Technomics
Better Decisions Faster

Using Technology Readiness Levels to Predict the Future of Nuclear Weapons

Michael Metcalf – mmetcalf@technomics.net

Abby Schendt – aschendt@technomics.net

Raymond Vera – raymond.vera@nnsa.doe.gov

International Cost Estimating and Analysis Association

May 2022

Abstract

Technology Readiness Levels (TRLs) are used to measure and assess technology maturity. This paper presents innovative research demonstrating how historical TRL data can be used to perform credible, data-driven schedule analysis for programs early in development. Originally designed to help schedule analysis for the nuclear weapons stockpile, the resultant methodology combines historical milestone data and statistical methods to generate a Monte Carlo simulation of a risk-adjusted schedule for complex programs.

Keywords: Technology Readiness Level, TRL, Monte-Carlo, Schedule Analysis, Risk-Analysis

Table of Contents

Table of Figures

Table of Tables

Introduction

Technology Readiness Levels (TRL) are used to measure and assess technology maturity. This paper presents innovative research demonstrating how historical TRL data can be used to perform credible, data-driven schedule analysis for programs early in development. Originally designed to help schedule analysis for the nuclear weapons stockpile, the resultant methodology combines historical milestone data and statistical methods to generate a Monte Carlo simulation of a risk-adjusted schedule for complex programs.

This novel research has been performed in support of our early program schedule analysis for the National Nuclear Security Administration (NNSA) weapons programs. These are large complex programs that often span over a decade or more in development and production.

TRLs: History and Overview

Technology Readiness Levels (TRLs) have been a part of government programs since around 1974, when they were first introduced by Stan Sadin at NASA. Until the 1990s, TRLs used a 7-point scale. Since then, TRLs have employed a 9-point scale. Today, "TRLs are the most common measure for systematically communicating the readiness of new technologies or new applications of existing technologies...to be incorporated into a system or program." [6] As the Government Accountability Office (GAO) and others note, TRLs can enable understanding of technology maturation risk; however, they are currently not maintained in a format to provide a consistent risk assessment to the program. This paper explores an approach to incorporate the TRLs into a risk-informed schedule analysis that allows program managers and evaluators, from system engineers to independent cost estimators, to consider technology maturation.

To understand a TRL, it is best to first start with a discussion on Technology Readiness Assessments (TRA). The GAO defines a TRA as "a systematic, evidence-based process that evaluates the maturity of technologies (hardware, software, and processes) critical to the performance of a larger system or the fulfillment of the key objectives of an acquisition program, including cost and schedule." [6] TRAs should be conducted as a standard part of the program life cycle. A successful program will rely on more than one-off assessments and will continually re-evaluate TRLs as a mechanism to monitor the health of the program and the program's schedule. "In general, a distinct TRA should be conducted at several points during the 'life cycle' of a new technology and of new systems. These might include (a) the completion of systems [analyses] and conceptual design studies, (b) the point [of] decision from among several competing design options, as well as (c) the point of decision to begin full-scale development." [2]

The authors collected and archived TRL values (i.e., TRL data) for a variety of systems. In order to use this data properly, we applied the definitions of each TRL using GAO's standard thermometer graphic and supporting table of definitions, which are presented in Figure 1 and Figure 2 and discussed below.

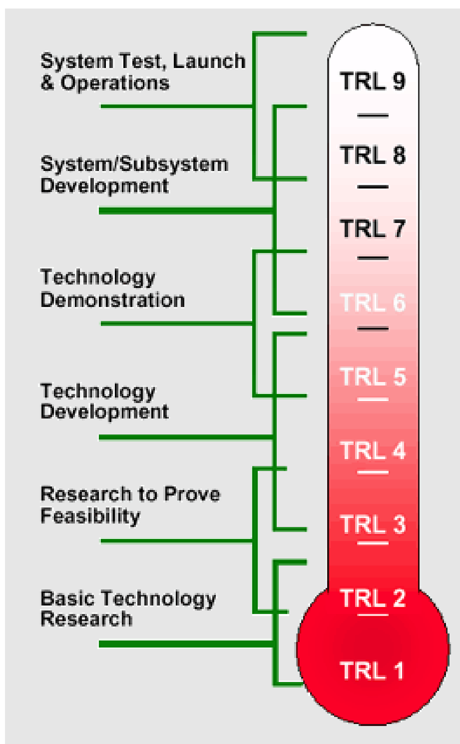

Figure 1: Popular TRL thermometer chart [6]

Technology readiness level (TRL)		Description
1	Basic principles observed and reported	Scientific research begins to be translated into applied research and development. Examples include paper studies of a technology's basic properties.
2	Technology concept and/or application formulated	Invention begins. Once basic principles are observed, practical applications can be invented. Applications are speculative, and there may be no proof or detailed analysis to support the assumptions. Examples are limited to analytic studies.
3	Analytical and experimental critical function and/or characteristic proof of concept	Active research and development is initiated. This includes analytical studies and laboratory studies to physically validate the analytical predictions of separate elements of the technology. Examples include components that are not yet integrated or representative.
4	Component and/or breadboard validation in laboratory environment	Basic technological components are integrated to establish that they will work together. This is relatively low fidelity compared with the eventual system. Examples include integration of ad hoc hardware in the laboratory.
5	Component and/or breadboard validation in relevant environment	Fidelity of breadboard technology increases significantly. The basic technological components are integrated with reasonably realistic supporting elements so they can be tested in a simulated environment. Examples include high fidelity laboratory integration of components.
6	System/subsystem model or prototype demonstration in a relevant environment	Representative model or prototype system, which is well beyond that of TRL 5, is tested in its relevant environment. Represents a major step up in a technology's demonstrated readiness. Examples include testing a prototype in a high-fidelity laboratory environment or in a simulated operational environment.
7	System prototype demonstration in an operational environment	Prototype near or at planned operational system. Represents a major step up from TRL 6 by requirement demonstration of an actual system prototype in an operational environment (e.g., in an aircraft, a vehicle, or space).
8	Actual system completed and qualified through test and demonstration	Technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation of the system in its intended weapon system to determine if it meets design specifications.
9	Actual system proven through successful mission operations	Actual application of the technology in its final form and under mission conditions, such as those encountered in operational test and evaluation. Examples include using the system under operational mission conditions.

Source: GAO analysis of agency documents. | GAO-20-48G

Figure 2: GAO Description of TRLs from TRL Assessment Guide [6]

The start of the TRL thermometer is what the authors consider symbolic TRLs. TRL 1 and 2 deal with a material/technology not in the final form, fit, or function. TRL 1 is assessed when a material or technology is tested or demonstrates a key physical attribute, often a basic property. This work to develop a potential new material or item is typically done by universities. At TRL 2, the developers start considering a range of applications for the material or technology, but it still exists in a hypothetical space.

By TRL 3 the material or technology of interest will include form/function of a specific application. Money, time, and research is being invested to determine whether an item is suitable in a unique and specific application. At TRL 3, there is no requirement for a prototype. From TRL 1 to 3, there are few test requirements and programs can quickly move through these levels. Certain items may migrate through early levels rapidly when the technology has already been used elsewhere. The time to complete these initial TRLs is not necessarily a meaningful measure of maturity progress.

The next two TRLs (4 and 5) are where the program will start producing something physical. TRL 4 is a prototype in a lab environment. The component should show proof-of-concept for critical capabilities and characteristics. This first deliverable/component is not required to be fully functional or in final form. TRL 5 looks at the component or item in its relevant environment(s). Some basic elements should be integrated with realistic supporting elements and testing requirements for the component will significantly increase.

The next two TRLs (6 and 7) address system level testing. TRL 6 requires proof/modeling/testing that the component will work within the subsystem or system. Testing should focus on system level and integration testing to ensure the component works in the relevant environment(s) and in the relevant system. “Upon entering product development and therefore having achieved at least TRL 6 (system demonstration in a relevant environment) the [critical technology] is now considered beyond the reliance of science and technology investment and is dependent on standard systems engineering development practices to achieve a fully mature status.” [6] TRL 7 is the final stage of system level tests. The system should be prototyped and demonstrated in the relevant environment. This prototype should be as close to the final version as possible.

TRL 8 is reached when the system or subsystem itself has completed all testing requirements. TRL 9 can only be achieved when the technology or component is successfully demonstrated an operational and fielded unit. Figure 2 displays the GAO’s definition of each TRL.

TRLs in NNSA

The NNSA is tasked within the Department of Energy to maintain the nuclear deterrent, advance nuclear nonproliferation, promote international nuclear safety and provide support to the Nuclear Navy. The NNSA has an extensive portfolio of infrastructure projects, weapons modernization programs, science missions, and research and development activities. Many areas of the NNSA would benefit from metrics on technology maturation, but the weapons modernization programs currently have the most rigorous TRL implementation.

The NNSA weapons programs are managed through what is called the 6.X Process, referring to Phase 6 (operations and maintenance) of the nuclear weapons lifecycle. The 6.X process provides a unique lifecycle specific to weapons modernization program (as opposed to building a new weapon from scratch, which the NNSA has not done since the end of the Cold War). Figure 3 shows the 6.X process in more detail.

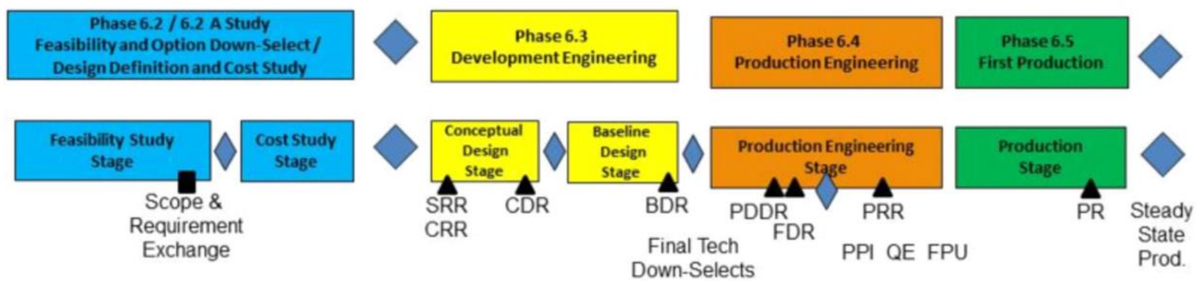

Figure 3: Phase 6.X Process

Once a weapon completes feasibility and initial design options (Phase 6.1 and 6.2), it enters detailed design engineering (Phase 6.3). The goal of this is to mature all technologies and components and begin to integrate into a system. A typical weapon modernization has 30 to 50 major components and many more subcomponents that are individually managed and tracked. Phase 6.4 includes final design reviews, engineering releases and completion of production engineering, safety basis reviews, qualification, and certification. Phase 6.5 authorizes the assembly of the first delivered nuclear weapon, called the First Production Unit (FPU). FPU is the most significant milestone, and achievement of system FPU represents the completion and confirmation of all technologies, components and environments.

As shown in Figure 4, many of the major milestones within the weapons modernization use inputs from the TRL process as exit criteria. These represent the ideal status of all components at each milestone, but in practice some component maturation or documentation bleeds into the next milestone.

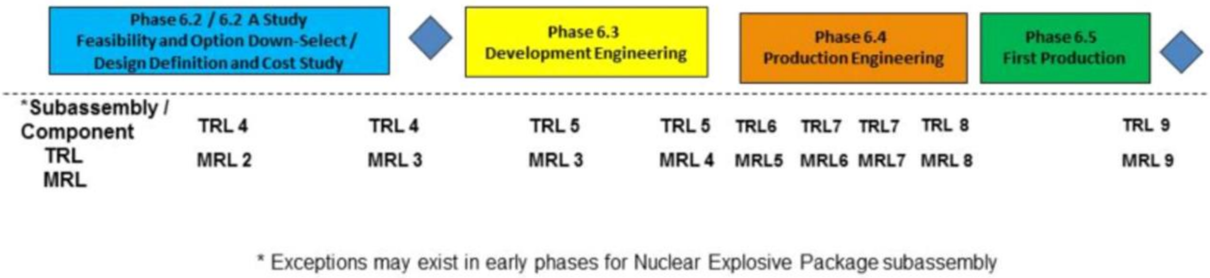

Figure 4: Alignment of the Historical TRL Data to the Phase 6.X Process

Similar to most agencies, the NNSA tracks its weapons programs (both for 6.X milestones and for TRLs) at the system level, but uniquely also tracks at the subsystem and component levels. The work below relies on detailed, component level TRL progression. The analysis relies on measurement of all TRLs, but a few key TRLs are most notable. NNSA policy recommends that all components reach TRL 4 prior to the start of development engineering (Phase 6.3). The authors use TRL 6 as a measure of system health; program risk increases greatly if the program enters production engineering (Phase 6.4) while many components have yet to demonstrate system integration. TRL 8 is used as a measure of component FPU, meaning the individual component can successfully be produced and works within the system in all environments. A weapon cannot reach system FPU until all components have completed TRL 8/component FPU. Within the NNSA, TRL 9 is reached for all components and the system one (1) year after system FPU. As the authors' primary interest was estimating system FPU, we ideally wanted a point between TRL 8 and TRL 9 to measure schedule. The results and discussion of how the authors determined a system FPU estimate are presented in the methodology section.

The NNSA has recognized inconsistencies with how NNSA contractors have managed the TRL maturation process. Defense Programs within the NNSA has recently implemented a more rigorous TRA process and calculator for upcoming programs to ensure that all programs and components are confirming maturity the same way. The new process will also realign TRLs to Phase 6.X milestones from what is shown in

Figure 4. Much of the work and historical data described below is based on the old process, so the authors will seek to understand the impact of the new process and implement adjustments as necessary as part of future work.

Author's Interest

The authors' interest in TRLs emerged during a discussion about an early-development NNSA program that had yet to establish a baseline schedule. In previous work we calculated a system FPU date by syncing with previous programs and then setting interim 6.X milestones to match. However, we knew that this method fails to account for various scope differences between the programs. This prompted further discussion of how programs establish early acquisition milestones prior to established baselines. This discussion highlighted the need for programs to have a strong understanding of the impact of technology decisions earlier in the program when scope details are still part of the trade space.

As part of a quick-turnaround report to Congress, the authors performed an analysis comparing initial TRL and Manufacturing Readiness Level (MRL) projections with actual/ongoing results. That analysis yielded a promising rough estimate for schedules that required further exploration.

In preparation for an Independent Cost Estimate, our second look at the data yielded even better results. We were no longer constrained to small steps between TRLs; we could provide averages and standard deviations for specific levels. We have thus moved forward with testing a methodology to generate an early-look schedule as a primary methodology for schedule estimates early in a program. This method can also be used as a cross-check for schedules once a schedule baseline has been achieved.

Mankins points out in his article, "Technology Readiness and Risk Assessments: A New Approach", that there is no explicit link between TRLs and risk assessment. Program managers and decision makers are often provided a TRL for a given technology and have very little understanding of the potential impact to a program. In his paper, Mankins proposed a technology readiness and risk assessment (TRRA) approach to evaluating technologies. In this approach he advocates that the ideal TRRA would have four key features: clarity, transparency, crispness, and usefulness in program advocacy. The process should involve “clear decisions criteria” and “be analytically grounded in a way that allows independent evaluation”. The process should be formal and “consensus based” and “easy for participants, managers, and independent observers to understand both the process, the interim steps in the assessment, and its results.” Decisions in this process should be “made by and/or with the ownership of senior management” and it must by “crisp, timely and keyed to annual [research and development] and system program budget planning requirements”. Finally, the processes used should “also produce the basis for advocacy of the result”. [3]

Like Mankins, the authors hope to understand the additional risk introduced into the schedule by integrating less technologically mature components. We feel this method might help bridge the gap between the current use of TRAs and what Mankins views as the future for TRRAs.

Advantages using TRLs for estimating Schedule

Within the NNSA, TRLs provide a consistent measure of progress by major component and provide comparability across components than other measures. Other milestone decisions for the system or component can be achieved “with risk,” where the item or system proceeds even if it is short of the full requirements for that milestone. This can provide an artificially optimistic understanding of the program’s health and schedule realism. However, TRLs have a more rigorous gate process and the system as a whole cannot proceed to FPU without all components having reached TRL 8.

Many schedule analyses rely on critical path analysis. This is only effective if the program is sufficiently mature to create a reliable detailed schedule and if the resulting schedule is completely integrated with proper predecessor/successor relationships. The NNSA schedules come from multiple sites and stakeholders that have challenges in integration. They are typically not baselined until well into the design phase. TRLs do not require significant integration and are created early in the program and updated regularly as each technology matures.

Similarly, other schedule assessment techniques (including task burn analysis and schedule performance index) require baselined schedules and reliable detailed projections that aren't available early in a program's development. Additionally, not all tasks have to be completed for a program to create a first production unit.

Much of the existing TRL analysis that the authors reviewed focuses on the total system maturity. By exploring the portfolio of major components, the authors seek to ensure the highest risk components are highlighted and their potential impact on the schedule is properly considered.

This detailed TRL analysis can also be used to compare programs, providing a contrast to complexity-based analogies that within the NNSA rely upon assessments from subject matter experts.

Disadvantages of TRLs in this application

TRL data is neither consistently collected nor widely available across the enterprise. Not every organization or program requires regular publishing of TRLs. Major sources of TRL data (cost reports, site schedules, Cost Analysis Requirements Descriptions [CARDs]) are maintained at different degrees of fidelity. Some sources report TRLs at the subsystem or major components and other sources report TRLs for all components below the subsystem.

In addition, approaches for defining TRL status and projections have historically been vague or inconsistent, though as noted above NNSA's Defense Programs is implementing new oversight to drive consistency going forward within the organization.

TRLs will not represent additional program schedule risks beyond technology maturity; within the NNSA these include producibility risks, supply or reliability issues, regulatory and safety concerns, requirements changes, or funding and staffing challenges. Other schedule measures could potentially account for these risks.

Our literature review also supported several limitations we saw in our work. Phrased best by Sauser, "TRLs were not intended to address systems integration nor to indicate that the technology will result in successful development of a system" [5]. Sauser continues, “systems often fail because attention is given to the technology while knowledge of the linkages/integrations is overlooked” [5]. To this end, the authors hope to also explore the system integration aspect of TRLs in the future, as most of the work to date has focused on component and subsystem maturity measures.

Methodology

Assumptions

The core of any analysis is understanding the assumptions that will encapsulate the modeling. We built our assumptions around ensuring that we could use as much of the data as possible. Because of the different reporting requirements between sites and programs, the raw data comes with gaps and small corrections allow the data to be more useable.

The first major assumption we made is that all the TRLs must progress in order. TRL 4 must come before TRL 5. This seems trivial, but occasionally due to when or how data is collected or faulty evaluations, components can have TRLs out of order. The largest cause of disorder is paperwork being processed on different dates. We addressed this by leaving the raw dates from the source document and creating a separate set of processed dates where we implemented our logic assumptions. We used the processed dates to force the TRLs to proceed in order. The second major cause of levels being out of order is when a technology does not work as expected or the program discovers that a legacy technology doesn’t meet requirements. This data is maintained in the original format with a negative progression so we can capture the risk of an erroneous assessment.

Another assumption we make is that all TRLs must be achieved. This doesn’t make a huge impact, but we put in place holders in the same quarter that all other levels will progress. This place holder does not appear in our raw data pull as the higher TRL is the level that is pulled for that quarter. As will be discussed later in the Methodology section, that level-skip will get filled in so we don’t lose information. That fill can also be required for components that stack a TRL progression, i.e., process multiple TRLs in the same day/month/quarter.

Raw Data

The majority of this effort has been targeted for NNSA's Life Extension Programs, the multi-billion-dollar modernization programs for the nation's nuclear weapon stockpile. The programs have two primary design agencies and up to seven production sites charged with developing component designs, an integrated system design, manufacturing capabilities, a test program, qualification and certification, and a safety and maintenance program. Most programs track 30 to 100 major components and subcomponents. Our data often examines each component separately and does not yet tie in dependencies for higher architecture within the system or subsystem.

Table 1 provides an overview of the sources of TRL data used to build our database. The first data source is cost reports provided by the NNSA. Formally known as the Weapon Design Cost Report (WDCR) and Baseline Cost Report (BCR), these reports document the program's estimates of record for cost, schedule, staffing, scope, and programmatic details that constitute the program's baseline. The WDCR is the first cost estimate of record produced by the program office and the first point in time when the NNSA sites provide a detailed, built-up estimate with near final design decisions.

Table 1: TRL Data Sources

Type	Source	Frequency	Description
Program Documents	Weapon Design Cost Report (WDCR)	6.3 Milestone	First program level definition of scope, schedule, labor, cost, and technical descriptions.
	Baseline Cost Report (BCR)	6.4 and 6.5 Milestone	Update to the WDCR as the program passes two milestones.
	Quarterly Program Review (QPR)	Ad hoc	Program office update on the status to leadership and decision-makers. Not all QPRs have TRL data.
	Cost Analysis and Requirement Document (CARD)	6.2A, 6.3, 6.4, and 6.5	The program office fills out all relevant program details to complete an ICE
Schedule	NNSA Integrated Master Schedule (NIMS)	Annually	The combination of site schedules
	Site Schedules	Annually	Each sites separately managed schedule

As part of the WDCR and BCR process the Product Realization Teams, the integrated teams charged with bringing components to design and manufacturing maturity, provide a writeup of scope and assessment of all major components. The WDCR also contains tables for the current TRL and MRL level for all major components outlined a forecast of when each future level should be achieved.

Since most WDCRs and BCRs provide this forecast by fiscal quarters, we now track all TRLs by quarter from all data sources. In the few cases where the data was provided on an annualized basis, we assumed that TRLs would be achieved the very last quarter of the year.

Our second major data source is NNSA Integrated Master Schedule (NIMS) and Site Schedules for each design agency. We identify TRLs within this data from the activity name and convert actuals and projections into quarters to sync with the WDCR and BCR data.

Our third data source quarterly program reviews (QPRs). These reports will occasionally include information on TRL status within descriptions of accomplishments or risks. Our fourth data source is the CARD, within which we require each program to submit TRL and MRL history and projections at each major milestone. The CARD provides the most detail and flexible format for our TRL analysis and will be our most robust data format going forward, though unfortunately prior program submissions were not in this format.

The resulting TRL database contains approximately fifty thousand datapoints across four hundred individually identifiable components. We spend significant time syncing the databases to remove duplicates and improve consistency between submissions. When uncertain of whether two entries align, we treat them as two separate entries. We then create a pivot table with each row being a unique component and each column being a unique fiscal quarter: the value is the maximum of the TRLs in that quarter. For all our analysis described below, we use TRL actuals, but do comparisons and cross checks using projections. Table 2 below shows a sanitized example of the formatted source data, with both actuals and projections.

Table 2: Sample TRL Source Table

Program	Component	Source	Source FY	TRL	TRL FY and Quarter
CW 1	Motor	Schedule	2033	3	2030Q2
CW 1	Pit	WDCR	2030	7	2033Q4
CW 1	Radar	Schedule	2033	4	2030Q1
CW 1	Test Equip.	Schedule	2033	3	2031Q1
CW 1	Motor	Schedule	2033	4	2032Q3
CW 1	Test Equip.	Schedule	2033	4	2031Q4
CW 1	Electronics	WDCR	2030	4	2035Q4

Raw Data Challenges and Limitations

Working with the raw data poses a few critical challenges. The main one is mapping across different units of time. Our data reflects differing timeframes – from daily (schedule) to annual (by fiscal year). The authors decided to consider all data in terms of quarter. To ensure traceability and appropriate documentation, we noted fidelity and original time units were maintained in separate raw columns. Everything was then translated to cleaned columns tracking quarters and fiscal year.

Another challenge was mapping components across time and documents. Often small changes in spelling, the use of acronyms, and changes in the underlying system design posed a challenge in tracking components. The authors compared TRL progressions and used acronym tables from other documents to attempt to map as many components as possible across sources. We maintained raw inputs for traceability. Where there was uncertainty in mapping, we left raw inputs as is. This results in a few cases where the authors believe we may be double counting components or are unable to track components.

Additionally, component data often contain gaps in development. For example, a component might track that it reached TRL 4 and not track again until TRL 6. The authors found a work-around to ensure that we would have use of all the data we could use; this method will be detailed below.

Data Analysis

Using the raw data sample shown back in Table 2 we report normalized dates for each TRL accomplishment to the number of quarters from the start of the database. In our sanitized data above, we set the start of the database as 2028Q1. From this sample we can see that the motor hit TRL 2 in the ninth quarter (of the database). With all dates normalized to the same time, we can create a timeline for documents and TRLs across both programs and the enterprise.

Each value in this table represents the time it takes that component to achieve that TRL from the previous TRL. We took these data points and looked at the mean time to get from one TRL, the standard deviation, and the number of components that we have data for moving from one TRL to the next. While we track over 400 total components, most components only show a few TRLs of progression or have missing dates for some TRLs; it was difficult to find statistically significant data for many TRL steps. An example of the raw data is shown in Table 3.

Table 3: Raw data with no processing (Quarters from database start)

TRL	1	2	3	4	5	6	7	8	9
Warhead Case			5						
Electronics		12		31					
Test Equipment			12	15					
Motor		9	18	22					
Radar			7		9				
Env. Controls				11		19			
Pit						9	23		

Because the authors were interested in the time it takes a component to complete all 9 TRLs, we tested (and ultimately used) a "filled" raw table of data to complete gaps in known TRL accomplishments. For example, if a component completed TRL 4 in quarter 24 and TRL 6 in quarter 31, we filled in quarter 28for TRL 5; this is the average of the two known times for the surrounding TRLs. For fills where the result would contain a decimal, we rounded to an integer. Table 4 builds on the prior example to show example fill operations.

Table 4: Data with filled values (Quarters from database start)

TRL	1	2	3	4	5	6	7	8	9
Warhead Case			5						
Electronics		12	22	31					
Test Equipment			12	15					
Motor		9	18	22					
Radar			7	8	9				
Env. Controls				11	15	19			
Pit						9	23		

Using this filled raw data table, we recreated the same analysis we had from the unfilled original raw data. We compared the timelines and, while there were some changes (the magnitude of less than a quarter across the board), we determined this change was acceptable as the total time to get from any TRL to TRL 9 did not vary by more than half a quarter. The authors are ultimately interested in predicting the time it takes to get from its initial TRL to the final TRL. For analysts looking at unique TRL completions, it is recommended to use the unfilled original data, which the authors often use as a cross-check on final results.

Treating the data this way gave us statistically large data sets for all but our first 3 TRLs. Table 5 shows the resulting time to complete each TRL.

Table 5: Time to complete TRLs (Quarters from prior TRL)

TRL	1	2	3	4	5	6	7	8	9
Warhead Case									
Electronics			10	9					
Test Equipment				3					
Motor			9	4					
Radar				1	1				
Env. Controls					4	4			
Pit							14		

Results

Analysis of Historical

After testing results from the formatted data, the authors decided that TRL 9 was not the right end level for analysis in the NNSA; because the NNSA defines component TRL 9 as demonstrated performance one year after system FPU, all components typically share the same TRL 9. This artificially lengthens the time from TRL 8 to TRL 9 for components that had an easier maturation process. The authors thus decided to use TRL 8, component FPU, as the final TRL. Once all components have reached their individual FPUs the system can formally reach its FPU (as long as all production engineering and certification requirements have been met).

Once we had distributions for every TRL, we created a set of combined TRLs. The authors ultimately created two “sets” of distributions. The first set of distributions covered the amount of time it takes each TRL to get to TRL 8. These values were created by summing the mean for each individual level starting from the level of interest. For example, to determine the time required from TRL 4 to 8 we would sum the means starting at the time from TRL 4 to 5, 5 to 6, etc., through TRL 7 to 8. The standard deviation for this same TRL progression is found by summing the variances (the square of the standard deviation) and taking the square root of the sum. This is a standard method for combining standard deviations.

Based on this data, we generated the bar graph in Figure 5. This graph shows the TRL maturation timelines to get from each TRL to TRL 8 using the extrapolated normal distributions from the historical data. For complex programs, lower TRLs can drive a high schedule variance, as schedules will be driven by the worst performers. A program might be able to deliver 80 percent of components on time, but the schedule slip will be driven by the last 5 to 20 percent of "outlier" components depending on the program's ability to adjust scope.

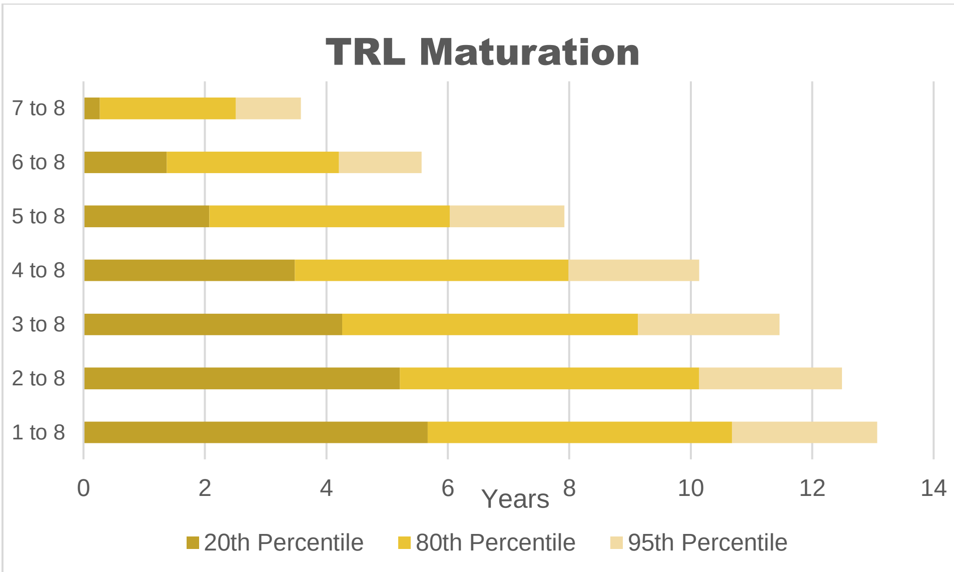

Figure 5: Percentiles from the Extrapolated Normal Distributions

Understanding Current Programs

For this exercise we have used generic data and warhead names. For ease of use, we have separated our warheads into three categories: completed warhead (CW), in progress warhead (IPW), and new warhead (NW). Additionally, we have removed the scaling from our chart as the raw data and results are not approved for public release.

The pie charts presented in Figure 6, Figure 7 and Figure 8 display the historical TRLs for the CW-1 and IPW-1 at the 6.3 Development Engineering milestone and where we forecast the NW-1 to be at the 6.3 milestone. We adjusted the data to show the type of results of our analysis without showing data tied to any actual program.

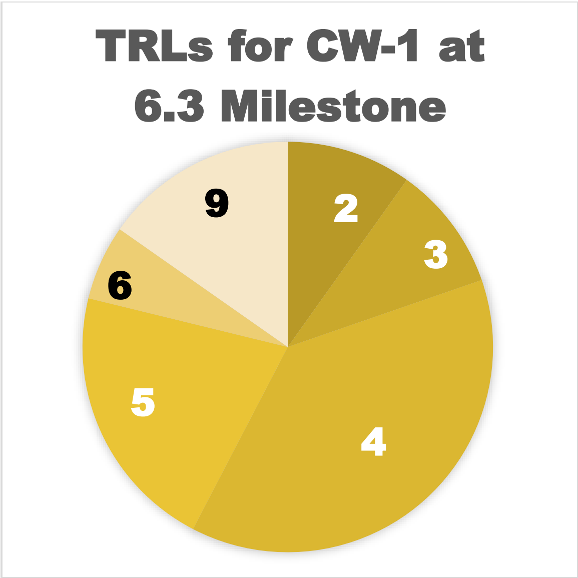

Figure 6: TRL Distributions for CW-1 at 6.3 Milestone

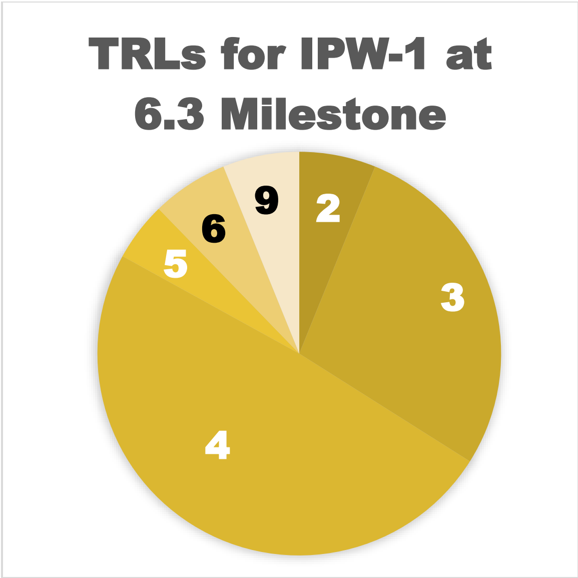

Figure 7: TRL Distributions for IPW-1 at 6.3 Milestone

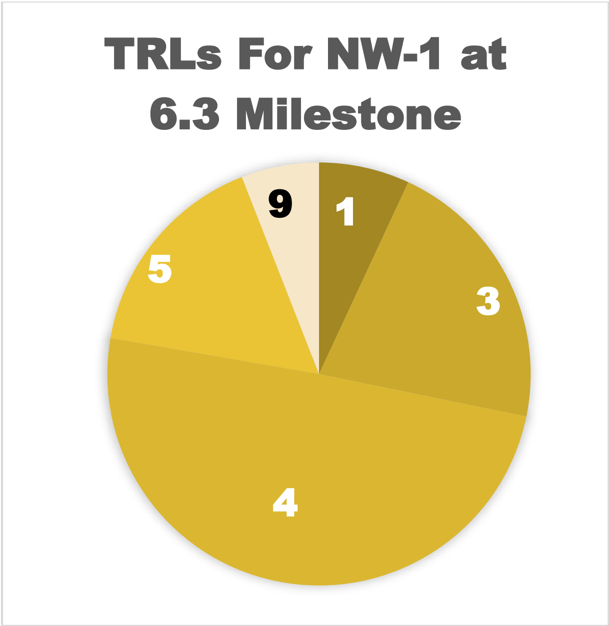

Figure 8: TRL Distributions for NW-1 at 6.3 Milestone

The pie charts show another interesting trend — upcoming programs are less technologically mature than predecessor programs. This may be due to more rigorous

TRL rubrics and more accurate assessments of existing technologies or could be an indicator of more ambitious program scopes.

Application to New Programs

After our first run with the data, we wanted to see if the forecast aligned with actuals for similar programs. This was particularly important as we did not have a complete view of the development process for our historical programs, primarily relying upon two programs still completing their development cycles. In addition, our target program was technologically less mature than its predecessors. As discussed above, internal program policy recommends all components meet TRL 4 at the start of Development Engineering (Phase 6.3) and TRL 6 at the start of Production Engineering (Phase 6.4). As noted previously, TRL 8 is widely considered component FPU and internal NNSA guidance recommends TRL 8 for all components at the start of the FPU phase.

We artificially recreated these milestones by assuming that 80% of components needed to achieve TRL 6 to progress into Production Engineering and 80% of components needed to complete TRL 8 to start the formal FPU phase; this reflects practice where major milestones progress while some components might be completing a TRL or awaiting paperwork before proceeding. The end of FPU was marked as when 95% of components had achieved TRL 8; this gives leeway for the program to resolve its most difficult component maturations via scope change or surge support.

A second item of note is that over 50% of components achieved TRL 6 two years prior to the Production Engineering milestone (Phase 6.4). By this metric that two-year period should really be considered a transition as various components and subassemblies transition from Development Engineering to Production Engineering at different times.

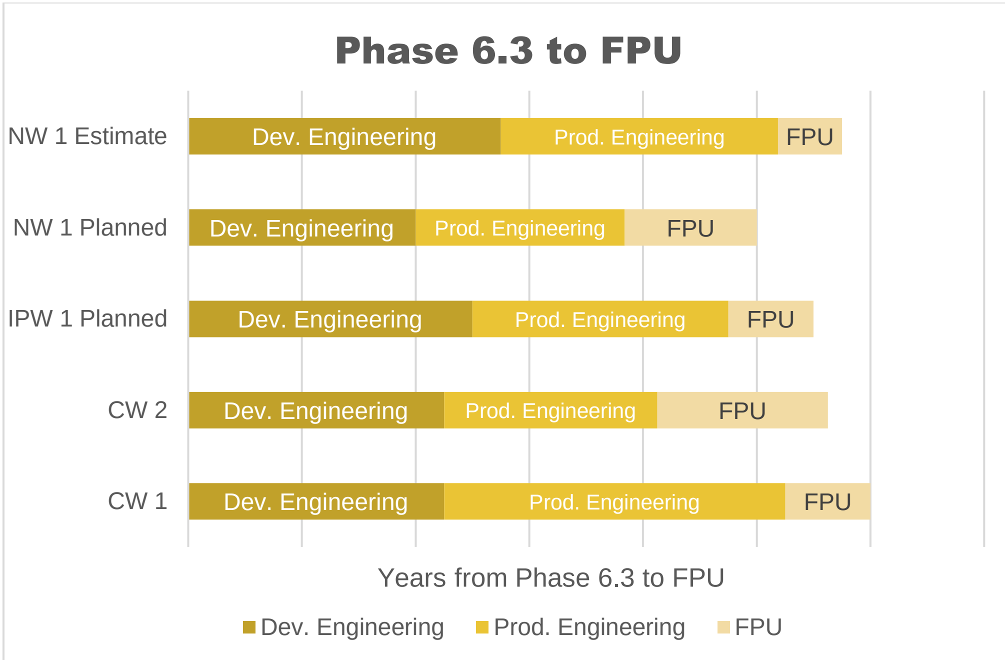

Figure 9: Extrapolated Milestone Durations

Figure 9 demonstrates the high-level milestones of a program and compares the approximate time in each phase to the time predicted our TRL methodology with the previously described cut-offs. We can see that one of the reasons our Development Engineering estimate is higher for NW-1 is that it started out with a lower maturity than predecessor programs (revisit pie charts in Figures 6, 7 and 8).

The NW-1 Estimate bar shows our prediction using this TRL methodology. The NW-1 Planned and the IPW-1 Planned show the current timeline that is the program of record. The CW-1 and the CW-2 display the actuals achieved by previous programs.

By aligning our development engineering and production engineering phases, we can cross check our results again historical programs and easily communicate our results to decision-makers. This allows the analyst to communicate whether the program is cutting short design or manufacturing time, while concisely explaining the impact to FPU.

Additional Areas Explored

The authors also explored several other system metrics, with success to-date than the TRL analysis. The GAO notes that "other readiness level measures, for example [MRLs], have been proposed with varying degrees of success and use throughout the life-cycle of a program." [6] The NNSA uses MRLs in conjunction with TRLs. In our early work, we performed the same analysis using MRLs. An interesting finding was the MRLs did not drive program timeline. We could not reconcile components with mature MRLs that were continually demonstrating manufacturing and producibility challenges. Further discussion with the production sites revealed that they often tied MRLs to milestones of the program whether or not a component was producible. Additionally, although these sites track producibility as a separate metric, this metric can also miss key problems with yield and repeatability. The authors hope to be able to explore other metrics like MRLs and producibility in the future.

Conclusion

Summary of Findings

The TRL database serves as a useful tool to: a) understand the schedule risk that can be driven into a program due to lower technology maturation in lower-level components and b) compare program health at various points in the program with predecessor programs. This allows a more holistic view of the program and its success outside of milestones, task burn rates, labor, and cost performance metrics. While not suitable for use as the only metric, TRLs provide a method for decisionmakers and analysts to view and better understand the technical risks of a program at a lower level. It also shows the schedule impact of pushing programs forward with less-mature technologies.

The resulting schedule analysis allows analysts to identify reasonable schedules early on in the program before system schedules, critical path, or other schedules can be established. This analysis can act as an early warning on programs that carry significant technical risk in the form of less mature technologies.

As the NNSA implements more rigorous TRL assessment standards across the weapons programs, we will pursue potential adjustments to the historical data or additional context for the future data. As of right now, the authors believe that the improvements that Defense Programs have made to the rubric will provide a long term stability to this analysis. In the short term, this change will require thoughtful approaches to normalize historical and incoming data.

Limitations

The small size of the current database limited the type of analysis the authors could perform. The database does not provide data by component category (mechanical, electrical, hazardous, software, etc.) and includes limited data by design and production agency. Although the authors can currently map the components to various categories (using a separate component/part database for that exercise), we do not have statistically large numbers to support analysis by component category.

Due to the nature, age, and size of our data, we must also note that this analysis skews to a shorter maturation timeline. Our first use of this data relied on TRL data from two programs nearing completion and one program midway through development. As a result, the data encompasses components that have successfully completed various TRLs. The components that have reached TRL 7 or TRL 8 are naturally going to bias towards the components that might have been easy, less risky, or more mature to at inception. Our data will miss the components that caused the program to slip schedule.

Furthermore, as the dataset has grown, we still have a small handful of components that haven't completed TRL 9 in the early programs. This is less a concern as we now measure to TRL 8 rather than TRL 9. Additionally, as we add data from the program that was mid-way through development and a new program just starting development, we will again start to skew towards seeing numbers from components that were successful. As our dataset grows larger (and we consistently receive more granular data from future CARDs), these concerns will be less prominent.

Future Work

Going forward, the authors hope to expand the analysis to increase fidelity, confidence, and scope of use. Our success will depend on collecting more data more often. For weapons systems, our current data collection policy requires periodic collection of NIMS and site schedules as well as QPRs. These sources will b1e used to populate and update the TRL database regularly. Additionally, when a program reaches a major milestone, the most recent CARD will provide more detailed updates and projections, and our work to require more detailed TRL data will be invaluable for future work.

Once we have more data, we hope to examine the risk profiles of different design and production agencies. Additionally, we will review the risk by component type (i.e., electrical, mechanical, structural). This insight will help provide an even better understanding of risk to the program driven by unique component decisions. Until we can break out this level of fidelity, we have made the assumption that the risk profile of different sites and component types is the same.

As we get more comfortable with the data, we plan to explore correlation or interdependencies among components or a hierarchy between components and subsystems. If a set of components are waiting for the same system-level milestone to meet their TRL goals, individual components may show different TRL progressions that don’t represent a true difference in maturity.

Additionally, the authors hope to expand the depth of our current TRL analysis by incorporating more information on component type (electrical, mechanical, etc.) and subsystem information into our database. This will allow us to perform more precise analysis, looking at the risk driven by component type and whether the component is a part of a subsystem that cannot mature without its constituent components having completed development.

For capital acquisition projects and other programs, our office has attempted to find a central source that reports TRLs in a consistent manner. We anticipate a real challenge in managing TRLs across various interfaces and organizations to ensure reporting consistency. Regardless, collecting the data to better understand the schedules of capital acquisition projects and other programs is essential work that will continue.

Finally, as mentioned previously, our research with MRLs produced unsatisfying results. Going forward, our office hopes to be able to understand these limitations and determine an effective metric for measuring the progress of manufacturing process maturation. We will also aim to influence improvements to the usage and definitions of MRLs.

References

[1] Fahimian, Mahi; Behdinan, Kamran. On Characterization of Technology Readiness Level Coefficients for Design. DS 87-2 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 2: Design Processes, Design Organisation and Management. 2009. Vancouver, Canada, 21-25.08.2017. https://www.designsociety.org/publication/39585/On+characterization+of+technology+readiness+level+coefficients+for+design

[2] Mankins, John C. Technology Readiness Assessments: A Retrospective. Acta Astronautica, Volume 65, Issues 9-10, 2009, Pages 1216-1223, ISSN 0094-5765, https://doi.org/10.1016/j.actaastro.2009.03.058.

[3] Mankins, John C. Technology readiness and risk assessments: A new approach. Acta Astronautica, Volume 65, Issues 9–10, 2009, Pages 1208-1215, ISSN 0094-5765, https://doi.org/10.1016/j.actaastro.2009.03.059.

[4] Nakamura, Hiroko. Kajikawa, Yuya. Suzuki, Shinji. Multi-level Perspectives with Technology Readiness Measures for Aviation Innovations. Springer. 15 September 2012. https://link.springer.com/content/pdf/10.1007/s11625-012-0187-z.pdf

[5] Sauser, Brian. Remirez-Marquez, Jose E. Magnaye, Romulo. Tan, Weiping. A Systems Approach to Expanding the Technology Readiness Level within Defense Acquisition. Stevens Institute of Technology. DTIC. 20 March 2009. https://apps.dtic.mil/sti/pdfs/ADA530242.pdf

[6] US Government Accountability Office. Technology Readiness Assessment Guide. January 2020. GAO-20-48G. https://www.gao.gov/assets/gao-20-48g.pdf

Acronyms

BCR	Baseline Cost Report
CARD	Cost Analysis Requirements Description
CW	Completed Warhead
FPU	First Production Unit
GAO	Government Accountability Office
IPW	In Progress Warhead
MRL	Manufacturing Readiness Level
NASA	National Aeronautics and Space Administration
NIMS	NNSA Integrated Master Schedule
NNSA	National Nuclear Security Administration
NW	New Warhead
QPR	Quarterly Program Review
TRA	Technology Readiness Assessment
TRRA	Technology Readiness and Risk Assessment
TRL	Technology Readiness Level
WDCR	Weapon Design Cost Report