Semantic Modeling

A New Direction for Knowledge Management

Edmund W. Schuster‡
Research Affiliate
The Data Center
Massachusetts Institute of Technology

David L. Brock
Principal Research Scientist
Auto-ID Labs and The Data Center
Massachusetts Institute of Technology

Stuart J. Allen
Professor Emeritus
Penn State University

Pinaki Kar
New York, NY

November 11, 2004

‡ Corresponding Author
email: Schuster@ed-w.info
Phone: 603.759.5786
23 Valencia Drive
Nashua, NH 03062

About the Authors

Edmund W. Schuster has held the appointment of Director, Affiliates Program in Logistics at the MIT Center for Transportation and Logistics and is currently helping to organize a new research effort involving the large-scale analysis of data. His interests are the application of models to logistical and planning problems experienced in industry. He has a bachelor of science in food technology from The Ohio State University and a master in public administration from Gannon University with an emphasis in management science. Ed also attended the executive development program for physical distribution managers at the University of Tennessee and holds several professional certifications.

David L. Brock is Principal Research Scientist at the Massachusetts Institute of Technology, and co-founder and a Director at the Auto-ID Center (now EPCGlobal, Inc. and Auto-ID Laboratories). The Center was an international research consortium formed as a partnership among more than 100 global companies and five leading research universities. David is also Assistant Research Professor of Surgery at Tufts University Medical School and Founder and Chief Technology Officer of endoVia Medical, Inc., a manufacturer of computer controlled medical devices. Dr. Brock holds bachelors' degrees in theoretical mathematics and mechanical engineering, as well as master and Ph.D. Degrees, from MIT.

Stuart J. Allen is professor emeritus, Penn State – Erie, the Behrend College. He works on design of decision aids for application in manufacturing environments. His educational background includes a bachelor of science in mechanical engineering from the University of Wisconsin, a master of mechanical engineering from Seattle University and a Ph.D. in engineering mechanics from the University of Minnesota. Stuart began his research career in the field of non-Newtonian fluid mechanics and has published over 50 journal articles in engineering and management science. He has also owned and operated three businesses in Wisconsin and New York State.

Pinaki Kar is currently an independent consultant working in the pharmaceutical industry on analysis and modeling to support strategic planning, business development, and marketing. He is interested in the application of operations research and statistical techniques for planning and decision support across a wide range of business issues. His experience spans multiple industries that include pharmaceutical, chemical, high-tech, and insurance. Pinaki's educational background includes a bachelor degree in mechanical engineering from the Indian Institute of Technology, Kanpur and a master's degree in logistics from MIT.

One of the main goals of knowledge management (KM) involves building computer systems for sharing information between dispersed human or machine entities within large, complex organizations. Implicit in this goal is the need to reduce search cost in finding relevant knowledge needed to address an important problem at hand. The capability of a KM system to locate, organize, and share information is the main gauge of practical performance.

The use of mathematical models raises an interesting issue for KM that is seldom considered. Models are highly structured representations of reality that utilize various types of mathematical approaches such as linear programming, simulation, and heuristics, to solve practical problems. In all situations, models depend on a set of assumptions that only partially capture an accurate representation of the physical or business world. At best, a model is a metal image of reality that achieves predictive success based on a limited number of assumptions and a set of input data.

For many organizations, mathematical models have become the linchpin of decision-making for a wide range of disciplines. This offers a mechanism to express tremendous amounts of creativity and ingenuity. Since models are written in machine understandable language, the opportunity exists for forming a comprehensive KM based system that would reduce search time in matching the proper model to a practical application. However, there currently are few intra or inter organizational links, technical or non-technical, that form any type of interoperability between dispersed models, or between models and data. While much of the focus of KM has concentrated on information, the management of models represents a challenging new frontier that will push the bounds of productivity.

Most would agree that mathematical modeling is a craft industry analogous to the production of automobiles prior to the advent of the assembly line. Although models are ubiquitous management tools, they are, for the most part, isolated from one another. In other words, a model from one knowledge domain, such as weather forecasting, does not interact with another, such as logistical systems.

The reason for this is obvious. Until very recently humans were the only ones who built, used, and shared models. Our limited cognitive ability naturally restricts the number and diversity of models we can accommodate. Current KM for modeling rests with skilled individuals that posses many years of experience combined with mathematical intuition and an intensive desire to stay current regarding new approaches documented in the literature.

Computers, on the other hand, have the ability to execute and communicate models with vast numbers of other computers. With ever increasing processing power, data storage and networking bandwidth, the computing grid is poised to revolutionize our ability to understand and manage the physical world. The Internet with its standards and languages provides the backbone for communication, but does not provide the mechanism for describing and integrating diverse models. In this regard, there are few assimilating structures needed for KM in modeling. The future is a form of modeling on demand similar to other efforts in establishing a computer grid that resembles electric power distribution (London 2003). In this way, knowledge about models is distributed within organizations or possibly Internet wide. To get true productivity from the practice of modeling, there can no longer be the common instance of isolated modelers working independently and communicating only by means of occasional technical journal publications.

In this article, we discuss a proposed standard for a language and protocol that will enable computers to describe and share models and to assemble new models automatically from a general repository (Brock 2003a; Brock 2003b). While this will substantially increase the Clockspeed (Fine 1998) of modeling, and the computational efficiency of applying models to perform the functions of "sense," "understand," and "do," that comprise the underpinning of creating smart objects, the computer languages also form the basis for a new KM approach. The balance of this article describes our work at MIT in designing a network for abstract objects like models. Networks share the premise that leaps in productivity arise from the free flow of information. Creating an *Intelligent Modeling Network* will accelerate the flow of information and create new forms of KM (Brock et al. 2004).

SEMANTIC BASED INTERNET SEARCH

The existing standards of the Internet do not provide any semantics to describe models precisely or to interoperate models in a distributed fashion. For the most part, the Internet is a "static repository of unstructured data" that is accessible only though extensive use of search engines (Fensel et al. 2003, p. 377). Though these means of finding data have improved since the inception of the Internet, human interaction is still required and there are substantial problems concerning semantics. In general, "HTML does not provide a means for presenting rich syntax and semantics of data (Fensel et al. 2003, p. 7)."

For example, one of the authors of this article recently did a search for "harvest table, oak" hoping to find suppliers of home furniture. Instead, the search yielded a number of references to forestry and the optimal time to harvest oak trees. Locating the URLs relating to furniture required an extensive review of a number of different web sites. This process of filtering can only be accomplished though human interdiction and is time consuming.

With inaccurate means of doing specific searches based on one semantic interpretation of data, information, or models, it is nearly impossible for the Internet to advance as a productive tool for KM of models.

Several Types of Webs

The problem of semantics arises from the fact that keywords are the means used to describe the content of web pages. Each keyword can have multiple meanings, creating a situation of great difficulty when attempting to accomplish an exact search. The difficulty increases by an order of magnitude when attempting to do phrase-based searches. Without exact search capability, it is impossible to create any sort of machine understandable language for the current *Web of Information*.

Even though the search engine issue has not been resolved, industry forces are pushing for a new type of Internet characterized as the *Web of Things*. Driven by developments in Auto-ID technology and ubiquitous computing, the *Web of Things* aims to link physical objects to the internet using Radio Frequency Identification (RFID) tags as real-time communication devices and to "shift from dedicated computing machinery (that requires user's attention, e.g., PC's) to pervasive computing capabilities embedded in our everyday environments (Fensel et al. 2003, p. 363; Weiser 1991)."

Aiding this effort is EPCglobal, Inc.,[1] an international standards organization formed by the Uniform Code Council (UCC), and European Article Numbering (EAN) Association (known in the industry as GS1). The group administers the Electronic Product Code (EPC) numbering system, which provides the capability to identify an object uniquely. With serial identification for physical objects, searches accomplished through Internet search engines or proprietary IT infrastructures will become much more effective in finding an exact match. This provides the ability to do track and trace across entire supply chains and other computerized functions important to logisticians. Linking the physical world, using Auto-ID technology and ubiquitous computing, will form the basis for a revolution in commerce by providing real-time information and enabling smart objects (Schuster and Brock 2004; Schuster et al. 2004a; Schuster et al. 2004b).

As impressive as the effort to create the *Web of Things* has become, it still does not address the question of semantics in describing objects beyond the use of a simple serial number. There exist a large number of abstractions, such as mathematical models, that cannot be characterized by a unique serial number no matter how sophisticated the syntax. Without the ability to provide unique identification of an abstraction, the Internet

[1] EPCGlobal, inc. http://www.epcglobalinc.org/

will serve little useful purpose in linking mathematical models together in a way similar to the manner that the *Web of Things* will eventually link the physical world.

In the future, the definition of a model and the sharing of models though a network will become as important as the model itself. To accomplish this higher goal, the Internet must become a *Web of Abstractions*, in addition to a *Web of Information* and a *Web of Things*.

Creating a *Web of Abstractions* requires a semantic definition of models that is precise and can be machine understandable. Given this capability models can be searched, organized, categorized and executed – sequentially and in parallel – creating multiple, large-scale synthetic environments. These synthetic modeling environments will exist only in virtual reality and offer the potential for creating a dynamic meta-structure for specific classes of models.

Through a *Web of Abstractions*, models can be matched much more quickly to practical problems, along with the available data, and shared beyond single end-user applications. This capability is of great value to both practitioners and researchers who are interested in gaining the maximum value in modeling for practical decision-making.

SEMANTIC MODELING

Our goal is to turn modeling into a mass production system based on standardization, scale, and interoperability. In summary, this means that a Semantic Modeling language capable of achieving this functionality must include:

1. "A formal syntax and formal semantics to enable automated processing of their content (Fensel et al. 2003, p. 8)."

2. "…a standardized vocabulary referring to real world semantics enabling automatic and human agents to share information and knowledge (Fensel et al. 2003, p. 8)."

Achieving this goal will mean that practitioners can produce models in a timely manner with greater productivity and relevance. This anticipates a new era for computers in terms of insight and awareness and it implies the ability to organize data, and define the inputs and outputs of models in a semantically precise way.

The mechanism we put forth to mass produce models and create interoperability draws inspiration from current efforts to improve the search capabilities for the *Web of Information*. The World Wide Web Consortium (W3C) is responsible for initiating select efforts to improve overall web search capabilities.[2] Some of the initial work conducted by W3C forms a reference base for our research in developing and implementing a *Web of Abstractions*.

Each abstraction (model) has unique elements that can be defined just as a language has a specific syntax and grammar. Defining these elements alone will be of no benefit unless there is a protocol, or computer language, to communicate and execute the elements of models across a large network like the Internet. Our efforts in establishing Semantic Modeling are grounded in the idea of having data and models defined and linked in a way that can be used by machines not just for display purposes, but also for automation, integration and reuse across various applications. Accelerating the reuse of model elements across vast networks of users will lead to the mass production of models and great benefit to practitioners. In addition, distributed modeling, a set of

[2] W3C Semantic Web, http://www.w3.org/2001/sw/

geographically separated model elements working simultaneously in parallel, adds additional prospects for large-scale parallel computing.[3] This capability will improve the utilization of desktop computers and provide grids of almost unlimited modeling power.

Though the W3C provides something called a Resource Definition Format (RDF) that defines the basics of representing machine processable semantics (Fensel et al. 2003, p. 9), no formal computer language has been put forth that enables the sharing of models or doing large-scale modeling in parallel. The next section gives an overview of our vision for a computer language and protocols that achieves Semantic Modeling.

SYSTEM ARCHITECTURE

The fundamental idea is to design a family of standards that enable the creation of models that integrate automatically into an executing synthetic environment. In this way, developers can formulate models within their particular areas of expertise and know that the resulting models will interoperate in a shared environment. We believe it is possible, with sufficient care in the definition, to create such a language that is both precise and expressive in its description yet shows constraint in its breadth to ensure compatibility.

The goal is to create synthetic environments that receive data from the physical world (for example through Auto-ID technology) and then produce inferences, interpretations, and predictions about the current and future states of the environment.

These interpolated or extrapolated state data are essential for any automated decision system. In other words, the estimated environmental states support networks of

[3] Software Agents for Distributed Modeling and Simulation, http://www.informatik.unirostock.de/~lin/AnnounceIEEE/node2.html

decision-making algorithms so that they can make informed decisions and deliberate plans (that feed back to the physical world.) This type of modeling is essentially the underlying basis for automated control, monitoring, management, and planning.

The proposed architecture is composed of four fundamental components: the Data Modeling Language (DML), Data Modeling Protocol (DMP), Automated Control Language (ACL), and Automated Control Protocol (ACP).

The DML is a semantic for describing modular, interoperable model components. Models written in DML should automatically assemble into executable model environments. Although a number of ways exist to depict a model component that is interoperable, we choose to focus on data inputs as the means of describing a model component. This concept is explored in detail as part of the example presented later in this article.

We assume any model can be executed across multiple, heterogeneous platforms, which is a computational grid.[4] The semantic that describes the communication between the computing machines that host the models is the DMP. The DMP exists for the sole purpose of coordinating the operation of two or more models running in parallel on different computer platforms. In some cases, this coordination might take the form of an algorithm that communicates the timing of a model run and the timing of information transmission that will be used as an input to another model running on a separate computing platform.

A faithful reproduction of reality is the first objective for a successful model. However, the central goal of our initiative is to provide a framework for intelligent decisions. Humans make most of these decisions, but increasingly synthetic systems

[4] The Globus Project. http://www.globus.org/

augment many of these decisions. The ACL is a specification for describing these decision-making elements. We envision these elements will exist within networks of many, perhaps millions, of other decision-making elements.

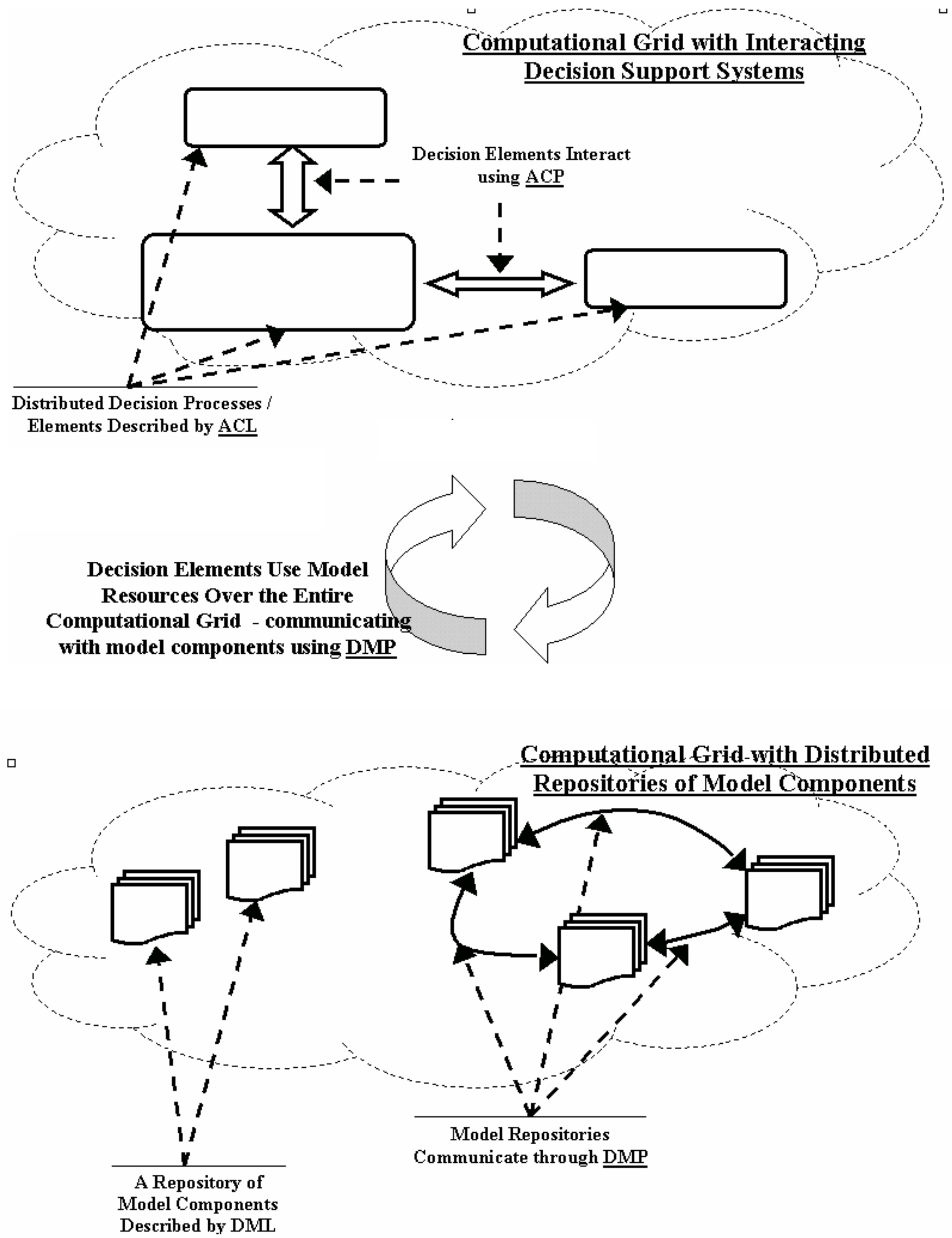
If we succeed in creating such networks, it is likely the relation between decision-making elements will be complex, that is, hierarchical organizations of specialized components dynamically creating and readjusting network topology and subordinates to match the needs of a particular task. In any case, we will need some standardized protocol to enable disparate elements to communicate with one another in a common language.

The ACP serves this purpose. Using the ACP, decision-making elements can locate one another, even though the individual models may exist in different host systems and organizations.

The combination of these languages and protocols, DML, DMP, ACL, and ACP, represents the foundation needed to construct general-purpose synthetic environments, as shown in FIGURE 1. The idea is that computers can construct a synthetic environment automatically, then modify it in real-time to analyze, manage and predict the states of a physical system.

FIGURE 1

PROPOSED DISTRIBUTED SYSTEM USING DML, DMP, ACL AND ACP



OTHER APPLICATIONS

The applications of such an architecture as we propose extend well beyond the formal discipline of KM, affecting nearly every aspect of industry and commerce. Any system that uses large amounts of data gathered from sensors could benefit from Semantic Modeling.

Hospitals could monitor and predict patient health based on real-time biometrics. Assisted living facilities and home care services could adjust medication in response to expected activity and individual metabolism.

Automobiles could dynamically adjust power, transmission, suspension, and braking given driving and road conditions. Trans-metropolitan traffic signal optimization could drastically reduce delays and improve network efficiency.

Agriculture and live stock management could use an entire range of diverse data to regulate day-to-day operations such as feeding and the harvest (Allen and Schuster 2004). Pesticides, fertilizer, and feed could be dispensed in complex patterns – optimized for individual efficiency.

The entertainment industry, particularly electronic games and motion picture visual effects, rely to a great degree, on complex physical models and engaging character behavior. These industries could not only benefit from an open modeling environment, but could also contribute to the technologies and modeling components across a broad range of applications. Furthermore, their ability to produce compelling visuals will help communicate abstract data sets and predicted physical environments within many application domains.

Environmental impact studies and public policy are dictated to a large degree by physical models and sensory data. A shared, open standard for simulation components could allow validation of these environmental projections with multiple independent models. Furthermore, the propagation of hazardous material, the dispersion of chemical agents and the flow of recycled material could be anticipated and controlled to a greater level with accurate analytic models.

Regulation of the financial services industry including securities, insurance, banking, and housing occurs almost entirely through analytic models and data projections. An open modeling infrastructure would allow exchange of economic models and enhancements in real-time to allow far greater precision in financial projection and economic efficiency.

Legal services, from corporate law to criminal defense, use models to form their language and plead their case. These models are created on an *ad-hoc* basis according to the needs of a particular case. An interoperable modeling environment, however, could allow the legal profession to share the physical and human behavioral models developed by other industries.

Engineering and the sciences use models in every aspect of their work. Clearly, the ability to create and share models in an open environment will have tremendous benefit in advancing these fields.

CONCLUSION

The prospect of sharing, through standard languages and protocols, the collective efforts of modelers throughout the world is an advanced application of KM. It has the potential to revolutionize nearly every aspect of human endeavor, as well as provide unprecedented benefit and savings across industry and commerce. Yet the challenges and difficulties are extraordinary, from theoretic achievability to practical implementation. Still the rewards make the journey well worth pursuing, which may lead to a true *Intelligent Modeling Network* in support of a comprehensive KM system.

REFERENCES

Allen, Stuart J. and Edmund W. Schuster (2004), "Controlling the Risk for an Agricultural Harvest," *Manufacturing & Service Operations Management,* Vol. 6, No. 3, forthcoming.

Brock, David L. (2003a), "The Data Project – Technologies, Infrastructure and Standards for Distributed Interoperable Modeling and Simulation," *MIT Data Project Workshop*, September.

Brock, David L. (2003b), "Beyond the EPC – Making Sense of the Data – Proposals for Engineering the Next Generation Intelligent Data Network," *The MIT Auto-ID Center*: Cambridge, MA.

Brock, David L, Edmund W. Schuster, Stuart J. Allen, Pinacki Kar (2004), "An Introduction to Semantic Modeling for Logistical Systems." *Journal of Business Logistics*, forthcoming.

Fensel, Dieter, James Hendler, Henry Lieberman and Wolfgang Wahlster (2003), *Spinning the Semantic Web*. Cambridge, MA: MIT Press.

Fine, Charles H. (1998), *Clockspeed*, Reading, MA: Perseus Books.

London, Simon (2003), "Big Blue Reinvents Itself Again," *The Financial Times*, Oct. 9.

Schuster, Edmund W. and David L. Brock (2004), "Creating an Intelligent Infrastructure for ERP Systems; The Role of RFID Technology," Falls Church, VA: American Production and Inventory Control Society, White Paper Published on Home Page.

Schuster, Edmund W., Daniel W. Engels, and Stuart J. Allen (2004a), *Raising the Bar (Code): The Value of Auto-ID Technology*, Cincinnati, Ohio: South-Western Educational Publishing, Forthcoming.

Schuster, Edmund W., David L. Brock, Stuart J. Allen and Pinaki Kar (2004b), "Enabling ERP through Auto-ID Technology," book chapter in *Strategic ERP Extension and Use*, edited by E. Bendoly and F. R. Jacobs, Palo Alto, CA, Stanford University Press, Forthcoming

Weiser, Mark (1991), "The Computer for the Twenty-First Century." *Scientific American*, Vol. 265, No. 3.