A Proposed Evolution of Validation Definition

D. Girardot, R. Jacquart

***Abstract* — The United-Kingdom and France started a common study in 2001 in order to produce a VV&A Framework. During working meetings, some issues appeared concerning VV&A process and nature of VV&A itself. This paper focuses on some particular questions relating to Operational Analysis when real world is not - or badly - known. After reminding general concepts of M&S, as the conventional VV&A thought process, it introduces some contributions previously written that have discussed the issues we have to deal with. Then, it comments on the definition of Validation with respect to the "real world" anchoring used in DoD and DMSO glossaries. The concept of Validation is progressively refined through examples coming from the Software Engineering community and discussed in order to introduce its relative and quantitative nature. Introducing the idea of Measures of Merit as a specification representation of both "intended purpose" and "just good enough" aspects does this. Finally, a definition of Validation is proposed based on two complementary views devoted to the "what" and the "how" of the Validation definition.**

***Index Terms* — VV&A, Real World, Validation definition, Credibility**

I. INTRODUCTION 1
II. VV&A logic and assumptions 2
A. Existing M&S Processes and their contribution 2
B. The VV&A Metaprocess 2
III. Discussion 3
A. Base Model, Referent, User Requirements and SME 3
B. Fidelity 4
C. Variability in trial approaches 4
D. Application Domain limitations 4
E. Variability in application context 4
F. To sump up with the first definition 5
IV. An IEEE Definition 5
A. Positive consequences of this definition 5
B. Problems and differences with the M&S context 5
V. Yet another definition 6
A. OA needs for a study process 6
B. Measures of Merit 6
C. Definition of validation 6
VI. Conclusions 6
VII. Acronyms 7
VIII. References 7
IX. Authors biography 7

D. Girardot is with the French Ministry of Defense, Délégation Générale à l'Armement, Centre d'Analyse de Défense, 16bis avenue Prieure de la Côte d'Or, 94114 Arcueil Cedex, FRANCE (phone: +33-1-4231-9179; fax: +33-1-4231-9026; e-mail: daniel.girardot@cad.etca.fr).

R. Jacquart is with the Office Nationale d'Etudes et de Recherches Aéronautiques, Centre d'Etudes et de Recherche de Toulouse, 2 Avenue Edouard Belin, 31055 Toulouse Cedex, FRANCE (phone: +33-5-6225-2877; fax: +33-5-6225-2593; e-mail:jacquart.@cert.fr).

I. INTRODUCTION

EARLY 2001, the United-Kingdom and France started a study in order to produce a VV&A Framework. This work is placed under the auspices of the *Anglo-French Defence Research Group* (AFDRG) supervising studies common between the two countries.

It was decided that the results had to be fitted and validated in two different application domains, UK people being involved in Synthetic Environment Based Acquisition (SeBA), while French people will treat Operational Analysis studies.

At that point of our thoughts, some important issues appeared coming as well from methodological advances as during discussions concerning the involved process and the nature of VV&A in itself.

The theoretical results can be found in four British and French reports [1].

Evolution of the VV&A concept

To start reasoning on Validation, we have to make some remarks about the permanence of VV&A concept. Already used in ALSP and DIS, this concept has evolved in parallel with the FEDEP [5]. The works about VV&A have permitted to refine – and in some sense to complexify – progressively the related concepts.

The software engineering foundation of the FEDEP allows bringing federations into operation in a coherent way. Many experiments and practical realizations have shown that, in a general way, the FEDEP can be adjusted to support different kind of developments.

They have already brought some issues to light. Different aspects of the question have progressively been explored, for example, fidelity, conceptual model definition, or recently substantive interoperability or credibility.

This enumeration shows that we have to deal not only with technical items but also more and more with fuzzier concepts related to technical or operational acceptance. In fact, the VV&A center of gravity is progressively shifting from verification to validation and accreditation. In addition, some important issues appeared, as for example those derived from the more important part dedicated to human decision making.

All this establishes that VV&A cannot be considered as an immutable concept.

The validation process and products are an essential basis of decision-making during accreditation. Therefore, we have

decided to focus our approach on the effects of the evolution of Validation concept.

The current paper does not cover all topics possibly related to technical or social acceptance, but only those linked to the way of addressing the relation to the real (or actual) world when conducting Operational Analysis studies.

During the development, these aspects are covered by a good understanding of references and validation framework - as discussed in this paper - supported by a formal methodology.

After, and during the whole life-cycle, it is necessary to set up an efficient information system to enriched continuously data important for validation.

II. VV&A LOGIC AND ASSUMPTIONS

The current official definitions of verification, validation and accreditation [15] are based on a global knowledge during simulation development - starting from the Real World to end with the Final Product. Real World and Final Product are in fact shortly linked in validation and accreditation processes, as the Validation definition shows it:

***Validation*: is the process of determining that a model or simulation is an accurate representation of the real world from the perspective of the intended use of the model and simulation.**

Accreditation assessment is then performed based on evaluation of V&V and tests results.

A. *Existing M&S Processes and their contribution*

Previous technical contributions have presented a global view of the Modeling & Simulation process [17,18,19]. Shortly presented, the classical M&S process implies some existence of a real world, a conceptual model, and a simulation (as result of the process). Many development processes, including the FEDEP, match that kind of logic and chronology:

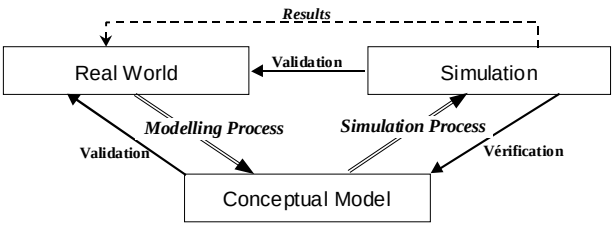

Fig. 1. M&S Process

The DMSO definition is directly translating this view. In some cases (low-level training, training with platforms, etc.), it fits well with the development process and the V&V relationships between the main involved elements.

The chronological representation of the M&S Process in terms of main VV&A products is the following:

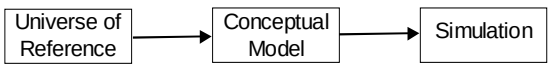

Fig. 2. Main Implementation Elements

Notice that the real world is to be considered as an "universe of reference".

To go further in the capture of the permanent knowledge coming from the real world, Hone and Moulding [9] have derived this basic M&S Process in a more sophisticated view of a VV&A metaprocess called "Y-Process". This is moreover in line with contributions found for example in CMMS or FDB. This schema is more powerful as it stresses that knowledge comes in parallel from application and problem domains.

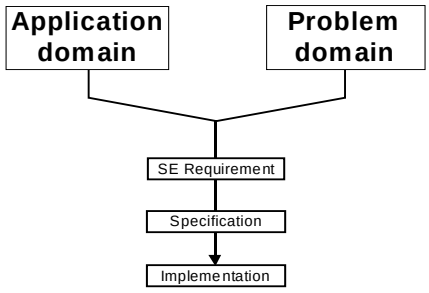

Fig. 3. The Y-Process

This "Y-Process" presents a method that insure a good traceability between needs (application domain), knowledge of problem context (problem domain) and simulation to provide.

Actually, the traceability is an important informal proof or justification that a process was followed in a correct manner. We could remark that this is more related to Verification than to Validation but acceptable as a reasonable contribution. In that way, it takes part of validation as an accreditation data.

The Y-Process brings up the contribution of knowledge modeling and its impact on different aspects of development, use and cost, in particular validation, reuse and capitalization.

B. *The VV&A Metaprocess*

The VV&A metaprocess is associated to the Y-Process in a frame intended to define clearly the related activities of development and VV&A.

These relationships have been presented in [2]:

The problem domain represents the battle space or environment which is to be modelled. This is the context in which the problem exists. The application domain, on the other hand, is what the SE is to be used for. This is why we want to address the problem. The Y-model approach suggests that a useful SE will be developed if the problem domain and application domain converge through a clarity of purpose.

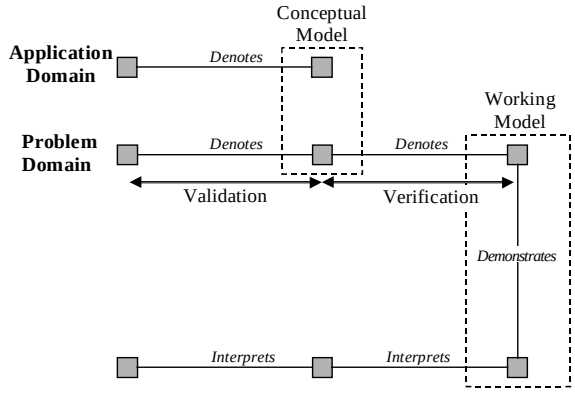

Fig. 4. VV&A Process

- Application domain defines the objectives and constraints of a specific purpose. When trying to extend the definition, we can show that the problem CM is derived of *intended purpose*.
- Problem domain is the set of data common to a whole kind of simulations. It offers many opportunities of sharing common experiences and reuse and is based on the *referent world*.

The Y-Process has been modified to clarify these different concepts and to introduce the validation links:

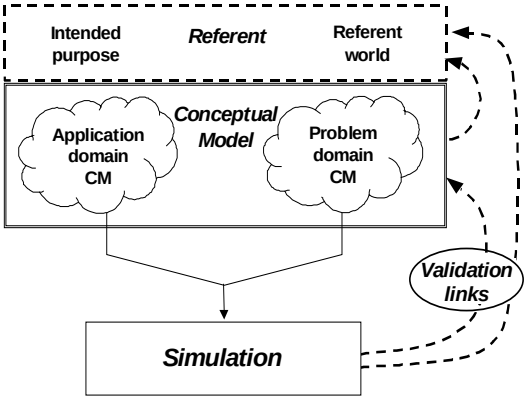

Fig. 5. Modified Y-Process

The intended purpose and referent world have their own peculiarities and they both contribute to the conceptual model design, which itself determines all the implementation process.

1) The Universe of Reference

As describe before, the universe of reference comprises both application and problem domain. That "**Referent**", in many cases, could be in an acceptable way the real world itself (when we are in data rich context to reuse the D. Pace terminology).

But, the definition can be likely extended to many other items: subject matter experts' (SME) opinion, other simulations results, theoretical studies, literature, previously validated models, empirical or experimental data. In any cases, the "real world referent" or its surrogates or substitutes is the basis of development and, therefore, of the validation process.

2) The Conceptual Model

As acknowledged in all technical contributions, the conceptual model is a pivotal point of the whole process.

A now well-know definition has been given and documented by Pace (for example [6])

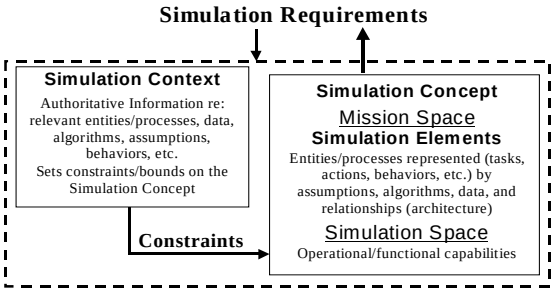

Fig. 6. Pace Conceptual Model

From the VV&A point of view, the CM is an obligatory step in a path between users and developers roles. Involved people have to understand it as a contract between all parts. Its validation is so critical for cost-effective development that, actually, the resources needed to establish the CM are of less importance than the consequences of a inadequate validation.

3) The Simulation

End-result of the development process, the simulation is derived from the previous products. But we have to remind strongly that it is NOT the end of the real and operational use from the perspective of a full life-cycle. This is specifically true if we refer to the data/results collection, exploitation and capitalization.

III. Discussion

It stems from the previous presentation that the application and domain knowledge have been formalized and represented as important components of a referent world. But many operational contexts could justify the assertion that the real world could NOT be reached: cost, complexity or even non-existence of this real world, in particular when studying future objets or situations. From this point of view, it is important to identify the domains where the real world is NOT the referent (data poor application and problem domain) and to propose alternatives.

In the following paragraphs, we discuss other contributions on this topic, their perspectives and limitations.

A. Base Model, Referent, User Requirements and SME

In its seminal work Zeigler excluded most of the time all strict comparisons with the real world [20]. The real world contains many data and properties without any relationships with the intended purpose and their representations are not useful. He proposed to define what he names a "base model", whose objects, relations and properties constitutes the "universe of discourse "of the intended simulation. From this base model a "lumped model" is derived to define what will, actually, be represented in the simulation.

One can try to take up Zeigler's model again with current usual terms. Here, we can assimilate the base model to Users Requirements or to the Referent. The lumped model is named in the present terminology the Conceptual Model.

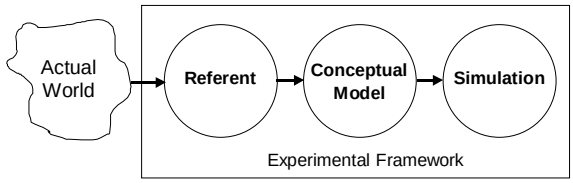

Fig. 7. Experimental Framework

The Zeigler's proposed framework is very useful, even 25 years later, as it clarifies the validation and verification relations. The construction of the base model is obviously a very complex task, and Zeigler, initially, has suggested that this model was not built explicitly. This position now could be changed in the perspective of all the results available from the Requirements Engineering domain and the way of managing

the knowledge acquisition phase from SME.

The role of SME in the M&S process and its related VV&A sub-process has many facets that have been finely analyzed and documented in the on-line RPG.

We could mention here the work of SME related to:

- initial acquisition phase, the analysis, elicitation, refinement and derivation of requirements which leads to the specification of the "intended use"
- early validation of this specification,
- prediction of expected/desired results,
- validation of Conceptual Model,
- validation of end-product,
- validation of simulation results,
- acquisition by SME of new expertise based on the current process and its direct or indirect results.

All the tasks previously listed are common to System or Software Engineering and M&S, even they have been mainly proposed and experimented in Synthetic Environments (SE). The M&S area generally and the Operational Analysis domain specifically have their own set of difficulties in extracting and validating high-level knowledge.

In these areas we could observe that *the validity of models designed in the context of defense is blurred by the combined facts that*

a) little data may exist, or it may be impossible to validate;

b) we may be considering capabilities and systems that are entirely conceptual (perhaps the future tank replacement, for example);

c) The profession of arms is (despite the best efforts of some very intelligent men) irreducible to a science, and the human aspect of war will always render it, to a certain degree, imponderable. This is certainly not true in most other domains where simulation is used to assist with decision making.

These observations could be seen as the most stringent of all the issues that we have to deal with in Operation Analysis (OA), especially when studying future systems or doctrines.

B. Fidelity

A second problem in using the real world as a referent is related to comparison between real world – or referent – data and simulation results.

That could be illustrated via the fidelity paradigm based on the following metaphor:

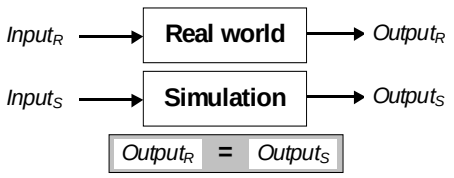

Fig. 8. Fidelity Paradigm

The Fidelity metaphor implies that:

- (some) knowledge of the real world is mandatory,
- "fidelity is good and that more of it is better" (see Fidelity ISG [7]) . This is stated in Fig. 8. by a representation of the "best" reachable situation, the one where the outputs are equal (for equal inputs).

These position statements exclude all the application where we need an abstract or aggregated representation of knowledge (what will be the fidelity expression for "information dominance" as well as for "doctrines of engagement". In these cases, the evaluation by measurements and comparisons should be very difficult, or weakly informative.

C. Variability in trial approaches

Jacquart, Bourret [10] have reused and adapted to M&S a framework proposed in Pfleeger [11] about the way of conducting trials or experiments and, then, inferring knowledge from them. This contribution was initially devoted to investigate consequences of the introduction of new methodologies in the Software Engineering area. It is quite easy and fruitful to reuse these concepts in the M&S area where, most of the time, we are trying to infer knowledge from trials.

Pleeger has introduced a set of criteria and constraints in order to formalize the implementation and control of trials. This led to the exhibition of two major ways of setting up trials:

- case studies, which are characterized by lose control on parameters and variables, and high cost for replication,
- formal experiments where we could have, conversely, quite strong control and cheap replication.

This discrimination is also a good approach to characterize low-level M&S (data rich-environments and formal experiments) against high-level M&S (data-poor context and case studies). In this last category the referent is hardly controllable and is costly to access.

D. Application Domain limitations

Miller, Allen |14] have shown some of these limitations in evidence. In fact, simulation approaches by themselves induce difficulties of non-realism for some application domains.

It is for example the case for training simulation if some representations of combat effects are inadequate. When credibility is weak, (non-lethal armaments, bad visual or sound effects, neglected human factor…), the outcome can result even in "negative training".

About Analysis, the process of studies generates its set of own difficulties. Actually, the need of multiple runs and scenarios excludes many live or virtual simulation for cost or organizational issues, and implies rapid loops involving interlaced steps of collecting data, models development and simulation runs. Consciously or not, mutual dependencies may be introduced among these steps, and the full process could be complex to master. We could say, even more strongly, that as the referent is NOT existing, the process IS the referent: it constructs the way to reach the desired level of knowledge.

E. Variability in application context

In our methodological proposal we can find an organizational view of a VV&A-oriented development process. It put the emphasis steps considered as important for

Validation than the ones related to development stages.

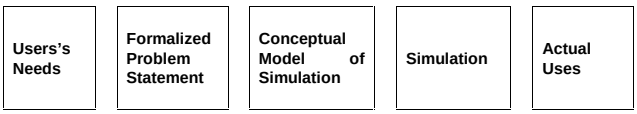

Fig. 9. Validation oriented development

This particular view is consistent with many development processes. It does not be seen as a chronological diagram but as an analysis vehicle in order to precise the validation organization.

The variability problemsmentionned in the title of this section are mostly associated to the actual weight of the different step and to the knowledge available at these stepss. For OA simulation, the first and last steps have a reduced contribution while the second one is magnified (the rigth way to state the problem is essential for OA , this defines the border of the study and consequently the level of the CMS).

This observation could be related to a very broad classification of simulation uses : the ones related to decision support (for every context) and the ones devoted to skills acquisition (training, formation,…).

The identification of these two classes has implication on the way of managing simulation programmes as well as their consequences (accreditation, return of experiment, capitalization of knowledge,..).

F. To sum up with the first definition

All the previous examples show that identifying the referent to the Real World holds difficulties of different kind: variability in evaluation capabilities, application limitation, M&S levels of aggregation etc.

We can observe that, precisely, for the OA domain the VV&A could be very complex and disputable. In many cases, the real world is not easy to hit, or its use involves some important bias.

Therefore, we have to try to search for a new definition of Validation

IV. An IEEE Definition

IEEE/ETA 12207.1 proposes a definition of Validation for the System and Software Engineering domain: .

***Validation*: is the process of determining that the requirements and the final as-built system or software product fulfil its specific intended use.**

A. Positive consequences of this definition

The reference to the Real World is absent in this definition and is replaced by the intended use. The requirements here are the mean of anchoring the development of the final product (for the M&S context the simulation) both have to satisfy the intended use, which appears here as a surrogate of the referent. This is a more precise and limited definition than the previous one if it is possible to specify (in a semi-formal or formal way) this intended use. The following schema tends to summarize the process:

interested in abstract notion of satisfaction or behavior equivalence. For instance, one can remark that variable of interest are not necessarily observable. In fact, such variables like doctrine of use, dominance, or capabilities - very important when evaluating systems efficiency or situations – are aggregated data. They can be handled by simulation without any equivalence in the real world.

Model validity and evaluation have been linked in [2]. As in the following excerpt:

Model validity is a relative matter. If we try to "prove" that a model is "right" in some sense we are doomed to failure. Model validity is all about establishing whether a model is good enough. From this perspective, we should perhaps judge the model not against the real world, but against what else we might use to aid our decision. Interestingly, this is very much in line with the ideas of Jay Forrester, who established the discipline of System Dynamics modelling in the 60s. "Model validity is a relative matter. The usefulness of a mathematical simulation model should be judged in comparison with the mental image or other abstract model which would be used instead ". In essence, he says we should not be comparing models with perfection but with what else might be available. In other words, we should use the best model available to inform our decision-making. Whilst this argument has some merit, we need to use it in the context of our need to demonstrate an acceptable level of validity in order to justify a Synthetic Environments fitness for purpose.

The validation of simulation in this context is perceived as a kind of multi-criteria decision-making.

V. Yet another definition

A. OA needs for a study process

If we ask OA practitioners about their preferred development process, they answer that it is such that different components are improved concurrently. We can observe that they are practicing a kind of zapping between requirements, CM, use, and exploitation of results. As the problem definition evolves regarding the first results, the knowledge of the real world is thus progressively increased as it could be observed in all activities of information gathering for decision making.

Nevertheless, with this kind of spiral/circle-like process we are possibly facing never ending developments with some risks of over-quality and bad uses of resources.

The quality that we intend to reach must be also a part of the specification of the intended use. This can be done by introducing Measures of Merit.

B. Measures of Merit

Measures of Merit are tools necessary in order to determine the quality of the produced simulation. The set of measures must express the fitness for purpose in a way that could be understood by customers, developers and analysts.

The set of MoM in that case acts as the referent. The MoM offers a frame to modeling and simulation. It has to translate the needs into measurable criteria intended for helping the final decision, to say accreditation. Customers and analysts have to use them and to interpret them for accreditation

They are the right vehicle to produce informed comparisons between expected/required/observed outcome values or process values of the simulation. The involve means could be graphical (such as Kiviat's diagram) or mathematical techniques coming from Constraints Satisfaction or Optimization

The MoM are also of great importance to tackle with cost-effective development and "just-good-enough quality" that permits costs savings and improvements in just-in-time supplying.

With this approaches using MoM, the validity level is adjustable in relation with intended purpose, available resources, and simulation means

C. Definition of validation

The previous discussion was giving elements to motivate an evolution of the concept, and then the definition of Validation. The current state of our understanding does not give easy ways to capture all the previously discussed connotations of Validation in a single definition. This is illustrated by the two definitions given below: one tackles the WHAT, the second one opens some ways about HOW doing Validation.

What-definition

Validation is the process of determining that a model and simulation has been evaluated as fit for purpose from the perspective of the intended use.

How-definition

Validation of models and simulations is a decision-making process based on their evaluation and assessment against a set of Measures of Merits (MoM). These MoM characterize (specification of) the intended use and the tolerance level for the satisfaction of this use.

VI. Conclusions

This article benefits of contributions produced by a UK-FR contractual co-operation MOU [1]. The aim was to produce an improved VV&A framework starting from the DMSO standards.

As a first step, this paper considered the concepts and terms used and their connotations in some particular application areas. Some points critical for domains such as Acquisition or Operational Analysis were discussed. For these application domains, the use, in the official definition of Validation, of a reference to a presumably existing "real world" was not considered as a simple matter to tackle.

A discussion on the concepts of Real World or Referent (as a surrogate) has derived towards the production of Requirements and the way of being convinced of their fulfillment by the delivered product. This connotation was precisely captured in the IEEE/ETA 12207.1 definition for Validation.

A discussion on some peculiarities of M&S was at the origin of a tentative definition of Validation integrating some connotation of evaluation and tolerance level of satisfaction, these being captured by Measures of Merit.

This work is still in progress and the definition could be improved to better take into account the different ingredients that are used to support actually the Validation as a decision-making process. This work will benefit as well of approaches as the one presented by O. Balci in a recently published article [21] , unknown by the authors when writing this article.

VII. Acronyms

AFDRG	Anglo-French Defense Research Group
ALSP	Aggregated Level Simulation Protocol..
CM	Conceptual Model
CMMS	.Conceptual Model of the Mission Space.
CMS	Conceptual Model of the Simulation
DIS	.Distributed Interactive Simulation.
FDB	Functional Descrition of the Battlespace.
FEDEP	Federation Development and Execution Process
HBM	Human Behavior Modeling
ISG	Implementation Study Group
M&S	Modeling and Simulation
MoM	Measures of Merit
OA	Operational Analysis
SDI	Space Defence Initiative
SeBA	Synthetic Environment Based Acquisition
V&V	Verification and Validation
VV&A	Verification, Validation and Accreditation

VIII. References

[1] [AFDRG-SG6, 01] See [John, Coombs, Bathe, Price, 01] and [Jacquart, Bouc, Zanon, Girardot, 01] below.

[2] Professor P John, M E C Coombs, M R Bathe, S N Price, "Validation & Verification of Synthetic Environment based Acquisition", Report 1 – Application (January 2001), Report 2 - Fitness for purpose, *Royal Military College of Science*, Engineering Systems Department, 2001.

[3] R. Jacquart, P. Bouc, G. Zanon, D. Girardot, "Validation des simulations technico-opérationnelles", Rapport 1.1 - Domaine d'application (May 2001), Rapport 1.2- Adéquation au besoin (August 2001), *ONERA*, *CAD*, 2001

[4] S. Youngblood, S. Harmon, D. Pace, "Federation Credibility Challenges", May 2001.

[5] [FEDEP, 99], "HLA, Federation Development and Execution Process (FEDEP) Model", version 1.5, *DMSO*, Dec. 1999.

[6] D. Pace,, "Development and Documentation of a Simulation Conceptual Model", *Fall SISO Interoperability Workshop*, Sep. 1999

[7] [Fidelity ISG, 99], from "Fidelity Implementation Study Group", *Spring SISO Interoperability Workshop*, 1999

[8] J. Graffignini, S; Youngblood, R. Lewis, "An Overview of VV&A process for the HLA FEDEP"

[9] Hone, G. N., Moulding, M.R., "Application Domain Modelling to Support the Verification and Validation of Synthetic Environments", *Spring SISO Interoperability Workshop*, Mar. 1998.

[10] R. Jacquart, P. Bourret, "État de l'art sur la validation de simulation", *ONERA/DTIM*, Sep. 1998

[11] S. Pfleeger, "Design and analysis in Software Engineering", *SIGSOFT vol. 19*,4 October 1994

[12] R. Jacquart, R. Miller, "The development of the VV&A process: a franco-british view", *16th DIS Workshop*, mars 1997

[13] D.S. Hartley, "Verification and Validation in military simulations", *IEEE Proceedings* 1997 WSC

[14] R. Miller, P. Allen, "VV&A of synthetic environments: proposed definitions", *14th DIS Workshop*, Mar. 1996

[15] DoD/DRE/DMSO, "VV&A Recommended Practices Guide", Nov. 96

[16] R. Burton, B. Obel, "The validity of simulation models in organization science - from model realism to purpose of the model", CIFE WP 30, *Stanford University*, April 1994

[17] J. Palmore, "Conceptual models for verification, validation and analysis in distributed interface simulations", *11th DIS Workshop*, September 1994

[18] R. Sargent, "An overview of verification and validation models", *IEEE WSC*, 1987

[19] B. Zeigler, "Multifaceted modeling and discrete event simulation", *Academic Press*

[20] B. Zeigler, "Theory of modeling and simulation", *J. Wiley*, 1976

[21] O. Balci "A methodology for certification of modeling and simulation applications", ACM Trans. MCS , vol 11,4, oct. 2001

IX. Authors biography

Daniel Girardot is a Software Quality Engineer of the French Centre d'Analyse de Défense (Center for defense analysis). After a master's degree in physics, he was involved in operational analysis studies. Then, he worked on different domains of information science (system, network, and graphics software…) and also software methods and quality. In charge of VV&A methodology since 1996, he is currently taking part of miscellaneous working groups, in particular with UK and NATO.

René Jacquart is a Research Engineer at ONERA, a french research institution for Aeronautics and Space. After a PhD degree in Computer Aided Design , he was involved in many projects related to software engineering, formal methods and their applications . He is currently involved in research devoted to VV§A methodologies , both for their definition and their application to DGA programmes.