ADAPTIVE INDEPENDENT VERIFICATION AND VALIDATION (IV&V) REDUCES RISK OF SOFTWARE IMPACTING SAFETY IN ARTEMIS MISSIONS

Gerek A. Whitman[(1)], Ryan P. Starn[(2)]

[(1)]SAIC, 100 University Dr, Fairmont, WV 26554, USA, Email: Gerek.A.Whitman@nasa.gov
[(2)] SAIC, 100 University Dr, Fairmont, WV 26554, USA, Email: Ryan.P.Starn @nasa.gov

ABSTRACT

The National Aeronautics and Space Administration (NASA) is asking more of its human spaceflight programs than ever before through the collective Artemis Missions. The NASA Independent Verification and Validation (IV&V) Program contributes to NASA's human spaceflight goals by providing IV&V services for NASA's critical spacecraft and ground software. The IV&V Program is tasked with providing assurance from both individual and integrated mission software perspectives. The Artemis IV&V organization is actively supporting six distinct development efforts: Orion, the Space Launch System (SLS), Exploration Ground Systems (EGS), Mission Control Center (MCC), the Lunar Gateway, and the Human Landing System (HLS), representing a wide diversity of developer organizations, management structures, and development approaches. With much of this extremely complex flight and ground software being essential to human safety both on the ground and in space, Artemis IV&V is likewise challenged to provide more value-added assurance to future Artemis missions within a constrained budget.

To meet this challenge, Artemis IV&V employs a variety of novel and evolving "Adaptive IV&V" approaches for planning and executing IV&V analysis to increase both the efficiency and effectiveness of the IV&V Program's assurance activities, and to address the difficulties imposed by assuring software for a large, highly integrated, multi-mission enterprise managed and executed by physically and organizationally distinct programs.

Instilling agile principles like iterative planning cycles, self-organizing teams, and regular retrospectives, into IV&V planning and execution has led to a more rapid turnaround of a minimum viable assurance product and allowed for increased alignment of assurance activities with development progress. Adopting an assurance case methodology has led to greater consistency and clearer communication of assurance design and provided a foundation for long-term maintenance of assurance plans, products, and results across missions. The IV&V-developed Assurance / Safety Case Analytical Network (A-SCAN) framework and tool has enabled the quantification and tracking of system/software risk and confidence. These confidence measures provide a means to repeatedly express the impact of planned and completed assurance work and the remaining residual risk. Applied as part of a "Follow-the-Risk" organizational ethos, this allows consistent rightsizing of analysis rigor and intensity commensurate with the perceived risk of defects, as well as appropriate targeting of the highest risk areas of the software to find safety issues before they can manifest. Finally, the development of the IV&V Advanced Risk Reduction Integrated Software Test and Operations Tri-program Lightweight Environment (ARRISTOTLE), an integrated software-only simulation of Orion, SLS, and EGS systems, has made it possible to independently test integrated pad and flight scenarios and inject faults to observe how the Artemis multi-program, mission software behaves in degraded modes and in response to hazards.

These adaptive IV&V investments have enabled Artemis IV&V to become more efficient and effective in IV&V planning and execution and respond more readily to changes in the risk landscape, increasing the breadth and depth of risk reduction possible within the available resources. Residual risk tracking allows IV&V to communicate more effectively with stakeholders, both internal and external at all levels, and inform key decision-making personnel. This evolving assurance design approach provides IV&V surety that work is performed in the highest risk, most value-added areas of the software, to keep our astronauts and ground crews safe and ensure mission success.

1. INTRODUCTION

With the successful completion of the uncrewed Artemis I flight, NASA's Artemis Program is on track to send humans back to the moon for the first time since Apollo XVII in 1972. Artemis II, the first crewed flight of the Orion crew vehicle and the Space Launch System (SLS), is scheduled for launch at the end of 2024, with the subsequent lunar landing of Artemis III scheduled for the following year. Each Artemis mission requires expanded capabilities in flight and on the ground, as well as the involvement of many NASA development and operational programs. These include the Orion program, the SLS program, the Exploration Ground Systems (EGS) Program, the Mission Control Center (MCC), the Human Landing System (HLS) Program, which will land astronauts, cargo, and equipment on the Moon, and, as the Artemis Program continues toward

sustainment and long-term presence on the Moon, the Gateway in lunar orbit and additional lunar surface infrastructure.

Software plays an increasingly important role in the success of these complex and ambitious space systems. EGS software is essential in preparing the SLS launch vehicle and Orion crew module on the ground. SLS software is crucial onboard the launch vehicle itself, controlling where the vehicle is flying. Orion's software commands, manages, and tracks vehicle capabilities during its journey to the Moon and back, including autonomous time-critical mission events during ascent and re-entry. Software will be crucial in carrying humans, cargo, and equipment to and from the extreme conditions of the lunar south pole. Software is used to control mission modes that will transition between autonomous, automatic, human tended, and human directed software and system operations, while addressing changes in operating environments, physical configurations, communication modes, paths, bandwidths, and latencies, all while subjected to complex potential faults.

The focus of NASA's Independent Verification and Validation (IV&V) Program is on assuring this safety- and mission-critical software. NASA's IV&V Program, located at the Katherine Johnson IV&V Facility in Fairmont, West Virginia, was established in 1993 as a direct result of recommendations made by the National Research Council and the Report of the Presidential Commission on the Space Shuttle Challenger Accident. Since its inception, the IV&V Program has been contributing to the safety and success of NASA's highest-profile missions in human spaceflight, robotic exploration, and earth science by assuring the software on those missions performs correctly. The IV&V Program falls administratively under NASA's Goddard Space Flight Center and operates under functional guidance from the Office of Safety and Mission Assurance.

IV&V has spent years, and in some instances, decades gaining system understanding in the extremely complex software across many of the platforms involved in the Artemis Program. In that time, the IV&V teams supporting the Artemis Program have understood the risk associated with all the critical mission capabilities that will ultimately make these missions successful. IV&V support is required by NASA on Orion, SLS, EGS, MCC, HLS, and Gateway, and has been executed throughout planning and development for Artemis I and will continue for future Artemis missions. To assist with coordinating IV&V support across all these inter-related software development efforts, we have established an Artemis IV&V organization that consists of approximately one hundred IV&V personnel who are responsible for adding assurance for the software that executes the highest risk mission capabilities within EGS, SLS, Orion, Gateway, HLS, and MCC. These Artemis IV&V project teams interface with each of the development programs and remain in-synch with development, focusing on how all the capabilities come together and make the mission a success.

With the increasing number of concurrent development programs and their complexity, there is a greater demand for quality software assurance. Like any organization, the resources of the IV&V Program are not infinite and must be justified as part of a value for cost analysis. It is necessary to find ways to increase confidence and decrease risk within a constrained resource pool by focusing on the most significant threats. To achieve the Artemis Program's aggressive schedule for putting the first woman and first person of color on the Moon within a few years, the agency will have to accept some risk. The Artemis IV&V organization's role is to help identify that risk throughout the development of these missions, as well as add assurance that the safety- and mission-critical software will do what it is supposed to do, not do what it is not supposed to do, and respond appropriately under adverse conditions. Fundamentally, we strive to find the high-impact software defects before they manifest so that NASA can keep our astronauts safe.

2. NASA IV&V ASSURANCE STRATEGY AND ASSURANCE DESIGN

NASA IV&V's assurance strategy has been honed over the lifetime of the IV&V Program to produce efficient and high-quality results. This strategy continues to evolve in response to new challenges. The NASA IV&V Program's approach to software assurance spans the software development lifecycle, from concept and requirements all the way through verification and validation, and into operations. IV&V defines assurance as the assertion and substantiation of positive declarations which give confidence. At its core, IV&V's assurance strategy is focused on what we call the Three Questions (3Qs):

Q1: Does the software do what it is supposed to do?
Q2: Does the software not do what it is not supposed to do?
Q3: Does the software respond appropriately to adverse conditions?

NASA IV&V is complimentary to other safety and quality assurance organizations focused on assuring adherence to standards. NASA IV&V performs practical, evidence-based analysis on software artifacts throughout the development lifecycle to assure the proper operation of the software and its capabilities in its expected environment, thus providing positive assurance, or confidence in the software and systems, as well as delivering Technical Issue Memorandums (TIMs) for any defects in software artifacts, and Risks

when there is an unresolved possibility for the mission software to not meet expectations.

2.1. Assurance Design in the IV&V Project Lifecyle

Fig. 1 depicts the lifecycle of an IV&V Project from inception to completion. The process starts with a period of building mission and system understanding, then using that accumulated knowledge in the form of a Technical Reference to begin risk assessments. Risk Considerations in the form of risk-based assessments of the mission, system, and software-level capabilities inform IV&V about the inherent risk in the systems that comprise the mission and help to better frame and focus IV&V effort to reduce those risks. Completion of risk assessments marks the beginning of the assurance design portion of the process, during which Assurance Objectives (AOs) are defined and decomposed.
Assurance Objectives represent targeted statements of what claims IV&V would like to make when analysis is complete, and are informed by our understanding of the mission, system, software, and inherent risk. A typical AO is posed and broken down into manageable and achievable portions attributed to software lifecycle phases and the respective IV&V Technical Framework (TF) objectives [1] using the following pattern.

> AO: Provide assurance that the thermal control system will prevent runaway heaters from damaging sensitive instruments.
>
> > Sub-AO1: Provide assurance that the thermal design has thermostatic control or runaway heating is below thermal limits.
> >
> > Sub-AO2: Provide assurance that the thermal system testing verifies the detection, isolation, and recovery from runaway heating.

IV&V starts with high level AOs for mission and system capabilities, but over time elaborates them in order to identify and target specific software capabilities for focused analysis. The subsequent AOs are defined and refined at a sufficient level of abstraction to support analysis of specific software artifacts. Assurance design continues with the identification of appropriate analysis tasks to apply the necessary rigor to address the identified risk. These analysis plans are informed by and dependent on staffing, schedule, and budget considerations. For example, the highest priority planned analysis may not be immediately completable because mature artifacts are not yet available to support analysis.
Eventually, as analysis is conducted and completed, IV&V accumulates evidence and results that contribute to the veracity of the initial AO. IV&V uses these results to generate an Assurance Conclusion (AC) that summarizes the evidence and notable findings, and articulates the confidence IV&V has in the corresponding capability.

2.2. Capability Based Assurance (CBA)

All software executes in support of system and mission capabilities. Sometimes, the role of software is essential for and inseparable from a capability, while other times, software may provide a supporting, secondary, or recovery role, but software always serves the actualization of capabilities within a larger context. Therefore, IV&V assurance design follows a Capability Based Assurance (CBA) approach. In CBA, the mission, system, and software capabilities and their identified risks are used as the basis for planning what analysis activities are necessary to satisfy an AO, while the IV&V TF objectives, IV&V analysis

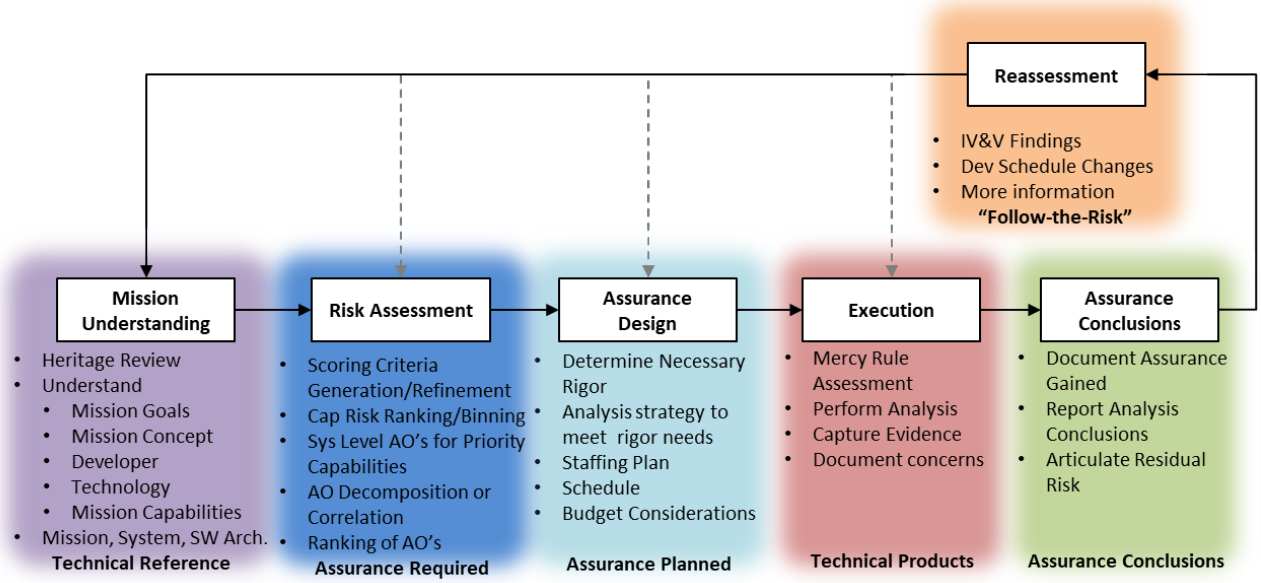

Figure 1. The IV&V Project Lifecycle

methodologies, and the available software artifacts are used as inputs to determine how this analysis should be conducted.

IV&V budgetary constraints coupled with growing software obligation on missions requires a top-down scoping and focusing scheme. In CBA, analysis often targets only a part of any software component, and often cuts across software components, as the goal is to assure the success of the capability, which may be realized by piecemeal contributions from various software elements. This creates additional challenges for IV&V to understand and track the inclusion and exclusion of software components, and to better understand how analyses from certain components contributes to or rolls up to parent capabilities without requiring separate, often redundant analyses. IV&V uses the results of these analyses to draw conclusions of the software's ability to meet particular mission objectives, and evolve our understanding of the software, system, and mission risk to further sharpen the assurance design using a follow-the-risk approach.

2.3. Follow-the-Risk (FTR)

Follow-the-Risk (FTR) is the approach by which IV&V understands, identifies, and prioritizes areas of risk within the projects' capabilities and software continuously, to focus effort in the areas of highest risk. The goal of FTR is to reduce residual risk efficiently and effectively across the entire risk landscape by prioritizing and addressing the most significant risk areas first. Rather than eliminating all identified risk in any one single capability or domain of the mission, a FTR execution strategy addresses and reduces risk in targeted areas, and then moves analysis effort to other high-risk areas. This allows IV&V to strike a balance between addressing critical software risk while also getting the best return on our investment. It is not possible to eliminate all software risk to any particular capability; there are always unknowns. Past a certain point, continuing to reduce risk in a specific area or project requires ever more resources to understand, identify, and assess that which has not already been understood, and diminishing returns in the quality, cogency, and/or value of the analysis results compared to other activities that could provide greater impacts to mission success. FTR posits that there is an acceptable level of residual risk, and when that is reached, further efforts are better spent buying down risk to acceptable levels in other areas. The existence of residual risk is driven by economic/programmatic realities (schedule, cost, and mission hazards), limitations of the systems analyzed, and the role of software within the system. It may not always be possible to reduce risk to the acceptable level.

2.4. Adaptive IV&V

Part of a successful FTR approach requires a continual assessment of risk as new information is discovered. IV&V assesses risk against system and software capabilities before ever performing analysis, to identify the most promising high-value assurance targets, and to identify the characteristics of that risk and how to best reduce the risk. Sometimes, those initial focus areas turn out to have less risk than was originally anticipated, or the strategy planned is not achievable given development maturity or other unforeseen hurdles. When this happens, rather than continue with the initial plan, IV&V applies Adaptive IV&V to change the plan in response to this new information and think critically about the correct focus and approach. We might reduce or defer our effort in that capability and apply it toward different analysis strategies, or shift effort entirely to different assurance targets. The inverse can also occur; we may learn new information about a capability which was not initially in focus because of previously perceived low risk. Upon gaining new insight that causes us to re-evaluate, we may elevate our assessment of the risk on that capability, and subsequently plan and execute some analysis to address that new risk.

Adaptive IV&V applies to changes in risk prioritization as well as analysis execution. If the analysis being performed is not producing or cannot produce the necessary evidence to support the AO, or the primary artifacts under assessment do not contain the anticipated information, the Adaptive IV&V approach identifies a need to change. IV&V analysis leverages our Technical Framework and analysis methodologies to plan and execute analysis efficiently and effectively. If an existing analysis approach is insufficient in producing evidence toward an AO, IV&V analysts are empowered and encouraged to develop or adapt new methods and techniques to obtain the necessary evidence. These new techniques are subsequently refined and shared across the IV&V Program so that others can take advantage of these novel methods when they run up against similar challenges. Over time, useful approaches to assurance strategy have been consolidated and formalized into "Threads," or patterns of techniques and methods that more effectively derive evidence for different types and levels of risk.

2.5. Assurance - Safety Case Analytical Network (A-SCAN)

IV&V has evolved its approach to assurance design by employing assurance case concepts and methodologies. An assurance case is a reasoned and compelling argument, supported by a body of evidence, that a system, service, or organization will operate as intended for a defined application in a defined environment. [2] IV&V's AOs are essentially the same as claims in an assurance case, and they are similarly related

hierarchically, in what IV&V refers to as "Assurance Networks", in which AOs for mission and system capabilities decompose down to AOs for software capabilities, which are supported by the evidence accumulated through analysis, much like how an assurance case is a hierarchical network of claims that are elaborated until they can be directly addressed by solutions. Fig. 2 shows how all the assurance design concepts covered so far are interrelated and lead to evidence and results, and demonstrates the similarities to the claims, strategies, and solutions one would find within an assurance case.

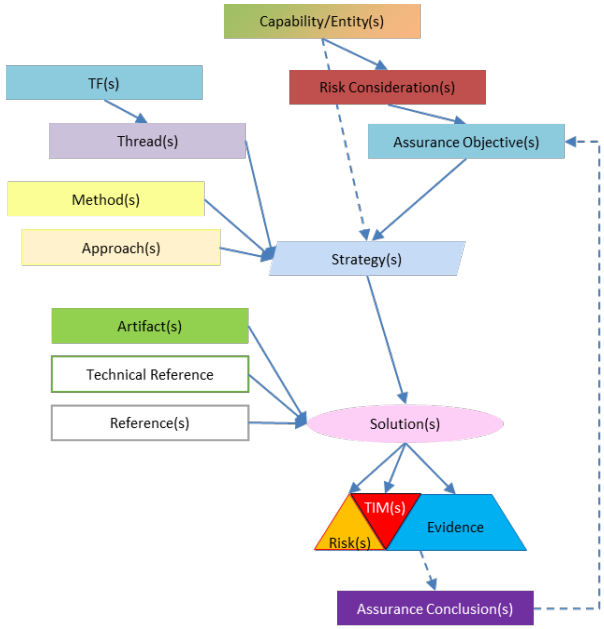

Figure 2: IV&V Assurance Design Concepts

The Assurance – Safety Case Analytical Network (A-SCAN) is a set of principles that define an approach to assurance design and assurance strategy, as well as a functional tool that implements those principles. A-SCAN is built upon assurance case fundamentals, and produces quantitative metrics to help understand risk, confidence, intensity, and rigor based upon the Dempster-Shafer theory of evidence. In short, Dempster-Shafer theory posits that for any claim, there exists a mass of belief which is a summation of the mass of belief for its subclaims. The belief is based on evidence supporting the claim, and limited by doubt, or disbelief, and uncertainty. [3] Following this theory, confidence, or belief, in a claim, or especially a network of claims, like an assurance case, can be modeled and expressed quantitatively. A-SCAN allows for calculating and expressing software assurance in terms of this confidence.

It is important to note that, in our usage of the A-SCAN framework, confidence should not be understood to be equivalent to reliability. Confidence is not a guarantee, nor is it intended to represent an actual success or failure rate. Other domains may be able to model assurance based on reliability; however, software reliability is notoriously difficult to quantify and thus "confidence" in software success is more subjective. Rather, it is better to understand confidence as a measure of IV&V's contribution to software assurance through the elimination of unexplored threats to software success, based on the bounds of all possible reasonable IV&V activities.

To establish this quantitative model of IV&V confidence, A-SCAN requires acceptance of the following simplified theorems and assumptions.

1. There exists some "inherent confidence" that software and software development products will be correct, complete, and reliable to a certain degree without any external intervention or influence by IV&V. Inherent confidence is derived from the standards, policies, and procedures used by the developer, as well as the quality of the resulting products.
2. There exists some acceptable level of risk and thus a notion of "enough" confidence, referred to as "target confidence."
3. The assurance that IV&V generates for a system or mission is a function of the type, number, and rigor of the analyses performed, the maturity of the artifacts used in those analyses, and the quality of the evidence produced.
4. Each IV&V TF objective, when satisfied, provides some quantifiable degree of confidence, and the complete set of TF objectives represents the complete amount of assurance possible on any given IV&V effort.

IV&V analysis will increase total confidence that the system is correct, complete, and reliable, above the inherent confidence already derived from the development characteristics. The total confidence in any claim is a combination of the inherent confidence and the added IV&V confidence accumulated through analysis. A relative deficit between the target confidence and inherent confidence drives IV&V analyses; this is referred to as the "required confidence." In areas of sufficient inherent confidence such that there is no deficit between the target confidence and inherent confidence, IV&V analysis is not required. Perfect confidence, or zero risk, is unrealistic, and therefore 100% confidence is the asymptotic maximum target. Fig. 3 depicts the relationship between the constituent elements of confidence in terms of undefined functions of various inputs.

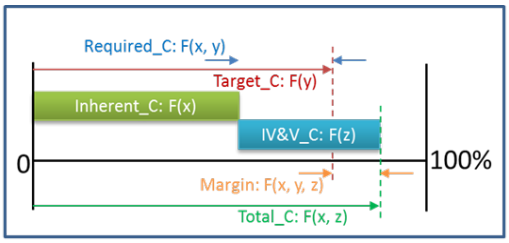

Figure 3: A-SCAN Quantitative Representation of Confidence

The factors impacting confidence are not static, and confidence will vary during the mission lifecycle. The dynamic nature of risk, confidence, and assurance are modeled such that IV&V can plan and achieve a level of IV&V confidence with sufficient and appropriate "Margin" to anticipate and account for the unknown unknowns.

2.5.1. Risk Considerations/Assessments

IV&V utilizes a risk assessment methodology to determine what portions of the mission software systems should be analyzed. One such method called System/Software Consequence, Obligation & Priority Evaluation (SCOPE) assesses risk based on three axes:

- Consequence – what is the worst-case scenario if this capability should fail?
- Software Obligations – how vital is the role of software in achieving, monitoring, or recovering this capability?
- Likelihood – how reasonable is it to suspect that software defects are present in the capability, or could be introduced?

Each of these axes are scored by considering various risk considerations and criteria, resulting in an aggregated risk assessment that provides a numerical score along each axis, as well as qualitative assessment rationale for developing AOs and targeted analyses to address the known risks. These quantitative risk assessment scores allow prioritization of capabilities by risk level and provide calculation of inherent confidence and target confidence thresholds. Inherent confidence is derived from the likelihood risk factors; for example, if increased complexity or inadequate developer practices suggest that defects are more likely, then our inherent confidence is lower. Likewise, target confidence is derived from a combination of the consequence and software obligation risk factors. If the worst-case scenario of failure is serious, like a loss of crew scenario, the target confidence is similarly elevated, and is higher in cases where there is significant software involvement in the capability. The difference, or required confidence, between the resulting target confidence and inherent confidence establishes the relative need for IV&V analysis to bridge the confidence gap. Fig. 4 provides a notional view of how we might prioritize our assurance targets based on the results of these risk assessments across separate 2D slices of a 3D software risk matrix.

2.5.2. Confidence Contribution from Solutions

Use of the A-SCAN framework continues throughout the execution of IV&V analysis, not just at the planning stage for determining priorities. As analyses are planned and conducted, those strategies and solutions can be modeled in A-SCAN. The coverage of individual TF objectives represents the breadth, or intensity of the analysis. A higher intensity analysis covers a greater number of Technical Framework objectives and therefore produces broader sets of distinct evidence, or, to return to the Dempster-Shafer theory of evidence, a wider array of individual beliefs to accumulate into the mass of belief ($b(x)$) or evidence from IV&V (e_{IVV}). The methods employed in the analysis, and the quality of the evidence, dictate the rigor (Rig_i). Methods of higher rigor produce more irrefutable evidence in support of their claims and objectives and leave less room for doubt and uncertainty. Thus, they provide a more

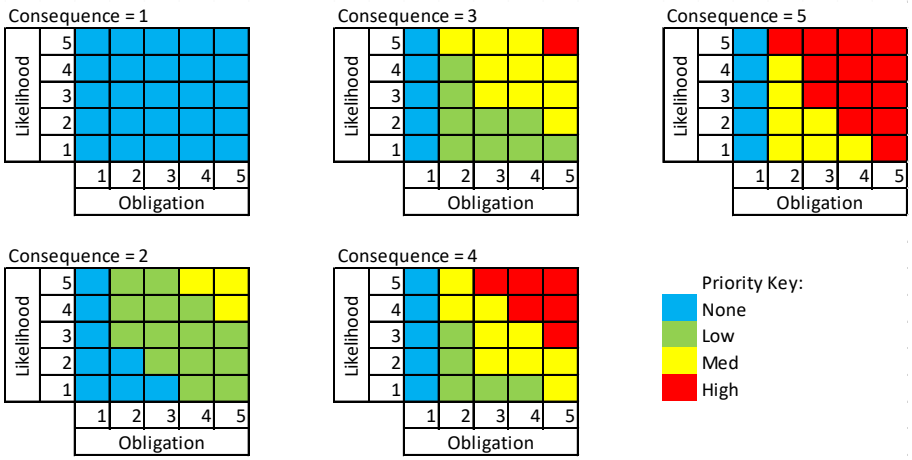

Figure 4. Notional Priority Levels at Each Consequence Score

complete contribution of confidence for the Technical Framework ($TFCC_i$) objectives those analyses address. Direct doubt and defeaters to belief/confidence can be represented by the presence of identified TIMs and scaled according to their number and severity represented by a TIM Scale Factor (TSF_i). Eqs. 1 and 2 demonstrate the relationship between these parameters in the form of the Dempster-Shafer theory of evidence.

$$\sum e_{IVV} = \sum (TFCC_i \times Rig_i) \quad (1)$$

$$b(x) = \sum (TFCC_i \times Rig_i) \times (1 - \sum TSF_i) \quad (2)$$

Confidence accumulated through IV&V analysis is therefore a function of both intensity and rigor. A-SCAN models accumulated confidence by assigning relative weights to each objective in the TF. Applying all the TF objectives represents everything IV&V can potentially achieve, maximizing the potential IV&V accumulated confidence. The relative weights of each TF objective vary, reflecting that not all objectives are equally beneficial in gaining confidence. As the analysis is planned and conducted, A-SCAN also models the rigor using factors based on the method, the quality of the evidence produced, and the quality of the artifacts inspected in the analysis. These rigor factors alter the total potential accumulated confidence from the TF coverage.

A-SCAN's confidence model can be used prior to any analysis being performed to help right-size the analysis effort to the level of risk. Using A-SCAN to input the TF objectives and planned methods for an upcoming analysis regimen indicates if that analysis might be insufficient to produce the necessary level of confidence, or overzealous in producing more confidence than necessary. In this way, A-SCAN helps avoid over-expending resources in areas when they would be better used elsewhere to address more serious risks.

A-SCAN's quantitative model is valuable in monitoring the assurance that IV&V gradually generates. As analysis progresses, the tool calculates the accumulated IV&V confidence, closing the gap between the inherent and target confidence for each interconnected node in the assurance case network. These metrics enable monitoring of assurance progress throughout the life of the IV&V project using consistent inputs and measures, allowing the team to gauge, at any point, how much confidence, or assurance, has been generated to date, and how much more is needed to reach the target level, across the entire risk landscape. This level of insight across an IV&V project empowers teams to make much more informed decisions about where to apply resources at any given time and adapt readily to changes in risk.

3. APPLICATION OF IV&V'S ASSURANCE STRATEGY ON ARTEMIS

Thus far, this paper has discussed NASA IV&V's approach to assurance strategy and assurance design in general, remarking upon concepts and practices that most, or even all IV&V project teams use in their approach to software assurance for the various missions supported by IV&V. The remainder of this paper will discuss in greater detail the IV&V approach for the Artemis missions specifically, and how the unique challenges posed by assuring Artemis software have shaped these adaptations.

Artemis presents unique and specific challenges for IV&V. Artemis missions involve many large and complex flight and ground systems that must integrate in evolving multi-program configurations. Each program is a system of systems with numerous software capabilities. Artemis is a long-term program with multiple missions planned, requiring comprehensible documentation and continued maintenance of assurance plans and results. Each Artemis IV&V project team interfaces with a different software development program, spanning many different contractors and NASA centers. Individual IV&V analyses, TIMs, risks, and program level confidence must be delivered to and understood by the associated program team. Our assurance products and results must also aggregate to produce a cohesive assurance message for each Artemis mission, involving all programs in concert. Assurance planning and execution must address these different targeted audiences, adequately assuring both the subordinate single program objectives and the aggregated multiple program objective.

IV&V analyses, management and reporting are constrained by the resources of the IV&V Program; it is impossible to assure everything. IV&V must be able to make consistent decisions about risk, scope, and priorities across IV&V teams. As a result, our application of assurance strategy on Artemis has taken on some unique characteristics and solutions in response to these numerous challenges.

These challenges surfaced during support of the Artemis I mission. In 2019, IV&V briefly paused analysis effort on Artemis and held a process improvement event to discuss possible changes to our approach and new ideas to bring consistency across the individual IV&V project teams, which resulted in the formation of an overarching Artemis IV&V organization. We also realized the potential value of adopting more formal assurance case syntax to model our assurance design, resulting in the creation of the Artemis Assurance Case. [4] Since that time, our approach has continued to evolve and develop, with consistency, comprehensibility, and adaptability being central tenets to ensure success.

3.1. Agile IV&V Practices

Adaptation must start with project management. Rigidity in project management only tends to stifle change and innovation. Therefore, when looking for ideas about how to both manage large volumes of work and knowledge across many analysts with distributed skillsets, without limiting the creativity and adaptability of the team, IV&V looked to agile software development principles. In 2016, the Orion IV&V team worked with a consultant from the Carnegie Mellon University Software Engineering Institute to better understand the agile software development practices in use by the Orion software development organization. As a result, the Orion IV&V team was able to adopt some of the agile principles and practices to enhance the management of IV&V assurance work. [5] Many of these practices became infused across other Artemis IV&V teams as cross-team collaboration escalated within the Artemis IV&V organization.

Agile IV&V is an application of agile and lean principles appropriate to the planning, management, and performance of IV&V, rather than an adoption of a branded framework or tool. Artemis IV&V uses a three-month schedule of "assurance releases." At the start of each assurance release, each Artemis IV&V project team reviews and presents on the completed assurance work from the previous release, plans the ready-for-work priorities for the next release, and coordinates across teams on integration assurance targets, slack or surge in resource needs, and any important watch items or risks that need attention across the Artemis IV&V organization. These regular, iterative planning cycles enable adaptability in responding to changing priorities based on the ongoing software development and the availability of artifacts in the recent past or near future. In addition, assurance releases promote collaboration and awareness across IV&V Project teams, and regular check-ins as to the gradual assurance progress over time. At the end of each assurance release, each Artemis IV&V project team holds a retrospective to discuss what went well and what did not go well during the release. The retrospectives also provide an opportunity for the teams to discuss things the analysts think need to be changed, and dive deeper into the day-to-day analysis processes and approaches and propose potential improvements.

Artemis IV&V project teams are internally self-organizing. There is a strong reliance on the analysts to identify, prioritize, assess, and select potential assurance targets directly. Analysts are encouraged to own areas of subject matter expertise, perform risk assessments, provide inputs to planning, and manage and track related assurance progress. Most IV&V project teams subdivide their team members into smaller focus areas, and these sub-teams generally have regular stand-up meetings, or the equivalent, to allow each individual analyst to talk through progress on their current task and get help from their peers with any blockers to forward progress.

Many of these analysis teams use Kanban-like approaches to manage and track their day-to-day tasks. Tickets are added to team Kanban boards to track upcoming analysis work and monitor the volume of planned and in-progress tasks. Triaging the incomplete tasks as the end of the assurance release approaches identifies what the team can complete in time for the next planning cycle. The Kanban framework's use of a work-in-progress limit challenges IV&V to define shorter and more manageable tasks, which has led to some improvements in the turnaround time of our assurance work for each AO.

3.2. The Artemis Assurance Case (AAC)

To promote communication, collaboration, and consistency of assurance design and related information, Artemis IV&V made the decision for Artemis II and beyond to use explicit assurance case formalization using Goal Structuring Notation (GSN) syntax. Using a GSN assurance case allows our Artemis IV&V teams and analysts to clearly capture their reasoning, argumentation, and desired evidence when decomposing and elaborating capabilities. In addition, underlying assumptions and justifications leading to assurance design decisions are explicit and apparent. An assurance case covering the entire Artemis IV&V effort enables collaboration across IV&V project teams on cross-system integrated capabilities. The Artemis Assurance Case (AAC) was architected and developed in a way that keeps it as mission-agnostic as possible, as to minimize the number of changes necessary when moving from one mission to the next.

The AAC is comprised of claims related to assurance of capabilities. These cover both a cross-scenario perspective (e.g., capabilities like Environmental Control and Life Support, Command and Data Handling, and Guidance, Navigation, and Control), as well as mission scenarios (e.g., aborts, separation events, and vehicle docking). Fig. 5 contains a theoretical fragment of the AAC that demonstrates the capability-centric claims and evidence Artemis IV&V produces. In addition, the AAC makes room for other types of arguments and assurance strategies that IV&V addresses, including cybersecurity and code quality. The GSN syntax standardizes the way our assurance arguments are captured, and its broad applicability to any domain allows these different strategies to all coexist in the same network of claims. In addition, there are helpful extensions included in the GSN Standard that deal with modularization and cross-reference of claims and evidence, which are especially useful for the large and complex AAC by enabling it to be more adaptable to additions, deletions, and other changes in assurance design over time. [2]

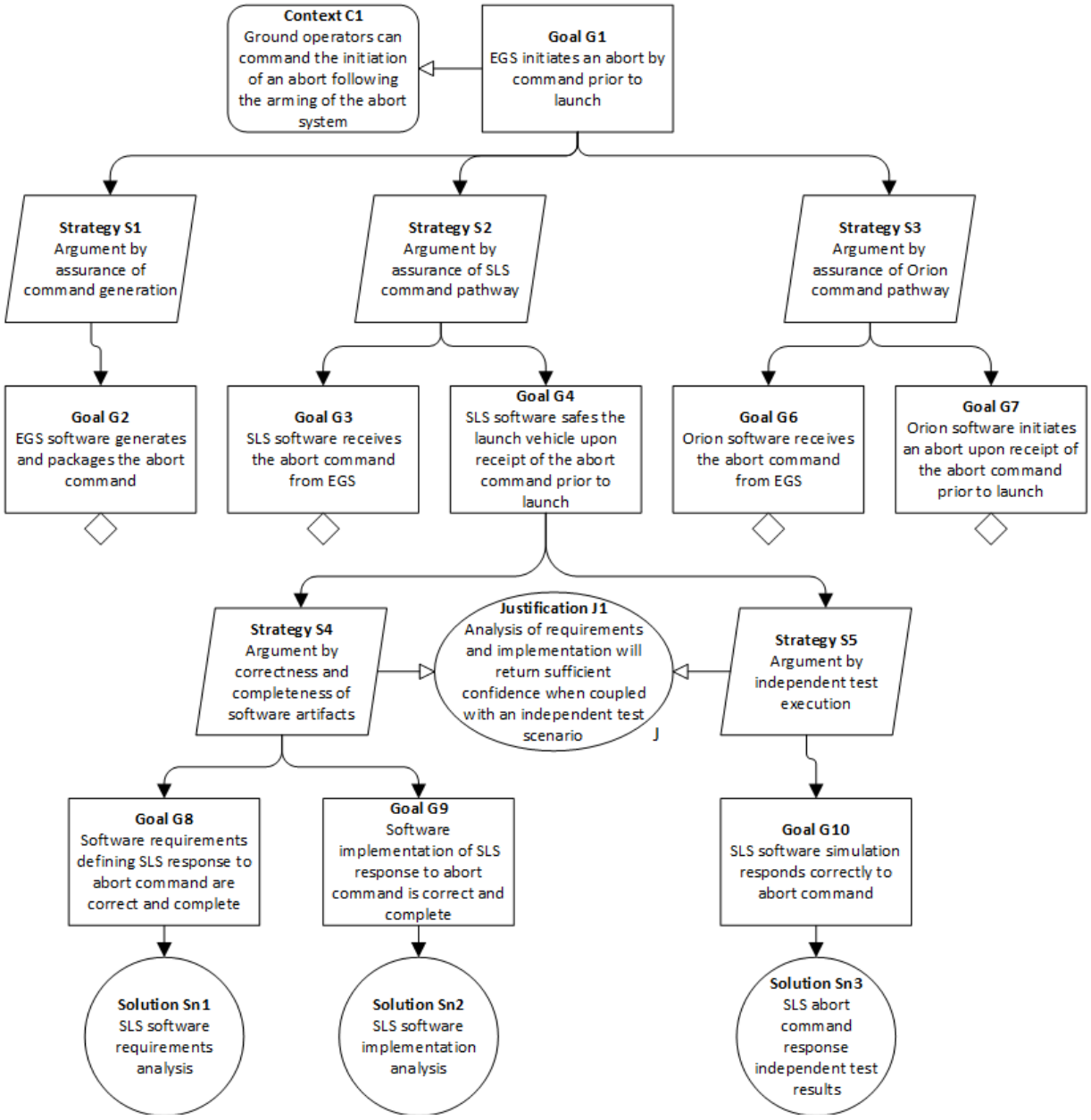

Figure 5. Theoretical Artemis Assurance Case Fragment

3.3. The Artemis Assurance Toolchain

The vast and complex network of assurance and evidence must be supported by the right tools for construction and maintenance. Tool selection, development, and integration has been, and remains, a critical part of ensuring the end-to-end IV&V process functions smoothly.

The Enterprise Architect (EA) model-based engineering tool was chosen as the platform for modeling the AAC, due to its ease of extensibility, support for collaboration, and very conspicuous and easily interrogatable database back-end. EA provides robust and tailorable search functionality using Structured Query Language (SQL) syntax, such that any of the data captured in the model can be retrieved in whatever form is desired. It also allows for customization of modeling language syntax and metadata, which enabled Artemis IV&V to define a GSN meta-model to implement the necessary GSN syntax and capture the data necessary to understand each model element and maintain the AAC.

The in-house developed A-SCAN tool integrates seamlessly with the AAC in EA via the back-end database. By mirroring the claim structure and solutions from the AAC in A-SCAN, we avoid the need to manage two separate but identical datasets. To facilitate this, we have built a synchronization process that occurs regularly between EA and A-SCAN. Thus, as the AAC is constructed in EA, records are automatically populated in A-SCAN so that corresponding risk

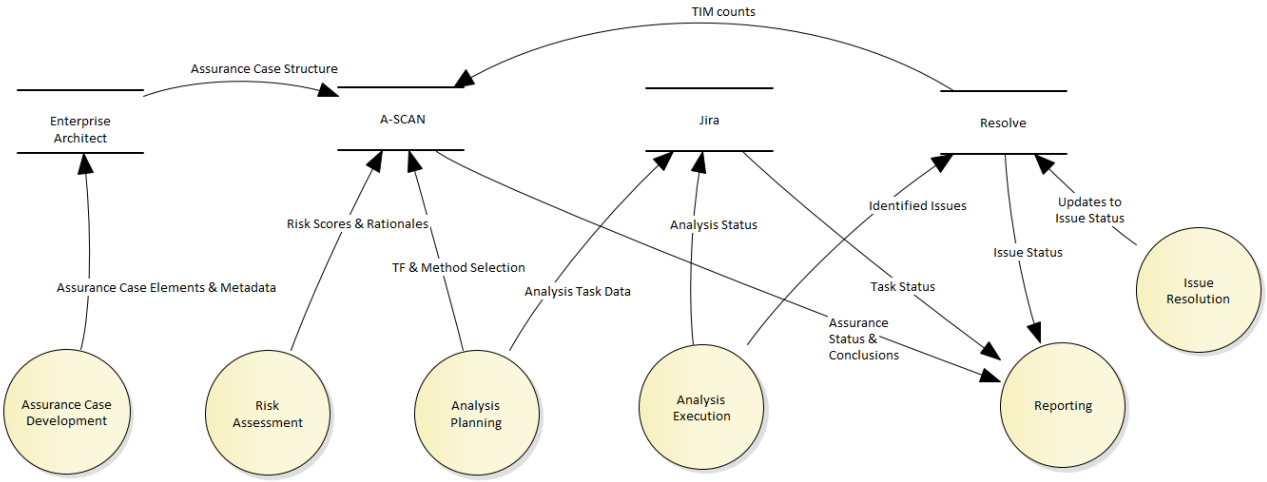

Figure 6. Artemis IV&V Assurance Toolchain

assessments and analysis plans can be documented. A-SCAN essentially sits on top of the AAC as an additional layer to capture the assurance status data, including AOs and ACs, and help us to manage the identified software risk and IV&V confidence.

For day-to-day task management, the Artemis IV&V project teams use Atlassian Jira. The assurance design and analysis tasks relate directly back to the nodes in the AAC and the analysis plans described in A-SCAN, so traceability is maintained between these related records so that all the information for a given node in the AAC can be easily retrieved. Analysts regularly record their progress on an analysis in Jira during the task, and at completion, close the Jira task and record the resulting assurance status data in A-SCAN so that the accumulated confidence can be calculated.

Resolve, an in-house developed tool, is used for tracking IV&V-identified issues in the form of TIMs. TIMs continue to be tracked to closure in Resolve throughout the lifecycle of the IV&V project. Because TIMs represent defeaters, or doubt, toward IV&V's desired assurance claims, Resolve integrates with A-SCAN so that issues can be related to the claims and evidence they directly impact. An accumulation of issues toward a particular capability has an impact in the IV&V confidence of that capability. Fig. 6 is a data flow diagram that shows how data moves between the tools in the Artemis IV&V assurance toolchain, and where and how the analysts and other Artemis IV&V team members interact with the tools to generate or retrieve data. The fundamental philosophy behind the development and use of the toolchain is to avoid duplication of data wherever possible and use the right tool for the right purpose. Artemis IV&V assurance data needs to remain comprehensible and manageable over future Artemis missions.

3.4. Measuring and Reporting Risk and Confidence

The adoption of A-SCAN across all the Artemis IV&V teams advanced the ability to capture, understand, and make use of consistent metrics for software risk and confidence. Prior to A-SCAN, Artemis teams had different means of determining and representing assurance status and risk levels, such that, as an example, what qualified as "high risk" from the perspective of the Orion IV&V team may not have been the same as what the SLS IV&V team considered "high risk." This became quite problematic when attempting to produce a cohesive and consistent message of assurance and risk for the Artemis I flight.

A-SCAN solves this problem by introducing a single, centralized risk assessment framework which produces an understanding of software risk that is consistent across all Artemis IV&V teams. Similarly, completed work and IV&V confidence is modeled the same way across teams, so that the amount of effort needed to reduce risk from (notionally) "high" to "low" is stable. Resource management across teams can be done much more readily because there are common measures; for example, if one Artemis IV&V project has addressed all their identified "high" risk capabilities, then those resources can be moved to help address remaining "high" risks on other projects.

A-SCAN's risk and confidence metrics allow for useful forms of reporting and tracking, like risk heatmaps. Because A-SCAN models the effect of IV&V analysis on confidence, which tends to increase over time as more analysis is done, data from A-SCAN can be used to show how the residual risk gradually decreases over the life of the IV&V project as a result of direct or indirect assurance. Visualizations of this change in risk over time have been especially useful in demonstrating and communicating the value of IV&V assurance. One such example can be seen I Fig. 7, which is the heatmap of Orion IV&V assurance targets for Artemis I. By the end of Orion IV&V support for Artemis I, the

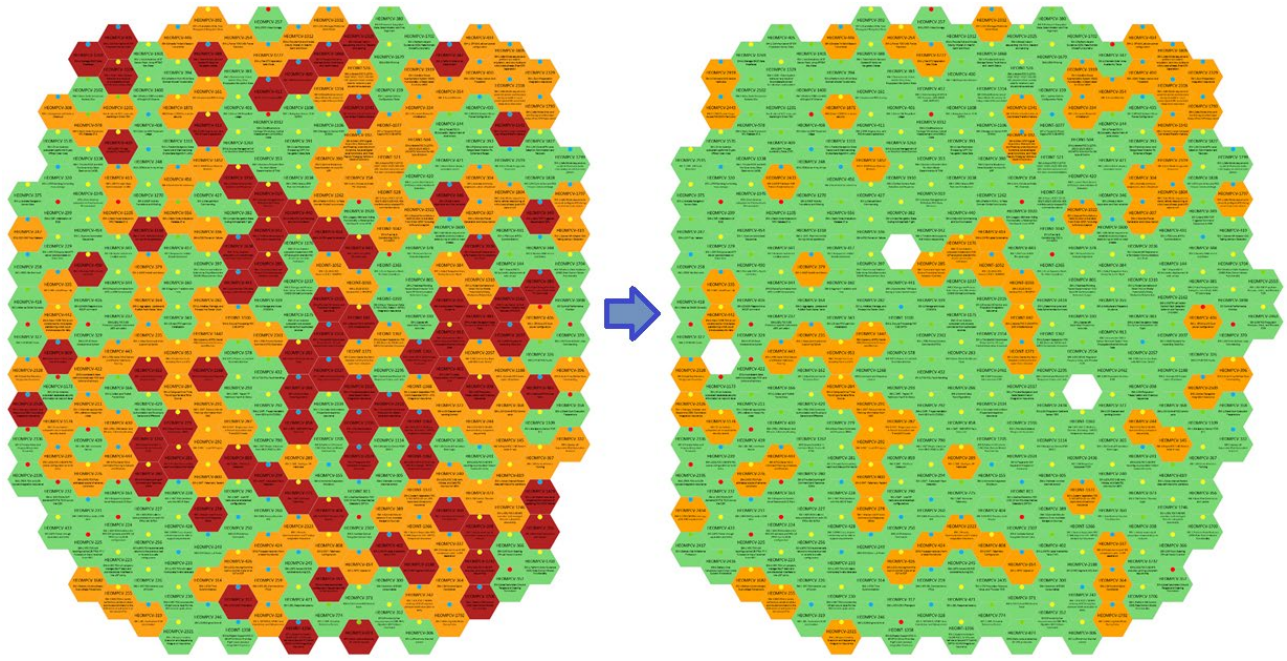

Figure 7. Orion Heatmap - Risk Burndown 2019-2021

burndown of high-risk areas in red is apparent as a result of the IV&V work completed.

Risk can be very dynamic, especially in more complex missions and projects. Risk is not limited to only decreasing over the lifecycle of an IV&V project. Occasionally, new information or understanding results in updates to risk assessments that elevate the level of risk, new capabilities evolve, or old capabilities are removed. When this happens, these heatmaps and other views of the A-SCAN assurance data are also useful in identifying and recognizing the need to potentially plan additional assurance work or bring into focus a capability that IV&V had not previously prioritized. In short, A-SCAN, and its resulting assurance data enables adaptability to the changing risk landscape.

3.5. Integrated Independent Testing Capability

The Artemis IV&V organization has a unique capability to conduct independent testing on integrated Artemis systems via the Advanced Risk Reduction Integrated Software Test and Operations Tri-Program Lightweight Environment (ARRISTOTLE). ARRISTOTLE was developed to facilitate risk-reduction testing of integrated Artemis systems and allows for the execution of flight and ground software in an operational, flight-like software-simulated environment, with capabilities for fault injection, mid-simulation pausing and analysis, and detailed post-processing logs and analysis tools. Currently, ARRISTOTLE is an emulation of the integrated Artemis launch system (SLS, upper stage, and Orion) and EGS, but there is a desire to integrate software simulations of Gateway, HLS, and MCC where possible for integrated testing on future missions. Because ARRISTOTLE integrates software emulations of these multiple systems, Artemis IV&V has the capability to test integrated scenarios using the actual flight and ground software that might be difficult or impossible to run on other test beds, allowing software integration testing to occur earlier and more often in the mission lifecycle.

The nature of assurance design is to build incremental assurance components that can be targeted at the lowest and most reusable levels and then roll that assurance up to higher level capabilities. Strong assurance evidence is built from more than the sum of its parts, and this is especially true for complex integrated systems. As such, independent testing in ARRISTOTLE is a key piece of the assurance design for Artemis IV&V. High risk off-nominal mission scenarios, especially those that involve interactions between Artemis systems, are prime targets for independent testing, due to the difficult nature of producing concrete evidence for software assurance when cross-system interfaces and off-nominal behavior are involved. Executing test cases can often generate stronger evidence for the correctness and reliability of data and command flows than other analysis methods, especially across complex system interfaces. In addition, typical verification and validation testing only addresses a limited set of off-nominal scenarios, so the ability for Artemis IV&V to run additional off-nominal tests adds a substantial amount of assurance value. The ability to introduce a failure or adverse condition in a simulation and inspect the software response produces confidence that the software is robust and reveals defects that are not encountered in nominal testing.

Independent testing is resource intensive and in a constrained budget environment, it is crucial to properly

target the necessary tests. A dedicated Artemis IV&V independent test team is identifying, prioritizing, and executing test cases in line with the assurance release planning cycle followed by the Artemis IV&V project teams. The team considers the capabilities of the integrated Artemis system, as well as the ARRISTOTLE emulation, and defines test cases that target aspects of risk that are difficult to address through other types of analysis. These test cases are producing evidence that integrate directly into the AAC and support Artemis IV&V's claims.

3.6. Ongoing Challenges

We have made outstanding progress in the last few years advancing our capabilities in planning and executing assurance efficiently and effectively on the areas of highest risk, yet there remain ongoing challenges. As a result of having so many team members working within the tool chain, it is difficult to achieve consistency within the data being generated, particularly when jargon and project-specific terms do not align completely across Artemis development projects. We are exploring solutions to both help identify and correct errors in GSN modeling syntax, as well as enhancing the tools' capabilities to search for and retrieve relevant information. To maximize the utility of the AAC, it needs to support the day-to-day use cases that result from analysts needing to find the information they are looking for.

With a new suite of data management tools, some of the reporting and tracking mechanisms our stakeholders used in the past are no longer viable. We are working to identify and build replacement solutions to meet those planning and reporting use cases. Visualization enhancements to enable better use of the data we are tracking, including new heatmaps and progress reports of various views of the AAC, are some of the next objectives for toolchain developments.

Artemis IV&V's sophisticated approach to carrying assurance from one Artemis mission to the next has yet to be proven. Because the AAC and related toolchain has been piloted on Artemis II, this transition will be tested for the first time during close out of Artemis II assurance activities. There will certainly be some pains associated with promoting such a large volume of assurance data to Artemis III for the first time using the AAC and A-SCAN, but we will continue to learn, adapt, and streamline the process so that it is easier the next time.

4. CONCLUSION

The assurance design and execution approach we are evolving at NASA IV&V for Artemis missions is a direct response to the challenges presented by the need to assure the software for a multi-mission program made up of large, complex systems of systems. Agile principles and practices enable teams to self-organize around the work they need to complete, promoting more adaptive task management and better turnaround cycles for assurance results. Assurance Case methodology makes our assurance design reasoning and logic explicit and interrogatable, promoting consistency and communication across teams. The AAC allows for distributed ownership of the assurance design, so that experts can develop and refine the argument and analysis plans in the areas they know best. A-SCAN's risk and confidence framework and metrics drive more consistency across the entire Artemis IV&V team in the understanding of risk levels and priorities, progress toward IV&V claims, and right-sizing analysis to fit the identified risk. The integrated assurance case toolchain allows us to take better advantage of the data we are capturing for reporting and progress tracking and enables better long-term maintenance of assurance data through common tools and processes. Independent testing capabilities via ARRISTOTLE open a vast array of possibilities for producing high-value, robust evidence toward the assurance of integrated capabilities and scenarios.

Artemis IV&V has been challenged to provide more value-added assurance to future Artemis missions within a constrained budget. The Adaptive IV&V investments made to date have enabled Artemis IV&V to become more efficient and effective in IV&V planning and execution and respond more readily to changes in the risk landscape as they manifest, increasing the breadth and depth of risk reduction possible within the available resources. Residual risk tracking allows IV&V to communicate more effectively with stakeholders, both internal and external at all levels, and inform key decision-making personnel. This evolving assurance design approach provides IV&V surety that work is performed in the highest risk, most value-added areas of the software, to keep our astronauts and ground crews safe and ensure mission success.

With such a long-term multi-mission program as Artemis, we do not expect to see all the benefits from these advancements immediately; however, improvements in efficiency and effectiveness of IV&V services have been observed incrementally. Many more will continue to be realized over time, especially in future Artemis missions, when our assurance data and reasoning is readily available and digestible to easily make sense of the assurance posture and the decisions and evidence that led to that state.

Each of the advancements and improvements the Artemis IV&V team has achieved so far has required some investment to bring it to realization, and future advancements will continue to do so. It is essential to pause and consider what will be necessary for success one, three, five, or more years into the future, and then set aside the resources to explore potential ways to meet

those needs before they come due. It is our firm belief that any successful safety organization needs some degree of flexibility, adaptability, and most importantly investment in the future. The safety of our astronauts is the top priority of Artemis IV&V as NASA looks to return humans to the moon with Artemis III and future Artemis missions.

5. ACKNOWLEDGEMENTS

The authors would first and foremost like to thank all the members of the Artemis IV&V team for their dedication to the Artemis mission, consistent excellence, and continuous improvement and adaptability as we have advanced the IV&V state of practice over the years, and especially those who helped contribute ideas and feedback on this paper. The authors would also like to thank the NASA IV&V Program leadership, as well as IV&V contractor leadership, for their continued support of the Adaptive IV&V practices laid out in this paper. None of this work would have been possible without the trust from leadership, and the support of the entire Artemis IV&V team.

6. REFERENCES

1. Independent Verification and Validation Technical Framework (IVV 09-1), Version P (2017). https://www.nasa.gov/sites/default/files/atoms/files/ivv_09-1_independent_verification_and_validation_technical_framework_-_ver_p_-_10-25-2017.pdf

2. The Assurance Case Working Group (2021). Goal Structuring Notation Community Standard, Version 3. https://scsc.uk/r141C:1?t=1

3. Shafer, G. (1976). A Mathematical Theory of Evidence, Princeton University Press, Princeton NJ.

4. Whitman, G., Amoroso, P., Black, G., et al. (2020). IV&V Assurance Case Design for Artemis II. https://ntrs.nasa.gov/citations/20200001646

5. Smith, J., Bradbury, J., Hayes, W., et al. (2019). Agile Approach to Assuring the Safety-Critical Embedded Software for NASA's Orion Spacecraft. https://ntrs.nasa.gov/citations/20190001434