

The Role of Credibility Assessment in M&S Verification, Validation, and Accreditation

October, 2007

Danny Thomas
Bobby Hartway

AEgis Technologies Group
631 Discovery Drive
Huntsville, AL 35806
256-922-0802
<http://www.aegistg.com/>

ABSTRACT: NASA's new Modeling and Simulation Standard requires a credibility assessment for critical engineering decisions based upon results of models and simulations. Credibility assessment is thus a "qualifying factor" in reporting results from simulation-based analysis. The NASA Constellation Program's Verification, Validation and Accreditation process for models and simulations accordingly contains an added, explicit credibility assessment step to the traditional process. This paper will explore the dimensions and criteria of credibility assessment and its relationship to the more traditional Accreditation Criteria. The role of credibility assessment in the context of a typical end-to-end engineering and design analysis process will be described.

Introduction

The primary goal of the new NASA-STD 7009, Standard for Models and Simulations, is to ensure that credibility of results from models and simulations (M&S) are properly reported to those making critical decisions [1]. The secondary goal of the standard is to assure that the credibility of the results from M&S meets project requirements. The standard addresses the key factors considered to be necessary for assuring this credibility.

Credibility management of modeling and simulation (M&S) results depends on two factors; 1) how well analysts using M&S understand the credibility of the M&S used in their analyses, and 2) how transparently that knowledge is represented to decision makers using their analyses reports, so risk-of-use for M&S credibility may be incorporated in their decisions. Satisfying these two related factors can be accomplished by establishing a "credibility chain-of-custody" lifecycle workflow to account for all the "touch-points" of activities impacting on M&S results credibility, from initial M&S requirements, through development, use, and results reporting. In this paper, we present the initial work on the development of such a lifecycle workflow, and how it "illuminates" the touch-points of credibility for M&S results used in analyses that support project and program decision-making.

We limited this paper to addressing M&S results credibility only. Factors that could affect credibility of analyses reports based upon M&S are considered characteristics of analyses, and therefore beyond the M&S credibility domain. The exception to this would be if the analyses used *only* M&S data, and added no other investigative data or manipulations of data. In this case, the analysis credibility is one and the same as the M&S results credibility.

First, we describe the overall "*credibility chain-of-custody*" workflow diagram, and describe its key features for quantifying and illuminating factors that affect M&S results credibility. A key distinction that shows up on this workflow is that accreditation is done *before* M&S use, and credibility of results is done *after* M&S use. Each workflow step is considered a credibility "touch-point" along the credibility chain-of-custody.

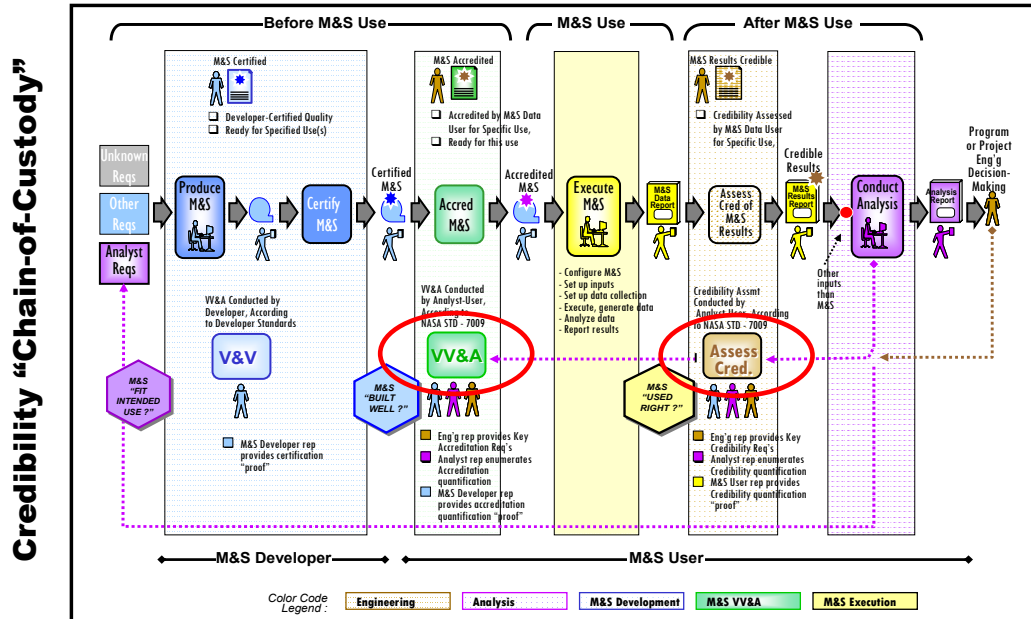
Second, we define credibility, and its interplay with the traditional VV&A accreditation process, and then the *top-level quantifying criteria* for assessing accreditation and credibility "proof".

Third, we expand the traditional M&S verification and validation workflow to show the individual component assessments, and how they are similar, but different for a developer's V&V and an analyst/users VV&A.

Fourth, we expand the "M&S use" workflow in order to show the key touch points for M&S results-credibility.

Finally, we introduce a larger-scope chain-of-custody, one that reaches back before M&S, into systems engineering and integration, and also forward, into analysis and project decision making. We use this larger scope in order to emphasize that credibility of M&S results is dependent upon prior, non-M&S work - and that project decision-making is dependent upon factors *within* analysis workflow, very similar to factors within M&S credulity workflow.

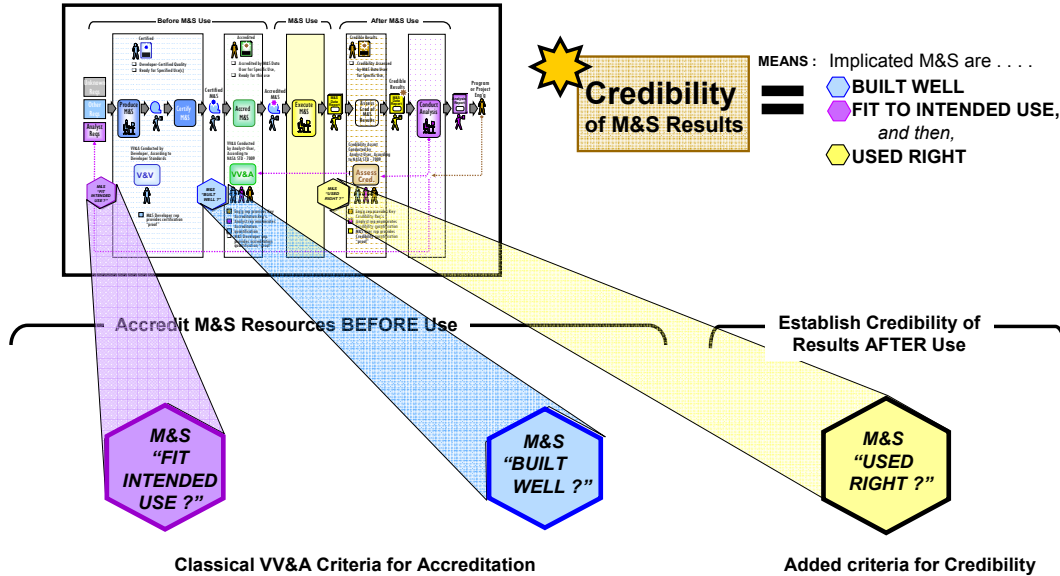
The "Chain-of-Custody" path for M&S Results Credibility is the end-to-end critical path of M&S results - - traced through sequential supporting and underlying data and operations, from M&S development, through M&S operations, to M&S results reporting, for subsequent use in Analysis Reporting. Notice that Credibility Assessment FOLLOWS VV&A.



This slide introduces the "credibility chain-of-custody" diagram, which is a workflow diagram of the key M&S development and use activities. The workflow progresses left to right, from initial requirements development, through production of an M&S Results Report, with assessed credibility, ready for use by analysts to conduct their analysis studies for the specified problem for which the M&S was developed or chosen. The diagram is divided into three major partitions, 1) activities before M&S use, 2) activities of M&S use, and 3) activities after M&S use. These activities are all color-coded consistently throughout this presentation material, according to the color-code legend shown at the bottom.

The key features of this workflow are explained in following slides, but here we simply want to emphasize the roles and context for traditional VV&A compared to the new "M&S Results Credibility Assessment" activity. A primary distinction between these two is that VV&A should be conducted before M&S use, and credibility assessment can only be conducted after M&S use.

Basic Chain-of-Custody of “M&S Results Credibility”



We define Credibility as the union of three key criteria 1) is the M&S built well?, 2) is the M&S fit to the intended use?, and 3) is the M&S used right?

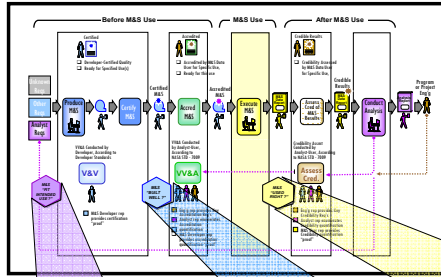
These three key criteria are defined as follows:

- 1) Built Well means that all M&S are verified, validated, and use good design and development practices,
- 2) Fit Intended Use means there is a good match between the M&S parametric capabilities and performance and the specified analysis problem requirements,
- 3) Used Right means the execution of the M&S is well orchestrated, managed, and documented; the users are properly qualified; and work is appropriately peer reviewed.

The Built Well and Fit Intended Use criteria are assessed as part of traditional VV&A activities, and should be done *before* the M&S is used. The Used Right criteria are assessed *after* the M&S has been used to produce an M&S results report. The credibility-assessed report is then handed over to the analysts for their use in conducting the problems/issues for which the M&S data was originally requested. Note that since the analysis activity may use data other than data from M&S, the credibility of the analysis is not 100% attributable to the M&S credibility.

The next slide defines the sub-criteria for each of these three credibility criteria.

Basic Chain-of-Custody of “M&S Results Credibility”



Credibility of M&S Results

MEANS : Implicated M&S are

- BUILT WELL
- FIT TO INTENDED USE, and then
- USED RIGHT

so . . . what is the “proof” of credibility ?

Accreditation “Proof” Established using Classical VV&A Criteria

- Should be done before M&S use -

Credibility “Proof” Established using added credibility criteria

- Can only be done after M&S use -



- Referent scope, scale, detail match intended use ?
- Referent parameters, limits match intended use ?
- Referent Parametric sensitivities match intended use ?
- Referent Functions, interactions match intended use ?
- Inputs and outputs match intended use ?
- Coded model validated against referent system ? M&S results correctly assessed & qualified with respect to limitations and caveats ?



- Requirements documented ?
- Requirements verified ?
- Requirements documented ?
- Conceptual model verified ?
- Coded model verified to rqmts ?
- Coded model tested to rqmts ?
- Coded model validated against similar systems ?
- Design & Development process managed well and documented (including configuration mgmt)?

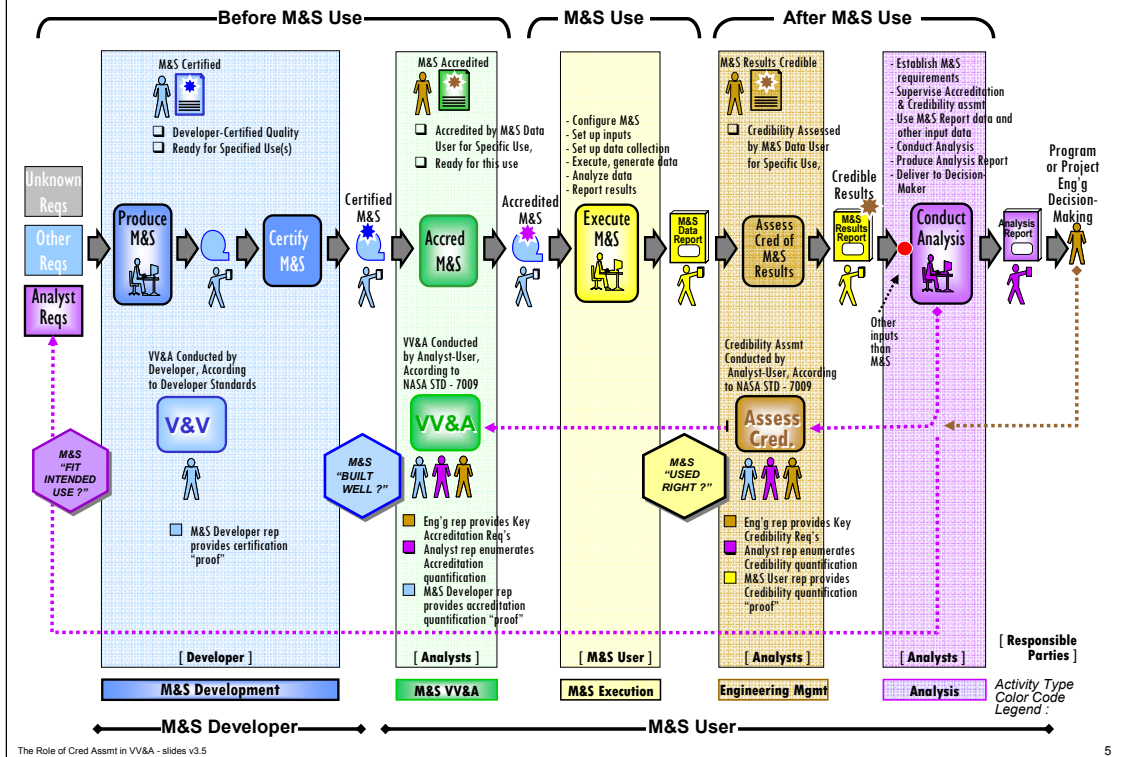


- M&S application & setup defined ?
- Inputs & outputs verified ?
- M&S correctly executed ?
- M&S data properly collected and analyzed ?
- M&S operator/analyst qualified ?
- All input and configuration data well defined & managed ?
- Appropriate peer reviews ?

Each of the three key criteria for assessing VV&A/Credibility have sub-criteria, as listed in this slide.

The logic behind this criteria is as follows:

Before an M&S should be used, it should be ascertained that it is fit to the intended purpose, otherwise there is no sense to continue. If it is fit to the intended purpose, then it should be built well – meaning a quality managed process and verified requirements and validated performance. If these first two criteria are satisfied, then the M&S may be accredited FOR use. The credibility of the results AFTER the use of the M&S must then be assessed against criteria that provides assurance of the proper use of the M&S. If the M&S passes this criteria, the M&S results are considered credible, with credibility traceable back to the M&S requirements.



The Role of Cred Assmt in VV&A - slides v3.5

The basic entities of the credibility workflow are explained briefly below.

Starting on the left side, with the M&S developer (blue column), the M&S is developed and quality-certified for use, using the developer's V&V methods, and based upon M&S requirements from one or more sources. [The detailed sub-activities inside this activity are described further in slides 6 and 7 following.] This developer activity results in a "certified for [specified requirements] use" status of the M&S.

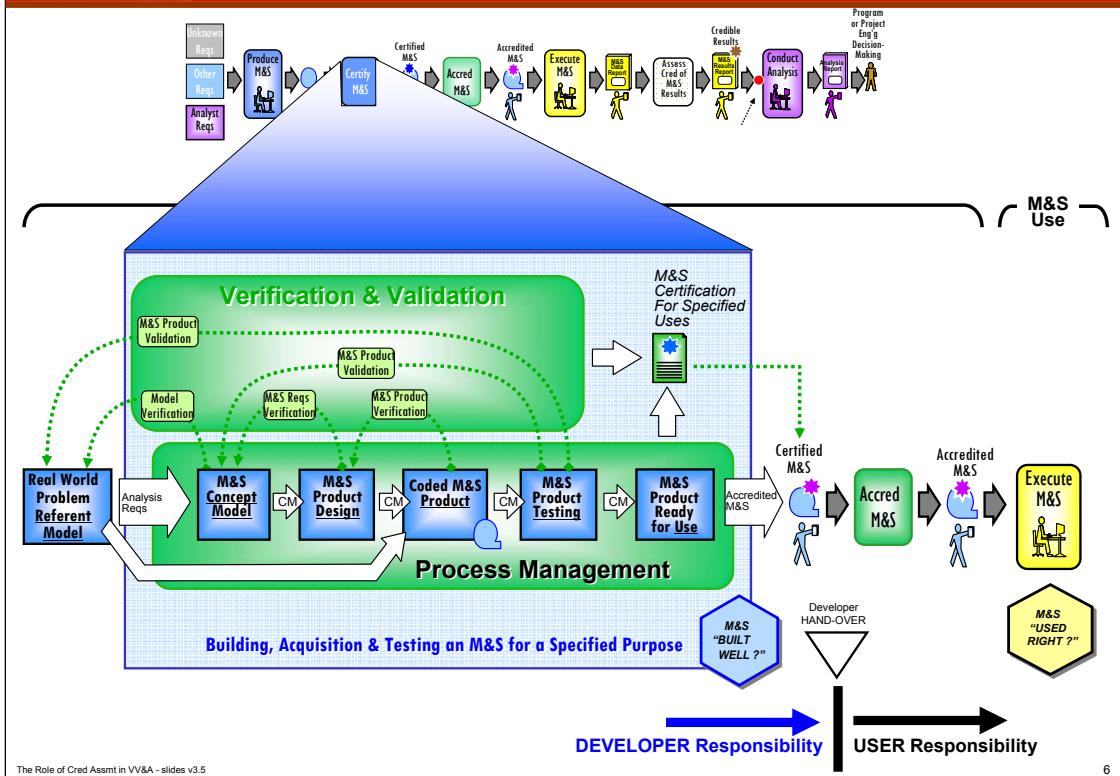
Regardless of who developed the original M&S, NASA then conducts its own, conventional VV&A, using its own criteria set (as described previously). This is the **Accredit M&S** activity (green column) in the workflow

The first step in **Accredit M&S** activity is to set up the VV&A responsible parties forming an accreditation team. The result of this is indicated by the small colored-coded person icons at the bottom of the diagram. A NASA project person (brown), representing the consumer of the analysis report that used the M&S results, acts as the accreditation authority, and sets the key accreditation requirements. These are next enumerated and quantified by the analysis report team [M&S using group] representative (purple), who acts as an accreditation agent. Finally, a developer's representative, called a V&V agent (blue), is assigned by the developer to be responsible for collecting evidence required for "proof" of achievement of the specified accreditation criteria.

Accreditation is then accomplished by this accreditation team. The accreditation agent guides and supervises the developer's V&V agent as necessary to achieve the development team's collection of evidence required for "proof" of achievement of the specified accreditation criteria. The accreditation agent then assesses this proof to determine the accreditation rating and any associated caveats and conditions on the use of the M&S, and puts this in an accreditation report to be signed by the accreditation authority. This activity culminates in the M&S being accredited by the project for the specified use under the specified conditions and caveats.

The M&S is then executed (yellow column) to produce the data required by the analysis activity (purple). However, before the M&S results report is used by the analysis activity, the M&S results must be assessed for credibility. This process is essentially the same as the activity previously described for VV&A, except that now the collection of "proof" for the *M&S Used Right* criteria is done by the M&S user team. This proof is assessed by the same accreditation agent as the previous VV&A. The result of this activity is that the M&S results are established as "credible", with possible conditions and caveats of data use for the specified analysis activity.

The workflow has been described in a serial sequence, but in fact sometimes many of these activities are performed in parallel, or even reverse sequence. The danger in out-of-sequence activity lies in discovering "after the fact" that some earlier work will have to be repeated, causing everything following that to require re-checking and possible rework – sometimes at great cost. Program schedule, cost, and issues drive the actual sequence degree of use.

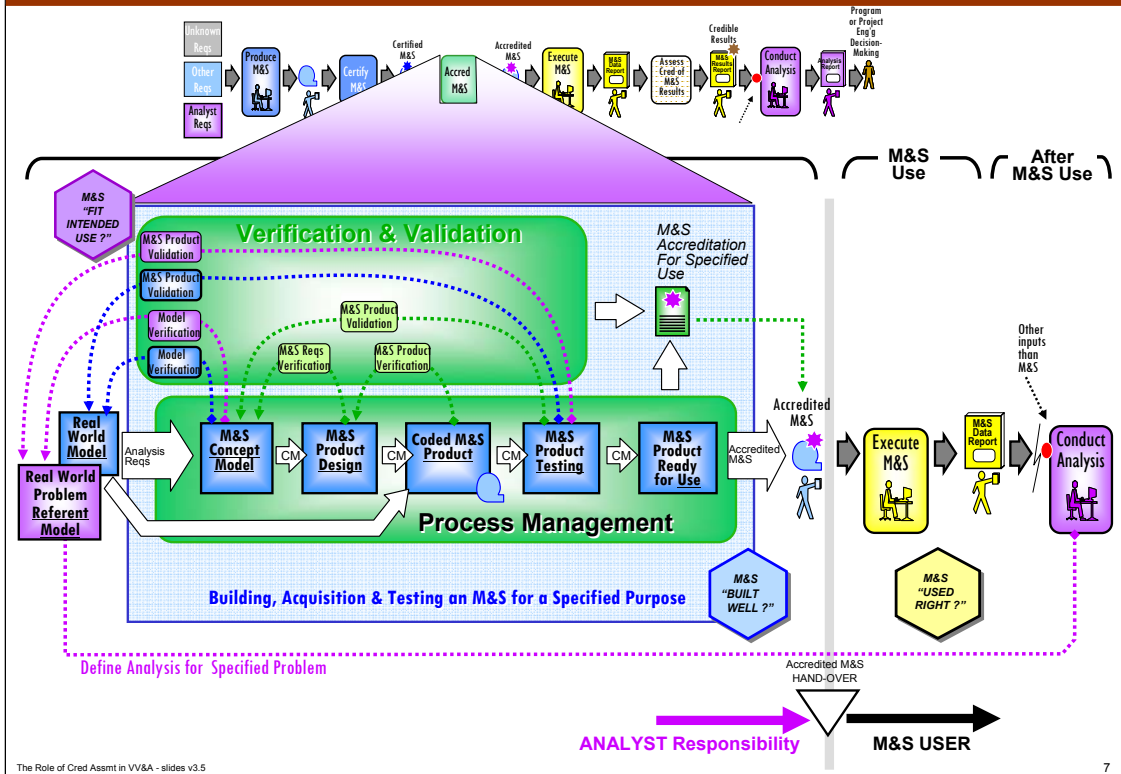


This slide expands on the first column of the roadmap, dealing with M&S development, acquisition, testing, and V&V for certification, which we refer to here as the “developer’s view, before M&S use”. These activities are all the responsibility of the developer, before handover to the using group.

Key components for V&V are the same for developer certification and user group accreditation, and this criteria was enumerated in the M&S Built Well criteria set described previously. “V&V” provides evidence for proof of M&S requirements verification and performance validation (small green activity boxes in middle left), and also that the M&S is built to standards and under quality process management sufficient to ensure the proper type and quality of data can be provided to the using group.

The illustration highlights the importance of end-to-end configuration management (CM), which is considered a part of process management.

An often overlooked but critical component of verification activities, is verification of the M&S concept model against the real world problem model, or “referent”, as shown on the left of the diagram. This verification is key to providing assurance that the correct model is being built, regardless of how good it is being built.



This slide expands upon the previous slide, but now from the viewpoint of the problem analysts needs.

The developer may use its own “V&V” methods for this certification, different than the sponsoring organization (NASA in this paper). The developer may also have used M&S requirements other than those now needed by the present investigator-analyst for the problem at hand (purple). This is why NASA conducts its own VV&A following the developer certification, to recheck that the M&S was developed and V&V’d to the correct requirements set.

To highlight these distinctions, we have added another set of verification and validation boxes to the diagram (in purple on the left), emphasizing the possible differences between requirements the M&S was originally built to and those now required for the problem at hand.

A note here about Fit to Intended Use.

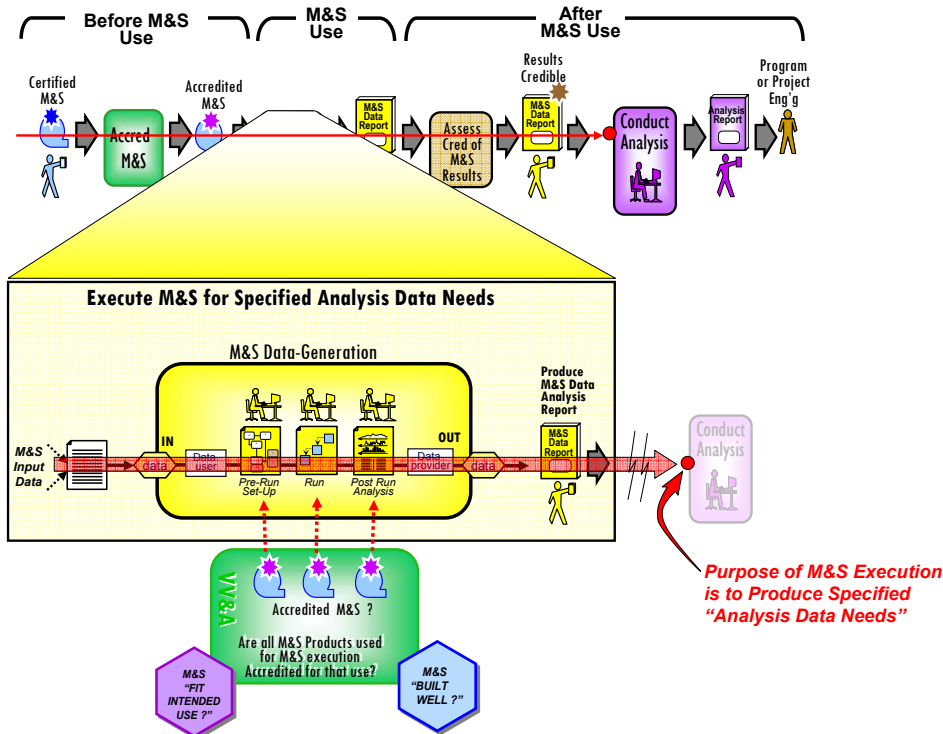
The main workflow, at the very top, shows there are three possible paths of entry to an M&S product development, 1) an M&S is built to unknown requirements (gray), 2) an M&S is built upon some other known requirements (blue), that may not be the present requirements, and 3) an M&S is built upon requirements explicitly set for the analysis problem at hand (purple).

Obviously, if an existing M&S product is to be chosen, AND if it has NO defined analysis problem “real world” concept model, it can not be determined if the M&S product matches the requirements, and is Fit to Purpose, and accreditation is not possible, and M&S credibility fails.

The only way around this is to reverse engineer the missing analysis problem concept model from design information for the chosen M&S product, and then compare that to the presently needed model.

If the M&S was developed to some specified real world model, but not the present specified problem model, and was verified and validated against it, it must still be determined if this model is the same as the presently needed problem analysis model, to satisfy Fit to Purpose criteria.

If the M&S was developed to the specified real world problem referent model, then all verifications and validations will be correct, and directly support accreditation.



This slide expands upon the “M&S Use” (yellow) column of the workflow diagram, and shows the key features of M&S use that are to be assessed against the “Used Right” criteria set.

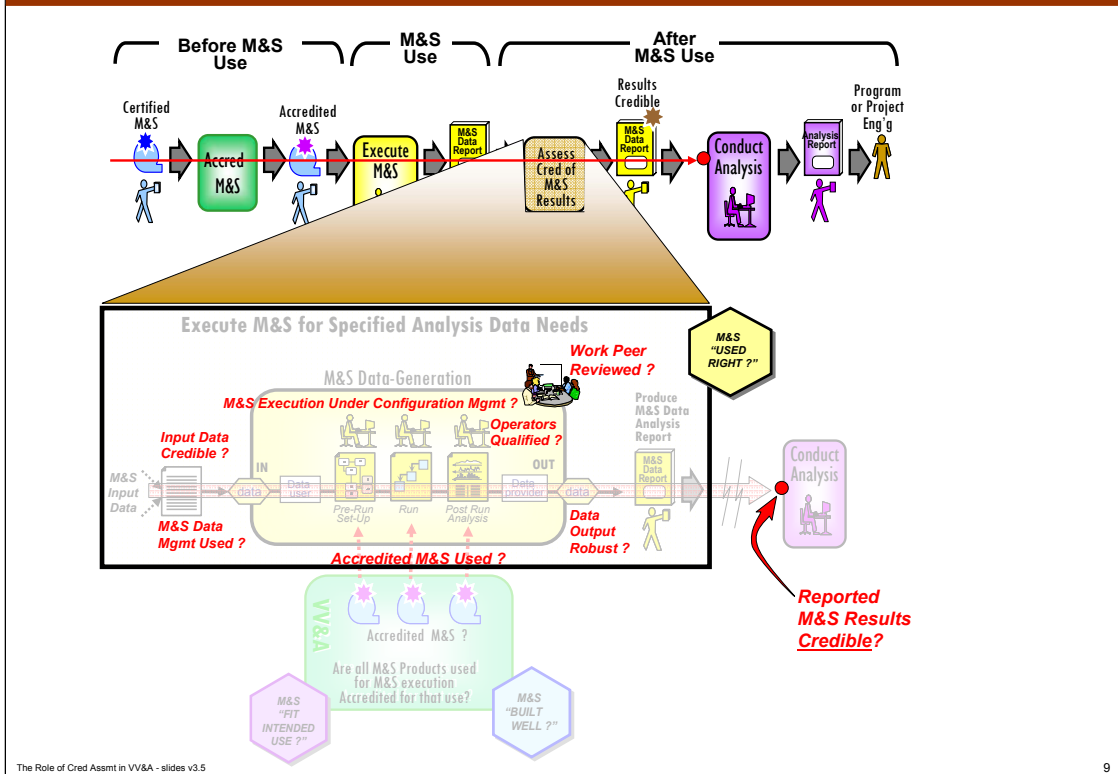
The bottom part of this diagram repeats an indication of the previous “Accredit M&S” activity, showing that all M&S resources used should have passed the previous accreditation activity before they are used in M&S execution for the problem at hand.

The focus of satisfying Used Right criteria should be that, given the use of accredited M&S resources, the M&S execution provides credible results for the specified analysis data needed from the M&S.

The red arrow through the middle of the M&S execution and reporting activities is used to emphasize that all data inputs, data preprocessing, M&S data generation, M&S data post-processing, and M&S data reporting are on a “critical path” that must be under data management, and provide traceability from reported data to all other data and processing used.

Key points for Used Right are that the input data is credible, operators are qualified, M&S resources are all accredited for this use, and the M&S operation and reporting is peer reviewed.

The detailed Used Right criteria were introduced in an earlier slide, and are not repeated here.



This slide expands upon the “M&S Use” (yellow) column of the workflow diagram, and shows the key features of M&S use that are to be assessed against the “Used Right” criteria set.

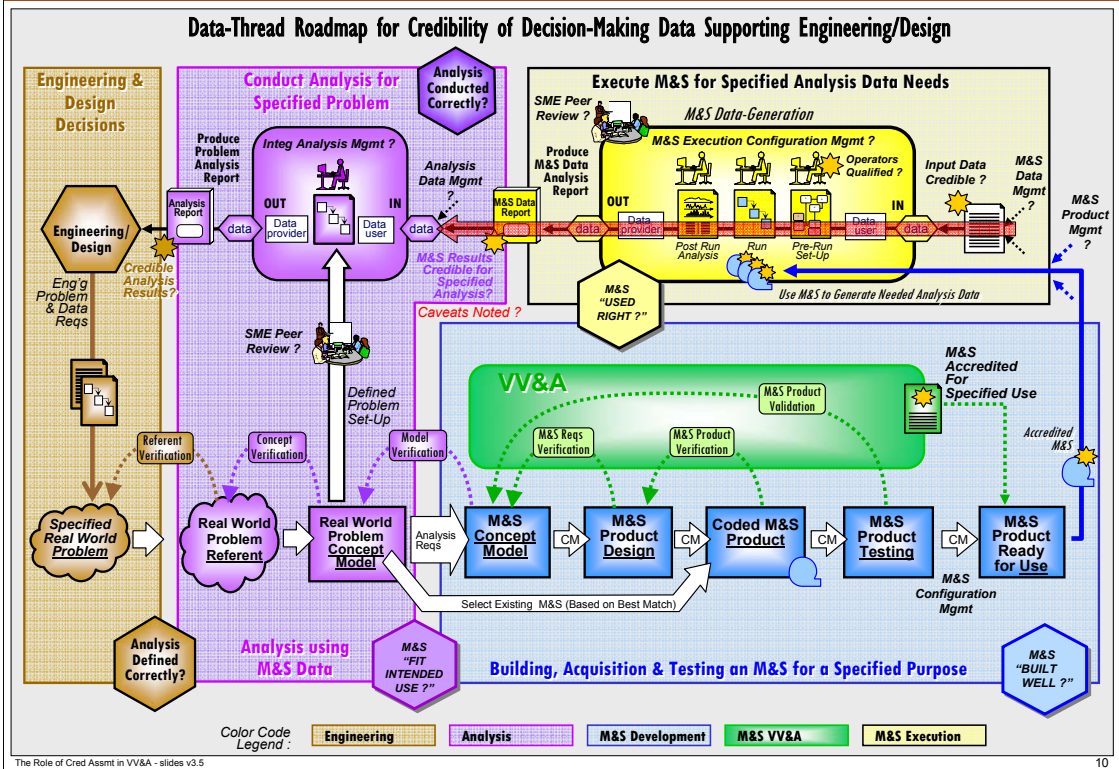
The bottom part of this diagram repeats an indication of the previous “Accredit M&S” activity, showing that all M&S resources used should have passed the previous accreditation activity before they are used in M&S execution for the problem at hand.

The focus of satisfying Used Right criteria should be that, given the use of accredited M&S resources, the M&S execution provides credible results for the specified analysis data needed from the M&S.

The red arrow through the middle of the M&S execution and reporting activities is used to emphasize that all data inputs, data preprocessing, M&S data generation, M&S data post-processing, and M&S data reporting are on a “critical path” that must be under data management, and provide traceability from reported data to all other data and processing used.

Key points for Used Right are that the input data is credible, operators are qualified, M&S resources are all accredited for this use, and the M&S operation and reporting is peer reviewed.

The detailed Used Right criteria were introduced in an earlier slide, and are not repeated here.



This slide presents the previously described M&S credibility workflow activities into a larger context workflow, showing the engineering design information that precedes definition of the real world analysis problem referent, as well as the integrated analysis activities that span the use of multiple M&S operations, and provide credible decision support information back to engineering and design.

In this case, the workflow starts at the upper left, goes counterclockwise down, to the right, back up to the top, and left back to the beginning in the upper left.

It is noteworthy that problem analysis and reporting do not typically rely totally upon any one M&S results report, and therefore M&S results credibility ends with the M&S data report.

A similar workflow for analysis report credibility could be developed for problem analysis credibility assessment as was done here for M&S results credibility. If that were to be done, there would more than likely be many similarities between the two processes. The main difference would be in the determination of the responsible parties.

REFERENCES

1. NASA, *Standard for Models and Simulations*, NASA-STD-(I)-7009, (December 2006).
2. DoD Modeling and Simulation Office (DMSO), *DoD Modeling and Simulation Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide*, Build 3.0, (September 2006).
3. DND SECO, *Modeling and Simulation Verification, Validation and Accreditation (VV&A) Guidebook* (DRAFT), (May 2003).
4. NASA - Exploration Systems Mission Directorate (ESMD), Constellation Program (CxP), Modeling And Simulation/ Data Architecture Office (MS/DAO), *Recommended Practices Guide (RPG) For Modeling And Simulation Verification, Validation, And Accreditation* (Baseline Draft), CxP 70148, (January 2007).