

Measuring Modeling and Simulation (M&S) Credibility

DAVID PHILLIP HENRY

B.S.E.E. Drexel University, 1979

MS Computer Science Drexel University, 1984

Portion of the thesis submitted in fulfillment of the requirements for the degree of Doctor of Modeling and Simulation at the University of Central Florida Orlando, Florida Spring Term 2011

Major Professor: Dr. Peter Kincaid

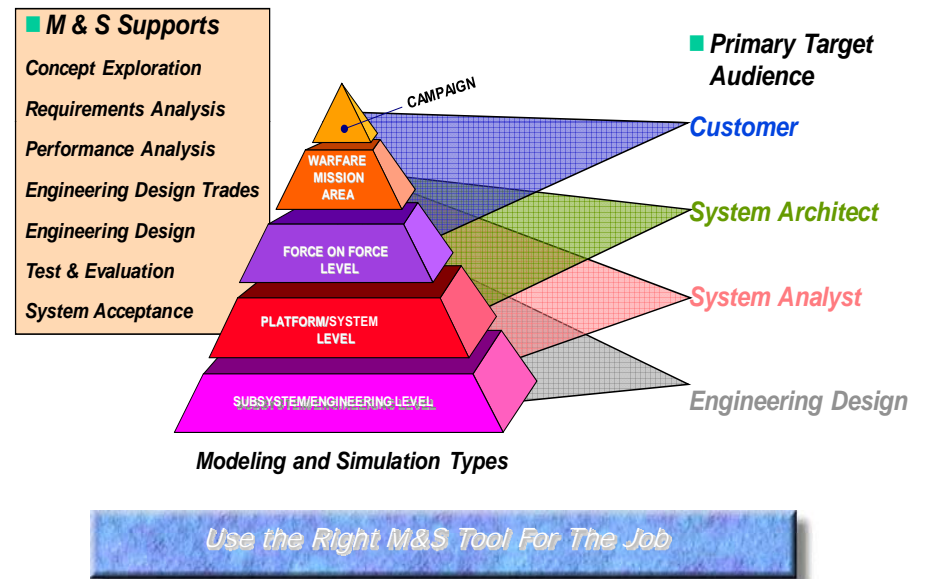
Statement of Problem

How does M&S developers/analysts offer reasonable grounds for being believed?

System Level M&S Context (Hierarchy)

“The M&S Challenge”

- Developing “acceptable” representations of reality
- Producing desirable and qualified results



This is the crust of the M&S issue of credibility which is no different for when a person is recognized as the expert or keeper of the knowledge in a particular field or domain up and down the M&S context pyramid.

Background

- Things to overcome: (the way it is)
 - Traditionally placed in the engineering tool category
 - The benefit is at the task level (one who can wield the hammer is the benefactor)
 - The “it-wasn’t-built-here” syndrome
- Things to reach for: (the way to go)
 - Increase to primary use (reduce cost, reduce risk)
 - Re-use (stop re-inventing a wheel)
 - A means to gain greater sophistication/complexity

Takes an incredible level of knowledge to transport the needed level of belief.
The more complex the more knowledge base required.

The Framework

Knowledge Base Level Concept

Development Life Cycle

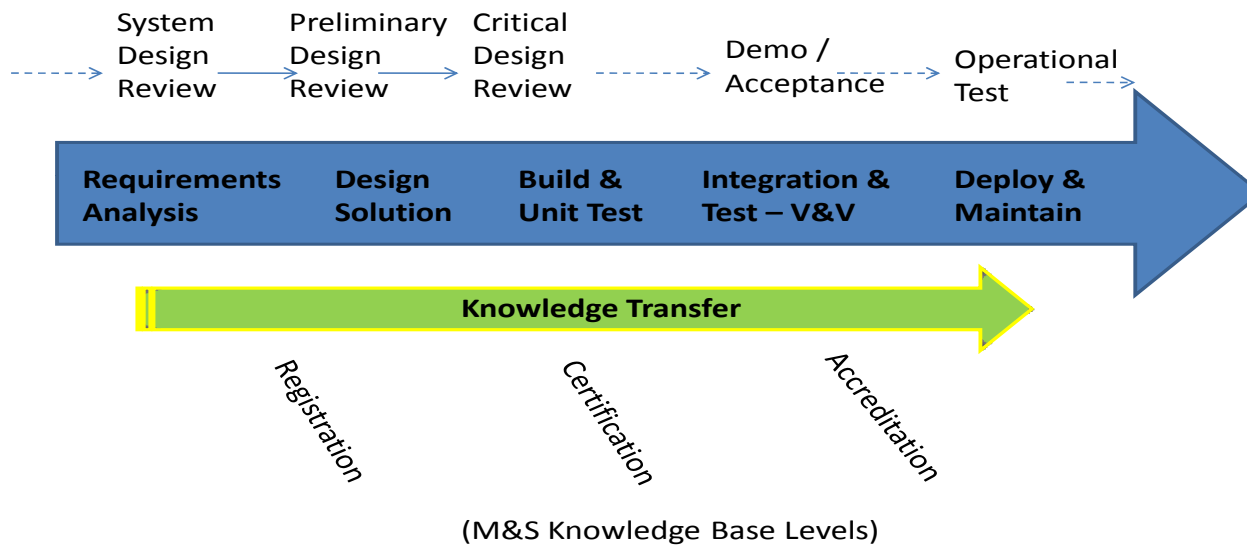


Figure 3

M&S knowledge transfer through three levels of representation:
-----Registration-----Certification-----Accreditation-----

Research Questions

(in addressing the M&S credibility issue)

- Can a framework be provided institutionalizing common approaches as a standard in order to represent consistent knowledge levels of M&S?
- **What measures will be needed (and do they exist) to support this knowledge base framework and provide the means to delineate between the knowledge levels?**

The evolution from an engineering tool concept to a knowledge base concept is called for especially if M&S is to grow as a representative academic field or engineering branch.

Quantifiable Measures

- Modeling and Simulation Qualification **(50)** – measuring M&S health of development
- Modeling and Simulation Readiness **(30)** – measuring M&S quality aspects
- Modeling and Simulation Producibility **(20)** – measuring M&S attribute value

Measures can be used to delineate between models and simulations under a defined weighted **(point)** system for measuring credibility

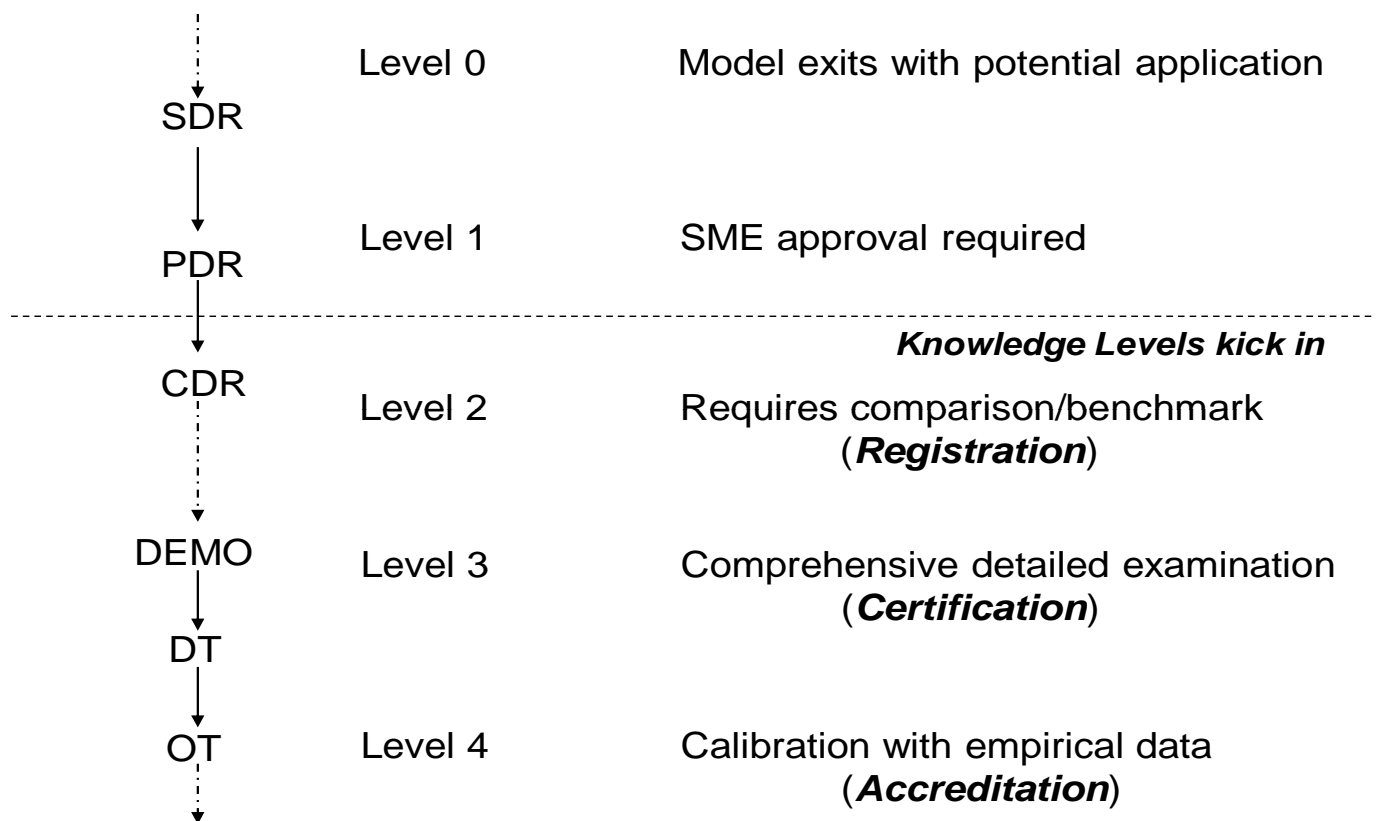
M&S Qualification

(Maturity Compliance)

- Addresses acceptance portion by management and/or client – a means to qualify acceptance
- Five (5) descriptive levels with regards to the knowledge level representation
- Development activity and rigor are the criteria (configuration management, technical reviews, verification & validation, integration & test)

A total score of 50 points maximum. Based on development products such as documented plans, procedures, reviews, memos and recorded results.

M&S Qualification Levels (0-4)



The new DoDI 5000.02 gives significance to this approach.

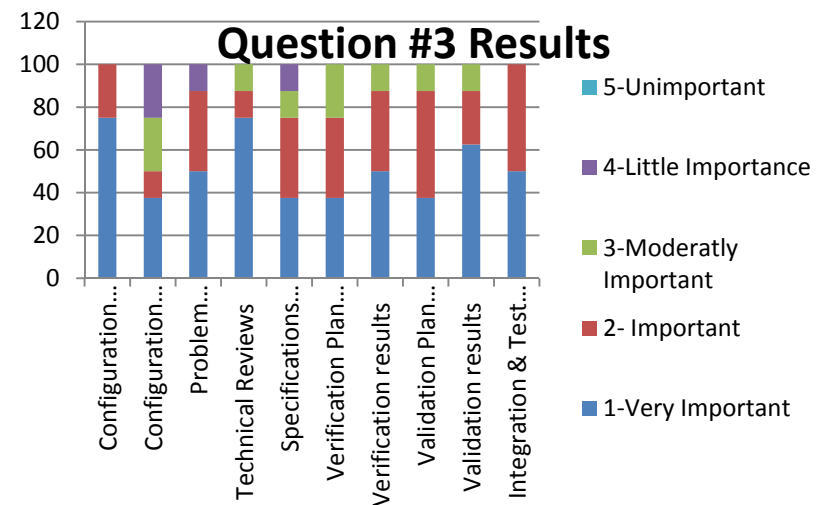
The levels of progression a model or simulation development could be planned for.

M&S Qualification Measure

Scoring Scale

Official Approval Letter	2
CM Plan (approved)	5
Read-Me Files	2
Rational – memos, studies	2
DM Plan (approved)	3
CCB – minutes, notes	2
Problem Reports/Tracking	4
Technical Reviews	5
Specifications (approved)	3
Verification Plan (approved)	2
Verification results	3
Validation Plan (approved)	3
Validation results	4
Overall Test Plan (approved)	3
Implementation Procedure	3
Integration & Test Procedure	4

Questionnaire used to survey regarded M&S community of interest resulted in assisting in the development of the scoring scale.



NOTE: No real scale but the higher the number the more important that product is to the essence of scoring this aspect

M&S Readiness

(Intended Use)

- Addressing quality aspects from 3 dimensions
 - Degree of readiness
 - Readiness criteria
 - Purpose complexity
- Each scored individually and then summarized
- Represents in addition a release-ability measure with regards to the purpose

A total score of 30 points maximum combined over the three dimensions based on the individual scores and specific criteria

Degree of readiness

- Readiness degrees - 4
 - 0 = Not ready
 - 1 = One time use only
 - 2 = Small group of users (community, single org.)
 - 3 = Wide use of many users (multiple orgs)
- Each degree is the value of scoring contribution

Wide use means some significant amount of documentation is available and supported by other things such as level of process compliance, V&V efforts.

Readiness Criteria

- Readiness criteria (6)
 - Completeness of VV&A
 - Completeness of Documentation
 - Level of configuration management control
 - Maintainability
 - Type of architecture
 - Coded in a commonly used and supported computer language
 - Level of fidelity
 - Maturity of algorithms and design representations

Verified as high (3), medium (2), and low(1) ratings for scoring.

Purpose Complexity

- Specific purpose identifiers
 - Back of the envelope analysis (1)
 - Concept evaluation (3)
 - Requirements flow down or allocation (3)
 - Final design evaluation (4)
 - Predicting lab test results (6)
 - Predicting live test performance (7)
 - Ready for use in a distributed simulation (8)
 - Suitable for direct use in a tactical system (8)

Scored as low to high in terms of complexity of use.

The Six Levels Of Release-ability

<i>M&S Readiness Level</i>	<i>Description</i>
1. First Order Simulation	Lowest level of modeling and simulation readiness. Suitable for back of the envelope analysis for quick analysis of specific issues. Does not have a high degree of V&V, documentation or software maintainability. High variability in the quality of the software design and architecture. Level of system fidelity and maturity of algorithms highly variable. Often limited to one time use. Use usually limited to one or very few users.
2. Low Fidelity Parametric Simulation	Start of a deliberate software design and architecture to address a specific set of issues. Suitable for use in concept evaluation studies. Some verification of code, low level of in line code documentation, low degree of software maintainability, and no formal configuration management. Low level of fidelity, parametric representation of design, and use of immature algorithms. Use usually limited to one or very few users.
3. Medium Fidelity Parametric Design Simulation	Increasing formalization and process in the development of software design and architecture to address a more generalized set of issues. Suitable for use in requirements flowdown or allocation analysis. Medium level of verification of code, low level of validation of the model, medium level of in line code documentation, some external documentation such as read me files, low degree of software maintainability, and some formal configuration management. Medium level of system fidelity base on parametric representation of design, and use of more mature algorithms. Used within a single organization by a small group of users.
4. Medium Fidelity Specific Design Simulation	Medium level of formal process used in the development of software design and architecture to address a specific set of issues. Suitable for use in final design evaluation. Medium level of verification of code, medium level of validation of the model, medium level of in line code documentation, more formal external documentation such as simple users manual, medium degree of software maintainability, and increasing level of formal configuration management. High level of system fidelity, specific representation of design, and use of more very mature algorithms. Used within a small group of organization by a large number of users.

The Six Levels of Release-ability (cont)

M&S Readiness Level

Description

<p>5. High Fidelity Parametric Design Simulation</p>	<p>Medium level of formal process used in the development of software design and architecture to address a specific set of issues. Suitable for predicting lab test results. High level of verification of code, medium level of validation of the model, high level of in line code documentation, formal external documentation such as a full set of users programmers manuals, medium degree of software maintainability, and SEI level 3 or better formal configuration management. High level of system fidelity, object oriented highly parametric representation of design, and use extremely mature algorithms. Extensive use by a wide range of organization and a large number of users.</p>
<p>6. High Fidelity Trusted Simulation</p>	<p>Medium level of formal process used in the development of software design and architecture to address a specific set of issues. Suitable for predicting live test results, use as part of a distributed simulation, or for direct use in a tactical system. High level of verification of code, high level of validation of the model based on previous test data, high level of in line code documentation, formal external documentation such as a full set of users programmers manuals, high degree of software maintainability, and SEI level 3 or better formal configuration management. High level of system fidelity, object oriented highly parametric representation of design, and use extremely mature algorithms. Formal federation agreements in place to include a FOM and appropriate SOMs. Extensive use by a wide range of organization and a formal users group conducting TIMs.</p>

The description for each level depicts the relationship of combining the portions of this measure to represent a release order ranking

Distributed Simulation

- Attribute representing a computing environment that models or simulations executes in
- Following are available mechanisms and/or infrastructures as check points to determine score
 - HLA (Higher Level Architecture)
 - CORBA (Common Object Request Broker Architecture)
 - RTI (Real Time Interface)
 - DISN (Defense Information Systems Network)
 - TENA (Test and Training Enabling Architecture)

As a guideline, it only takes one to get the 4 points for this portion.

M&S Producibility

(Multiple Use/Application)

- Deals with attributes that are available or comprise of some value of the M&S
- Measure is a weighted subjective parameter of a roll-up of the identified acceptable attributes
- Attributes are special categories with respect to models, simulation, and software development in general

Producibility Attributes

- Random Number Generation (4)
- Distributed Simulation (4)
- Platform Dependency (4)
- Real Time Sensitivity (4)
- Application Dependency (4)

A total score of 20 points maximum with guidelines for individual attribute assessments for scoring each.

Random Number Generation

- Following are check point properties to determine a good arithmetic random-number generator
 - Distributed uniformly and no correlation with each other
 - Fast and low storage
 - Produce exact streams
 - Produce several separate streams

Distributed Simulation

- Representing a computing environment that models or simulations executes in
- Available mechanisms/infrastructures as check points to determine a score
 - HLA
 - CORBA
 - RMI
 - DISN
 - TENA

NOTE: It only takes one to get the 4 points for this portion.

Platform Dependency

- The listing of platform availability will have a value setting
- The typical platform list (could be more added over time)
 - PC
 - UNIX
 - MAC

More than one gets the 4 points for this portion.

Real Time Sensitivity

- Attribute that a model or simulation has the ability to maintain/support real time representation.
- Triggers would look such things as
 - Hardware-in-the-loop (HIL)
 - Computer-in-the-loop (CIL)
- This portion of measure also goes hand-in-hand with the Distributed Simulation attribute.

It only takes acknowledgment that this capability exists in some fashion to get the full 4 point score for this portion.

Application Dependency

- Attribute in that a model or simulation is built using a general purpose language and runs under a common operating systems
- The following are check points to determine a score:

- Languages

- C/C++
- Visual Basic
- FORTRAN
- Pascal
- Java
- Smalltalk

- Operating Systems

PC - Windows
MAC
UNIX
LINUX

These are examples of some of the common languages and OS for determining a score. The more the better, but any two from each category would get a maximum score of 4

Backup Slides

Literary Review

Good stuff but one must read between the lines!

- Standards, Policies, and Procedures
- Education and Academia
- Industry, Societies, and Associations
- Other Contributors

A far reaching breath and depth of general literature on M&S but this dissertation undertakes new ground in defining the platform for a true credibility measure.

Standards, Policies, & Procedures

- U.S. Congressional Modeling & Simulation Caucus is the pre-essential hope of establishing firm and concrete policies
- Department of Defense an early champion of standards (MIL-STD-3022, DoD Instruction 5000.02)
- Department of Navy forefront runner for M&S accreditation. Other DoD offices have picked up the proactive attitude in the same communities of practice.
- The top formalized standard go-getter is the National Aeronautics and Space Administration (NASA) with their NASA Technical Standards for Models and Simulations (NASA-STD-7009)

NASA Key Aspects of Credibility Assessment Levels

Level	Verification	Validation	Input Pedigree	Results Uncertainty	Results Robustness	Use History	M&S Management	People Qualifications
4	Numerical errors small for all important features.	Results agree with real-world data.	Input data agree with real-world data.	Non-deterministic & numerical analysis.	Sensitivity known for most parameters; key sensitivities identified.	De facto standard.	Continual process improvement.	Extensive experience in and use of recommended practices for this particular M&S.
3	Formal numerical error estimation.	Results agree with experimental data for problems of interest.	Input data agree with experimental data for problems of interest.	Non-deterministic analysis.	Sensitivity known for many parameters.	Previous predictions were later validated by mission data.	Predictable process.	Advanced degree or extensive M&S experience, and recommended practice knowledge.
2	Unit and regression testing of key features.	Results agree with experimental data or other M&S on unit problems.	Input data traceable to formal documentation.	Deterministic analysis or expert opinion.	Sensitivity known for a few parameters.	Used before for critical decisions.	Established process.	Formal M&S training and experience, and recommended practice training.
1	Conceptual and mathematical models verified.	Conceptual and mathematical models agree with simple referents.	Input data traceable to informal documentation.	Qualitative estimates.	Qualitative estimates.	Passes simple tests.	Managed process.	Engineering or science degree.
0	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.
	M&S Development		M&S Operations			Supporting Evidence		

(Snap shot from NASA-STD-7009 document)

Education and Academia

- Support available to champion M&S along professional/career path, but not much in terms of post educational development
- Belief is that being certified (M&S developer) should be enough to give credence to a development effort but not the result/product

University of Central Florida (Orlando) and Old Dominion University (Norfolk, VA) are two higher educational institutions offering officially recognized graduate level degree programs in M&S (masters and PhD).

Industry, Societies, Associations

(some key components)

- National Training and Simulation Association (NTSA) supports the Certified Modeling & Simulation Program (CMSP) and is also the right-hand organization to the Congressional M&S Caucus in supporting the execution of the Caucus's agenda.
- The International Council on Systems Engineering (INCOSE) is one domain group that tends to these special areas (i.e. M&S topology, MBSE)
- Private sector/industry are leveraging M&S – product-wise, R&D in terms of new development efforts

There are also steady conferences and working groups centered around and focused on M&S – but none directly specific to the topic at hand.

Other Contributors

- **Dennis P. Shea**, CNA Analysis & Solutions; Director of Information, Technology and Operations Team Advanced Technology and Systems Analysis
- **Dr. Averill M. Law**, President of Averill M. Law & Associates (Renown expert in simulation modeling)
- **Dr. David W. Kelton**, Director, Master of Science in Quantitative Analysis (MSQA) Program
Department of Quantitative Analysis and Operations Management University of Cincinnati
- **LCDR Harry M. Croyder**, SPY-1D(V) Models and Simulations Support Operational Testing In The Cornfield; co-author
- **William R. Ervin**, Models and Simulations Support Operational Testing In The Cornfield; co-author (recently retired LCDR); Director of Military and Marine at Dover Corporation – Pump Solutions Group; Norfolk, VA
- **David S. Mazel**, Models and Simulations Support Operational Testing In The Cornfield; co-author (Research Analyst at the time for COMOPTEVFOR USN); Department Manager at Technology Service Corporation; Washington, DC
- **LCDR Cris Miller**, DoN Area TBMD M&S Engineer
- **Michael Lee**, NTW M&S Engineer
- **Joe Uzdinski**, Lockheed Martin MS2; LM Fellow – M&S
- **John Shelby**, Systems Planning and Analysis Incorporated (SPA)
- **Phil Kyle**, Systems Planning and Analysis Incorporated (SPA)
- **Ellis Sutter**, Lockheed Martin – VA
- **Dr. Regina M. Greigo**; National Nuclear Security Agency (NNSA) – Technical Advisor and Sandia National Laboratories – Principal Member of Technical Staff

Research Method

Mainly a qualitative method using participant observations approach accompanied with selective open-ended interviews of key subject matter experts(working groups/associations). In addition, relying on intensive and specific literary search of the subject matter itself.

The plan of attack for this effort is to develop the specific meaning, the process significance, and the interconnection of each knowledge base level and develop the aspect measures to match the framework.

Results

- A knowledge base level framework defining the knowledge transfer markers along with the specific meaning, the process significance, and the interconnection of each knowledge base level was created.
- A set of quantifiable measures that are aligned with the knowledge base levels and assist in delineating between models and simulations in such things as development rigor, release-ability, and value was developed.

Registration

- Implement common practice to recognize models (subsystem/engineering) by a specific group/body of experts
- Institute formal process (as simple as):
 - Cataloguing models
 - Generating Read-Me files
 - Performing comparative analysis (similar algorithms/models)

Easy way to capture/present models of various types in multiple categories
THE KEY – Common Taxonomy

Certification

- Clearly and fully document characteristics of models/simulations (platform/system) beyond Read-Me files/cataloguing
- Must meet verification and validation requirements (customer satisfaction)
- Demonstrates adherence to a credible CM process

Most cases certified M&S will be built from registered M&S

Clear contribution to growth of maturity of M&S at this level

Larger community commitment and interest – Process Bar Raised

Accreditation

- Most widely recognized/implemented quality assurance method (Well Documented)

“Accreditation: the official determination that a computer model is acceptable for a specific purpose. [MORS]”.

- M&S (platform/system to warfare/campaign) must convey reality (as much as possible)
- M&S intent of use is clear and agreed to (official approval process)

Certification efforts direct contribution to this level

Knowledge level boundary crossed from “what” is represented to “how well”