WEAG THALES JP11.20 - Final State of the REVVA Methodology

Dirk Brade
KTH/FOI/FMV
SE-10044 Stockholm, Sweden
dirk.brade@foi.se

René Jacquart
ONERA
2 avenue Edouard Belin
31055 Toulouse, France
jacquart@cert.fr

Jeroen Voogd
TNO
P.O.Box 96864, 2509 JG
The Hague, The Netherlands
voogd@fel.tno.nl

Choong-ho Yi
Swedish Defence Research Institute (FOI)
SE-172 90 Stockholm, Sweden
Choong-ho.yi@foi.se

ABSTRACT: *Under the umbrella of the Western European Armament Group's THALES Memorandum of Understanding, a "Common Framework for Verification, Validation, and Accreditation of Simulations" (nickname "REVVA") was developed and stabilized during the period March 2003 through October 2004. Within the Joint Program between Denmark, France, Italy, Sweden, and The Netherlands, a VV&A methodology was developed, which includes: (1) The REVVA Generic Process, a stand-alone VV&A process, which can be easily linked to numerous types of model development processes, including such for the development of distributed simulations (e.g., the FEDEP); (2) The concepts of Target of Acceptance (ToA) and Target of Verification and Validation (ToVV), which put a strong emphasis on V&V requirements definition, formalization of V&V planning, and traceability from the individual Items of Evidence acquired during V&V implementation up to the intended purpose of model use; (3) The explicit distinction between the properties "correctness" and "validity", which are inherent to a simulation model and its intended use on the one hand, and the processes of "verification" and "validation" to reveal them on the other hand; (4) Hooks for the integration of methods for estimation of uncertainty introduced because of the current inability to prove neither correctness nor validity; (5) A definition of roles, which clearly distinguishes technical and managerial roles and sides from which the actors may come; (6) A demarcation of "worlds" related to or associated with modeling and simulation; And (7) an extensive review of techniques to support the making of a more objective acceptance decision. The ToA and the ToVV are consistent with the documentation requirements defined in the International Test Operations Procedure on V&V (WGE7.2), and the proposed methodology creates a stable framework for activities such as the NATO MSG019/TG016 "VV&A of HLA Federations". This paper presents the building blocks of the REVVA methodology, and motivates the objectives of the follow-on program.*

Keywords: WEAG THALES JP11.20, REVVA, Verification, Validation, Accreditation, REVVA Generic Process, Target of Acceptance, Target of Verification and Validation

1 Introduction

Modern simulation models tend to get more and more complex, thus becoming themselves complex systems, requiring a more formal engineering process, including validation of the model. Among the major deficiencies in the field of Verification, Validation and Accreditation (VV&A) of Simulation Models and Simulation Results is the lack of an internationally recognized and accepted common methodological framework for VV&A. Today VV&A efforts are usually conducted differently in different organizations and different nations based on their own methods, processes and policies, with varying degrees of maturity. This heterogeneity does not only hinder the sharing and reuse of models and simulations between industrial or governmental bodies, but also may result in the rejection of an accredited simulation model by a potential user unfamiliar with the used VV&A method.

With the objective to develop the basis for a common methodological framework for VV&A of simulation models and simulation results, the European REVVA project [1] was implemented from March 2003 through September 2004. The project was run under the auspices of the Western European Armament Group (WEAG) according to the THALES Memorandum of Understanding as Joint Program 11.20, which provides a mechanism for multinational defense research collaborations among Western European Union nations. REVVA was funded by five nations: France (lead nation, ONERA), Denmark (UNI-C), Italy (DATAMAT), the Netherlands (TNO) and Sweden (FOI and FMV).

The REVVA research effort relied on past and existing efforts coming from many institutional sources, including the US Defense Modeling and Simulation Office (DMSO) [2], the NATO [3], the International Test Operation Procedure (ITOP) on V&V [4], and the AFDRG MEVAS project [5], as well as commonly known scientific contributions, such as [6] through [12]. In the follow-on project of REVVA, the VV&A methodology will mature in such a manner that the result can be submitted to an appropriate, internationally accepted standardization body.

This paper summarizes the results of the REVVA project and gives an overview of its achievements. For reasons of brevity, however, the results are here only scratched at the surface. Those interested in more information concerning the REVVA methodology are referred to [13] and [14]. Technical details, background information, and state of the art analysis are given in [15] through [24].

The paper is organized as follows: Section 2 introduces and justifies the underlying assumptions, main objectives, and key concepts for the preparation of the methodology. The REVVA methodology itself, addressing organizational, process- and product-oriented aspects, is presented in section 3, with the first lessons learned from case studies documented in section 4. The paper concludes with a summary and future objectives description (section 5).

2 Main Objectives, Underlying Assumptions, and Key Concepts

In this section the scope of the REVVA methodology is outlined. The three pillars of the methodology are introduced and the relationship between simulation model development and VV&A briefly discussed. The concepts of behavioral indistinguishableness are discussed as well as the need to define clearly the intended purpose of use. Acceptance is distinguished from Accreditation, and a foundation laid for uncertainty analysis.

2.1 Main Objectives

The vision for the scope and focus of a VV&A methodology is founded on an attempt to answer the basic questions “why?” and “for whom?”. Among existing VV&A methodological endeavors in science and defense convergence concerning the answers to the “why?” question can be observed, but the answers to the “for whom?” question vary significantly. These differences in the targeted audience have important consequences for the allocation of the developmental efforts.

In the context of REVVA, VV&A is considered as the set of methods, techniques and standards, which are used to assess the justifiable confidence that may be put in the use of M&S products. This assessment is a many facetted activity dealing with the evaluation of the M&S products and the evaluation of M&S organizations and processes.

The objectives of the REVVA methodology includes supporting the focused, integrated and well balanced analysis and evaluation of all these facets under explicit consideration of their relevance for the M&S uses. It also identifies the source and impact of remaining uncertainties (and to provide an estimate of their degrees), to facilitate an overall assessment of the trustworthiness of an M&S product.

2.2 Scope of the Methodology

When defining a VV&A methodology, there is a risk of getting lost in the apparently limitless field of VV&A. To border our investigations, we first discuss the position of M&S inside the larger setting of a problem-solving approach [2]. In REVVA, a clear distinction is made between M&S VV&A from questions concerning the utility of model use within the application domain on the one hand, and hardware/software quality assurance in the System/Software Engineering (S/SE) domain on the other hand. This is represented by the “Four Worlds” diagram in Figure 1.

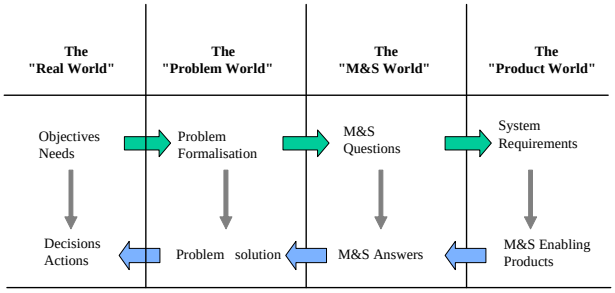

Figure 1: The "Four Worlds" diagram indicating the position the M&S World.

The distinction of “worlds” helps to separate concerns: For example, in the “Real World” a high level need may be to preserve the free democratic governments of Europe from military invasion, resulting in the action to maintain military forces, which are prepared for their mission. In the “Problem World” the task appears to keep those forces prepared, with an appropriate training program as problem solution. How this training program can be supported by means of simulation is answered within the “M&S World”, providing as answers forces, which are fit to react on simulated threats. The technical equipment required to place the forces in the simulated combat situations is created in the “Product World”.

Figure 1 clearly visualizes how the “M&S World” is framed by the “Problem World” on its left and the “Product World” on its right. This view helps to understand that M&S has a limited scope – and consequently a limited impact and responsibility – on decision making. It

helps to clarify, as well, the distinction between M&S uses (the "M&S World") and potential M&S developments and implementations (because we observe that development of new M&S products instead or reusing existing ones is the exception more than the rule).

2.3 Three Pillars: Process, Product, and Organization

The above introduction to the scope of VV&A motivates the definition of the three pillars the REVVA methodology is based on: products, organizations and processes. These dimensions have to be explored for their impact on the resulting justifiable confidence as well as on their impact on the way of doing the actual work.

REVVA assumes that the potential value of an M&S VV&A effort strongly depends on the organizational context. In general, the quality of the organization and the way of allocating and sharing the work and responsibilities is of primary importance. The principles recommended for the implementation of a VV&A project are based on our 3-pillars model. The three pillars can be described as:

(1) *the organization*, which involves different groups with different, sometimes conflicting, interests. This pillar builds on identified *parties* and *roles* (see section 3.2);
(2) *the process*, which directs the flow of activities and products during VV&A. The *REVVA Generic Process*, which builds this pillar, is a stand-alone VV&A process which can be mapped to standard modeling processes via the M&S intermediate and final products made available for VV&A (see section 3.3); and
(3) *the products*, which document the findings of the VV&A effort. This pillar is mainly built out of *Items of Evidence*, which are the basic results of the application of V&V Techniques. It is structured according to the semi-formalized Acceptability Criteria (documented in the *Target of Acceptance*, ToA) and the chosen V&V approach (*Target of Verification and Validation*). The basic results are integrated for the acceptance decision into an overall picture (see section 3.4).

This three-pillars model is the meta-process of the REVVA methodology. It captures the dependencies of and flow of information between the methodology components. It is expected that making these relationships explicit should be beneficial for the comprehensiveness, focus and balance of the VV&A project.

2.4 Relationship to M&S Development

To keep the REVVA methodology as flexible as possible, only the developmental products (such as the Simulation Conceptual Model or model documentation) are used. No assumptions or prescriptions are made concerning the approach to development. This way the VV&A methodology is not bound to a particular modeling paradigm and independent from the chosen model development process or current state of model development. It exclusively concentrates on available intermediate, final, or supplemental M&S products. Because it is most likely that especially for legacy M&S products there is no more process information available anyway, and because conclusions for product quality based on process assessment are unreliable, the M&S development process followed is not assessed during VV&A. For new M&S products this leaves maximum freedom to the developers concerning their development procedures; however, they will be confronted with detailed requirements concerning the documentation and delivery of (interim) developmental products, to ensure that all the information that is required to perform efficient V&V and to reduce the residual uncertainty associated with the use of the M&S product is available.

2.5 Behavioral Indistinguishableness

In various explanations or definitions of the term "validation", a relationship to the context of model use is created. Already early authors concentrate on the correctness of the inference about a system derived from the simulation [7], later others address the accuracy of a model's behavior within its application domain under consideration of the study objectives [10], or the impossibility to distinguish between the system and the model within the experimental frame of interest [8]. Under the precondition that (dynamic) simulation is always about behavior, the REVVA methodology continues this path and exclusively concentrates on simulation model behavior, with the desired objective to demonstrate sufficient behavioral indistinguishableness.

2.6 Well-defined Intended Purpose

If one is supposed to demonstrate something, one should have a clear idea of what this something looks like. In [25] a convincing rationale is given for why this statement also holds true for M&S VV&A. The "intended purpose" statement itself usually is too vague for objective demonstration. A good starting point for derivation of the Acceptability Criteria (AC) is given by the requirements used for the development process of the M&S asset. In the situation that these requirements are not available or not sufficiently precise for VV&A purposes (which is typical), or if the M&S asset needs to be validated for a different purpose than it was developed for, the criteria need to be refined or derived anew from the intended purpose of use.

First, the AC are considered to be related to validity aspects. Other requirements may be important for the over-

all success of the M&S asset (e.g., hardware and operating system under which the simulation software is expected to run), but will not be used in the evaluation of the correctness and validity. Therefore the most important type of requirements for deriving AC are the validity related functional requirements. For REVVA it is assumed that the determination of the AC is a top-down activity, which starts with questions about details of the intended purpose, and continues with the consequent development of objectives and sub-objectives, as appropriate. Finally, when the sub-objectives hierarchy is developed, from the lowest sub-objectives directly the AC are derived, which constitute the leaves of the hierarchy and address the desired correctness and validity properties of the M&S product.

The risk associated with the use of the M&S product is considered to be the driver for V&V. Impact domains have been identified by [26] and [27], which may serve as criteria for the estimation of the worst case impact. The potential consequences of the use of erroneous M&S product are, for reasons of pragmatics, qualified in four distinct classes for all identified impact domains, ranging from "negligible", to "marginal", "critical", and "catastrophic. For each Acceptability Criterion, the impact of using the M&S product for its intended purpose despite of failure of this Acceptability Criterion needs to be determined. The more critical the impact of an individual Acceptability Criterion is, the lower should be the residual uncertainty associated with its assessment. The REVVA methodology provides an appropriate structure to capture both the refinement and its rationale in the form of the Target of Acceptance (ToA) and the Target of Verification and Validation (ToVV) [15].

2.7 Accreditation, Acceptance, and Assessment

In the defense VV&A community, accreditation is defined as "the official certification that a model, simulation, or federation of models and simulations is acceptable for use for a specific purpose" [28]. However, this definition is not consistent with the use of the term "accreditation" in other domains, where accreditation is not associated with products, but organizations [29]. Also the official authoritative accreditation procedures vary from nation to nation. To avoid conflicts, the concept of acceptability for the intended purpose is introduced, assuming that acceptability is an indispensable prerequisite for accreditation or certification, whatever it is called.

Acceptance or rejection here is judged based on *Acceptability Criteria (AC)* addressing the behavior of the simulation model and its indistinguishableness from the behavior of the System of Interest are explicitly referred to as *Validation Criteria* which are a distinguished subset of AC). It further is assumed that all software quality related issues are covered by the appropriate software quality standards, and that the examination of "traditional" software AC is covered by the appropriate test procedures. In the following, validation criteria exclusively address the M&S product's validity and correctness with respect to its chosen representation of the System of Interest; non-functional requirements and functional requirements not addressing the validity or correctness of the chosen representation of the real world are considered to be AC.

2.8 Uncertainty in Decision Making

A simulation model or simulation result *is* or *is not* factually valid and correct, regardless of how much time is spend on V&V. V&V facilitate the perception of the presence or absence of these properties. When making an acceptance decision for an M&S product, *uncertainty* is introduced because feasibility of a proof is an exception rather than the rule. With a lack of proof, there may be a discrepancy between the perceived validity and factual validity, and the perceived correctness and the perceived correctness. Table 1 taken from [13] visualizes the uncertainty associated with the decision to accept or reject a simulation model for a particular intended purpose in form of type I and type II error.

Table 1: Factual validity vs. perceived validity

Unknown fact / Perception (Action)	**Factually valid**	**Factually invalid**
Perceived as valid (and accepted)	Ok	Type II Error
Perceived as invalid (and rejected)	Type I Error "False Alarm"	Ok

These errors can occur in various refinement stages of the examination objectives: On the highest, most vague level ("intended purpose"), the M&S product can be perceived as "(un-)fit for purpose", although it factually is not (or is, respectively). A (more precise) acceptability criterion may be perceived as passed (failed), although it factually is not (or is, respectively). A set of Items of Evidence might be perceived as substantiating that an acceptability criterion is met (although they are not necessary and sufficient). And an Item of Evidence might be much higher (or weaker) substantiation than perceived.

In the REVVA methodology, the following sources of uncertainty are addressed:

- the rationale, why a given set of AC is necessary and sufficient to assess a simulation model's fitness for the intended purpose;
- the rationale, why a particular set of Items of Evidence is necessary and sufficient to consider a given acceptability criterion as passed or failed;
- the way how the Items of Evidence were created.

The degree of residual uncertainty depends on these factors. A deeper elaboration of uncertainty can be found in

the project report [17], and in previous conference publications of the authors [30] and [31].

3 The REVVA Methodology

The methodology is based on the assumption that the VV&A activities are implemented by an organization or organizational sub-unit collectively named "VV&A Agent". This VV&A Agent should have, with respect to the M&S intended purpose, the adequate level of independence from the M&S customers and from the M&S suppliers, ranging from being an organizational sub-unit in the M&S supplier's organization to independence as defined in [30]. The VV&A organization has to be engineered both to satisfy this level of independence and to provide the personnel ("actors") fulfilling the tasks identified in the process view.

The process defines the logical and chronological flow of activities and products to facilitate the transition from the initial intended purpose, through a series of intermediate steps and products, to the final product, namely a report recommending/rejecting the use of the M&S product for the intended purpose. The process proposed here focus on the technical components of the M&S and VV&A products and activities.

The products result from the process steps. Products are identified as they correspond to expected stable states and also as they are associated to decision steps. Due to iterations in the process, products can be available in different versions and need to be configuration controlled. The methodology identifies a very limited number of intermediate products, namely those that are important for detailed VV&A planning and decision making.

The REVVA methodology is based on assumptions about what the most complex issues are when dealing with a problem to solve. The guidance to reduce the level of complexity given includes:

- independence,
- incremental nature of the VV&A,
- focus (of VV&A on the intended purpose),
- specification of observable outcomes,
- critical thinking and balanced evaluation,
- tailoring.

An elaboration on these hints is documented in [14].

3.1 Terminology

The definitions for Verification and Validation recalled here in order to give the feeling of the road explored by REVVA introduce an important distinction among properties of products (i.e., correctness and validity) and the way to perceive these properties, the processes and their constraints (which introduces some relativism).

Correctness: The property of a simulation model to comply with formal rules and bodies of reference information for its content and representation, and for the transformation into another representation.

Validity: The property of a simulation model to have, within a specific experimental frame, a behavior which is indistinguishable from the behavior of the System of Interest.

Verification: The process which is used to construct, under a set of time, cost, skills, and organizational constraints a justified belief about model correctness.

Validation: The process which is used to construct, under a set of time, cost, skills, and organizational constraints a justified belief about model validity.

This separation between properties and processes stresses the current situation that V&V cannot guarantee absolute correctness and validity. More VV&A related definitions from the context of REVVA can be found in the Appendix of [14].

3.2 Parties and Roles

To implement the pillar "organization", the REVVA methodology identifies *roles* and *parties*. A *role* is characterized by the skills required to accomplish a particular task or set of tasks, and the responsibilities that are taken. Also groups with different interests, including those who are going to acquire a simulation model or simulation results (and are likely to pay for it), and those who deliver the requested M&S product, are distinguished. These interest groups are called *parties*.

3.2.1 Parties

A party is assumed to be an organization or organizational unit. With the situation that somebody provides a simulation model or simulation results, which will be used by somebody else, there exists a "customer-supplier relationship":

- *Customer*: A customer is an organization or organizational unit which plans to use or is using an M&S product (such as a SEM, simulation results, or data) developed by another party.
- *Supplier*: The supplier is an organization or organizational unit which provides the M&S product.

A relationship of trust between the customer and the supplier is desirable, but it must be always kept in mind that the supplier is trying to sell something to the customer,

with all its implications. Thus, the REVVA Methodology introduces the

- *3rd Party VV&A Agent:* The 3rd Party VV&A Agent is an organization or organizational unit external of the customer and the supplier parties. Its degree of independence is assessed based on managerial, technical, and financial factors.
- *Acceptance Authority*: The Acceptance Authority is an organization or organizational unit external of both the customer and the supplier parties, officially entitled to accept M&S products, and trusted by the customer. Its degree of independence is assessed based on managerial, technical, and financial factors.

The roles introduced in section 3.2.2 are played by actors from the above parties. The decision, which party an actor comes from, must be made carefully and deliberately. If a V&V role is played by an actor from the supplier party, communication and information exchange between the M&S developers and those doing the V&V would be simplified, but then it must be assumed that the V&V results can be influenced by interests of the supplier party. It is assumed that the V&V effort will be more critical, if the actor comes from the customer party, but legacy solutions also might bias this activity. The highest degree of objectivity is achieved, if the actor comes from an independent 3rd Party VV&A Agent, but on the expense of an increased communication overhead.

3.2.2 The Roles

The assignment of tasks to persons or individuals (management of human resources) should be based on an agreement of the parties involved. In the following, roles interacting with and responsibilities within the VV&A process are identified [13]. Each role is outlined by

- the required knowledge and skill to complete the assigned tasks,
- the authority given and responsibility taken in the process, and
- its interaction with other roles.

A role does not determine, whether it is played by one actor, or shared by several actors, which even might come from different parties. However, particular roles require a sufficient distance between the individuals or teams performing them, while others are likely to be played by the same, single individual.

VV&A core roles are directly involved in the VV&A endeavor by using, planning, conducting, evaluating, or assessing the substantial VV&A work.

- The *Contextual User* defines the contextual objective. It is assumed that the Contextual User always is in the customer party.
- The *Acceptance Leader* is a user representative (trusted by the Contextual User), who is responsible for the assessment of the M&S product. The role also finally judges the success or failure of the V&V effort.
- The *V&V Leader* knows approaches to V&V, techniques, and tools. This role is responsible for developing an appropriate V&V approach to substantiate the AC with the information about System of Interest and simulation model available.
- The *V&V Executioners* is a composite of roles including Simulation Model Operators, System Analysts & Subject Matter Experts, M&S Experts, and HW/SW Engineers. It consists of a number of actors playing several roles that actually implement the analysis and test activities required to provide the Items of Evidence specified by the V&V Leader.

Affected roles take advantage of the REVVA methodology, but are not directly involved in the technical planning and implementation of VV&A. Often they are decision makers outside of the process, are responsible for the smooth organizational flow of the VV&A effort, and control the flow of information among all parties involved.

- The *M&S Promoter* sees an advantage of having (usually inferior) people within his organization use a simulation model or simulation results, and desires to benefit indirectly from the consequences of using M&S products.
- The *M&S Sponsor* creates the financial foundation for the development and VV&A of the M&S product. The actor of this role is member of the customer party.
- The *M&S Project Manager* organizes and controls a particular use or series of uses of an executable model (i.e., model selection, experiment design, and experiment evaluation).
- A *VV&A Project Manager* organizes the managerial aspects of the V&V endeavor, when it is managerially separated from model development.

Please refer to [13] for a significantly more detailed roles description.

3.2.3 Choosing Actors

Whether an actor or group of actors is appropriate to play a particular role depends on organizational aspects, including the desired degree of independence and required transfer of information, and on her/his educational background and experience. The assumption here is that if technical activities should be shared, responsibilities on both the customer side and the supplier side have to be clearly identified to prevent conflicts of interests. In accordance to [30] the REVVA methodology distinguishes:

- Dependent V&V (DV&V): The V&V is conducted by the M&S supplier according to the customer's V&V requirements (i.e., the actors for V&V Leader and V&V Executioners are members of the supplier party), and accepted "as is" by the customer.
- Independent Assessment (IA): The V&V work is conducted by the M&S supplier, but is assessed by an independent Acceptance Leader (from and independent 3rd Party) trusted by the customer,
- Independent V&V (IV&V): V&V activities are planned and conducted independently from both the supplier and the customer by the independent 3rd Party VV&A Agent.

Table 2 gives an overview over cost-effective assignment of actors to roles, considering independence from the customer's perspective.

Table 2: Actors, Roles, and Independence

	Acceptance Leader	**V&V Leader**	**V&V Executioners**
DV&V	Not explicitly assigned	Supplier	Supplier
IA	Customer or 3rd Party VV&A Agent	Supplier	Supplier
IV&V	Customer or 3rd Party VV&A Agent	3rd Party VV&A Agent	3rd Party VV&A Agent

3.3 The REVVA Generic Process

The REVVA Generic Process implements the pillars "process" and "product" of the REVVA methodology. As shown in Figure 2, it supports product-oriented VV&A during or after model development (e.g., as required for reuse for another related intended purpose), and can be used as guidance for planning a VV&A effort. The "V-Form" for the process representation was deliberately chosen, mirroring the preparation for V&V and the execution of the V&V activities on the left trunk ("\") of the "V", against the evaluation and the integration of the V&V results for the purpose of assessment on the right trunk ("/") of the "V".

3.3.1 Phases and Products

Each phase description contains a summary of activities, lists the input and output products, and points out the involved roles and their type of involvement. The REVVA Generic Process is no waterfall process, but iterative, which means that especially those products close to the bottom of the "V" become available in several versions. More detailed description is found in [13].

Develop ToA (phase 1): Based on the intended purpose of model use, a detailed set of AC is developed in such a manner that passing the AC implies fitness for purpose. All AC and the rationale for their derivation are recorded as the "Target of Acceptance" (ToA).

AC should be prioritized. For simulation-based endeavors with a low impact on real world decisions or actions, some superficial indicators that the AC are passed may be sufficient, while safety critical aspects might require an unmistakable proof.

Target of Acceptance (product): The Target of Acceptance (ToA) contains a precise specification of the AC and the rationale for their derivation from the intended purpose, and documents "what needs to be demonstrated" during the V&V effort. On top of a refinement hierarchy stands the vague intended purpose, which is refined into a set of sub-purposes, which again is decomposed, until AC related to the M&S product's correctness and validity can be derived directly from the lowest sub-purposes. The set of AC does not imply any methods or techniques how to assess them. If it is decided to have a 3rd Party VV&A Agent to do the V&V, the ToA defines the technical objective of the contract with the 3rd Party VV&A Agent. An abstract ToA is shown in Figure 4.

Acquire Information (phase 2): Under consideration of the intended purpose of model use and the detailed AC (documented in the ToA), knowledge about the System of Interest, its structure and behavior, its subsystems and their structure and behavior, or related systems is collected and filed (in related work this body of real world knowledge is referred to as "referent").

Model information and system knowledge (product): This information will be used as foundation of the approach to demonstrate the model's correctness and validity. The product identifies all sources of information and knowledge and all bodies of information and knowledge that are available or will become available during the V&V effort. The acquired information and knowledge about both the M&S product and the System of Interest is ideally stored in (an) appropriate remotely and securely accessible data base(s).

*Develop ToVV (phase 3):*For each Acceptability Criterion a rationale is developed, which points out how with the information at hand and the available technical means it can be demonstrated that the Acceptability Criterion is passed or failed. To substantiate that the Acceptability Criterion is met becomes a *V&V Objective*. Developing the ToVV usually includes the decomposition of a V&V Objective into more easily assessable V&V sub-objectives.

Target of Verification and Validation (product): The Target of V&V (ToVV) documents the approach taken to the substantiation of the AC. It elaborates on the "how to

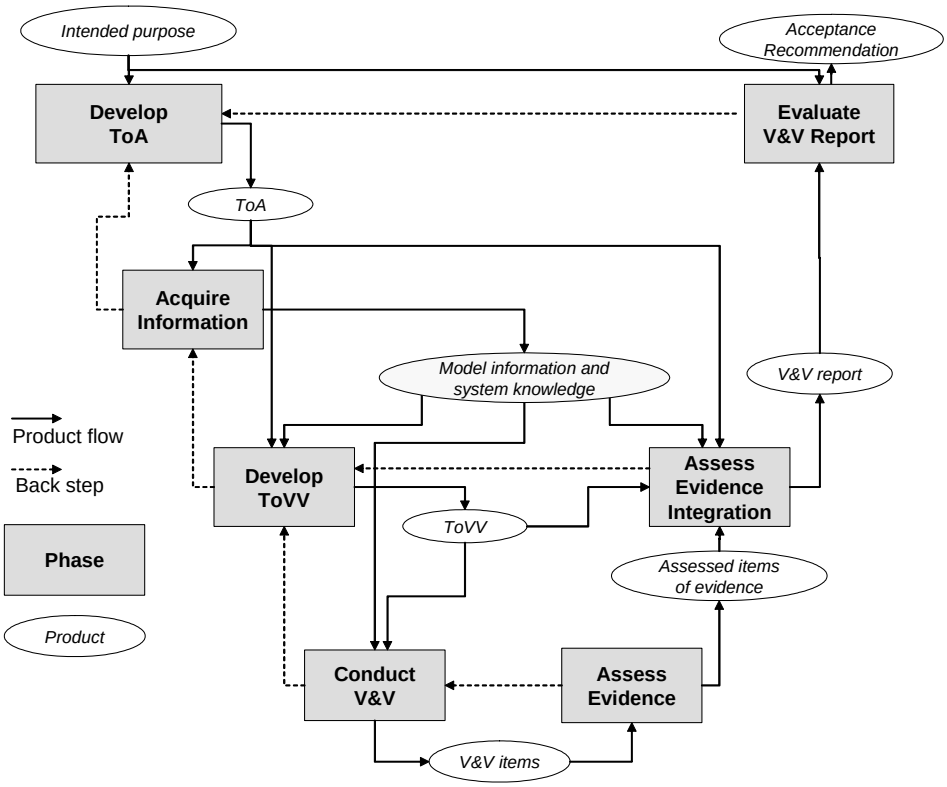

Figure 2: The REVVA Generic Process

demonstrate that the AC are passed or failed", identifies the Items of Evidence required to substantiate the AC contained in the ToA, and documents the rationale for the necessity and sufficiency of these Items of Evidence. The rationale for this decomposition includes the justification, why passing the lower V&V sub-objectives also implies passing the Acceptability Criterion from which they were derived. Besides the required information within the Items of Evidence, the ToVV also identifies their individual desired probative forces needed to consider them "strong enough". In Figure 4 an abstract ToVV is shown and its connection to the ToA revealed.

Conduct V&V (phase 4): V&V is conducted to provide the V&V items required by the ToVV. If, due to, e.g., missing or insufficient information about the model, missing knowledge about the System of Interest, or unavailability of the required tools, a particular Item of Evidence cannot be acquired, or if an elementary V&V objective is demonstrated to be failed, a step back to "Develop ToVV" is made.

V&V items (product): Each test result, analysis report, or proof outcome is documented as V&V Item, which as a set, constitute the "atomic building blocks" of V&V. A V&V Item consists of some piece of information about the Simulation Model, the evaluation objective, reference information, an evaluation technique, and the evaluation result. For validation, the reference information consists of knowledge about the System of Interest. For verification, the reference information consists of, e.g., representation rules, model information in a different representation form, or formalism. V&V Items have different probative forces, depending on the method or technique used for their creation, and the reference information or knowledge used.

*Assess Evidence (phase 5):*The key issue of this phase is to assess the probative force of the V&V items, to accept the individual V&V items as Items of Evidence, or to reject them. If the probative force of an Item of Evidence is considered to be unacceptably low, the Item of Evidence needs to be strengthened by repeated conduction of V&V activities or discarded. Otherwise, the Item of Evidence is added to the evidence pool, which its perceived probative force annotated. The *probative force* of each individual Item of Evidence is assessed based on the repeatability of the associated V&V activity. Criteria for the assessment of the probative force of an Item of Evidence add objectivity to this currently subjective procedure. The probative force of an Item of Evidence is considered to be high, if the V&V result is reproducible, independently from its subjective elements (human beings). It is considered to be low, if it strongly depends on its

subjective elements and its various results depend on the different individuals involved.

Items of Evidence (product): The Items of Evidence document the individual executions of single V&V techniques and their outcomes, as conducted or acquired by the V&V Executioners. The assessed Item of Evidence includes (in addition to the information contained in the V&V item from which the Item of Evidence originates) the assessment statement, and a judgment of its probative force.

Assess Evidence Integration (phase 6): A single Item of Evidence will usually not allow the conclusion that a particular Acceptability Criterion is passed, but several Items of Evidence are assembled according to the (most recent version of the) ToVV. The key issue of this phase is to build and accept or reject the rationale of supporting the AC with the available Items of Evidence. Under reconsideration of the ToA, the assembly of the evidence is reviewed and it is judged how sufficiently the evidence substantiates that the AC are passed (convincing force). The convincing force of the V&V sub-objectives hierarchy as documented in the ToVV is an expression of the preciseness and coverage of the AC. An Acceptability Criterion is considered to be completely covered, if the rationale for the derivation of directly succeeding V&V sub-objectives makes clear that meeting the V&V sub-objectives automatically implies meeting the parent Acceptability Criterion, too. If the available evidence leaves unacceptable gaps or loopholes for the substantiation of the AC, the ToVV needs to be adjusted and the additional V&V activities conducted to provide the missing Items of Evidence.

V&V report (product): The gathered or otherwise created Items of Evidence assembled and integrated by the V&V Leader to substantiate the AC in the ToA according to the most recent version of the ToVV, build the substance of the V&V report. The V&V report links the rationale why the referenced Items of Evidence substantiate the claim that the AC are passed with the Items of Evidence made available.

Evaluate V&V Report (phase 7): Based on the probative force of the evidence, the convincing force of the ToVV, and the selection of AC as motivated in the ToA (all documented in the V&V report), the residual uncertainty associated with the statement that the M&S product actually is fit for its intended purpose is estimated. If the residual uncertainty is considered to be too high, either the intended use must be modified in such a manner that invalid simulation results have a less critical impact, or the V&V effort must be partially repeated with an extended ToA.

The level of residual uncertainty needs to be identified for each Acceptability Criterion and each relevant set of AC individually. While for particular AC a high degree of uncertainty is acceptable (criteria which may be failed without serious consequences), for others only very low uncertainty is acceptable (criteria whose failure will have serious impact).

When no disproving evidence has been acquired or created, when the affirmative evidence is considered to be "strong enough", and when the strategy according to which the affirmative evidence is assembled to substantiate the claim that the AC are met is considered to be "sufficiently convincing", then the M&S product is perceived as correct with respect to all relevant specifications and constraints, and as valid for its intended purpose (as represented by the ToA) with sufficiently low residual uncertainty. To prepare a responsible acceptance or rejection decision, an upper bound for this residual uncertainty is estimated.

Acceptance Recommendation (product): The final recommendation whether to accept or reject the M&S product for its intended use, considering the uncertainty that is left even after V&V was successfully conducted, is documented in form of the acceptance recommendation. The acceptance recommendation confirms that the acceptability for the intended purpose is demonstrated by the Items of Evidence gathered to substantiate the AC, and states a reasonable degree of confidence in this confirmation.

The achieved level of uncertainty, regardless how low, must never imply that the results of the use of the M&S product can be blindly transferred to the real world.

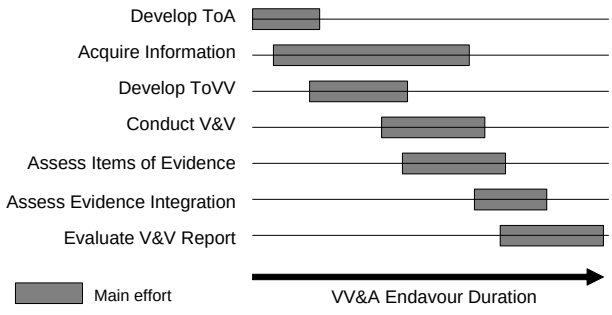

Figure 3: Sketch of the temporal overlap among the phases of the REVVA Generic Process

3.3.2 Temporal and Causal Dependencies

The REVVA Generic Process is rather an iterative process than a waterfall process. The expectable overlap among the phases during the whole duration of the VV&A endeavor is sketched in Figure 3. With the ToA nearly completely stabilized, based on the information

available so far (or scheduled to become available) the ToVV can be developed. V&V activities can be conducted, when the ToVV is sufficiently developed, but infeasibility of some planned V&V activities might require changes in the ToVV. Small iteration circles between the assessment of the V&V items and their production are essential for high efficiency.

3.4 Transparent and Traceable V&V Approach

The Target of Acceptance [19] and the Target of Verification and Validation implement together with the Items of Evidence the "products" pillar. The integrated structure of ToA, ToVV, and Items of Evidence holds all information required for an informed acceptance decision. The basic structure of the ToA stems from the hierarchical way the objectives are elicited, yielding a directed acyclic graph: It can happen that two V&V objectives both need the same sub-objective. The objectives that are refined must have an additional piece of information associated with them. This extra information is the argumentation on why the sub objectives together are necessary and sufficient to constitute their parent objective. This argumentation is named the "Decomposition argument", which is the glue in the hierarchical structure of the ToA.

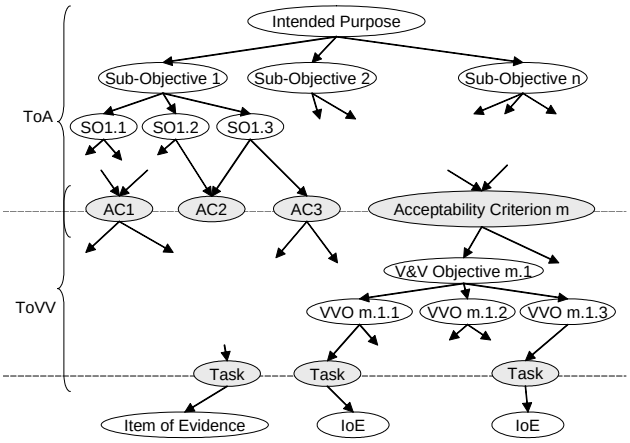

Figure 4: ToA and ToVV

The ToVV documents the approach how to substantiate the AC with the available model information, the available system knowledge, a choice of V&V techniques, and a suite of tools at hand. Its the leaf-nodes indicate performed V&V tasks with acquired Items of Evidence attached to them, as sketched in Figure 4. This hierarchical structure facilitates tracing between the Intended Purpose on top of the hierarchy and the elementary Items of Evidence at its leaves.

4 Case Studies

During the REVVA project life-time two case studies have been used to illustrate the application of the methodology's concepts and products, as documented in [21] and [22]. They have been used as well to test the empirical adequacy of these concepts and to adjust the definitions of some of them. The two case studies were chosen from different application domains: One was related to a SBA context (the so-called Predator weaponization case study) and the other one to the education domain (education of captains using an adaptation of JANUS). The two case studies rely on an existing quite well documented body of knowledge for the simulation model, for the uses, and for the organizations which are running or supporting these simulations.

From this work on case studies, it could be said, on the positive aspects, that:

- the argumentation framework and its related "case" structure have been very fruitful to structure at macroscopic level the way of formulating the VV&A problem, and
- the product and process defined in the REVVA reports were applied and the proposed structuring has provided the expected guidance.

The case studies have revealed also some limitations, weaknesses or holes in the methodological guidance (with respect to the exploitation of the present versions of the methodology definition). These limitations are listed here:

- the organizational aspects, resources, costs and V&V planning are mentioned but not actually integrated with the technical aspects of the methodology,
- the topics mentioned just above are not integrated in the argumentation framework, neither directly nor through their consequences,
- the argumentation framework (and its related products) is quite easily applied for the top-down decomposition phase of the V&V process. However, the recomposition phase is less easy. The present definition of the methodology does not help to structure the way of arguing and documenting the recomposition steps.
- the re-composition phase presented here is a quite simplistic one: it does not allow to use some multicriteria analysis technique, being in a situation where each evidence is indeed satisfied.

These first level conclusions have to be deepened but they could be used as return of experiment for an improved version of the Reference Manual [14] and User's Manual [13] of the REVVA methodology.

5 Summary and Future Work

This paper gives an overview over the VV&A methodology developed within the WEAG THALES JP11.20 "A Common Verification, Validation, and Accreditation Framework for Simulations" (nickname "REVVA"). The REVVA methodology is built on a set of underlying assumptions and concepts, including (1) a clear demarcation of M&S VV&A from problem formulation and HW/SW development, (2) the need for the definition of a VV&A organization, a VV&A process, and VV&A products within the methodology, (3) the desire to minimally constrain M&S development by VV&A, but to increase the quality of the M&S product, (4) the justification for exclusively behavior oriented validation, (5) the indispensable prerequisite of clearly defined Acceptability Criteria, (6) the distinction between formal accreditation and informal acceptance for use, which also should be done deliberately, and (7) the inherent uncertainty associated with an acceptance decision. The presented building blocks of the REVVA methodology include (1) parties, roles, and guidance on how to assign actors, who are members of those parties, to roles in an efficient manner, (2) the REVVA Generic Process, which is a seven-phased stand-alone VV&A process, and (3) the products Target of Acceptance (ToA), Target of Verification and Validation (ToVV), Items of Evidence (IoE), and their integration into a combined ToA-ToVV-IoE structure. Also lessons learned from the case studies were briefly reported. Although the work conducted during REVVA stabilizes numerous aspects of VV&A, the methodology is not yet complete. Open issues include the further investigation of the comprehensive integration of the methodology components (organization, products and process); the impact of this integration on the argumentation framework and on the evaluation of remaining uncertainties; a specific work on the tailoring of the methodology; and specific technical zooms on the support of the argumentation framework by formal techniques. Those open issues will be addressed in the REVVA follow-on program, which is arranged under the Western European Armament Group's EUROPA memorandum of understanding, and is expected to start in spring 05.

Acknowledgements

The results presented in this paper have been produced within the consortium of the WEAG THALES JP11.20 project, including as members Curzio Batini, Pierre Bouc, Emiliano Dall'Acqua, Martijn van Emmerik, Jean-Francois Gabard, Lars Horn, Fredrik Jonsson, Håkan Lagerström, Jørgen Moth, and the authors. The 18 months project was sponsored by the Ministries of Defense of France, Denmark, Italy, The Netherlands, and Sweden. The International Test Operations Procedure on Verification and Validation, the members of the NATO Modeling and Simulation Group 019 / Task Group 016, and the DMSO's website served as invaluable sources of inspiration. All reports produced by the THALES JP11.20 consortium can be freely accessed and downloaded under www.jp1120-revva.com.

References

[1] Jaquart, R., P. Bouc, and D. Brade. 2004. A Common Validation, Verification and Accreditation Framework for Simulations: Project JP 11.20, REVVA. In Proceedings of the 04 Spring Simulation Interoperability Workshop, Orlando, Simulation Interoperability Standardization Organization (SISO).

[2] Defense Modeling and Simulation Office. 2000. Verification, Validation and Accreditation (VV&A) Recommended Practices Guide (RPG). Alexandria, VA. http://vva.dmso.mil/

[3] North Atlantic Treaty Organization. 1998. NATO Modelling and Simulation Master Plan. Document AC/323 (SGMS)D/2, Version 1.0.

[4] ITOP. 2004. General Procedure for Modeling and Simulation Verification and Validation Information Exchange. International Test Operations Procedure 1-1-002, WGE 7.2.

[5] R. Jacquart, P. Bouc, D. Girardot. 2003. A socio-technical view of a VV&A methodology. Proceedings of the European Simulation Interoperability Workshop 2003, Simulation Interoperability Standardization Organization (SISO).

[6] Banks, J. (ed.). 1998. Handbook of Simulation, Wiley, New York.

[7] Shannon, R.E. 1975. Systems Simulation and the Art of Science. Prentice Hall, Eaglewood Cliffs, N.J.

[8] Zeigler, B.P., H. Praehofer, and T.G. Kim. 2000. Theory of Modeling and Simulation. Second Edition. Academic Press. ISBN: 0-12-778455-1

[9] Oberkampf, W.L. and T. Truncano. 2002. Verification and Validation in Computational Fluid Dynamics. Progress in Aerospace Sciences 38 (2002) 209-272. Elsevier Science Ltd.

[10] Balci, O. 2001. A Methodology for Certification of Modeling and Simulation Applications. In ACM Transactions on Modeling and Computer-based simulation 11 (4), pp 352-377.

[11] Sargent, R.G. 1999. Validation and Verification of Simulation Models. In Proceedings of the 1999 Winter Simulation Conference, pp. 39-48, Society for Computer Simulation International.

[12] Davis, P.K. 1992. Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations. Report published by RAND, Santa Monica, CA. ISBN 0-8330-1298-3.

[13] PROSPEC. 2004. VV&A Process Specification: User's Manual. THALES JP11.20 Technical Report. JP1120-WE5200-D5201-PROSPEC-V1.3.

[14] METHGU2. 2004. VV&A Methodological Guidelines Reference Manual. THALES JP11.20 Technical Report JP1120-WE5100-D5101-METHGU2-V1.2.

[15] TAXO. 2003. VV&A Global Taxonomy. THALES JP11.20 Technical Report. JP1120-WE1100-D1101-TAXO-V1.1.

[16] CRIT. 2004. VV&A Criteria Definition. THALES JP11.20 Technical Report JP1120-WE1200-D1201-CRIT-V1.1.

[17] LEVELS. 2004. VV&A Levels Definition. THALES JP11.20 Technical Report JP1120-WE1300-D1301-LEVELS-V1.1.
[18] PROC. 2004. VV&A General Process. THALES JP11.20 Technical Report JP1120-WE1400-D1401-PROC-V1.1.
[19] TOAGUID. 2004. TOA Guidelines. THALES JP11.20 Technical Report JP1120-WE2200-D2201-TOAGUID-V1.1.
[20] TECH2. 2004. VV&A Techniques. THALES JP11.20 Technical Report JP1120-WE3100-D3102-TECH2-V2.0.
[21] CASESTU1. 2004. Case Study Report. THALES JP11.20 Technical Report JP1120-WE6200-D6202-CASESTU1-V1.0
[22] CASESTU2. 2004. Case Study Report. THALES JP11.20 Technical Report JP1120-WE6200-D6202-CASESTU2-V1.0
[23] XPCAP. 2004. Experience Capitalisation. THALES JP11.20 Technical Report JP1120-WE4100-D4101-XPCAP-V1.1
[24] OPTIM. 2004. Experience Optimisation. THALES JP11.20 Technical Report JP1120-WE4200-D4201-OPTIM-V1.1
[25] Balci, O. 2000. Well-Defined Intended Uses: An Explicit Requirement for Accreditation of Modeling and Simulation Applications, Proceedings of the 2000 Winter Simulation Conference, pp. 849-854, Society for Modeling and Computer-based simulation International.
[26] Muessing, P., D. R. Laack, and J. J. Wrobleski, Jr. 1997. Optimizing the Selection of VV&A Activities: A Risk/Benefit Approach, Proceedings of the 1997 Summer Computer Simulation Conference, Society for Computer Simulation International.
[27] Mugridge, C. 1999. Verification, Validation and Accreditation of Models and Simulations Used for Test and Evaluation – a Risk/Benefit Based Approach. Internal Report, Technical Development Group, Defense Evaluation and Research Agency UK.
[28] Defense Modeling and Simulation Office. 1996. Department of Defense Verification, Validation and Accreditation (VV&A) Recommended Practices Guide, Defense Modeling and Simulation Office, Alexandria, VA. Co-authored by: O. Balci, P. A. Glasgow, P. Muessing, E. H. Page, J. Sikora, S. Solick, and S. Youngblood. USA.
[29] Rae, A., P. Robert, and H.-L. Hausen. 1995. Software Evaluation for Certification: Principals, Practice, and Legal Liability. McGnaw-Hill, London, UK
[30] Brade, D. and M. Lemaître. 2004. Assessment Criteria and V&V Levels in JP11.20. Spring Simulation Interoperability Workshop, Simulation Interoperability Standardization Organization (SISO), Arlington, VA.
[31] Brade, D. 2004. Quantitative Uncertainty Metrics in M&S VV&A. Foundations ´04: A Workshop for VV&A in the 21st Century, Tempe, AZ.
[32] IEEE Std-1012. 1998. Annex C.

Author Biographies

Dr. Dirk Brade is employed as guest researcher at the Royal Institute of Technology (KTH) in Stockholm on a position sponsored by the Swedish Defense Research Agency (FOI) and the Swedish Defense Materiel Administration (FMV). He received a Doctor Degree in Natural Sciences from the University of the German Federal Armed Forces Munich (UniBwM) in 2004, and a Diploma Degree in Computer Science from the Technical University Darmstadt in 1998. He participates in several international VV&A working groups on behalf of and in close cooperation with FOI and FMV. His research interest focuses on VV&A, decision making, component based simulation, the HLA, and formal modeling languages.

Dr. René Jacquart is a senior research scientist of ONERA – Toulouse research centre. He is involved into VV&A research and studies for defence simulations since 1995. His major interest topics are dealing with formal foundations for M&S and VV&A, requirements formalisation, formal specification and validation by formal techniques.

Dr. Jeroen Voogd is a member of the scientific staff in the Defence, Security and Safety Division at TNO. He holds a Ph.D. (1998) in Computational Physics from the University of Amsterdam in the field of modeling and simulating biophysical systems on parallel and distributed computing platforms. Currently he is involved in projects on simulating group behavior, operational analysis studies of army operations and connecting C4I infrastructure to simulators. A recurring theme in his work of the last years is quality. This involves issues round fidelity and VV&A of simulator assets, as well as quality assurance within TNO in the form of e.g. audits.

Dr. Choong-ho Yi is a Scientist at the Department of Systems Modelling, Swedish Defence Research Agency (FOI). Currently, he is leading a project “VV&A of simulation models” at FOI. His research interests include knowledge representation, conceptual modelling, and formal methods. Before joining FOI in 2001, Choong-ho worked as a university lecturer for over six years. In 2002, he obtained his Ph.D. in Information Systems Modeling, from Linköping University, Sweden.