The Unique Aspects of Simulation Verification and Validation

Danny Thomas
Alexia Joiner
AEgis Technologies Group
631 Discovery Drive
Huntsville, AL 35806
256-922-0802
ajoiner@aegistg.com
d.thomas@aegistg.com
Wei Lin
Michael Lowry
Tom Pressburger
NASA Ames Research Center
Mail Stop 213-13
Moffett Field, CA 94035
650-604-5000
w.lin@nasa.gov
michael.r.lowry@nasa.gov
tom.pressburger@nasa.gov

Abstract—Models and simulations (M&S) will be employed to support important design decisions and verification of system requirements in the development of NASA's Orion Crew Exploration Vehicle. Most simulations are implemented in software. For developed software, NASA's software engineering procedural guideline NPR 7150.2 and safety standard NASA-STD-8719.13B apply. Recognizing the need for critical M&S to be validated to be credible for their intended uses, NASA developed a Modeling and Simulation Standard, NASA-STD-7009[1]. This paper analyzes the requirements specified by these standards and their role in test, validation and certification of modeling and simulation software. It discusses simulation validation as a distinct instance of software validation with corresponding unique requirements. Simulation-specific validation concerns include fit to intended use, validation against experimental data, uncertainty quantification, and sensitivity analysis. The paper also describes the Orion M&S verification, validation, and accreditation (VV&A) process.

TABLE OF CONTENTS

1. INTRODUCTION

Software verification is confirmation by examination and provisions of objective evidence that software meets its specifications. Software validation is confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled.[2] Simulation validation in particular is the process of determining the degree to which a model or simulation is an accurate representation of the real-world from the perspective of the intended uses of the model or simulation.[3] Simulations are by definition approximations or abstractions of the real world. Specifying them involves determining the set of entities of interest to be simulated and the accuracy of representation.

For verification, simulation software can be tested against its requirements by a variety of mature software methodologies not covered in this paper. The simulation community has made considerable advances in defining uncertainty, fidelity and validity of simulations. When these parameters can be quantified and measured, traditional software testing techniques can be profitably applied.

[1] IEEEAC paper#1390, Version 3, updated 2009:12:30, contains U.S. Government work not protected by U.S. copyright. Unless otherwise noted all references are to NASA's Modeling and Simulation Standard, NASA-STD-7009, July 2008 http://standards.nasa.gov/released/NASA/NASA_STD_7009_APPROVED_2008_07_11.pdf

[2] Code of Federal Regulations Title 21, Volume 8 Revised as of April 1, 2008 Subpart A--General Provisions Sec. 820.3 Definitions.

[3] Committee on Standards American Aeronautics and Astronautics (AIAA) from Obercampf 2002

However, even if a simulation is certified to implement its requirements exactly, that does not measure the credibility of simulation results for a particular use. When NASA analysts could not state the credibility of simulations used to predict the success of the Shuttle launches, the Chief Engineer commissioned a standard for models and simulations. The resulting NASA-STD-7009 recognizes the importance of measuring the credibility of the final results of a simulation study and reporting these results unambiguously with a Credibility Assessment Scale. The assessment scale measures, among other things, the credibility of the verification and validation performed on the software that implements the simulation.

This paper describes how these considerations were taken into account in developing the Orion Modeling and Simulation VV&A process, where accreditation refers to the process of certifying that a model and simulation is appropriate to a particular intended use.

2. RELATIONSHIPS

The relationships between Systems Engineering, Software Engineering and M&S are shown in Figure 1. The figure illustrates that while there are commonalities and overlap between systems engineering, M&S and software engineering, there is also uniqueness. Each discipline has its own Body of Knowledge. Each has its own methodologies.

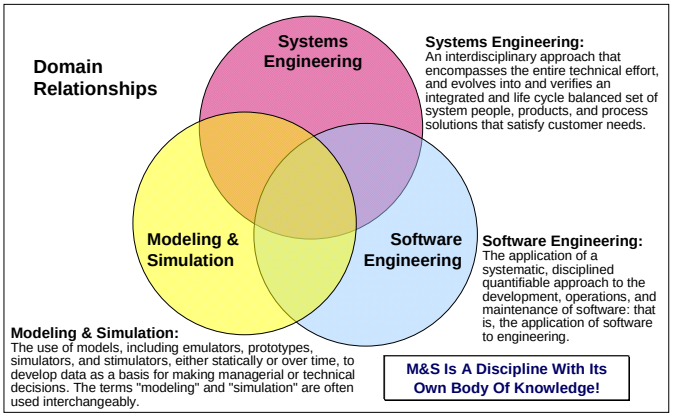

Figure 1: Three Distinct Domains

Systems Engineering is an interdisciplinary field of engineering that focuses on how complex engineering projects should be designed and managed. Issues such as logistics, the coordination of different teams, and automatic control of machinery become more difficult when dealing with large, complex projects. Systems engineering deals with work-processes and tools to handle such projects, and it overlaps with both technical and human-centered disciplines such as control engineering and project management. [4]

Software engineering is the application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software, and the study of these approaches; that is, the application of engineering to software.[5]

Universities have recently added advanced and undergraduate degrees in Modeling and Simulation. M&S education is fast becoming a recognized discipline within academe and industry. And as a discipline it has experienced a top-down approach with regard to academic programs that have their origins at the graduate level. Programs have grown from graduate M&S courses in such diverse fields as Engineering, Operations Research, Economics and Sociology.

3. SIMULATION AND SOFTWARE COMPARED

A simulation is a representation of the operation or features of one process or system through the use of another.[6] The system or process being represented is the referent. Many modern simulations are implemented in software.[7] These may profit from good Software Engineering practices and modern software testing techniques, but the focus of this paper is the unique aspects of testing M&S. Simulations are by the very definition a representation of the referent. They are abstract or they would be copies. In fact the very abstraction is critical to their use - but how much abstraction? How well does the M&S need to match the referent? In what aspects of behavior or constituency? How accurately? What fidelity? These terms are difficult to define much less specify as testable requirements.

A classic example from the aerospace industry is the difference between a simulation of a rocket engine and the software used to control the engine. The rocket engine may be simulated with a system of partial differential equations. It will be a representation of the operation of the engine. The inconsistencies caused by assumptions like laminar flow or inaccuracies in the equation solver may combine to produce an uncertainty in the estimate of performance, but that uncertainty may be perfectly acceptable. On the other hand the software used to control the engine must be predictably error free.

[4] http://en.wikipedia.org/wiki/Systems_engineering

[5] SWEBOK executive editors, Alain Abran, James W. Moore ; editors, Pierre Bourque, Robert Dupuis. (2004). Pierre Bourque and Robert Dupuis. ed. *Guide to the Software Engineering Body of Knowledge - 2004 Version*. IEEE Computer Society. p. 1-1. ISBN 0-7695-2330-7. http://www.swebok.org.

[6] www.thefreedictionary.com

[7] Notable exceptions include process simulations, table-top exercises, mock-ups, training, and electrical emulators.

Software is verified and validated by rigorous processes. Software verification is confirmation by examination and provisions of objective evidence that software meets its specifications. Software validation is confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled. NASA Software Independent Verification and Validation (IV&V) is an Agency-wide strategy to provide the highest achievable levels of safety and cost-effectiveness for mission critical software.

Likewise M&S are verified and validated by rigorous processes that are adapted for the unique nature of M&S. M&S verification is the process of determining that a model [or simulation] implementation accurately represents the developer's conceptual description and specification. M&S validation is the process of determining the degree to which a model [or simulation] is an accurate representation of the real-world from the perspective of the intended uses of the model or simulation. The subtle difference deals with that tricky aspect of fidelity.

4. NASA'S M&S STANDARD

The genesis of NASA's recent emphasis on assuring the validity of simulations and the credibility of simulation studies is depicted in Figure 2.

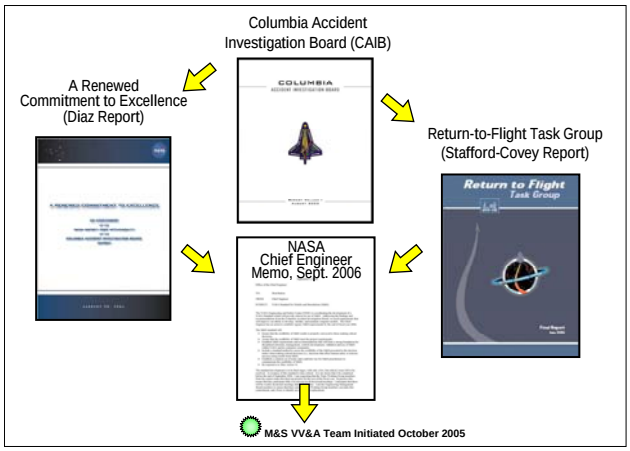

Figure 2: Drivers for the Simulation Standard

The Columbia Accident Investigation Report (CAIB) called for NASA to "develop, validate, and maintain physics-based computer models to evaluate Thermal Protection System damage from debris impacts. These tools should provide realistic and timely estimates of any impact damage from possible debris from any source that may ultimately impact the Orbiter. Establish impact damage thresholds that trigger responsive corrective action, such as on-orbit inspection and repair, when indicated."

The Diaz Report broadened the scope beyond STS. Action Item Number 4 called for NASA to "develop a standard for the development, documentation, and operation of models and simulations; documentation, configuration management, and quality assurance; verification and validation, operational data and trending; tool management, maintenance, and obsolescence; training requirements, best practices for user interfaces; and user feedback when results appear unrealistic"

The Stafford-Covey report enjoined "formal development, verification and validation, and outside review plans." It said that assumptions should be written down and consistently applied. Sensitivity analysis and careful analysis of uncertainty was to be performed.

The Chief Engineer's Memo required that the credibility of M&S results is properly conveyed to those making critical decisions, that analysts should assure that the credibility of M&S meets the project requirements. NASA was to establish M&S requirements and recommendations that will form a strong foundation for disciplined (structure, management, control) development, validation and use of M&S within NASA and its contractor community, include a standard method to assess the credibility of the M&S presented to the decision maker when making critical decisions (i.e., decisions that effect human safety or mission success) using results from M&S, and establish a common set of terms and a uniform way for M&S practitioners to communicate the credibility of M&S.

5. REQUIREMENTS FROM NASA-STD-7009

NASA-STD-7009 requires that the presentation of any results from M&S to a decision maker include (1) the best estimate of the results, (2) a statement on the uncertainty in the results, (3) the evaluation of the results on the credibility assessment scale, and (4) any explicit caveats that accompany the results. (An example of such a caveat would be use of the model in violation of its assumptions.) The decision maker then makes his/her own assessment of credibility based upon all four pieces of information in the context of the decision at hand as shown in Figure 3. This is intended to provide a standard method to assess the credibility of the models and simulations presented to the decision maker when making critical decisions (i.e., decisions that effect human safety or mission success) using results from models and simulations and to assure that the credibility of models and simulations meet the project requirements.

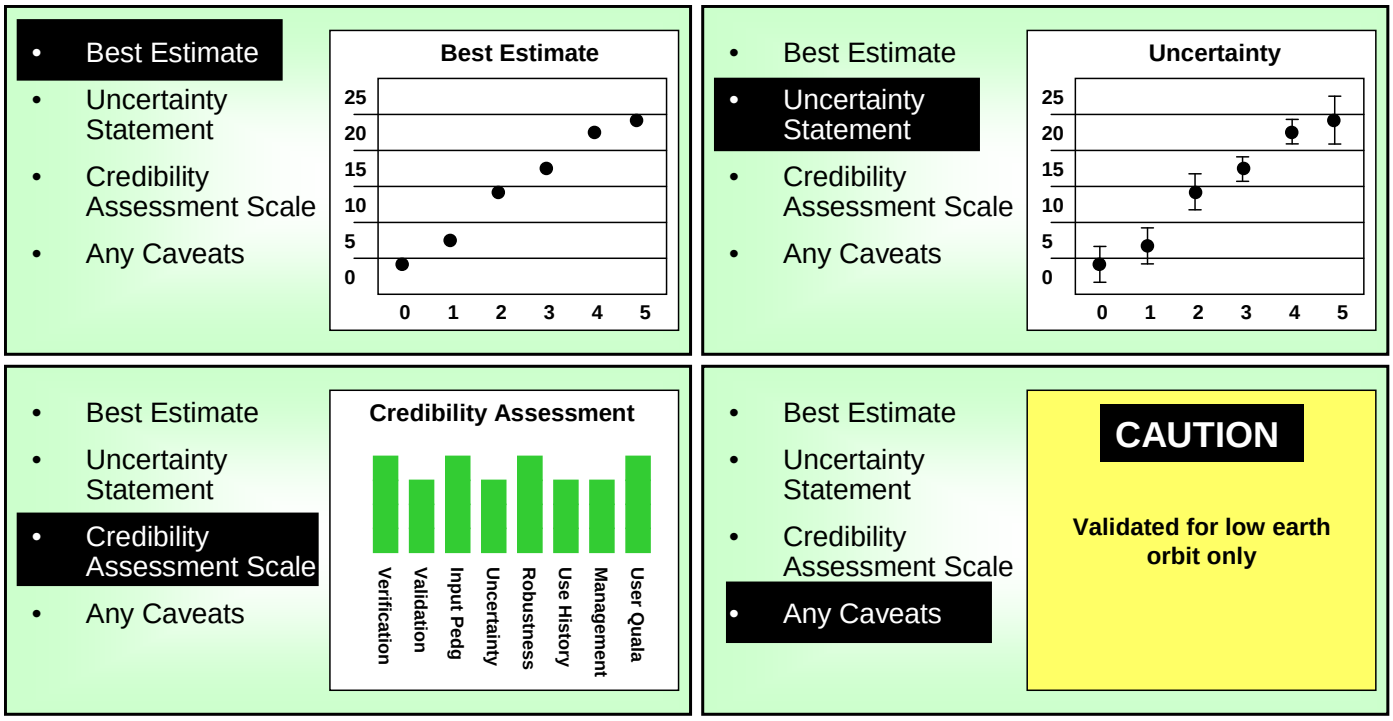

Figure 3: The Four Elements of Credibility Reporting as Prescribed by NASA-STD-7009

The credibility assessment scale (CAS) is shown in Figure 4. This CAS consists of eight factors grouped into three categories. The assessment process involves evaluating the M&S results on each of eight factors, and then rolling up these eight factor results into a single number that represents the summary credibility assessment. The M&S Development category captures those aspects of the M&S that pertain to the general assessment of the credibility of the M&S for their broad intended use; the M&S Operations addresses the aspects relevant to the current application of the M&S to generate the particular M&S results under assessment; and the Supporting Evidence category addresses three cross-cutting factors.

The credibility assessment scale does not purport to measure credibility; rather, it assesses the M&S results, and the rigor of the processes used to produce them, against key factors that affect the credibility judgment. The fundamental premise of this approach is that as a general rule, the more rigorous the key processes used for generating the M&S results, the greater the credibility of the M&S results, all else (including the estimated uncertainty) being equal.

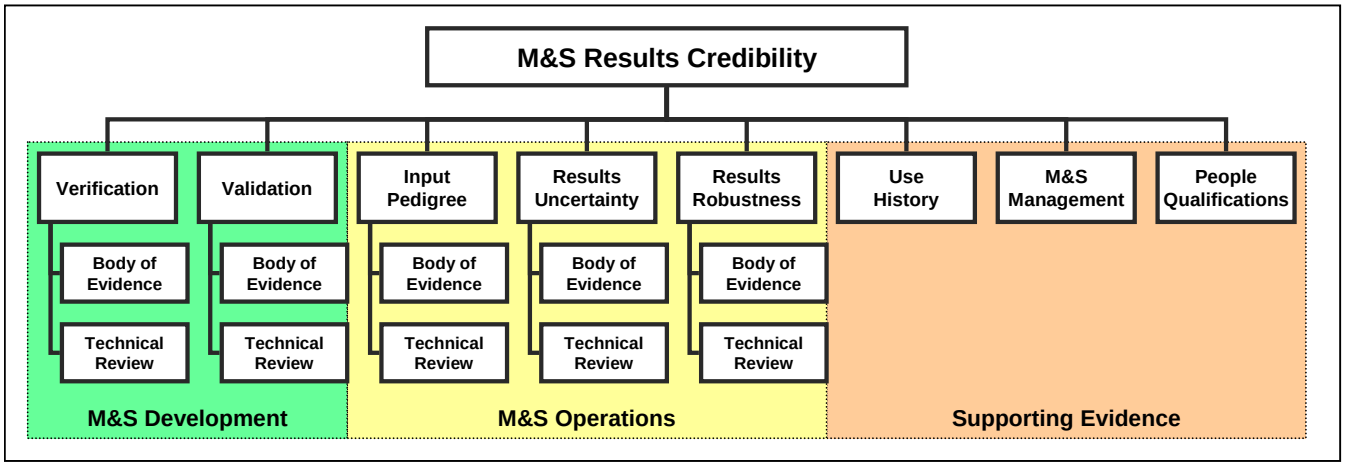

Figure 4: The Credibility Assessment Scale from NASA-STD-7009

The M&S Operations category deals with the credibility factors for the application of this particular computational model in the generation of the current M&S results. This includes the conduct of the present simulation and the analysis and reporting of the results. The Supporting Evidence category covers the use history of the particular computational model employed in the M&S; the overall management of the M&S processes; and the qualifications of the people involved in the development, operation, and analysis of the computational model.

6. Constellation Implementation

The Constellation Program developed the process shown in Figure 5 to meet the requirements of NASA-STD-7009. The Three-Phase VV&A Process was developed to work within NASA's dynamic and diverse environment. Designed specifically to accommodate each unique situation, this process allows the practitioner to evaluate and determine which activities are relevant to their needs. It also allows for necessary information gathering and subsequent planning before committing to V&V activities that can be cost and resource-intensive. To determine the types and amount of evidence needed to ensure credibility, it is critical to understand the program need(s) that the M&S is supporting, as well as the ability of the M&S to fulfill those needs. The phased approach addresses this, rather than assuming that a full VV&A effort is necessary in every situation, and provides management with the information necessary to make critical decisions. Breaking the VV&A activities up into discernable phases allows the practitioner or manager to determine whether to accredit the M&S based on evidence in existence at that time, or proceed to the next phase.

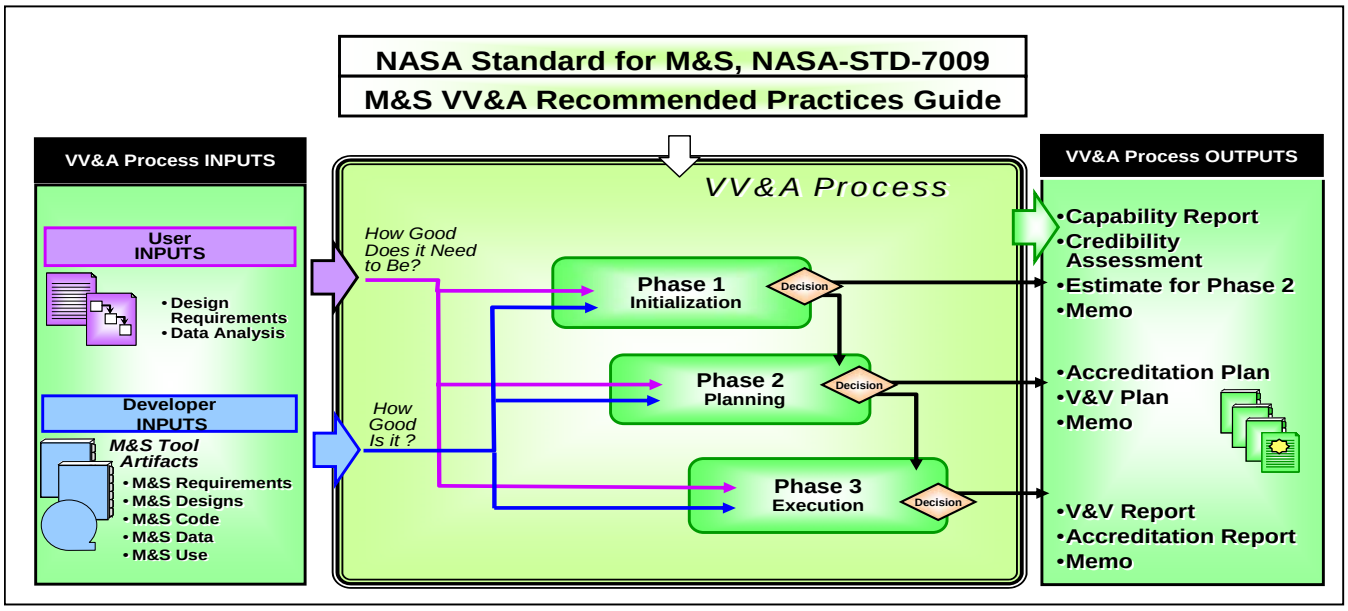

Figure 5: The Constellation Program's Three-Phase VV&A Process

A never-before-assessed but existing M&S may already have all the V&V evidence needed to accredit. When the evidence is assessed and perhaps documented in a capability assessment, the M&S may be accredited following initialization. If not, additional verification and validation activates may be planned in the second phase. After planning, which includes an estimate of required resources, it may be discovered that the program lacks the resources to accomplish the new V&V activities at the time. Management may elect to use the M&S with the understanding that additional V&V was needed but not yet performed. Finally there is the decision after the execution of the new V&V activities.[8]

7. Orion Implementation

The mission of the Orion Modeling and Simulation Office (OM&SO) is to ensure that the Orion Project has adequate modeling capabilities to support spacecraft design, implementation, test, and operations activities: safely, accurately, and efficiently. The OM&SO developed an approach to implement M&S standards to support important Orion design decisions and verification of system requirements. NASA is using the process for developing in-house M&S products. Example products are M&S used in

[8] A more complete discussion of this process is in the NASA VV&A Recommended Practices Guide https://ice.exploration.nasa.gov/confluence/pages/viewpage.action?pageId=9803356

docking system emulators and aerodynamic databases generated by M&S. For the Orion Design and Analysis Cycles, NASA streamlined the M&S VV&A process, which proceeds as follows and shown in Figure 6.

- The Project Leader documents the M&S, their intended use, and test procedures.
- The customer user of the particular M&S lists its requirements.
- The developer implements the M&S according to the customer's requirements.
- The tester tests the M&S against the results of physical experiments, other simulations, or spot calculations, as appropriate, and documents the results in a V&V report...
- The technical team (consisting of the lead, developer, tester, subject matter experts, and customers) review the test results, and document the four items required by NASA-STD-7009 (1) the best estimate of the results, (2) a statement on the uncertainty in the results, (3) the evaluation of the results on the credibility assessment scale, and (4) any explicit caveats that accompany the results.

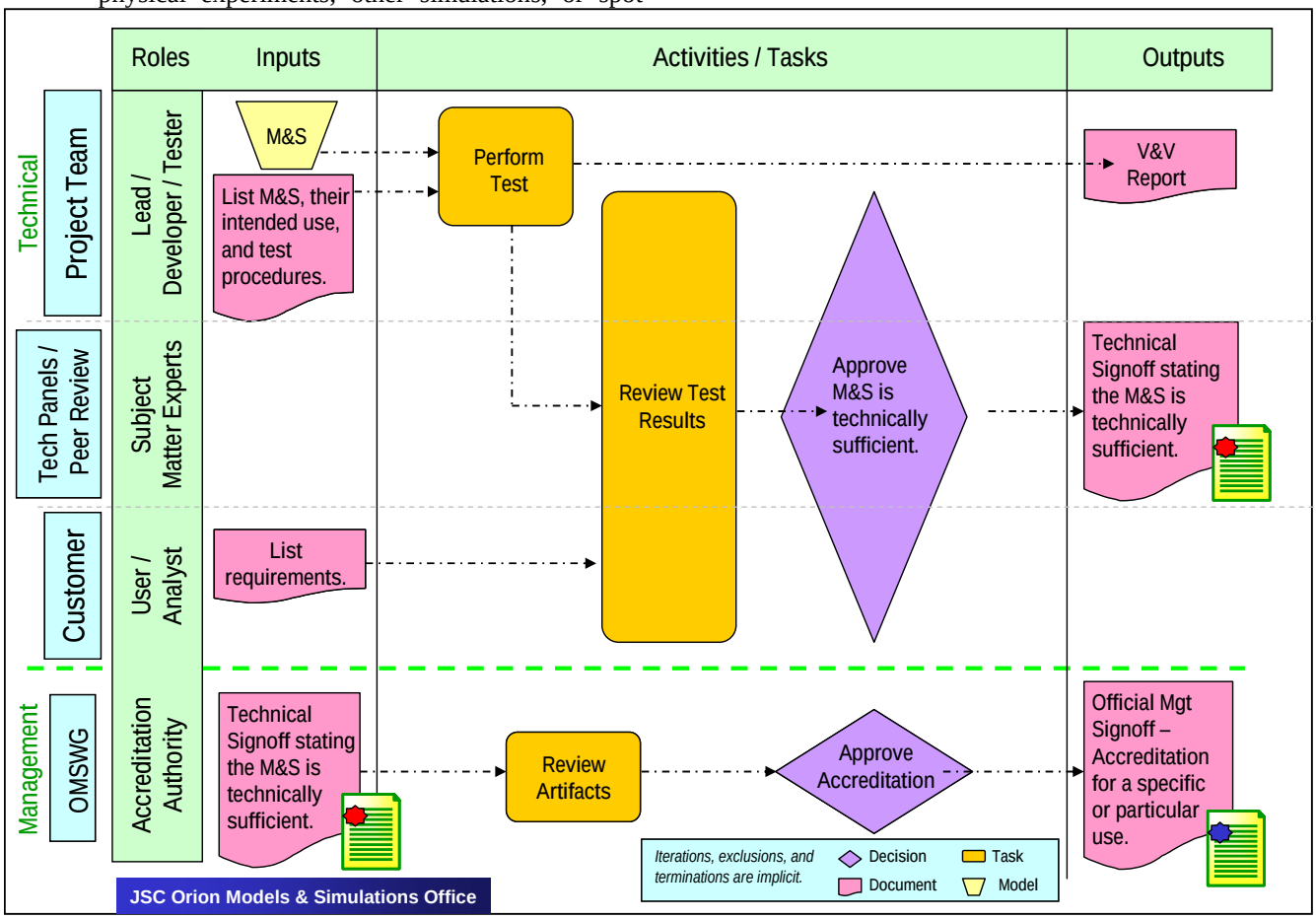

Figure 6: The Orion VV&A Process for Government Furnished M&S

If the technical team deems the M&S is technically sufficient for the intended use, the technical team will officially sign off on the M&S product. For accreditation, the Orion M&S Working Group (OMSWG) will audit all of the artifacts. The purpose of the OMSWG is to coordinate and oversee M&S activities within Project Orion, including M&S development tasks, contractor deliveries, process & standards definition, VV&A reviews, and model and data reuse across Orion. If the M&S is to be used to verify a specific Orion requirement, then the Accreditation Authority, who is the lead of the Orion M&S office, will officially sign off on the product.

8. CONCLUSIONS

M&S software is a unique kind of software. Simulation-specific validation concerns include the M&S being fit to its intended use, validation against experimental data, uncertainty quantification, and sensitivity analysis. Traditional software testing techniques must be augmented with credibility assessment techniques that address these unique concerns. NASA recognizes the unique aspect of M&S software through agency level standards, program requirements and project requirements.

9. REFERENCES

NASA's Modeling and Simulation Standard, NASA-STD-7009, July 2008
http://standards.nasa.gov/released/NASA/NASA_STD_7009_APPROVED_2008_07_11.pdf

Code of Federal Regulations Title 21, Volume 8 Revised as of April 1, 2008 Subpart A--General Provisions Sec. 820.3 Definitions.

Committee on Standards American Aeronautics and Astronautics (AIAA) from Obercampf 2002

SWEBOK executive editors, Alain Abran, James W. Moore; editors, Pierre Bourque, Robert Dupuis. (2004). Pierre Bourque and Robert Dupuis. ed. Guide to the Software Engineering Body of Knowledge - 2004 Version. IEEE Computer Society. p. 1-1. ISBN 0-7695-2330-7. http://www.swebok.org.

10. BIOGRAPHY

DANNY THOMAS is a Senior Research Scientist with AEgis Technologies Group in Huntsville, Alabama. He is currently supporting NASA's effort to institute consistent management practices for simulation development and use.

ALEXIA JOINER is the NASA Program Manager with the AEgis Technologies Group in Huntsville, AL. Ms. Joiner has 10 years experience in various NASA systems engineering and operations related areas. During her career Mrs. Joiner has served as a Payload Operations Lead and has conducted Astronaut Payload Training.

WEI LIN is a Software Systems Engineer in the Systems Engineering Technical Area at NASA Ames Research Center. She is responsible for Verification, Validation, and Accreditation (VV&A) in the Constellation Orion Level III Modeling & Simulation Office.

MICHAEL LOWRY is the NASA chief scientist for Reliable Software Engineering. After receiving his BS/MS from MIT and PhD from Stanford, all in computer science, he joined the Kestrel Institute as PI working on program synthesis. In 1993 he joined NASA Ames as group lead then area lead, and was promoted to chief scientist in 2008. Dr. Lowry is the editor of MIT Press "Automating Software Design" and serves on the editorial board of the journal Automated Software Engineering. He has published numerous papers principally on the topics of program synthesis and software V&V. He is currently the software production tools lead for NASA Orion, as well as the PI for NASA's research in advanced software engineering for exploration systems.

THOMAS PRESSBURGER is a Computer Scientist in the Robust Software Engineering area at NASA Ames Research Center. His expertise is in design and use of automatic code generation and advanced verification technologies. His current focus is Orion, specifically software development tools and process and the application of verification technologies. For more detail, see http://ti.arc.nasa.gov/profile/ttp/.