

# Data Verification and Validation (V&V) for New Simulations

RPG Special Topic

11/30/00

## Table of Contents

<b>Introduction</b>	1
<b>Data V&amp;V Activities During M&amp;S Development</b>	1
<b>Determine M&amp;S Requirements Phase</b>	2
V&V Activity: Verify Requirements	3
<b>Plan M&amp;S Development Phase</b>	3
V&V Activity: Develop V&V Plan	3
<b>Develop Conceptual Model Phase</b>	4
V&V Activity: Validate Conceptual Model	4
<b>Develop M&amp;S Design Phase</b>	6
V&V Activity: Verify Design	7
<b>Implement and Test Phase</b>	8
V&V Activity: Verify Implementation	8
V&V Activity: Validate Results	9
<b>Prepare M&amp;S for Use Phase</b>	10
<b>References</b>	10
External Links in this Document	11
RPG References in this Document	11

## Introduction

Data verification and validation (V&V) activities are performed to ensure that [data](#) are appropriate for use in a particular simulation for a specific application.

- Data verification is conducted to ensure that the data selected are the most appropriate for the application and are properly prepared for the model.
- Data validation is conducted to ensure that the data accurately represent of the real world to be simulated.

In simulation, it is virtually impossible to separately evaluate the model being executed from the data used because it is the interaction of data and code that produces simulation results, making both responsible for simulation credibility.

This interdependent relationship between a simulation and its associated data dictates that data V&V activities be considered part of the overall M&S verification, validation and accreditation (VV&A) process [DMSO, 1998; IEEE 1278.4, 1997]. However, because of the specialized nature of data V&V, and particularly because of the large varieties of data subject areas, subject matter experts (SMEs), in particular the data producers themselves, are frequently called upon to assist in the data V&V process.

**Example:**

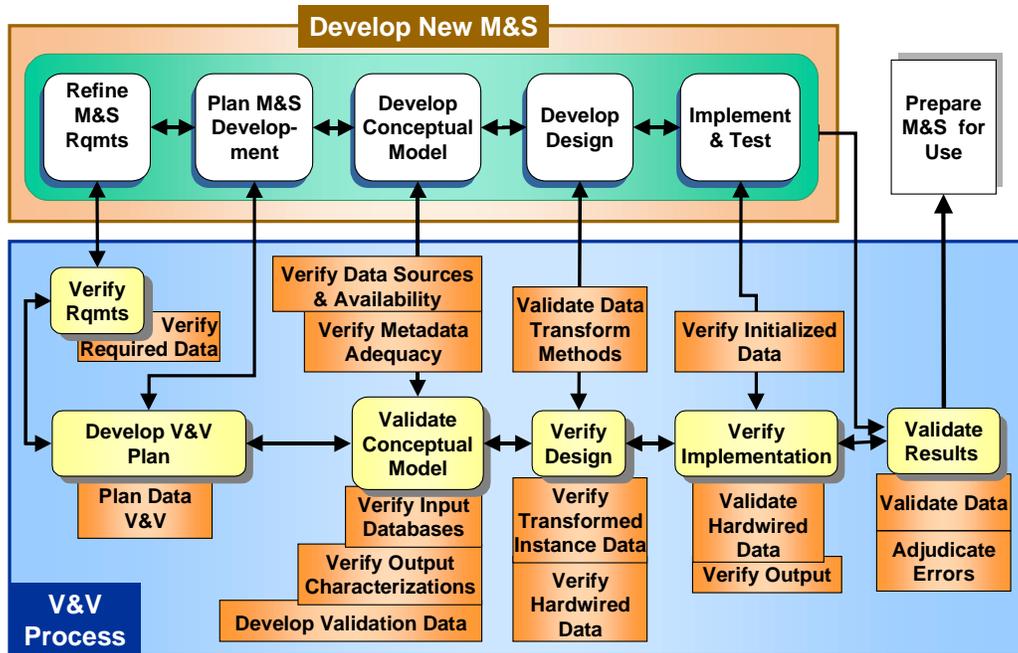
A military simulation will typically involve instance data describing the simulated natural environment, man-made obstacles, weather, force structure, system characteristics, system performance, behavior, command and control, etc.

Regardless of who conducts data V&V activities, they should work closely with the Developers and those performing M&S V&V activities.

## Data V&V Activities During M&S Development

The following sections discuss data-related activities and issues that should come up during the development and V&V assessment of a new model or simulation (M&S). A visual depiction of typical data V&V activities, overlaid on the V&V process for new M&S development, is provided in the [figure](#) shown below. Although data-related activities are each shown and discussed in connection with specific M&S development phases and V&V activities, these associations merely mark an appropriate opportunity for the data activity to take place. The sheer magnitude and complexity of data needed by a simulation suggest that data-related activities frequently begin earlier than recorded here and continue into later development phases in an iterative fashion. The approach recommended in this document is to discuss specific data activities as early as possible

(e.g., when sufficient information is available to perform the specific activity) rather than to tie data activities to specific themes. This approach is taken to emphasize the importance of early action on data V&V.



Data V&V for New M&S

## Refine M&S Requirements Phase

Data should be an important consideration early in the development of a model or simulation. Once the [requirements](#) of the application have been determined, the developer, user, and program manager should work together to establish the M&S environment by

- identifying constraints based on the simulation development, such as resource availability, timelines, etc.
- refining requirements, including identification and definition of the data needs of the simulation for the given application, such as the data subject areas, the needed level of fidelity, the types of data to be collected during execution of the simulation, etc. Data subject matter experts ([SMEs](#)) may be needed to help identify and define these data needs.

Example: One method used to capture data needs of the simulation is to prepare a Characteristics and Performance (C&P) Specification. The C&P Specification is a technical document that provides a complete description of needed data, including a set of sample instance data, and can be used for validation of the conceptual model.

Once the data needs have been identified and defined, the User should begin locating appropriate authoritative sources (e.g., Authoritative Data Sources (ADS) [[ADS website](#)]) and collecting candidate data and metadata [[DMSO](#), 1998; IEEE 1278.4, 1997]. Like the M&S requirements, data needs are subject to refinement and change as plans are finalized and the conceptual model is developed.

### ***V&V Activity: Verify Requirements***

#### **Verify Required Data**

As part of requirements verification, data needs should be reviewed to determine

- if the specified input data (as described by metadata) will support the M&S concept
- if the output data will support the needs of the application

The V&V Agent or appropriate SMEs should carefully review the required data to ensure they are sufficient to support the application. The review should determine if the appropriate data varieties and levels of aggregation, accuracy, fidelity, and [data quality](#) have been identified to address the needs of the application. This assessment should begin as early as possible to avoid difficulties with planning and development of the conceptual model.

## **Plan M&S Development Phase**

Data should be a major consideration when planning an M&S development. The types and quantities of output data to be collected can impact the design and implementation of the simulation. Input data availability, quality, and appropriateness for the application impact both the timeline and the cost of the overall program. Candidate input data sources should be identified as early as possible (e.g., as soon as the required data have been defined). When suitable data exist, they should be evaluated with respect to their compatibility with the simulation and their appropriateness for the application. If appropriate data do not exist, development planning should include additional time and resources for developing new data and the subsequent impact on development program milestones.

### ***V&V Activity: Develop V&V Plan***

Data V&V activities should be carefully planned to complement the different M&S development phases and M&S V&V activities. Early detection of data problems (e.g., nonavailability, inappropriate fidelity) can have a large impact on the model design. During V&V planning, the V&V Agent should work closely with the Accreditation Agent

to identify data-related assessment priorities and appropriate acceptability criteria. This work facilitates the selection of data V&V activities most suited for providing evidence to support the accreditation decision.

### Plan Data V&V Activities

- **Data Verification.** The affinity between model algorithms and their associated data should be of primary concern because of the direct impact such affinity has on simulation credibility; however, appropriateness and sufficiency of all data associated with the simulation -- [reference, hard-wired, and instance](#) -- should be considered in verification planning.
- **Data Validation.** Planning the data validation effort includes identifying appropriate validation activities and expected outcomes as well as identifying and evaluating appropriate validation to be used in the results validation. In addition, validation planning should address the impact of obtaining and evaluating validation data on development program timelines and resources.

## Develop Conceptual Model Phase

The conceptual model serves as a vehicle for transforming requirements into functional and behavioral capabilities and provides a crucial traceability link between the M&S requirements and the design implementation. Many data selection decisions take place during the development of the conceptual model. As the conceptual model is being built, the Developer should identify

- what input instance data are needed for specific individual algorithms, functions, and behaviors
- which input data should be hard-wired instead of used dynamically
- what specific data are needed to support the desired levels of fidelity for the application
- what output instance data need to be collected to support the application

Candidate data sources and databases for the input instance data and candidate sources for the hard-wired data. Once appropriate data have been found, their descriptions should be included in the conceptual model.

### ***V&V Activity: Validate Conceptual Model***

The [conceptual model](#) is validated to ensure that it adequately specifies both physical and behavioral aspects of the problem domain and appropriately traces operational requirements in the emerging design. Data availability and data appropriateness are key considerations during this phase because of their impact on model design. Several

data-related tasks that can be done during this phase are described in the following paragraphs.

### **Verify Data Sources and Availability**

As soon as data source candidates are identified, data source metadata should be reviewed to ensure the sources are authoritative, appropriate for the application, and able to provide the required data in a timely manner.

**Examples:**

The authoritative source for helicopter data may be different depending on whether the simulation will be used by the Army or the Navy.

The authoritative source for tank data may be different depending on whether the simulation is to be used in an unclassified training exercise or in a classified, force-on-force combat analysis.

### **Verify Adequacy of Metadata**

Once a data source has been verified (i.e., shown to be appropriate and available), the metadata should be reviewed to ensure that they provide the information needed to satisfactorily characterize the data for the given application (e.g., they should be reviewed for data currency and availability, quality assessment history, usage in similar applications, fidelity, and other quality characteristics).

### **Verify Input Databases**

The input databases (i.e., aggregated sets of all the input instance data) should be reviewed to ensure that they are adequate and complete and mapped to the algorithms, models, or simulation components in which they are to be used to confirm that they are appropriate for the application.

**Example:**

For a combat simulation, the input databases should be checked to make sure that attrition data are available for all possible weapon system/target combinations used in the application, that terrain maps include all features specified in the application, that systems characteristics are available for all systems, etc.

A sample input database might be generated to prove that the data producer can provide appropriate data in the proper formats and within the required timeline of the application.

If data or data source problems are discovered, the User may need to decide whether to modify requirements (and adjust the conceptual model), use available data that are not totally appropriate (i.e., accept the risk of reduced credibility), or undertake a data production effort to fill the void. The earlier such a decision is made the better. Once

the simulation is designed and built, changes to resolve data problems become much more costly and time-consuming.

## Verify Output Data Characterizations

The algorithms and models identified for use in the simulation should be examined to ensure they can provide output data to support the needs of the application. This review should include data characterization (e.g., fidelity, format, completeness) as well as methods of collection and preparation.

**Example:**

For an application concerned with missile lethality, the missile fly-out model selected for use in the simulation should be able to produce outputs for probability of kill (Pk) values as well as the probabilities of detection, track, engage, and closest point of approach.

## Develop Validation Data

Because results validation normally involves comparison of the results of a simulation to a referent (a codified body of knowledge about the thing being simulated based on M&S requirements), data describing the referent should be identified and collected or developed. Real-world empirical data are best (e.g., physical measurements, test range results, historical records) when available. However, when real-world data are not available, appropriate test scenarios should be developed and SMEs asked to provide reasonable, expected outcomes for the scenarios or use cases to be executed in the simulation. [Rothenberg et al., 1997]. These validation data, both empirical and expected outcome, will be compared to the actual output instance data of the simulation during results validation. For more information on referents and results validation, see the special topic [Validation](#).

## Develop M&S Design Phase

During the design phase of the M&S development, the Developer uses information and relationships articulated in the conceptual model to determine how to implement them in code. Often this is done iteratively. The preliminary design allocates the requirements derived from the conceptual model in hardware and software configuration items. This design then evolves into a more detailed design that includes the allocation of requirements to software components, the identification of the specific data sets and elements to be used, and the definition of interfaces among components and external systems.

Although data identification and selection processes should be initiated as early in the development process as possible (e.g., during conceptual model development), they

can continue throughout this phase. Additional considerations focus on how to prepare the selected data for use.

- For hard-wired data, considerations include selecting an appropriate format (e.g., floating or integer) and degree of accuracy (e.g., should  $\pi$  ( $\pi$ ) be represented by 3.14, 3.1416 or 3.141592?) to meet the needs of the simulation and then documenting the rationale.
- For input instance data, a major concern is determining what form the data are available in, what form the model needs the data to be in, and the best way to make the transformations. In most situations, instance data will have to be transformed from their original state.

Examples:

- Converting all rates of speed to kilometers per hour, all ranges to kilometers
- Converting the probabilities of acquisition, shot, hit, and kill into a single-shot kill probability
- Aggregating kills of individual classes of targets by individual classes of weapon systems over time into a overall kill probability

Appropriate transformation techniques should be selected (or developed) and validated.

- For **output instance data**, a major concern is ensuring the design (i.e., algorithms, code, interfaces) can produce, collect, and store the desired output data. In many situations, output data have to be transformed (e.g., aggregated, combined) to produce usable results. Whether data collection occurs external to the model or not, appropriate transformation techniques should be selected (or developed) and validated during design to make sure that proper collection is possible.

### ***V&V Activity: Verify Design***

The focus of design verification is to ensure that all features, functions, behaviors, and interactions defined in the design can be traced back to the requirements expressed in the conceptual model and that all requirements expressed in the conceptual model are articulated in the design. The primary data-related V&V activities associated with design verification are described in the paragraphs below.

### **Validate Data Transformation Methods**

The techniques used to prepare data for use are examined to ensure the data maintain their accuracy, fidelity, and integrity. Algorithms and techniques used to prepare each input instance database should be assessed to ensure that the transformation occurs correctly. Similarly, the methods used to collect and prepare output instance data, whether internal to the model or not, should be assessed to ensure that they can collect

and transform data properly. Generally, this activity does not investigate the validity of the output instance data; rather, it focuses on the validity of what has been done to the data.

### **Verify Transformed Input Instance Data**

This activity assures that transformed input instance data correspond to the original intent and are in appropriate form for use in the simulation. The producers of the original input instance data may serve as SMEs during this activity. Transformed output instance data should be traced to their origins to ensure they have been properly collected. Their characterizations should be verified to ensure that they provide appropriate, usable results.

### **Verify Hard-wired Data**

Hard-wired data are considered separately because typically they are independent, single instances of data or fixed constants to be used in specific algorithms. They should be verified in conjunction with the algorithms in which they are placed. Specific tasks include

- ensuring that the data have been obtained from an authoritative source
- verifying that the data, when transformed into the form required for use in the simulation, are appropriate for their intended purpose
- ensuring that any deviations from the required data are corrected or at least documented

## **Implement and Test Phase**

During this phase, the M&S design is realized in code using the actual hardware and data. The components or modules are built, tested, and integrated. The actual input instance data sets are initialized and tested.

### ***V&V Activity: Verify Implementation***

Requirements are traced to the implemented software components, individual algorithms and components are tested to ensure that they perform as designed, and data/code relationships are reviewed for appropriate operation.

### **Verify Initialized Data**

The initialized data sets (i.e., the aggregated sets of transformed input instance data, in their initialized or start-up state) are checked to ensure that they continue to correspond

to the original data, have been transformed as intended, and have maintained the accuracy, fidelity, and integrity required for the intended use.

### **Validate Hard-Wired Data**

Hard-wired data are evaluated separately because they typically consist of individual fixed constants used in specific algorithms or formulas. They should be validated along with the algorithms into which they are placed. Hard-wired data can be checked by executing the associated individual algorithms to ensure that they execute properly and provide appropriate output when compared with [validation data](#) (i.e., the expected output). Any deviations should be assessed to determine the cause (i.e., statement or execution of the algorithm, hard-wired data, or validation data) and recommendations made to resolve the problem.

### **Verify Output Data**

The code implementing individual algorithms and models should be examined to ensure that these algorithms and models provide output data to support the needs of the application. This review should include data characterization (e.g., fidelity, format, completeness) as well as methods of collection and preparation.

**Example:**

For an application concerned with missile lethality, the missile fly-out model used in the simulation should be executed and the output examined to ensure that it includes the necessary data categories (Pk, closest point of approach, etc.) and that the values produced for each category are accurate and usable.

### **V&V Activity: Validate Results**

Results validation is conducted to determine the extent to which the simulation addresses the requirements of the application. Because the data and the simulation are inextricably intertwined (i.e., if one is not valid, then the validity of the other cannot be demonstrated), their validations are usually conducted in concert. It is perhaps better to think of the simulation as being calibrated - its performance observed, known, and understood within the range of data values under which it is intended to operate.

This activity examines the extent to which the simulation, driven by valid input instance data, can provide appropriate responses when exercised in the context of the application. Beginning in the implementation phase, individual components or modules are executed to ensure appropriate performance and output. Additional testing is done as the components are integrated. Ultimately, the complete simulation is executed and the resulting data are compared to the [validation data](#) to determine if the simulation is producing credible, appropriate answers.

### **Validate Data**

Also beginning in the M&S implementation phase, the impact of the input data on the performance of individual components and on the integrated simulation is assessed. For each model validation test, key data elements should be tracked to ensure appropriate output. Sensitivity excursions can be run to test boundary conditions on key data elements to assess the impact of data ranges on model output. Data validation can also be conducted incrementally. For example, the terrain database for a battle simulation can be validated before battle entities and objects are added.

Data validation is performed to ensure that input data are appropriate for use in a particular simulation for a specific application. All data used to drive a model are subject to validation; however, the sheer quantity of data may make this impractical given cost and schedule constraints. Careful planning is needed to identify and prioritize key data components -- those data that most directly impact the performance of the model for the application.

### **Adjudicate Errors**

Discrepancies between simulation outputs and the validation data are examined to determine probable cause. Errors can result from problems with the code, input data, output data, validation data, or a combination of any of these factors, and it is important to determine the cause. The divergent output should be retraced through the code, key algorithms, and input data. Sensitivity excursions may be needed to isolate the error. When the culprit has been identified, the information is recorded and recommendations made to eliminate the problem.

## **Prepare M&S for Use Phase**

When all errors identified during testing are adjudicated and/or eliminated, the User accredits the simulation. At this point, the integrated model is ready for its intended use.

## **References**

- DoD Data Administration, *DoD Modeling and Simulation (M&S) Data Administration Strategic Plan (DASP)*, DMSO, April 1996.
- IEEE 1278.4: *Recommended Practice for Distributed Interactive Simulation -- Verification, Validation, and Accreditation*, Annex C, "Data Verification, Validation, and Certification," 1997.
- Rothenberg, Jeff; Walter Stanley; George Hanna; Mark Ralston, *Rand Project Memorandum PM-710-DMSO*, August 1997. This report offers an outstanding theoretical foundation for data verification and validation. It includes a data VV&C (verification, validation and certification; i.e., data V&V) process model

and lists considerations for structuring individual data V&V efforts with different kinds of data. It also provides a guide for planning both producer and user V&V activities.

### ***External Links in this Document:***

Authoritative Data Sources (ADS) website: <http://ads.msrr.dmsomil/>

DoD 5000.59-P: Modeling and Simulation (M&S) Master Plan, October 1995,  
<http://www.dmsomil/briefs/msdocs/policy/msmp.pdf>

DMSO DoD Data VV&C Tiger Team White Paper, 1998  
<http://www.dmsomil/briefs/war/vva/prod/whitepaper-final.pdf>

### ***RPG Links in this Document:***

*RPG Diagram:* "Data V&V for New M&S"

*RPG Reference Document:* "Data V&V Concepts and Terms"

*RPG Special Topic:* "Conceptual Model Development and Validation"

*RPG Special Topic:* "Requirements"

*RPG Special Topic:* "Subject Matter Experts and VV&A"

*RPG Special Topic:* "Validation"

*RPG Template:* "Data Quality Templates"

*The appearance of hyperlinks does not constitute endorsement by the DoD, DMSO, the administrators of this website, or the information, products or services contained therein. For other than authorized activities such as military exchanges and Morale, Welfare and Recreation sites, the DoD does not exercise any editorial control over the information you may find at these locations. Such links are provided consistent with the stated purpose of this DMSO website.*

§ § § § § § §