MSIAC's M&S Journal Online

Volume 4, December 2006

# *CAPABILITIES-BASED COSTING: APPROACHES TO PRE-MILESTONE-A COST ESTIMATING*

Office of the Deputy Assistant Secretary of the Army for Cost and Economics (ODASA-CE)
Arlington, VA 22202, United States

POC: Martha Roper
Senior Analyst, ODASA-CE
martha.roper@hqda.army.mil
(703) 601-4177

The issue of early, rigorous evaluation of program costs is becoming more important as defense funding comes under greater scrutiny. Often at this point in the life cycle, a requirement or desired capability is known, but the manifestation of the solution is unknown or described only at a high level. Can capabilities alone be used to produce a cost estimate? If so, how can we link the proposed solution to existing systems if only a particular solution's general capability set is known?

This work submits that better strategic decisions within fiscal constraints could be made if rough order of magnitude (ROM) estimates were available for proposed materiel or non-materiel solutions, based on that solution's capability set. This project further proposes the use of a knowledge base to provide support for these estimates; it is known as the joint Capabilities Knowledge Base (CKB). By using the relevant entities extracted from CKB, a ROM cost estimate may be developed using a wide spectrum of techniques.

According to Department of Defense (DoD) guidance dated June 19, 2006, the [2006] Quadrenniel Defense Review (QDR) report called upon senior departmental leaders to "better integrate the processes that define needed capabilities, identify solutions and allocate resources to acquire them in order to enable corporate decision-making that cuts across traditional stovepipes". In response to this directive, DoD leaders are evaluating a new early lifecycle decision-making framework that includes a Concept Decision (CD) Review (supported by an Evaluation of Alternatives or EoA). The CD Process has been set forth as a way to combine requirements, capabilities portfolio evaluation, and resource allocation considerations in the pursuit of joint, efficient, and well-informed decision-making early in the acquisition life cycle. The Concept Decision will either replace or occur in conjunction with Milestone A to decide which of the prospective solutions provided by the EoA will best enhance overall US defense capability while balancing priorities of cost, schedule, and risk management.

The issue of early and rigorous evaluation of program costs becomes more and more important as defense funding becomes more scrutinized. Clearly, decision-makers need high-fidelity cost information at this key decision point, but more often than not, it is scant. Providing reliable, useful cost estimates very early in the acquisition life cycle is challenging for several reasons. Often at this point in the life cycle, a requirement or desired capability is known, but the manifestation of the solution is unknown or described only at a high level. This is certainly a challenge, given that defense cost estimating is usually performed given a detailed system description. Given the changing face of the battlefield and warfare, proposed solutions are often unlike anything presently in existence.

As any cost estimator can confirm, there exists a spectrum of situations in which a cost estimate may be prepared. One theoretic extreme is creating a cost estimate in a situation where there is very little information about the item being estimated and no supporting data. The other extreme is when the entity being estimated is fully understood, and all data exists to estimate the cost exactly. In this case, the data are actual costs after the item has been developed, constructed, or bought. Figure 1 shows these extrema along with all points in between.

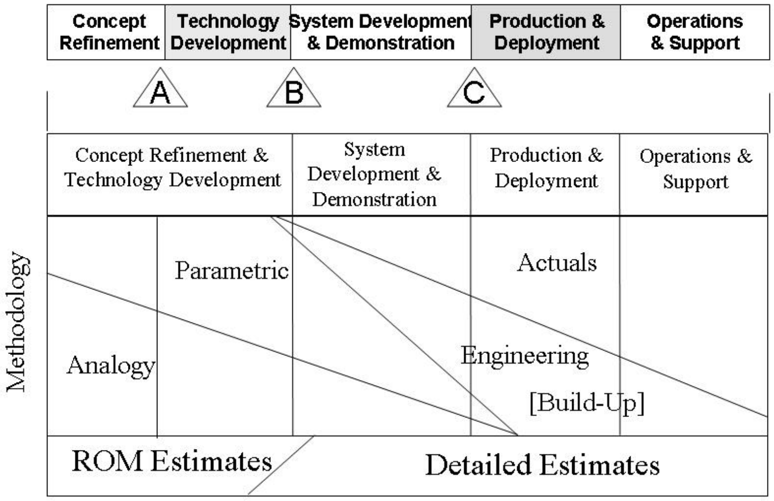

Figure 1: The Cost Estimating Spectrum

As we progress from the point of no information to the point of perfect information, our cost estimating methodology changes to suit the information climate. For instance, when information about the item or service is higher-level and/or data is not readily available (Figure 1, left), cost estimators tend to rely upon analogies and parametric methods to produce their estimate. However, as we move toward the right, estimates tend to utilize more “data-hungry” methodologies such as engineering builds and projections using actual costs to date. It is also clear to the casual observer that as we move along the spectrum from left to right, we may expect our estimate to be more reliable and closer to the actual cost at project or acquisition completion.

The pre-Milestone A costing environment is particularly challenging. This is the stage in which information is often extremely scarce. Figure 2 illustrates the “sub-spectrum” of pre-Milestone A data availability.

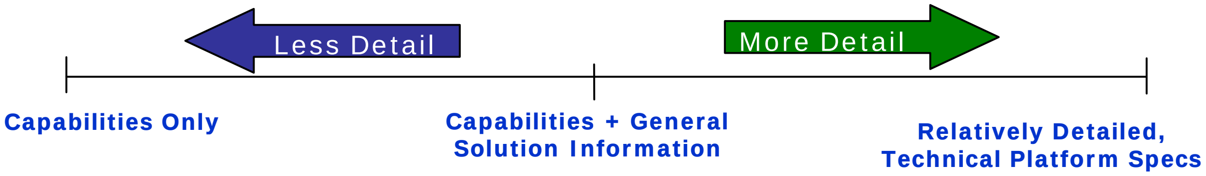

Figure 2: Spectrum of Information Availability at Milestone A

Depending upon the situation, there may be one or many proposed solutions to a set of capability gaps before Milestone A. These solutions could be a materiel system such as a vehicle or software package, or it could be a non-materiel solution, such as a policy change or a training curriculum change. As one can see in Figure 2, the information regarding the proposed solution(s) could range from simply the desired capability expressed in very qualitative terms to a relatively detailed, well-developed concept with some technical platform specifications. The most commonly-occurring scenario, however, is nearer to the middle where there exists high-level capabilities information along with some very general solution information.

Since every cost estimate of an item or project must be based on some type of past experience, pre-Milestone-A cost estimating is no exception. How can we link the proposed solution to existing systems (our past experience) if we know only a particular solution's general capability set? Can capabilities alone be used to produce a cost estimate? If so, could that cost estimate be used in decision-making with any degree of confidence?

Suppose we made the assumption that a system's capabilities have a relationship to its cost. To the casual observer, this assumption seems rather logical. If we buy something that can do more, do it quicker, or do it better, then it should cost more. However, one can identify situations in which this assumption might not hold; if a particular computer technology is maturing at an accelerated rate, the cost to acquire that capability might not be correlated to the cost of acquiring a similar capability five years ago. Yet, even this example has a relationship between capability and cost upon closer inspection; to arrive at an acceptable cost estimate one must understand the rate of technology maturation (and this maturity information may or may not be available to the analyst). The question at hand, however, is whether or not capabilities can predict cost within some acceptable level of percentage error to provide decision makers with data that helps avoid decisions that would yield negative future cost effects. In theory, these decisions could be avoided if a rough order of magnitude (ROM) estimate is available that is based on the proposed materiel solution's set of capabilities.

The capabilities costing team at the Office of the Deputy Assistant Secretary of the Army for Cost and Economics (ODASA-CE) is currently tackling the challenging pre-Milestone-A costing environment. Our approach includes the use of a knowledge base that records current system cost information and capabilities. In fact, the Joint Capabilities Knowledge Base (CKB) is presently under construction. By using the relevant entities extracted from the CKB, a ROM cost estimate may be developed using a wide spectrum of techniques.

Numerous costing approaches are being examined and developed as this project evolves, one of which follows: Let us assume that the set of capabilities requiring a cost estimate is rank-ordered; in other words, we know which capabilities among the group are most critical, somewhat necessary, or only slightly needed. Depending on whether the entities (that will be extracted from the CKB) have exact or partial matches in capabilities, we can then apply an appropriate weighting factor for certain combinations of capability matches; exact matches would receive a higher weighting than partial matches, for example. Next, relevant entities are extracted from the knowledge base that can be used in our cost estimate. The assigned weightings are applied. If a particular system entity is deemed to be even more relevant to the solution being estimated, it may be further emphasized in a variety of ways.

Larger-scale case studies using realistic scenarios are under development to test the usefulness and strength of the methodology frameworks being considered, which include simplistic techniques like that described above to more intricate parametric and data mining approaches. It is important, however, to emphasize that cost estimates at this point in the life cycle are highly situation-specific, and thus methodologies under development are only recommended strategies. The analyst's judgment is a key component.

Martha Roper is an Operations Research Analyst (ORA) with the Office of the Deputy Assistant Secretary of the Army for Cost and Economics (ODASA-CE). Ms. Roper is leading early-life cycle cost analysis initiatives in conjunction with analysts Nancy Houdek and Katherine McCormack.

# *STANDARDIZING ON ONTOLOGY OF PHYSICS FOR MODELING AND SIMULATION*

Dr. Joseph B. Collins
Naval Research Laboratory
4555 Overlook Ave., SW
Washington, DC 20375
202-404-7041

**Keywords:**
ontology, physics, modeling, markup language

## ABSTRACT

Interest in creating various scientific markup languages has been stimulated by the advent of XML and OWL. For example, there is a need for an ontology of physics for representing physics-based model semantics in Modeling and Simulation (M&S) applications. While basic principles have been outlined to proceed towards creating such an ontology, the difficulties in creating a standardized ontology lie in the magnitude of the task and the diversity of communities interested in having such an ontology.

Significant application areas of interest outside of M&S for scientific markup languages and their ontologies lie in the creation and accessing of electronic libraries and document archives. Other application areas include computerized science education, engineering education, and electronic navigation of technical manuals. The creation of credible standards in this area will require the pooling of resources from these distinct communities. We describe the variety of requirements of these communities and outline an approach for building a consensus towards standardization.

## 1. Introduction

In a prior paper [1] we initiated an examination of the requirements that Modeling and Simulation (M&S) would create on an Ontology of Physics for describing physics-based models. One of the key conclusions was that an Ontology of Physics would need to be a widely agreed-upon standard in order to achieve wide interoperability. While expanding a search for consensus may conceivably add new requirements for an ontology, the effort is nevertheless worthwhile in order to achieve broad interoperability. It is difficult to anticipate the interests and consequent requirements of a broader community without directly engaging with them in the development of standards that they would have an interest in using. We believe that for an Ontology of Physics we should work towards a standard for the Web as a whole, engaging the talents of many contributors, rather than to only support a Department of Defense (DoD) M&S intranet.

Our operating definition of the term ontology is as follows. An ontology is a formal, explicit description of concepts, their properties, relationships between the concepts, and the allowed values that they may take. An ontology together with a set of individual instances of classes constitutes a knowledge base [2]. An ontology provides a semantic reference frame useful for automating the communication of abstract information. The purpose of an ontology is to enable the communication of meaning for purposes of understanding, where understanding is achieved through common usages. It allows the addition of descriptive tags to existing terms, describing assumptions, contextual and other information that often goes unexpressed due to the lack of a formal structure for making such expression.

Note that our definition of the term ontology does not specify or recommend a particular computer language mechanism to be used. In order to proceed forward from the definition, we will need to choose some specifics of this kind, i.e., languages and tools. In the last several years the development of the Standardized General Markup Language (SGML) and, more particularly, the

Extensible Markup Language, (XML) has led to the realization of a capability to capture the ideas embodied in an ontology and put them to use in elucidating semantics within documents and data. These constructs and an associated set of ontologies and knowledge bases are being developed to create the Semantic Web. We will need to use these tools in developing a standard ontology of physics.

## 2. Tools

The Semantic Web is an idea conceived by the World Wide Web Consortium. Notable among these is Tim Berners-Lee, inventor of Hyper-Text Markup Language (HTML) and the first web browser, and currently director of the W3C. Whereas HTML allowed the creation and easy access and display of text-like documents, the semantic web consists of a set of constructs that will support the representation of layers of semantic descriptors, or metadata. These metadata promise to lessen ambiguity and even support intelligent automated processing of documents on the web.

A variety of tools have arisen due to efforts of the W3C [5]. Recently, on February 9th, 2004, the W3C released the Resource Description Framework (RDF) and the OWL Web Ontology Language (OWL) as W3C Recommendations. RDF is used to represent information and to exchange knowledge in the Web. OWL is used to publish and share ontologies, supporting advanced Web search, software agents and knowledge management.

Another tool, RDF Schema describes how to use RDF to build RDF vocabularies. RDF Schema defines a basic vocabulary and conventions for use by Semantic Web applications.

The DARPA Agent Markup Language and Ontology Inference Layer (DAML+OIL), another tool, is a semantic markup language for Web resources. It builds on earlier W3C standards such as RDF and RDF Schema, and extends these languages with richer modeling primitives. DAML+OIL provides modeling primitives commonly found in frame-based languages. DAML+OIL (March 2001) extends DAML+OIL (December 2000) with values from XML Schema data types. DAML+OIL was built from the original DAML ontology language DAML-ONT (October 2000) in an effort to combine many of the language components of OIL. The language has a clean and well-defined semantics.

Ontology Inference Layer OIL is a proposal for a web-based representation and inference layer for ontologies, which combines the widely used modeling primitives from frame-based languages with the formal semantics and reasoning services provided by description logics. It is compatible with RDF Schema (RDFS), and includes a precise semantics for describing term meanings (and thus also for describing implied information).
A DAML+OIL knowledge base is a collection of RDF triples. These triples represent a subject-predicate-object triple, where the predicate is a relationship between the subject and the object. DAML+OIL prescribes a specific meaning for triples that use the DAML+OIL vocabulary. This document informally specifies which collections of RDF triples constitute the DAML+OIL vocabulary and what the prescribed meaning of such triples is.

Finally, other tools developed under the coordination of the W3C are the Mathematics Markup Language (MathML) and its extension, OpenMath, which we describe in more detail later.

## 3. The Subject of Discourse: Physics-based Models

The subject of our effort is to represent a vocabulary with which to express the physical concepts that may be used to describe the mathematical statements that comprise physics-based models. We include dynamical models as well as data: a language that includes verbs as well as nouns is much more expressive than one that only includes nouns.

We intend an ontology of physics to capture the concepts of physical theories in a formal language so as to support various forms of automated information processing that are not currently supported. The current primary use of computers for physicists is as calculation devices, to estimate predicted values of observables. A secondary use, not formally coupled to the primary use, is for supporting documentation and communication of collected data and models. We intend for an ontology of physics to connect together in a more formal way the conceptual physics, its mathematical expression, and the consequent numerical evaluation procedures to better support documentation and communication.

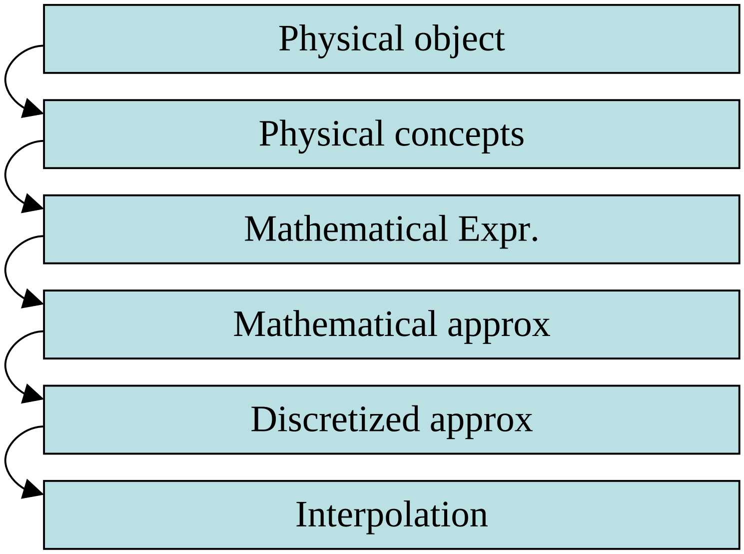

Figure 1. Model Ontology “Layer-cake”

To illustrate the structure of physics-based models, we diagram a description that we made earlier [1] in Figure 1. This hierarchical description of physics-based models illustrates the conceptual layers of the models and their relationships.

At the top is the actual physical object that is modeled. Its representation in an ontology is as a vocabulary of real-world objects, such as tanks and missiles. Each subsequent layer in the layer-cake provides a representation of the layer above it.

The physical concept layer is used to represent the physical attributes of the physical object. The mathematical expression layer represents the formalized statements of the physical concepts, i.e., the laws of physics in symbolic mathematical form. We believe that of these layers, that which is least developed from the perspective of a formal description is the physical concept layer. We describe below some of the fundamental concepts that belong in the physical concept layer.

Physics is a discipline for modeling real-world, physical objects and their effects upon each other as systems of physical objects. These objects and effects are characterized by model parameters called observables. Observables are measurable quantities that are counted in appropriate units, which are defined by standard reference units. The collective values of these observables comprise the state of a physical object. A defining concept is that if something is not observable, it is not physical.

All physical phenomena exist in space and time. Three parameter values are required to specify a spatial position and one parameter value is required to specify a temporal position. A physical object's spatial position may vary in time, giving it a velocity and a trajectory. Physical objects are often defined over a spatial region, and therefore have spatial volume. Physical objects have a spatial orientation that may vary by rotation.

Time is the dimension of classical conservation laws. By "conservation" we mean that conserved quantities are invariant as a function of time. Physical objects have multiple attributes that are conserved, among them being mass, energy, momentum, and charge. Conserved quantities may be exchanged between physical objects, but these quantities are not created or destroyed. The conserved properties of physical objects are what define physical objects and their persistence. A primary conserved observable in physics is energy. All physical objects possess energy and energy exists in multiple forms. Much of physical theory describes how energy is exchanged by physical objects and transformed from one form to another.

Causality is a temporal relationship between events that occur when physical objects are interacting. An interaction is defined by the exchange of some conserved quantity, for example, energy. Two events, whereby energy is emitted from one object and received by a second object, are causally connected and separated by a time lag. The notion of an asymmetric "cause - effect" nature of a causal relationship is that in two events that are causally related, one, the cause, precedes the effect. That one of two causally related events precedes the other is something we observe, a consequence of the fact that physical objects only progress in one direction in time. Since the fundamental conservation laws are statements of time invariance, they give no distinction between forward or backward time. Our experience that time goes forward is only explained by the law of ever increasing entropy, the Third Law of Thermodynamics. A non-intuitive property of causality, from the Special Theory of Relativity, is that if two events cannot be causally connected, the time-order of their occurrence depends upon the observer.

Various physical effects are exploited in the construction of transducers, devices created to transform energy from one form to another. These transducers are either used as sensors, which indicate to human sensory organs the values of observables, or as effectors, such as engines or machines, which are used to control and make changes in the state of the physical world. Direct knowledge of the physical world, i.e., of observables, comes to us only through our sensory organs and sensor devices. We cause intentional changes to the physical world through our bodies and the machines they manipulate. Experiments are conducted by a combination of manipulation and observation of the effects that the manipulation causes. The abstracted knowledge of collected observation and experimentation comprise physical theories.

The explanation of how physical objects interact is most succinctly stated in terms of mathematical theories of physics. Physical theories state mathematical relations between model variables that represent observables. What we call theories are model schemas, which define the ways to build models. The quality of physical theories are judged by considering: how well they predict the future states of physical objects given their past states; how few parameters and assumptions are required for the theory; their correspondence and consistency with past theories; and the ease with which they are applied.

In order to predict future states of physical objects with precision, physical theories are formal, i.e., they are mathematical in nature. Mathematics is commonly thought of as a pure, abstract discipline, independent of physics. This belies the observation that without attempting to describe and understand the physical world, most mathematics would not have been developed. A different way of considering the relationship between physics and mathematics is that bodies of mathematics often result from attempts to formalize descriptions of the physical world, and in a process of abstraction, lose reference to the physical world and become independent. These mathematical concepts usually maintain value when applied back to describing the physical world that inspired them.

Observed physical objects, their effects upon each other, and corresponding models have been categorized into sub-disciplines, or branches, of physics. These branches of physics are in part traditional, a product of an historical evolution, and in part de-compositional, partitioning physical theory into coherent components. There is currently no single, coherent and self-consistent model for the whole of physics. As a research discipline, the practice of physics is focused on the continued evolution and development of physical theory and the discovery and explanation of new physical objects and effects. Physics, as an applied discipline, is a body of theory that currently provides an incomplete, though useful, description of the physical world. A standardized ontology can only describe the concepts of physics as an applied discipline, since only those are settled matters.

## 4. The Rest of the Layer-Cake

In the mathematical expression layer of Figure 1, the physical concepts are represented with precise statements that can be used to provide, for example, predictions of the values of future states of the physical object. From a physicist's point of view, when the physical concepts have been laid out and the mathematical expression of those concepts written down, the model is complete save for a solution. Note that the model is considered physically incomplete or incorrect if the solutions are not functional, i.e., they must have a single solution for predicted values of observables as functions of space and time.

Since many mathematical expressions resulting from physical models may be difficult to solve, for example, due to inefficient or poorly developed mathematical methods for finding solutions, mathematical approximations are commonly made. These approximations are often made by neglecting terms of a mathematical expression that are considered to have a small effect on solutions. These approximations have consequences with respect to the physical concepts tied to the neglected terms. These approximations and the physical interpretation of the consequences are a common source of "hidden assumptions" in physics-based models, making their characterization particularly important.

Next, there is the discretized approximation layer. Subsequent to making mathematical approximations, lack of analytical solutions often forces us to resort to numerical, or discrete, approximation methods, in order to get an estimate of the answers we seek. The application of these methods to providing solutions to mathematical expressions is often ad-hoc with only weak formal justification for their use. It is common that there is only vague comprehension of the accuracy of these methods.

Finally, we have the interpolation layer. The choice of discretization of the domain, e.g., space and time, are frequently made for the convenience of obtaining solutions rather than per the request of a specific user of the resulting output. Consequently, the answers delivered to a recipient are frequently interpolated from those computed in the discretized approximation. This results in a final interpolation layer, the business end of the model, which provides the answers needed by other models.

We note two important properties of this hierarchy of conceptual modeling layers. First, there is generally a one-to-many relationship between each layer and the one below it. Each physical object may be modeled in multiple ways. There may be variations in the mathematical statement of a set of physical concepts. There are many approximations that may be made for a given mathematical representation, and so on. The second important property is that much of this conceptual layering is shared by other applied mathematical modeling disciplines, with the exception of the top two layers, which are specific to physical objects and the physical nature of those objects. We can imagine that logistics models, economic models, routing models, search and optimization models may also have similar conceptual frameworks, specifically where they are mathematical models.

The hierarchical layer-cake description also illustrates the nature of metadata needed. The physical object provides a context for the physical concepts that are used to model it. If we merely were to state the physical concepts and neglect to indicate what we are in fact modeling, the reader of our model is left to guess what our meaning is. While it is true that an educated reader of the model can often guess correctly what the intent of the model's author was, why leave the reader guessing? This is not acceptable in print documents, nor should it be acceptable in model representations.

The way that the Ontology Layer-cake illustration helps to understand how metadata may be utilized is that information from each layer may be used to "tag" the information from the layer immediately beneath it. This is because each layer gives the context and describes the thing that is being represented in the lower layer. The relationship between the layers exposes the assumptions made in constructing the model.

## 5. Approach to Standardization

We note that communities of appropriate technical expertise, not the W3C, must define vocabulary semantics. This means that physicists must make the substantial contribution to our effort . However, since the W3C coordinates the formation of web standards relevant to the development of an Ontology of Physics, in order to have broad impact, coordination with the W3C is desirable if not necessary. The W3C can act as a coordinating mechanism for bringing together the various communities of interest for a given topic. This coordination may be effected by having status as an advisory committee representative to the W3C in order to submit proposals for new activities, such as working groups and interest groups [6]. Currently, the Department of Defense has the Defense Information Systems Agency (DISA) and the U.S. Navy as member organizations, and therefore having advisory committee representatives for each organization.

While coordination through the W3C may help stimulate interest in the development of discipline-specific ontologies, there are already organizations that have indicated an interest in the development of markup language specific to physics-based documents. In particular, organizations representing large professional memberships and that produce physics publications would be interested in document metadata. Among them are the American Physical Society (APS), the American Astronomical Society (AAS), the American Institute of Physics (AIP) [3], and the International Union of Pure and Applied Physics (IUPAP) [4]. We have begun communications with these organizations. We note, though, that the primary expressed interests of these organizations at present are to create electronic document repositories. Nevertheless, a standardized representation of physical concepts is the common ground they share with the modeling and simulation community.

## 6. MathML

A key element of a complete framework for describing physics-based models is the language for describing the mathematical layer. For that we expect MathML and its extension, OpenMath to provide the ontology. MathML 2.0, a W3C Recommendation was released on 21 Feb 2001 [7]. MathML is a low-level specification for describing mathematics as a basis for machine-to-machine communication. It provides, for example, an interchange format between Computer Algebra Systems (CAS), such as Mathematica, Maple, Scientific Workplace, and MathCAD. The impact of the W3C on the development of this technical concept representation language seems clear, since MathML arose due to the efforts of the W3C Math working group.

The Math Activity of the W3C has been re-launched as the Math Interest Group of the W3C and has a charter to continue the task of facilitating the use of mathematics on the Web, both for science and technology and for education. The effort to build a standardized Ontology of Physics, whether or not it becomes an independent activity, will need to coordinate with the Math Interest Group for any overlapping areas of activity.

## 7. Summary

We have described how the development of a standardized Ontology of Physics should proceed, reviewing the structure of comprehensive model descriptions, the tools and organizational mechanisms available to implement the ontology, and existing standards that affect its development. We have received interest in the effort from the Modeling and Simulation community as well as the Physical Science membership and publications community. We expect to proceed with efforts to engage with the W3C in order to effectively build a standard for comprehensively describing physical modeling concepts.

## 8. Acknowledgments

The Defense Modeling and Simulation Office (DMSO) have supported the work described. The author would like to acknowledge useful discussions with Mr. Doug Clark, Drs. Terry Foreman, Ranjeev Mittu, and Katherine Morse.

## 9. References

[1] Collins, Joseph B., and Clark, Doug "Towards an Ontology of Physics", In *Proceedings of the European Simulation Interoperability Workshop*, July 2004, 04E-SIW-044.
[2] N. F. Noy and D. L. McGuinness, ``Ontology Development 101: A Guide to Creating Your First Ontology'', Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.
[3] http://xml.coverpages.org/gen-apps.html#physics
[4] http://publish.aps.org/IUPAP/ltaddp_report.html
[5] http://www.w3.org/TR/daml+oil-reference
[6] http://www.w3.org/2004/02/Process-20040205/activities.html
[7] http://www.w3.org/Math/