific decision milestones (D), compare measured knowledge of the system and its aspects (N) to necessary knowledge to make the decision

$$K_{S,D} = \sum_{1}^{N} \Delta \left(K_{D,N,measured} - K_{D,N,expected}\right)$$

Evaluate specific tests or sets of tests to support decisions against requirements (R_N) and performance factors (P_N)

$$\Delta K_S = \sum_{1}^{N,test} K_{S,N,test}\, P_N R_N$$

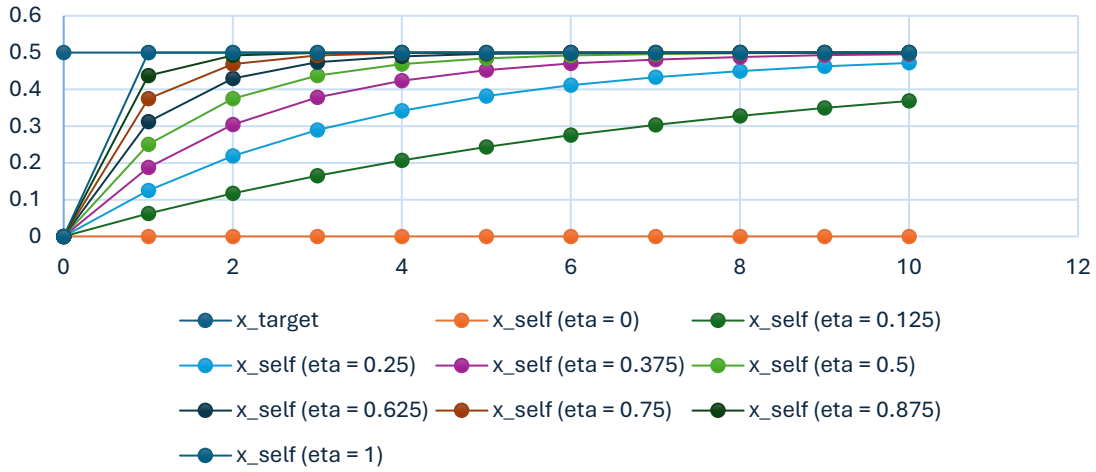
Risk can then be associated to knowledge of the system ($K_{S,R,N}$)

$$Min\ Risk = Max \sum_{1}^{N} K_{S,R,N}$$

Knowledge Accumulation / Trade-Off Analysis

- For knowledge growth to be meaningful it needs to be compared with reasonable expectation for knowledge accumulation on the development and acquisition program in question.
- Deviations (delta knowledge (ΔK_T) from expected knowledge accumulation impacts risk and opportunity in the excision of the program.
  - Risk of successful development
  - Opportunity to acceleration the program do to grater knowledge of program success.

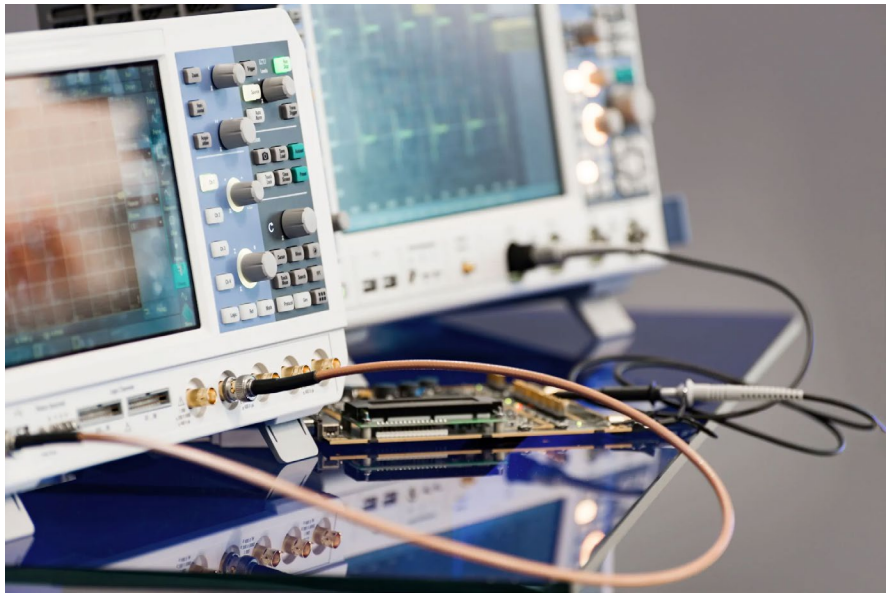
Notional Knowledge Increase with Respect to Target



- Each source of knowledge of the system (legacy data, design data, modeling and simulation, developmental test, operational test) contributes differently to the total.
- The total integrated test program can be optimized to maximize knowledge and reduce risk at critical times in the program.
- Alternative test (knowledge) approaches (M&S, and design data) can be used based on to optimize knowledge and test programs.
- The trade space analysis will trade different test program alternatives against risk and schedule

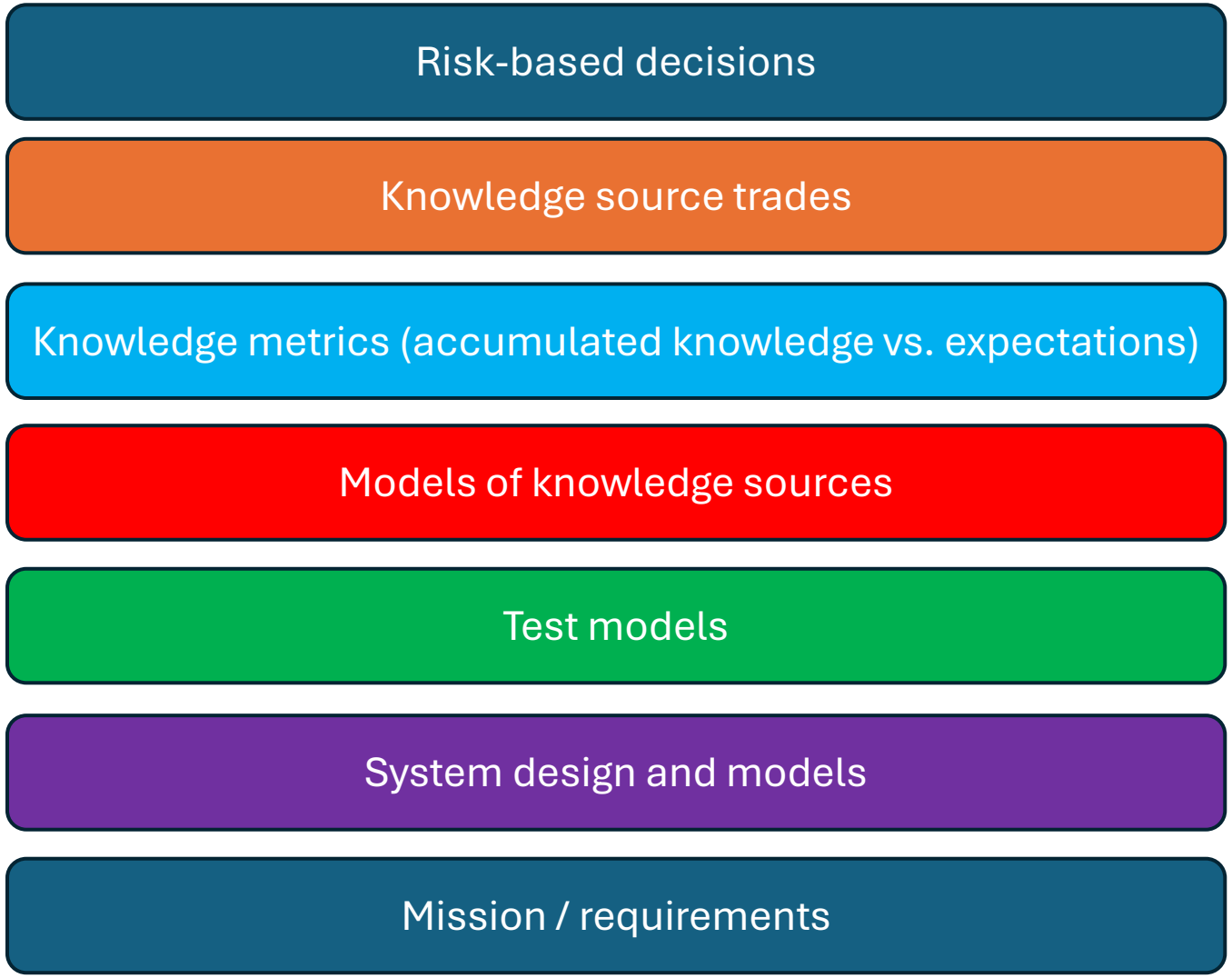
Test program									Risk
1						x	x		60
2			x			x		x	75
3		x		x	x		x		
4		x					x		
5			x	x					
6						x		x	

Knowledge Sources

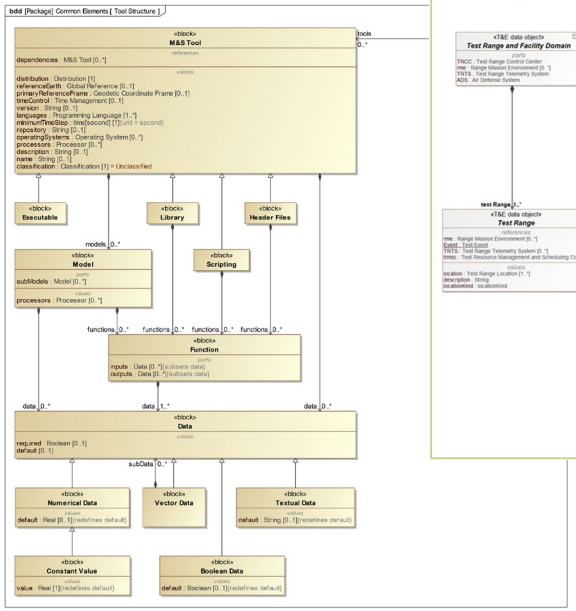


- Knowledge Sources
    1. Legacy system data
    2. Design data
    3. Modeling and simulation
    4. Developmental test
    5. Operational test
- Characteristics to Model
    1. The knowledge source class
    2. The requirements that this knowledge source is linked to
    3. The design sub-systems 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
    10. The required inputs and predecessor event(s) to execute the knowledge source.

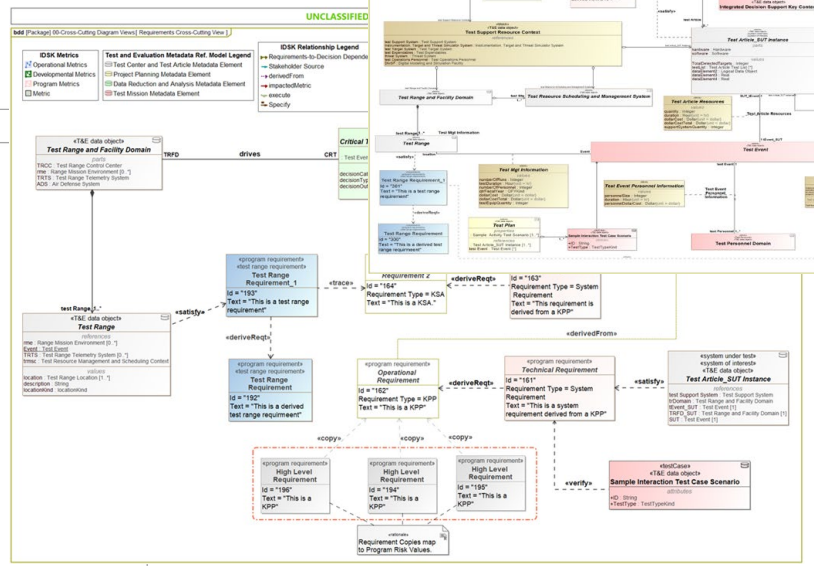
Knowledge Representation Stack



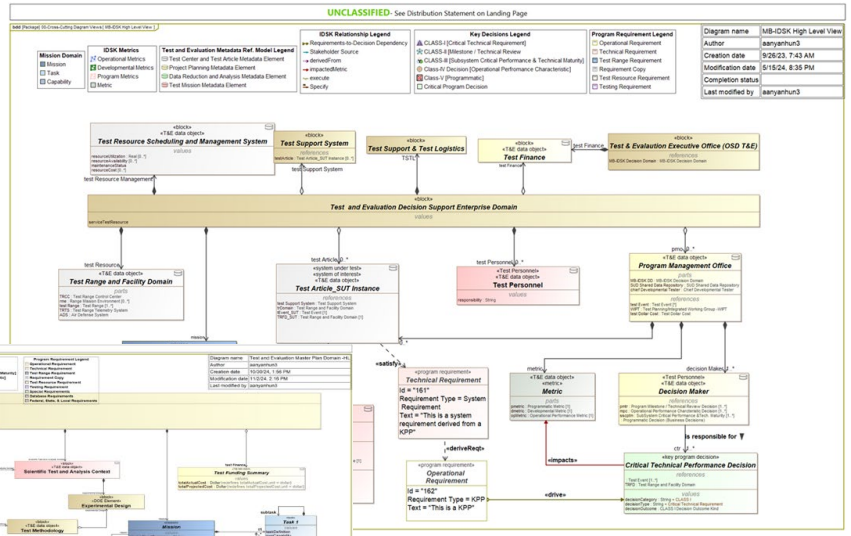
Modeling Interphase



Model of knowledge sources design / M&S / test

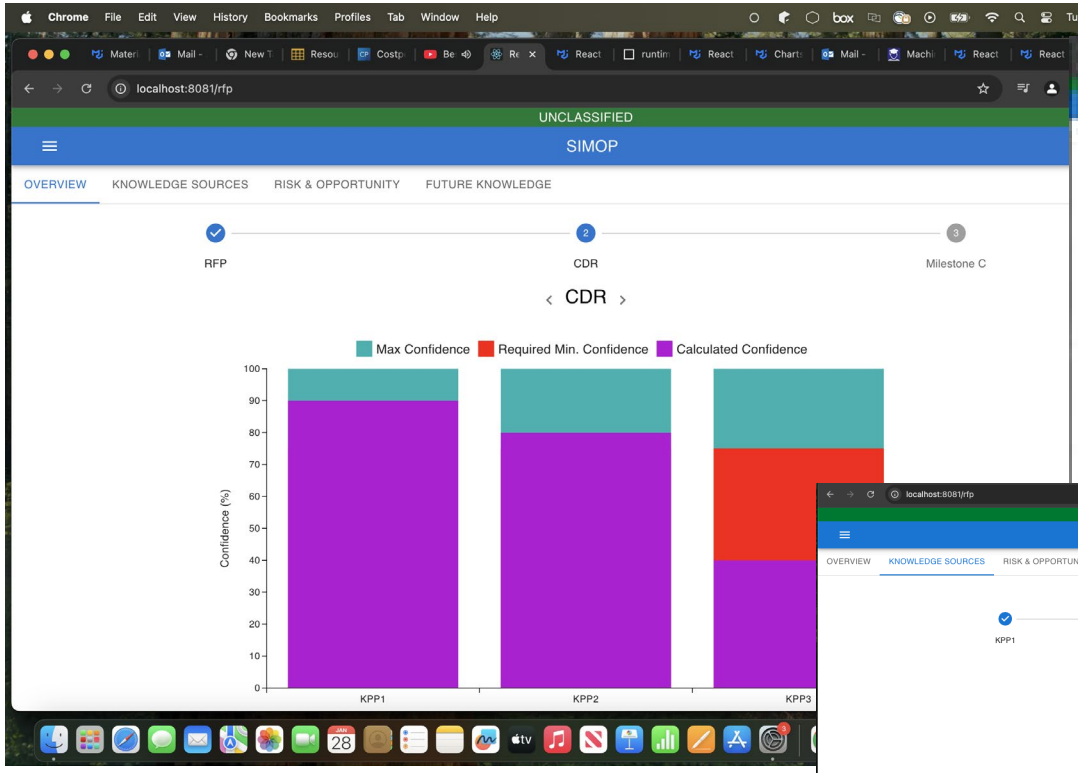


Model of test planning and test program TEMP

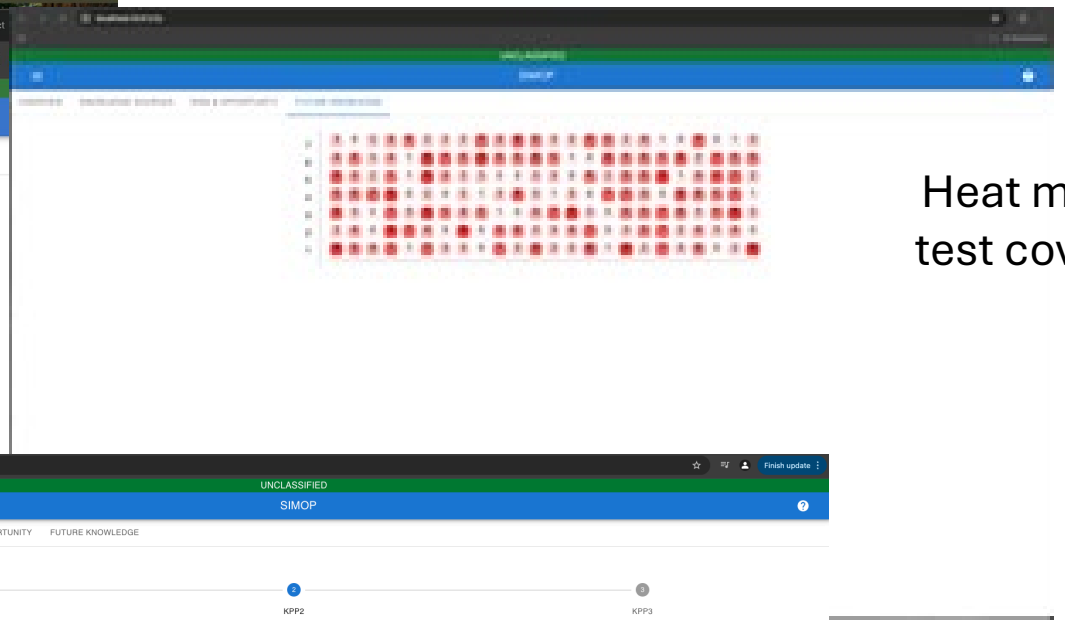


Model of decision-making IDSK

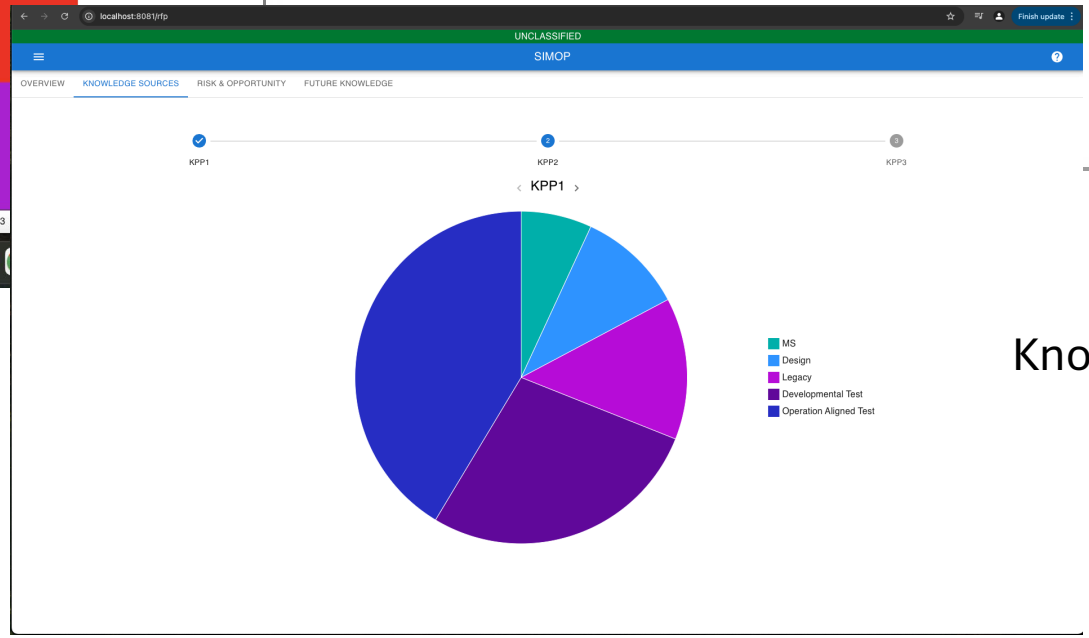
Decision Support Prototypes



Risks and opportunities by KPP, by time



Heat map of test coverage

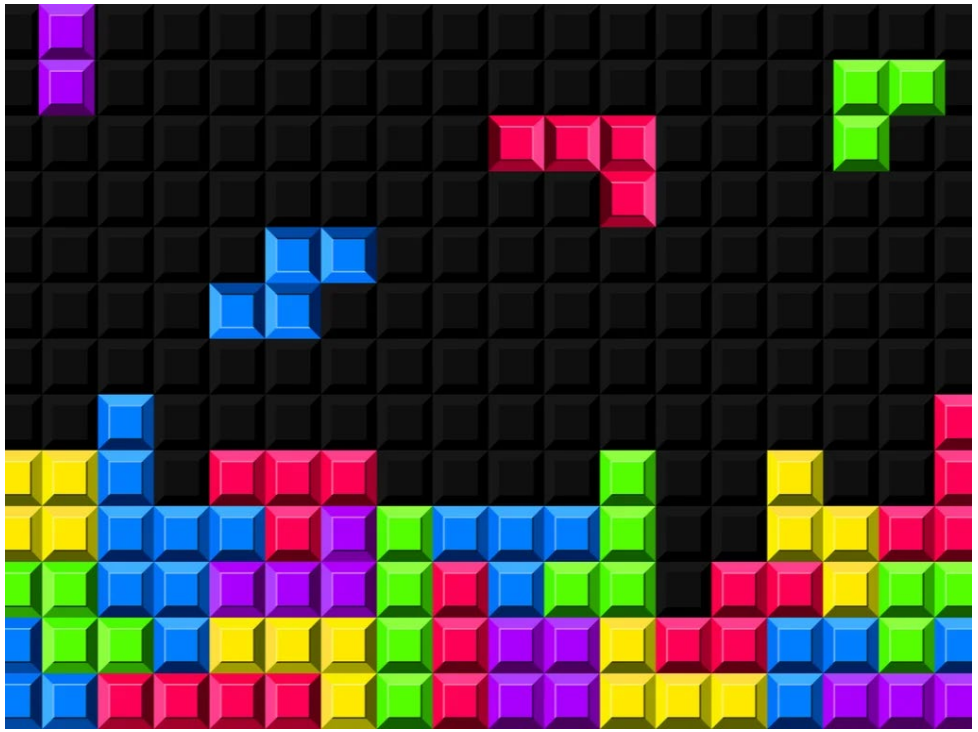


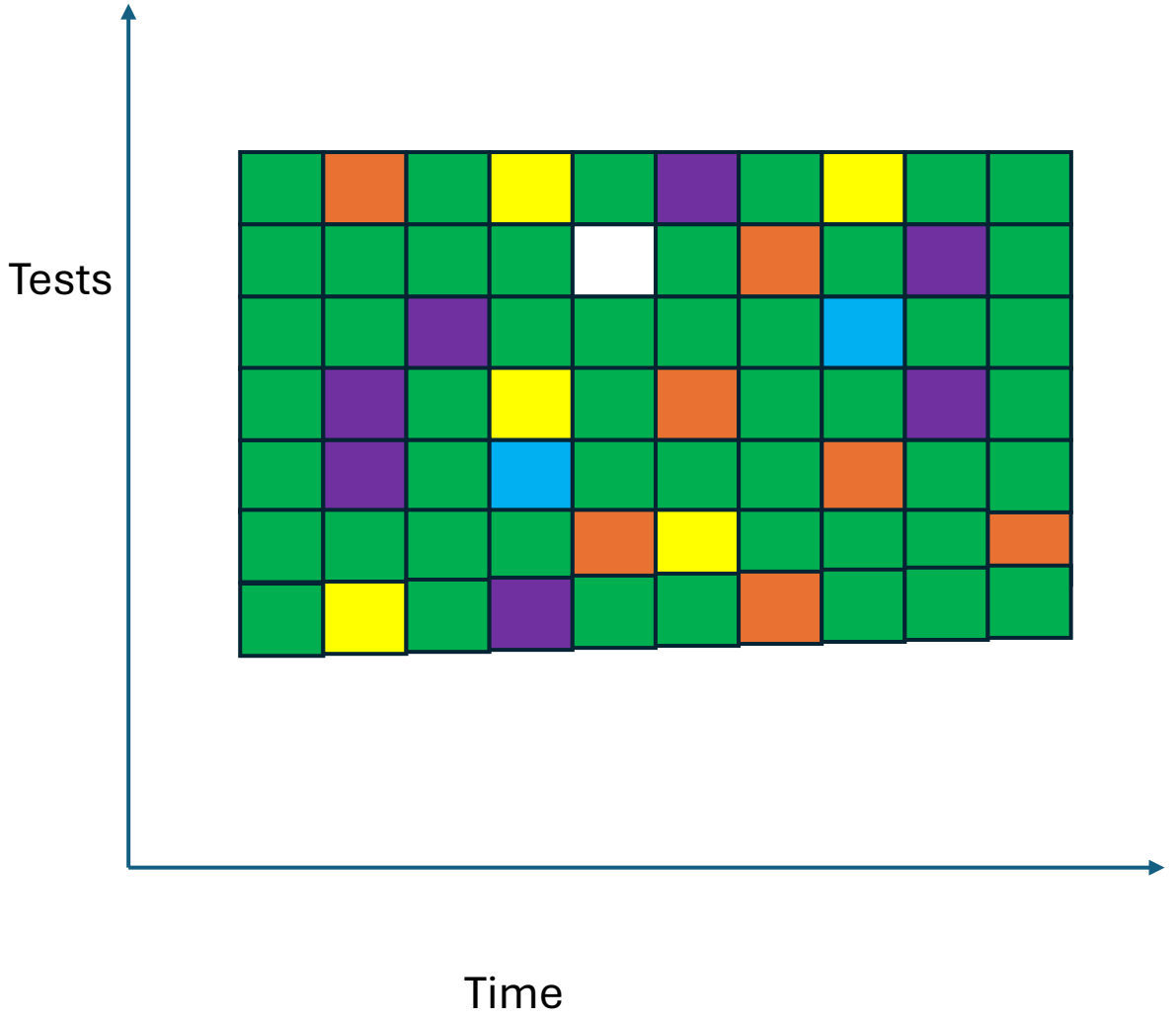
Knowledge sources

Key Use Cases

- Use Case 1: High risk, technically challenging development
  - Need: To understand when, where and how great the technical challenges are on complex programs
  - Better and objective mathematical representation of risk based on current and needed knowledge
  - Goal is to avoid huge schedule and cost overruns in high risk programs
- Use Case 2: Opportunity for accelerating test and programs
  - One of the key goals of the DoD is to accolated the development of programs
  - Knowing that testing is redundant with existing knowledge
  - Clasping test programs, relying on other knowledge gained earlier in the program

Tetris-Based Model

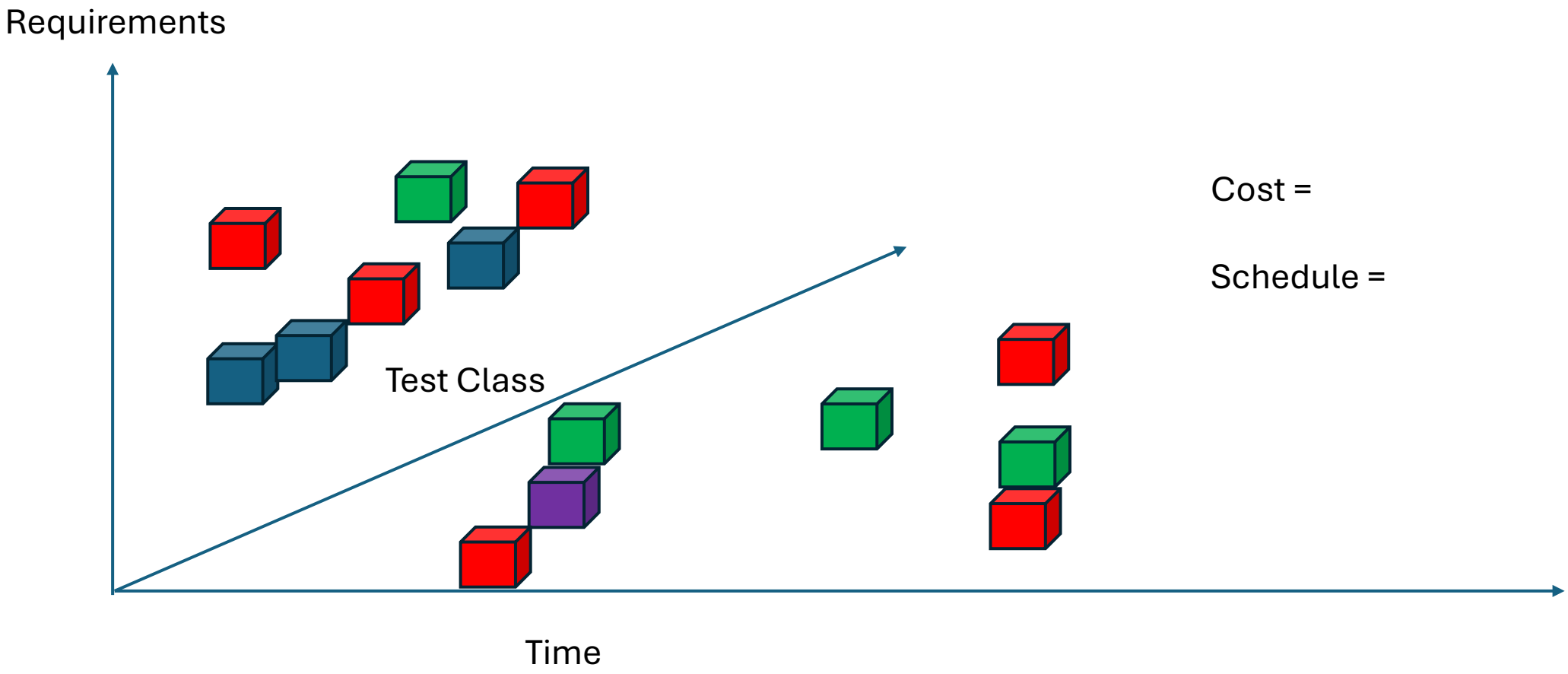




Trade Space Analysis

- Test programs include a wide range of different tests and models and simulation efforts for some or all of the system
- As we determine based on the knowledge surpluses and deficits are discovered, the process is then to look for ways to reshape the test program
- However, test programs are not that simple, test programs can be best represented a complex three dimensional matrix of time, test class and the requirements that the test (knowledge source) can provide insight into
- As we see risk (the need for more knowledge), or opportunity (a surplus, or redundance of knowledge) we can then look to add or delete tests as needed
- Because tests and test resources are complex, interdependent and interconnected we get to a complex game of 3D Tetris to add tests in some areas, and delete them in other areas and overall create a test program that used less resources in time and money

Trade Space Analysis / Test Plan



Value Proposition

- In response to the stated need to increase the transparence and understanding of risk in the development of new technologies this project will:
  - Provide the means of evaluating the ability of the program to accelerate the test program at specific level of risk
  - Provide the means of understanding the needed level and amount modeling and test for a program in order to reduce risk to an acceptable level
  - Provide the means of evaluating the mix of M&S, developmental, and operational test that will best support a specific program, given the knowledge captured by legacy and design

Conclusions and Recommendations

- The knowledge metric construct takes a significantly new perspective on the purpose of test
  - In the past test has been a method to validate compliance to requirements
  - The new construct see's test as one of many knowledge sources that can be used to predict the future performance of a system under development
- This new outlook and associated metrics gives us the ability to better manage risk and opportunity on programs, and to manage portfolios of programs
- To make these metrics and analysis work, will requirement programs to model and characterize knowledge source and work more closely with venders who maintain the design knowledge