Cutter IT Journal

The Journal of Information Technology Management

Vol. 22, No. 9
September 2009

"The emerging Semantic Web will require us to dramatically rethink traditional notions of how business, data/information, application, and technology architectures are conceptualized and realized within an enterprise."

— Mitchell Ummel, Guest Editor

The Rise of the Semantic Enterprise

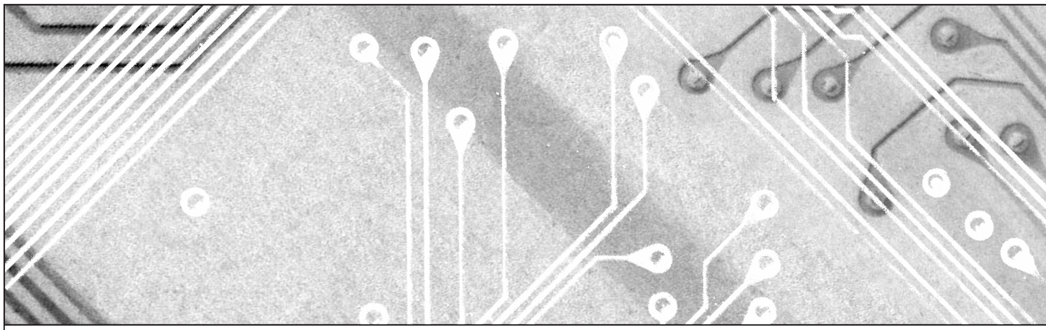
Semantic Enterprise Today

The Semantic Web offers us a compelling architectural framework upon which to build next-generation, Internet-ready applications within the enterprise. Now is the time to embrace the possibilities.

Semantic Enterprise Someday

Enterprises should continue to build out their traditional, proven architectures for the foreseeable future. It may be a decade or more before Semantic Web-based technologies are ready for mainstream enterprise applications.

CUTTER CONSORTIUM



Cutter IT Journal

About Cutter IT Journal

Part of Cutter Consortium's mission is to foster the debate of, and dialogue on, the business technology issues challenging enterprises today, to help organizations leverage IT for competitive advantage and business success. Cutter's philosophy is that most of the issues that managers face are complex enough to merit examination that goes beyond simple pronouncements. Founded in 1987 as *American Programmer* by Cutter Fellow Ed Yourdon, *Cutter IT Journal* is one of Cutter's key venues for debate.

The monthly *Cutter IT Journal* and its weekly companion *Cutter IT E-Mail Advisor* offer a variety of perspectives on the issues you're dealing with today. Armed with opinion, data, and advice, you'll be able to make the best decisions, employ the best practices, and choose the right strategies for your organization.

Unlike academic journals, *Cutter IT Journal* doesn't water down or delay its coverage of timely issues with lengthy peer reviews. Each month, our expert Guest Editor delivers articles by internationally known IT practitioners that include case studies, research findings, and experience-based opinion on the IT topics enterprises face today — not issues you were dealing with six months ago, or those that are so esoteric you might not ever need to learn from others' experiences. No other journal brings together so many cutting-edge thinkers or lets them speak so bluntly.

Cutter IT Journal subscribers consider the *Journal* a "consultancy in print" and liken each month's issue to the impassioned debates they participate in at the end of a day at a conference.

Every facet of IT — application integration, security, portfolio management, and testing, to name a few — plays a role in the success or failure of your organization's IT efforts. Only *Cutter IT Journal* and the *Cutter IT E-Mail Advisor* deliver a comprehensive treatment of these critical issues and help you make informed decisions about the strategies that can improve IT's performance.

Cutter IT Journal is unique in that it is written by IT professionals — people like you who face the same challenges and are under the same pressures to get the job done. The *Journal* brings you frank, honest accounts of what works, what doesn't, and why.

Put your IT concerns in a business context. Discover the best ways to pitch new ideas to executive management. Ensure the success of your IT organization in an economy that encourages outsourcing and intense international competition. Avoid the common pitfalls and work smarter while under tighter constraints. You'll learn how to do all this and more when you subscribe to *Cutter IT Journal*.

Cutter IT Journal®

Cutter Business Technology Council: Rob Austin, Ron Blitstein, Christine Davis, Tom DeMarco, Lynne Ellyn, Tim Lister, Lou Mazzucchelli, Ken Orr, Mark Seiden, Ed Yourdon

Editor Emeritus: Ed Yourdon
Publisher: Karen Fine Coburn
Group Publisher: Chris Generali
Managing Editor: Karen Pasley
Production Editor: Linda M. Dias
Client Services: service@cutter.com

Cutter IT Journal® is published 12 times a year by Cutter Information LLC, 37 Broadway, Suite 1, Arlington, MA 02474-5552, USA (Tel: +1 781 648 8700; Fax: +1 781 648 8707; E-mail: citjeditorial@cutter.com; Web site: www.cutter.com). Print ISSN: 1522-7383; online/electronic ISSN: 1554-5946.

Subscription rates are US $485 a year in North America, US $585 elsewhere, payable to Cutter Information LLC. Reprints, bulk purchases, past issues, and multiple subscription and site license rates are available on request.

Opening Statement

by Mitchell Ummel, Guest Editor

Our two-decades-old World Wide Web architecture is long past due for an upgrade. During what we might call the "Web 1.0-2.0 epoch," demand for computing has grown across every enterprise, in every sector, around the globe. We continue to struggle to meet this demand using our traditional approaches to building and managing enterprise information systems. Mounting barriers of complexity and scalability continue to hinder business agility, increase costs, and constrain overall productivity. The Semantic Web (also often referred to as Web 3.0) is emerging as the prescribed solution, and it offers us a compelling architectural framework upon which to build next-generation Internet-ready applications.

THE SEMANTIC WEB ARRIVES

Today's Internet consists of a coarsely woven fabric of hypertext links among billions of largely unstructured Web pages, all generally designed to be read by humans. The data behind these Web pages is selectively shared, but the semantic definitions either (1) don't exist, or (2) are locked away behind firewalls within our enterprise systems. This is so much the case that, across the Internet, we've evolved an entirely new architectural integration layer, with point-to-point semantic interfaces and translation exchanges based on services realized through service-oriented architecture (SOA).

Therein lies the problem — unstructured information overload, a very low signal-to-noise ratio, data subjected to human interpretation, and an associated deficiency of the required semantic precision needed to achieve a higher level of machine-interpreted cognition.

With the Semantic Web, what we're striving for is to add a layer of cognitive power to the Internet's digital gray matter — consisting of a finely woven fabric of semantically precise, linked data that can be processed automatically, by machine-based agents, on our behalf. This is to be achieved through adoption and use of a set of emerging standards promoted by W3C for Semantic Web technologies (SWTs), including Resource Description Framework (RDF), Web Ontology Language (OWL), and SPARQL. These foundational technology standards allow for the semantic linking of data and, therefore, semantic application interoperability among semantically aware applications (SAAs).

As I've suggested in my previous research on this topic,[1] the Semantic Web is indeed already here:

- SAAs are now growing "in the wild," generating and consuming semantically linked data. Leading search vendors are embracing annotation through semantic tagging constructs (typically using RDFa or microformats). In addition, a number of automated translation techniques are emerging for RDF-izing current structured — and in some cases, even unstructured — data that exists in legacy formats on today's Internet.
- Open source communities are embracing the emerging W3C standards, and major commercial vendors are now adding "semantic" features or extensions into their mainstream IT management products.
- The W3C's Linked Open Data initiative is growing very quickly.[2] The US government is jumping on the bandwagon as well, with an increasing volume of real-time government data now being released and translated to RDF.

Thus, we may describe the era paralleling — but also closely trailing — today's rise of the Semantic Web as the era of the semantic enterprise (SE). A semantic enterprise can be defined as one that exploits SWTs for applications within the enterprise. We can call the underlying architecture necessary to enable the SE a semantic enterprise architecture (SEA).

SETTING SAIL TOWARD A NEW "SEA"

Although a full exploration of the principles of SEA is well beyond the scope of this opening statement, it's becoming clear to me that a revolutionary mindshift will be needed in the way we've traditionally approached enterprise architecture. The emerging Semantic Web will require us to dramatically rethink traditional notions of how business, data/information, application, and technology architectures are conceptualized and realized within an enterprise.

This is far more than a simple terminology shift (e.g., to ontological engineering from what we used to refer to as information or data architecture). SEA will require architects to think about their enterprise from an "outside-in" perspective, where (1) open-world assumptions can apply, (2) enterprise data is largely self-describing through a tight coupling with federated ontologies (many of which are outside the direct control of the enterprise), and (3) our applications are able to effectively infer, deduce, and calculate across both private and public linked data in ways not specifically envisioned at the time they were originally designed.

SEMANTICALLY AWARE APPLICATIONS

SAAs are Internet-ready applications that generate or consume Semantic Web data — regardless of whether the applications and semantically linked data live inside or outside the firewall. In either venue, SAAs are designed, developed, and implemented in much the same manner and will use data/information, governed by ontologies, that cross-cut both the enterprise (private) and Internet (public) domains. I suggest (and our authors concur) that in the very near future, we will find beneficial applications of SAAs within specific business problem domains where traditional enterprise systems have fallen short, including:

- Business intelligence (BI)
- Data mining (across structured and unstructured data stores)
- E-discovery (across centralized and distributed data stores)
- Customer relationship management (CRM) systems, leveraging the entirety of the social networking fabric
- Dynamic business rules (inference engine) optimization
- Ontology-based security/trust credentialing, private social networks, and public referral/viral product marketing using FOAF (Friend of a Friend), POWDER (Protocol for Web Description Resource), and other maturing standards
- Extract, transform, load (ETL) services (RDF-izers), which automatically transform and publish enterprise data into private and/or public ontological stores
- Embedded control, telemetry, and data acquisition systems relating to devices, equipment, and sensors, including (but not limited to) enablement of smart grid[3] energy management systems

CHALLENGES FOR THE SEMANTIC ENTERPRISE

Of course, no amount of hand-waving, or cheerleading, will induce SWTs to solve all our enterprise computing challenges overnight. While the foundation for real solutions to today's most pressing information management problems is now here, there are certainly gaps and weaknesses in Semantic Web–based standards and technologies that will need to be addressed. These include:

- Semantic data governance, provenance, trust, and ownership across shared semantic data ontologies[4]
- Security, confidentiality, privacy, and appropriate use of data in a highly federated, distributed semantic data architecture
- Opportunities (as well as threats) related to potential monetization and commercialization of parts of the Semantic Web
- Rapidly evolving standards; vendor products to support the SE still in the inception phase of the product development cycle

I see three top action items for CIOs interested in pursuing the semantic enterprise vision:

1. Build general awareness and understanding across the enterprise for the practical application of an SEA roadmap. Invest in training and skills development in order to leverage SWTs in your specific environment.
2. Develop potential use cases for SAAs, based on your own organization's strategic business and technology opportunity areas. Charter at least one SWT-based pilot project (or proof of concept) in the coming year.
3. Begin thinking about your data/information architecture in terms of ontological engineering in order to semantically link to existing or emerging domain or industry ontologies.

IN THIS MONTH'S ISSUE

In this issue, each of our expert authors contributes to our understanding of this broad subject area through insights and unique domain knowledge. We begin with an article by independent researcher and consultant Paola Di Maio, who level-sets us in the semantics of the term "semantic" across the several different contexts in which we're seeing it applied today. She adds significantly to our understanding of what potential influences the Semantic Web will have within the enterprise.

Next, John Kuriakose of the Center for Knowledge-Driven Information Systems (CKDIS) labs at Infosys discusses some of the many enterprise applications enabled by SWTs. Kuriakose introduces us to the semantic technology stack, presents a roadmap for SWT adoption, and suggests ways to bridge the divide between traditional object/relational data stores, RDF-based triple stores, and OWL-based ontological definitions.

Our third article is by data integration expert Shamod Lacoul, who drills down into the technical case for adopting SWTs for data integration. Lacoul argues that SWTs show promise in addressing the top data integration barrier we've been struggling with for decades — specifically, the lack of an overarching shared semantic model. He also gives us insight into graph theory, which serves as the foundation for how Semantic Web data is linked, and points out the advantages a distributed semantic data model offers in its ability to adapt to the ever-changing requirements of a business.

Next up, Cutter Senior Consultants Bhuvan Unhelkar and San Murugesan explore the overall business case for adoption of SWTs within the enterprise. Included in their article is a wonderfully succinct statement (definitely tweet-able, at just around 140 characters!) that effectively sums up the central thesis of this month's *Cutter IT Journal.*

> By exploiting the technologies of the Semantic Web, an SE can create a people-machine continuum that enhances business agility.

Finally, MIT researchers Ken Lee and Ed Schuster introduce their Lee-Schuster Semantic Enterprise Architecture (LSSEA) in a fascinating case study in semantic engineering for the ERP problem space. While LSSEA is not strictly an SWT-based solution (absent W3C's OWL, RDF, and SPARQL standards), they argue that it is an effective way to deliver mathematical models to users quickly and cheaply, enabling experimentation and rapid adaptation.

At this moment in time, the new, transformational, Web 3.0 "cold front" is on a collision course with the muggy, evolutionary "stationary front" of challenged enterprise information and application architectures of the past 20 years. We face a potential storm of opportunity, and in this month's issue of *Cutter IT Journal,* our expert authors have faced this topic head-on. So let's batten down the hatches and together set sail for that new and not-so-distant shore — the truly semantic enterprise.

ENDNOTES

[1]Ummel, Mitchell. "The Semantic Web 3.0 Mashup Universe: Coming to a Browser Near You." Cutter Consortium Business Intelligence *Executive Update*, Vol. 9, No. 4, 2009.

[2]W3C Linked Open Data Project (http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData).

[3]"Smart grid." Wikipedia (http://en.wikipedia.org/wiki/Smart_grid).

[4]See "Protocol for Web Description Resources (POWDER): Primer." W3C, September 2009 (www.w3.org/TR/2009/NOTE-powder-primer-20090901).

Mitchell Ummel is a Senior Consultant with Cutter's Enterprise Architecture practice and President of UmmelGroup International, Inc., a US-based business and technology management consulting firm. Mr. Ummel's IT experience spans 25 years and includes service in a variety of CIO, executive management, training, coaching, mentoring, and consulting roles for government agencies and companies in a wide variety of industries. Earlier in his career, he served as a US state of Kansas agency CIO, where he contributed to the inception of the Kansas statewide technical architecture, helped define next-generation project and portfolio governance processes, and led numerous large-scale technology initiatives. Mr. Ummel holds advanced degrees in mathematics and computer science from Fort Hays State University. He is a graduate of, and contributor to, Kauffman Foundation's Entrepreneurial Leadership and New Tech Venture Development Programs, a frequent speaker at national business and technology conferences, and a member of the Project Management Institute. He can be reached at mummel@cutter.com.

When Semantic Enterprise Means Rethinking Everything

by Paola Di Maio

> Progress has not followed a straight ascending line, but a spiral with rhythms of progress and retrogression, of evolution and dissolution.
>
> — Johann Wolfgang von Goethe

The prospect of a Web of information- and knowledge-supporting systems capable of automating reasoning, inference, and decision making has been challenging Web researchers, developers, and "real-world" Web users since the publication of the Semantic Web vision.[1] In most countries Internet adoption is growing steadily,[2] and people's professional and personal lives are increasingly reliant on the Web. The main challenge for businesses is to learn how to leverage the Web's full potential, which is expected to be exponentially higher in "semantically enabled" environments.

Businesses are warming up to the idea of semantics and are becoming impatient to leverage the "magic," but few understand what semantics really means. There is a bit of a joke in the IT community that any old idea presented with a "semantic" twist (semantic enterprise, semantic search and retrieval, etc.) is going to benefit from the buzz and contribute to the mushrooming of "Semantic Webware," which includes software, toolkits, platforms, environments, standards, protocols, and a lot, a *lot* of talk, sometimes with limited appreciation of the core issues. Semantics in relation to information technology and data models is nothing new. It is even fundamentally something rather simple, but when it comes to semantics and the Web, with everything that the Web represents, the challenges inevitably become significant. In this article, I discuss and clarify some assumptions in relation to the semantic enterprise and the firewall and place "semantics" in the context of an evolving Web.

WHAT DO YOU MEAN, ER ... "SEMANTIC"?

In relation to the Internet and the Web, the notion of semantics remains very open to interpretation, probably because the overall concept of semantics in W3C is deliberately broad (read: a bit vague).

The distinction between syntactic, semantic, and pragmatic "dimensions" in linguistic and communication theory was credited to American semiotician and philosopher Charles Morris, who defined these dimensions as follows:

> Pragmatics "deals with the origins, uses, and effects of signs within the total behavior of the interpreters of signs,"[3] and thus has the widest scope of any semiotic study. Semantics concerns just the relations between signs and the objects they signify, narrowing semiotic study to the strict literal meaning of signs and propositions. Syntactics concerns the formal relations between signs themselves, further narrowing semiotic study to the logical and grammatical rules that govern sign use.[4]

Morris based much of his work on earlier studies of semiotics by Charles Pierce and the "social behaviorism" of John Dewey and George Herbert Mead. In common everyday IT language, semantics refers to the study of the meaning of words, as opposed to syntax, which refers to the structure of words and languages, such as grammars (be they natural language or computer grammars). It is worth remembering that in the logic of discourse, the distinction between semantics and syntax is sometimes fuzzy.

In the context of organizational information, semantics is represented by the relationship between information objects, in their various degrees of granularity (such as data and knowledge), to other information — about the objects themselves, about other objects, or about relations. Such relations, when properly elicited and structured, contain the "intelligence" that we seek from our systems. Cognitive systems (knowledge systems, information systems, intelligent systems, etc.) and the IT infrastructures that are designed to support them should be developed along all three dimensions to be able to communicate intelligently. Today, in relation to the Web, "semantics" is generally defined in terms of what has been established by the W3C (or, by those who may disagree, in contrast to it).

The W3C glossary states that the Semantic Web is "the Web of data with meaning."[5] It refers to a vision of advanced knowledge capabilities — automatic querying and retrieval, reasoning — that can be carried out on the open Internet by agents (i.e., software that has been programmed to do something) thanks to a pervasive "web of content." For example, a semantic search for a term would not return an unsorted list of relevant and less relevant results that a human must sift through in order to select the most appropriate response. Instead, the results would be presented in the context of a knowledge schema the user has defined.

The current Linked Open Data (LOD) model serves as the best working example of how this is done using Resource Description Framework (RDF) constructs. However, the notion of "semantics" started to trickle into data modeling in the 1970s, when IT researchers and scientists started to understand the limitations of the relational data model (which was then becoming dominant) and the practical implications of constraining the data view of the world to tuples and rows.[6]

The Business Relevance of Semantics

Now let's move away from pure technological issues and consider the relevance of semantics to business. From an enterprise viewpoint, the most important thing in defining semantics is making sure that the intended meaning and formalisms are agreed to (more or less), explicit, and declared (in case someone asks). This becomes especially important when it comes to discussing capabilities and functionalities, which, for CIOs and IT, is generally the main appeal of semantics in relation to business. So much of the notion of the semantic enterprise depends on how it is defined and its relationship to what surrounds it. To many, these are somewhat volatile and difficult things to pin down.

If by "semantic enterprise" we mean an enterprise that operates under closed-world assumptions (i.e., all decisions can be made based on data known to, and managed by, the enterprise), then the architecture and tools do not need to stray far from the traditional relational model approach. This is because the data whose semantics the enterprise needs to leverage is stored in a database, and as such its semantics can be extracted using conventional relational database systems. Surely RDF and OWL can be used in a closed system, where the data is stored in a relational database, but there would be little advantage in doing so.

However, if we define a semantic enterprise as one that operates under open-world assumptions (i.e., facts are assumed to be incomplete and will generally never be fully known) and thus needs to leverage the semantics of unstructured data not stored in a database, then we start thinking along the lines of the LOD model and perhaps RDF. Either way, we can forget about the firewall — in the traditional sense of dividing the world into two domains, public and private — for reasons I discuss in more detail below.

Generally speaking, in any given business ecosystem, a "semantic enterprise" is likely to be characterized by information systems capable of leveraging the relational associations between objects, data, and information so as to optimize the achievement of business goals. Today the technologies are in place to support that, and the main limitations are in the business models and in the ways organizations are managed and operated.

From an enterprise viewpoint, the most important thing in defining semantics is making sure that the intended meaning and formalisms are agreed to (more or less), explicit, and declared (in case someone asks).

RETHINKING THE FIREWALL

These days more than ever, there is a limit to what can be defined with clear-cut boundaries. The rapid speed of information exchange increases the rate of transformation of every domain and accelerates coevolution and change. Progress, as Goethe is reported to have said, does not happen in straight line. As tempting as it is to consider the expression "Web 3.0" as a synonym for the Semantic Web, such an approximation is likely to remain superficial, as the Web is continually evolving, and technical definitions benefit from some degree of formality.[7] While the media tends to bank on buzzwords such as Web 3.0, serious discussions are better based on formal definitions.

The fact is that the whole notion of the firewall needs to be reconsidered in a Semantic Web context. Semantically aware applications (SAAs) are typically designed to operate on the open Web and are constrained by the firewall configuration of the systems they run on. As researchers Simon N. Foley and William M. Fitzgerald[8] observe, configurations that are too tight may prevent optimal interaction of Web resources, resulting in SAA failure, while configurations that are too loose may leave the system vulnerable to attack (e.g., across open ports):

> While the Semantic Web may provide applications with security services that are domain-knowledge aware, it is argued that firewalls still have a role to play in securing the low-level infrastructure. Not only do firewalls protect services that do not provide built-in application-level security, it is considered best practice to rely on multiple

layers of security, providing "belt and braces." In practice, deploying a firewall for a Web server or Web client is not simply about opening port 80 on the server for all traffic; one may wish to deny certain nodes (IP addresses, etc.), only accept HTTP traffic from some nodes, require other nodes to use HTTPS, and also deal with HTTP traffic that is tunneled through proxies available on other ports. Furthermore, Web services do not necessarily communicate on port 80. In addition, firewall content sanitation (application layer) provides fine-grained access control that may cut across the host-based access controls; for example, certain content may be permitted (or denied) only to/from particular nodes.[9]

In essence, this means that a semantic enterprise is intrinsically defined by the degree of transparency and openness of its data model, and that the notion of semantics "inside the firewall" is probably handled by the semantic capabilities already supported, at least in part, by relational databases.

> **Information systems capable of supporting the automatic exchange of data over open or dedicated networks can only be made to work if clear data models and well-formed "semantic structures" are already in place.**

ENTERPRISE TRANSFORMATION

The impact of the Web on social and organizational boundaries is one of the most fascinating subjects of our times. Semantic systems architectures marketed under the "SOA" (service-oriented architecture) label are sometimes criticized because the term "service" in the acronym can appear arbitrary and misleading. A service — or "unit of logic" — is information exchanged automatically by two systems on the Internet using an agreed convention. What is being heavily marketed by the IT industry as SOA is, in essence, a semantically enabled architecture. Therefore, "SOA" could equally well stand for "semantic-oriented architecture."

In terms of information architectures, SOA models point clearly to the opening up of boundaries, not only in terms of knowledge domains (convergence and interdisciplinarity), but in terms of organizational structures and processes. In extended organizations, the boundaries of the enterprise become ever more flexible. For example, consider outsourcing, which is becoming increasingly important in the world economy and where all processes supporting the extended supply chain must allow for some degree of external collaboration.

In the earlier days of computers, data had to be encoded before it could be used by an electronic system. This meant that information that lived in its "natural" state — unstructured, scattered around, written or spoken language — could not be used by computers. It also meant that each system and program would work with information designed according to its own internal architecture, which often would not be compatible with other systems. Semantic Web technologies (SWTs) can help liberate information trapped in closed data structures by supporting "external integration," the ability to bind information and knowledge directly from unstructured sources by leveraging the logical structures of natural language. In business process terms, this can be both a blessing and a curse.

In theory, a semantically enabled network (a network whose nodes are equipped with semantic capabilities) can capture and infer logical relations directly from the structure of the information. It can do this in real time and with little or no intermediation, automatically making logical associations, provided the information is appropriately wrapped in "schemas" configured to capture "relations."

The "semantic" dimension of information, which initially has been the focus of interest mainly for knowledge architects, now is acquiring relevance in terms of organizational planning, as real-time information exchange is starting to affect the nature of operations. The benefits of semantic architectures derive from simple yet critical advances provided by real-time information exchange between networked applications: no more outdated information thanks to automated real-time synchronization of updates, automatic inventorying and supply chain management, and associations and relationships automatically inferred based on predetermined user profiles. However, information systems capable of supporting the automatic exchange of data over open or dedicated networks can only be made to work if clear data models and well-formed "semantic structures" are already in place.

FROM ENTERPRISE TO "SEMANTIC ENTERPRISE" ARCHITECTURES

Enterprise architectures (EAs) depend heavily on robust data models. Like EAs, semantic enterprise architectures (SEAs) follow such principles as uniqueness, independence, appropriate "representation," and the right level of abstraction. What is different, however, is that SEAs will be supporting the application of these principles not only to structured data, but also to nonstructured or semistructured information.

Ontologies are being used in support of the most disparate BI functions.[10] Following are some examples (although most of them are indicative of trends and are still at the research/prototype stage):

- RDF notation can be used to express the basic structure of an OLAP cube by applying SWTs to define the content of data sources.
- Semantic business intelligence (SBI) ontologies are used to support the navigation of enterprise data conforming to a definition of enterprise concepts as specified by business users.
- A kind of "semantic middleware" can be adopted that extracts and integrates data from heterogeneous information systems and compiles a functional ontology.

Newer generations of Web-enabled software are being designed to automatically support a semantic layer, both internally within the organization and externally at the exchange level. However, semantically enabled software will require vocabularies and well-mapped-out data and information schemas in order to function. If CIOs are serious about developing a semantic enterprise, they should start asking themselves what kind of organization they envisage for the future: a "closed-world" organization, an "open-world" organization, or — as is more likely — a hybrid of the two? A semantic enterprise is likely to be a hybrid organization, and that can be intimidating to many "traditional" business owners.

The first step toward developing a semantic enterprise is to ensure that the terminology used in all existing organizational documentation — from E/R diagrams to data flow charts to key policy documents — is consistent and optimized. It must also be adopted correctly in the metadata and all critical information structures, such as rule catalogs. Before even thinking of becoming a semantic enterprise, an organization must have good vocabularies (keyword catalogs) and schemas to represent the organizational knowledge and processes, with some mapping the "operational logic" using "normalized" natural language form, making sure to label clearly the schema elements and their values. In the future, a policy change published on some regulatory authority's Web site could well prompt a modification in a regulated organization's process flow, but this can happen only without disastrous consequences if the terminology and information schemas are valid, harmonized, and consistently implemented, with due support and provision for security measures.

When it comes to planning for the semantic enterprise, CIOs and CEOs are going to have to rethink where to draw the lines between public and private. How these lines are drawn will determine the shape of things to come. The ability to balance opportunities and risks is going to be reflected in how technologies are configured and used, which in turn is going to open up new organizational perspectives.

ENDNOTES

[1]Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web." *Scientific American*, May 2001.

[2]Internet World Stats (www.internetworldstats.com/stats.htm).

[3]Morris, Charles W. *Signs, Language, and Behavior*. Prentice Hall, 1946.

[4]"Charles William Morris: 1901-1979." Pragmatism Bibliography Center (www.pragmatism.org/genealogy/morris.htm).

[5]"Glossary." W3C (www.w3.org/People/Berners-Lee/Weaving/glossary.html).

[6]Di Maio, Paola. "Semantic Data Models." Cutter Consortium Business Intelligence *Executive Report*, Vol. 7, No. 12, 2007.

[7]Some use the terms "Web 3.0" and "Semantic Web" interchangeably. Tim Berners-Lee, founder of the WWW, describes the Semantic Web as a component of Web 3.0 in the following: "People keep asking what Web 3.0 is. I think maybe when you've got an overlay of scalable vector graphics — everything rippling and folding and looking misty — on Web 2.0 and access to a Semantic Web integrated across a huge space of data, you'll have access to an unbelievable data resource."

[8]Foley, Simon N., and William M. Fitzgerald. "Semantic Web and Firewall Alignment." *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering Workshop*. IEEE Computer Society, 2008, pp. 447-453.

[9]Foley and Fitzgerald. See 8.

[10]For a discussion of the kinds of functionalities that can be provided, see Di Maio, Paola. "Ontology-Supported BI." Cutter Consortium Business Intelligence *Executive Update*, Vol. 8, No. 22, 2008.

Paola Di Maio is a systems analyst and engineer who studied knowledge engineering and expert systems design before focusing professionally on content and knowledge management tools. She has a strong interest in online collaboration and knowledge networks, speaks four languages, works internationally as an independent researcher and consultant, and is a thought leader in the the practice of organizational innovation. Ms. Di Maio holds an MSc in information systems and is currently involved in doctoral research and innovation programs in the UK. She can be reached at paola.dimaio@gmail.com.

Understanding and Adopting Semantic Web Technology

by John Kuriakose

The industrial model that promoted manufacturing and engineering is slowly being phased out to usher in a new digital economy powered by developments in networking and information management. The old model placed controls on everything, including information flow, and permitted sharing of knowledge only through hierarchies of roles within the organization. The new business context places an emphasis on information availability, integration, and analytics and aims to foster collaborative communities whose shared contribution helps the organization respond to business opportunities in an efficient and effective manner.

Information and communication technology plays a vital role in this new context to enable these collaborative groups to interact and make sense of information from diverse data sources. The new context requires the effective deployment of technology for capture, storage, and dissemination of knowledge to aid analysis and interpretation.

BUSINESS VALUE: WHY CONSIDER SEMANTIC WEB TECHNOLOGY?

Business strategy and action are very much based on specific knowledge about the business context layered above general knowledge about a business domain and customer preferences. One aspect of strategy development also involves gleaning precise insights from existing data and previous results. This function completes the lifecycle shown in Figure 1 by providing feedback for further strategy development based on actual data. This assessment and feedback function is primarily characterized by information aggregation, integration, and summarization.

Figure 1 — Role of SWT in deriving insights from existing data for assessment and feedback.

The same abstract principle is also applicable for IT strategy and management. In both these contexts, the feedback loop requires a high-level description of the entities, their states, and the relations between them. This description acts as a "lens" that represents how a decision maker chooses to see and interpret the dynamics between various stakeholders, the business entities, and their relationships with the organization and between each other. This model of the business world forms the basis for all interpretations that guide business thought and investment. Much of this model and its underlying general context are either implicit or at best expressed in natural language.

I see an opportunity to apply Semantic Web technology (SWT) to realize and accelerate this vision of a connected collaborative enterprise to promote business agility. Agility in this sense will be achieved through the deployment of smart solutions that can bring about deep integration across diverse data sources and also offer methods for deriving abstract or summarized views from this massive data store based on stakeholder concerns. These relevant views of actual data will give key decision makers the support needed to adjust and align execution plans for better results.

SEMANTIC WEB TECHNOLOGY

The Semantic Web[1] is a recent research project that aims to build the infrastructure to support the creation of a machine-readable Web (Web 3.0), where agents will collaborate to exchange information and make useful inferences to support and enrich human activity on the Web. SWT has evolved out of the Semantic Web research project and provides the right primitives for building the next wave of interesting enterprise applications with the ability to relate, integrate, abstract, and reason on data.

The use of metadata to achieve data exchange,[2] interoperability, and discovery[3] has been known and applied previously. According to ISO/IEC 11179-1, metadata is:

> ... the information and documentation which makes data understandable and shareable for users over time. Data remains useable, shareable, and understandable as long as the metadata remain accessible.[4]

The ISO standard also states that it is the obligation of organizations producing data to make available the metadata that supports the formal interpretation and use of this data.

Creating computer-based systems that can understand the meaning of data requires a formal representation of the meaning and context of data using "semantic metadata." Formal representations of knowledge have been the focus of knowledge representation and reasoning research[5-7] for more than 20 years within the artificial intelligence community.

For our discussion, "ontology" is the semantic metadata formally represented in the form of the concepts, relations, and rules that constitute a shared understanding of a domain.[8] An ontology language describes the general or background knowledge in a domain in terms of concepts and relations, while a data language such as the Resource Description Framework (RDF) describes specific instances or individuals in that domain. Ontology languages such as the Web Ontology Language (OWL)[9] and data languages such as the RDF[10] form the core of SWT. RDF relies on a simple triple-based data model to capture specific data related to individual elements in the domain.

A triple, or a fact, is a basic unit of knowledge storage and is at the core of a semantic data model. It is composed of three elements: a subject, a relation, and an object. The subject and the object in the triple refer to some individual elements in the domain. Each element is an instance of a concept in the domain terminology. These facts along with the elements constitute specific knowledge that varies, while the concept constitutes the background knowledge in the domain and is applicable to all instances. For example, "Account" and "Balance" are concepts in the Banking domain, and 256665:Account and USD 995.25:Balance are instances of those concepts. "256665:Account hasBalance 995.25:Balance" represents a fact using the *hasBalance* relation.

SWT provides the interoperable standards-based infrastructure to build, query, and use semantic metadata to augment our understanding of data within the enterprise. This application of ontologies and reasoning to make sense of structured and unstructured data is what we refer to as the semantic enterprise within the enterprise space.

Meaning and Consequence in Semantic Models

Current languages and technology for storing and managing data, whether XML- or RDBMS-based, capture the structure of data in some syntax. However, what is missing is the meaning of the data element, which is absolutely required for automated interpretation of data. SWT describes the precise meaning of data elements by associating them to semantic metadata. The semantic metadata expresses concepts that are interconnected by various kinds of relations between them (see Figure 2). The meanings for concepts (e.g., Customer, Account, Invoice, Payment) in a semantic model are derived from these relations to other concepts in the context. Thus each concept has a semantic space that includes a set of related concepts. The key relations

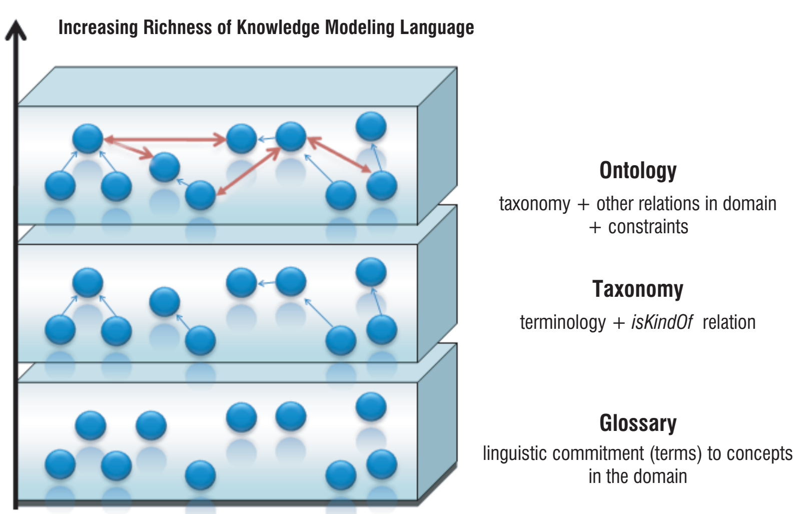


Figure 2 — Varying richness of metadata relations.

used to define the semantic space of a concept are as follows:

- *isKindOf*. Some concept is a subconcept of, or fully contained within, the semantic space of some other concept.
- *disjointWith*. Some concept is completely separate from some other concept (does not share semantic space at all).
- *equivalentTo*. Some concept completely overlaps the semantic space of some other concept.

There may be other relations in the domain of knowledge as well, such as *marriedTo*, *childOf*, *employedAt*, and so on.

Current languages and technology for information management also lack the ability to compute consequence. An assertion in semantic models is what has been explicitly stated to be true. Consequence is what follows or what is computed (by deductive reasoning) from existing assertions or other consequences. The key requirement is that the system always maintains logical consistency. The process of computing the consequence from what has been stated earlier is known as inference. This ability to compute consequence is the key differentiator between SWT and standard data technology. It is this inference that is applied to either verifying data against the rules described by the ontology or even making new inferences on the basis of relations that were not expressed in the triples but could be concluded from them.

Figure 3 shows an example of a simple semantic model and the consequences derived:

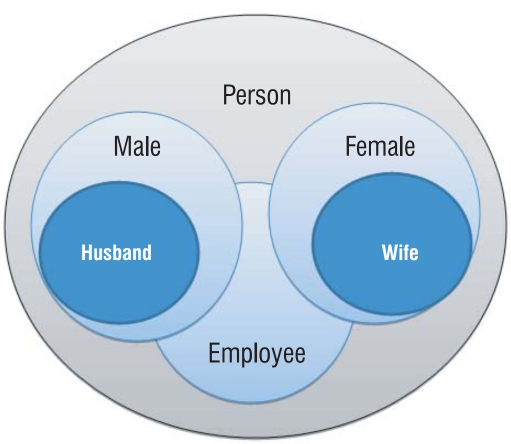


Figure 3 — Semantic space: the precise meaning of concepts based on context.

- A Person is always either a Male or Female (in other words, both Male and Female are kinds of Persons) — *isKindOf* relation.
- Male is completely distinct and separate from Female — *disjointWith* relation.
- Some Males are Husbands and some Females are Wives — *isKindOf* relation.
- Some Persons are Employees. Therefore some Males and some Females are Employees — *logical consequence.*
- Some Husbands and some Wives may be employees — *logical consequence.*

Key Characteristics

SWT is characterized by five major features or capabilities:

- **Flexible data model.** This is based on triples that supports modeling multidimensional relationships between data elements in a schema-independent manner. The triple-based data model is structured and does not require a schema, though it may use one.
- **Computational reasoning.** SWT provides the ability to compute logical conclusions from existing data (what is already known) by combining the ontology or semantic metadata as a precise and explicit expression of the background knowledge of the enterprise concepts.
- **Information integration.** SWT can relate and integrate data elements from diverse data sources and domains into a single model using bridge relations between multiple ontologies. The ontology constraints also serve to derive equivalence relations between individual data elements.
- **Information summarization.** The SWT stack includes a declarative rules language used to define abstractions (high-level concepts) as the precise decomposition from other detail concepts (low-level concepts) across multiple ontologies.
- **Information query.** SWT also defines an ontology-based query language that enables users to ask questions based on the concepts and relations in the ontology. The query focuses on the domain entities and relations without any reference to how the data is actually stored and organized at the physical level. A query constructed using terms in the user's vocabulary is answered by using the ontology to translate it to other concepts and relations that have actual data or instances associated with them.

Major Components of SWT

Figure 4 shows the major components that will be required in deploying SWT solutions in the enterprise. The key parts of this stack — the basic triple-based data model, the ontology language, and the query language (based on triples) — have been standardized by the W3C.

Key Tasks Involved in the Deployment of SWT

As shown in Figure 5, enterprise adopters of SWT must understand and plan for three distinct tasks. *Semantic modeling* involves the creation of semantic metadata in languages such as OWL, describing various units or aspects of the business. *Data population* involves translating structured enterprise data (RDBMS, XML, etc.) to RDF triples associated with the semantic metadata. This task is more complex when dealing with unstructured text, since it requires information extraction (extracting structured triples) using natural language–processing techniques and then refining the results with assistance from a domain expert. The final step is conceiving and implementing specific *information applications* that will embed and use the SWT components and to search, query, and reason on this data annotated with semantics.

BLUEPRINT FOR A SEMANTIC ENTERPRISE

A semantic computing platform is the architectural realization of the application of SWT to the enterprise context (see Figure 6). The platform provides the foundation on which applications that exploit semantic metadata will be built. SWT evolved out of the need to explicitly describe meaning and context for existing data. This implies that it has application within the enterprise wherever data is currently stored and used. Typical applications of a semantics platform to the enterprise are outlined in the next section.

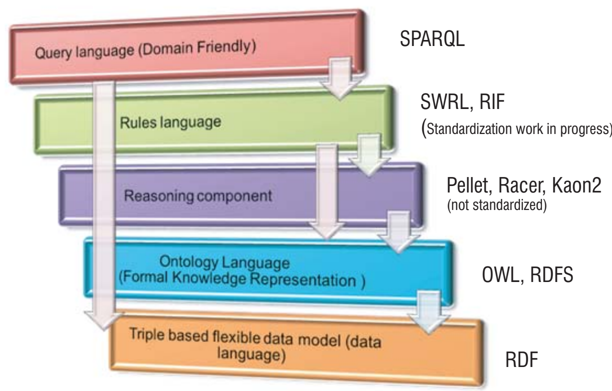


Figure 4 — Semantic technology stack showing components and languages.

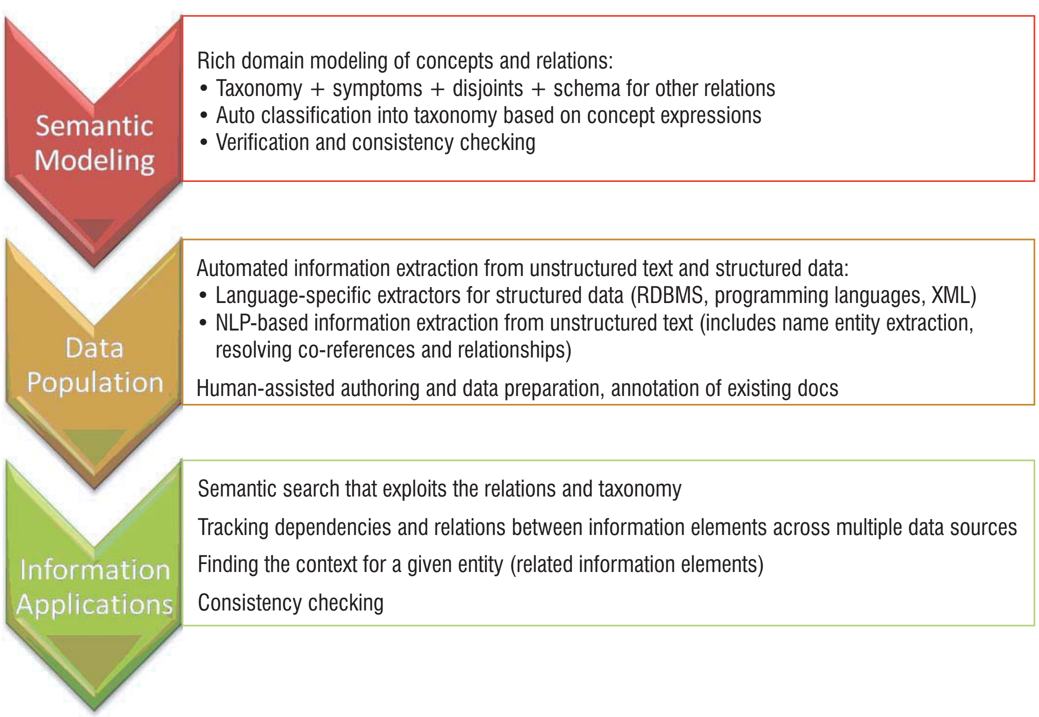


Figure 5 — Process for SWT adoption.

USEFUL APPLICATIONS

Enterprise Information Management

SWT can deliver significant value when applied to business intelligence (BI) and information search solutions in enterprise information management. Ontologies support better communication, explanation, and prediction, as well as better mediation between data representations. Current technology for managing data is mature in handling the operational requirements of scale and performance within the enterprise context. RDBMS and XML together cover a large portion of structured organizational data. These technologies are known to scale and support querying on the underlying data store.

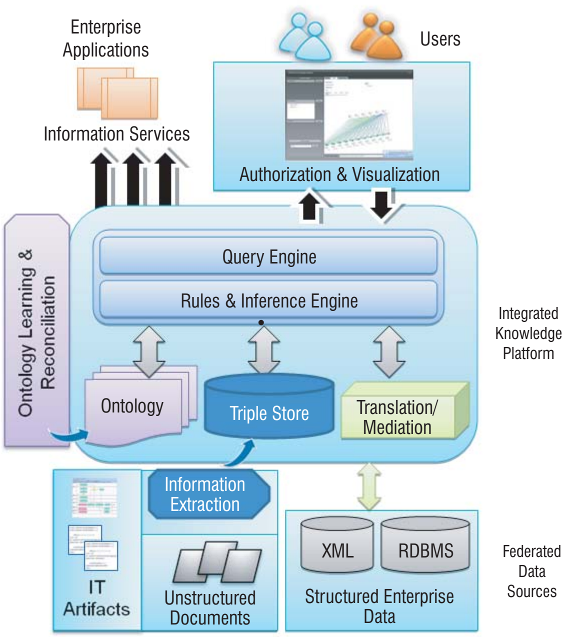


Figure 6 — Semantic enterprise architecture.

However, the scale and complexity of the current business context demand solutions beyond simple storage and retrieval — solutions such as mining and discovery of new relations from existing data and flexible views on an integrated IT portfolio database. Customer relationship management (CRM), fraud detection, compliance management, and many mining applications require this additional capability to reveal new insights from existing data.

Semantic Search

Current search technology is based on statistical occurrences of search keywords in a document corpus. This approach suffers from two primary drawbacks: it has low precision and is extremely sensitive to the actual words entered by the users. It has no understanding of the context of the keyword. Further, current search technology returns documents and not information content.

Semantic search within our scope is search over a formal knowledge base that includes the ontology and the individual data elements in RDF. Population of the knowledge base involves converting existing structured data into triple-based representation (RDF) and extraction of structured information triples from unstructured text. Once the knowledge base is populated, semantic search exploits the ontology relations to find related information content. In the example shown in Figure 7, a knowledge base is populated with a micro-ontology for the movie and mobile content domain. The search term used is "DiCaprio ringtone"; however, there is no ringtone associated with Leonardo DiCaprio, and so occurrence-based search simply cannot be considered. The ontology-powered search first looks for paths between the concepts (Person and Ringtone) involved in the ontology and then retrieves the correct RDF triples based on those paths.

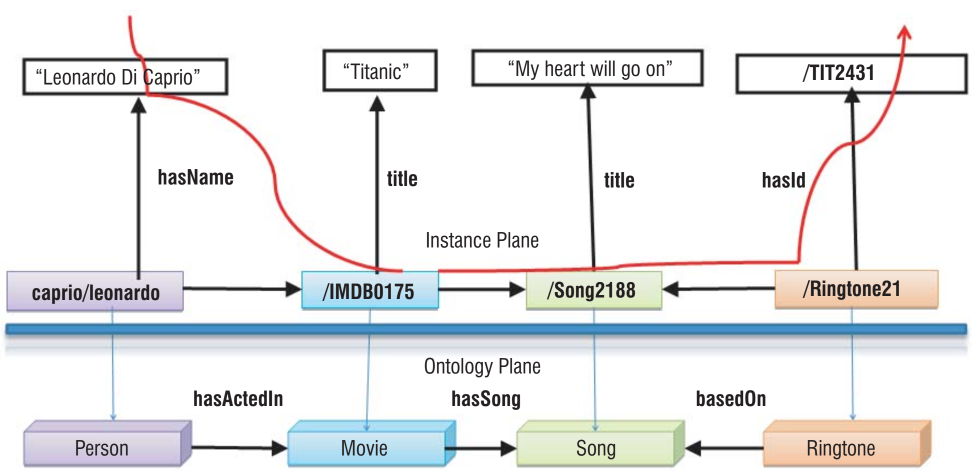


Figure 7 — Semantic information search.

Semantic Access to Structured Enterprise Data

Existing methods to access structured data are intrinsically coupled with the technology choice at the physical data level. For example, we use SQL for access to relational databases and XQuery/XPath for access to XML data. What we want is to enable information access that is free from the data management technology used in the physical layer. SWT permits analytic queries on data to be expressed against the conceptual data model as opposed to expressing queries against a physical data model (using tables, columns, joins, etc.).

Ontologies as rich information models provide the required expressivity and concrete language for conceptual information modeling in the enterprise. This implies that any existing database (relational, object, etc.) schema can be mapped and transformed into an OWL-based ontology definition. Ontology-based query language satisfies the exact requirements for semantic data access against the semantic data model expressed in ontologies.

Once the ontology has been extracted and mapped to the schema, there are two major options for dealing with the data in databases. One can either transform the data as triples into the semantic RDF store or simply use the mapping between the ontology and the physical data to mediate data access by translating semantic queries into the physical layer (in this case, SQL). It is possible to accommodate both these approaches within a single platform based on specific considerations for each database. The ontology defines the business terminology, while the data is maintained in the RDBMS. By establishing mappings between the ontology concepts and the physical schema, one can enable semantic querying. A useful feature of semantic querying is that it presents a unified business model across multiple databases and other data sources.

Smart Business Intelligence

The next generation of BI platforms will be powered by SWT and distinguishable by their use of semantic data models and schema-less data warehouses with greater flexibility and adaptability. This in turn will result in lower TCO and improved ROI. These systems will exploit SWT's information integration feature, establishing the relations between data elements that refer to the same concept but are expressed in distributed data sources, possibly using different languages.

Current BI platforms primarily employ a data warehouse based on a customized unified schema designed based on specific analytic needs. Existing data has to be cleansed and transformed (if necessary) before being moved into the warehouse. The RDF-based data model does not require a schema, and all attributes are explicitly modeled using relations. Information integration

Table 1 — A Comparison of Semantic and Relational Data Models

Semantic Data Model	**Relational Data Model**
Data represented at conceptual level.	Data represented at physical level.
Data is based on a flexible data model that can express any relationship — the schema is captured using expressive concept or ontology language.	Data is constrained by a rigid schema. Schema language has primitive expressivity compared to ontology language.
Stores information — data with context. Meaning is formally expressed and explicit.	Stores data — implicit informal meaning. Relations are translated as columns or constraints.
Captures subsumption (*isKindOf*) relations between concepts and relations.	No support for *isKindOf* relation — subsumption hierarchy is missing.
Domain-friendly language is used to express queries; everything is explicit.	Queries are bound by the schema, and explicit low-level joins have to be specified.
Rules language expresses new concepts and relations as expressions over existing ones.	Concepts and relations are limited to what is defined in the schema. No derivations — absence of rule support to define intentional concepts and relations.
Capable of making inferences on existing data by leveraging ontology.	No inference capability.
Information access is domain dependent and schema independent — ideal for information integration across diverse sources.	Information access is bound by the schema — difficult to merge and reconcile. Ideal for managing controlled data. Reasonable scalability and query performance.

from multiple operational databases is achieved by triplifying the data into an RDF-based triple store or mediating the data through the ontology. Further, by describing precise mapping rules between various stakeholder ontologies, semantics-powered BI will deliver automated information summarization from low-level operational data stores that will provide relevant information support while enabling drill-down into details.

Enterprise Application Integration

The information modeling, integration, and query features of the platform can be leveraged to improve IT-business alignment and enterprise application integration by integrating models and artifacts from business process management (BPM) and service-oriented architecture (SOA) into a single knowledge repository.

Enterprise Architecture

SWT can be an effective means of creating a machine-processable description of the various entities, functions, and relations between elements across all layers in a traditional enterprise architecture (EA) model. The technology permits slicing and dicing through the functions and layers to compute various views into the EA model from multiple stakeholder perspectives. For example, the platform can be deployed to create an integrated knowledge repository of IT and business artifacts and models, including business process models, entity models, physical database schema, use case models, application source code, and configuration files. The information integration and summarization capability of the platform will help multiple stakeholder concerns to be satisfied at varying levels of detail and scope from the single repository.

Software Engineering and Information Systems Development

SWT provides the infrastructure to create integrated knowledge repositories that import information from requirements, architecture, and design and from application programs and databases. This repository and the features of the technology then form the basis for semi-automatic traceability and impact analysis in software engineering.

Active Repositories in IT Management and Software Engineering

The current software engineering landscape, characterized by distributed teams, aggravates problems with informal knowledge management. Enterprise architects and software engineering teams struggle to cope with multiple scattered representations of data. Within the enterprise, we have data duplicated across multiple databases, IT applications, and business units. The software engineering context complicates this further by scattering business concepts across programming languages, modeling artifacts, and XML documents. This duplication of data across databases, artifacts, and languages increases overall IT costs, affects customer service, and increases maintenance effort. The main pain points are:

- The problem of semantic scatter (lack of integration between artifacts)
- Too much effort in impact analysis and system appreciation
- Poor knowledge management and reuse
- The challenge of knowledge transition across geographies
- Knowledge lost in employee turnover

Hidden or misunderstood relationships in the IT portfolio also lead to error-prone decision making. The following questions represent stakeholder concerns regarding an IT portfolio in a financial services organization:

- Which service returns the current balance of a Trading Account?
- What business processes rely on the historical price query service?
- What use cases in a specific IT application deal with Foreign Currency Accounts?
- How many customer-facing applications will be affected (directly or indirectly) if the payments server is down?

In the current scenario, these questions can be answered only by employees who are deeply involved in the design and implementation of the databases and applications in question. In order to support the proper scale and transition in an enterprise, we need an explicit representation of the implicit knowledge that is now restricted to a few experts.

Semantic technology can be employed to build an integrated knowledge repository for better insight into the IT portfolio. Multiple and diverse concerns and vocabularies from various stakeholders are represented using multiple ontologies that are bridged and reconciled. Process models, use case models, application code, and version history of software artifacts are primary data sources that are extracted into the RDF store using custom-built extractors to provide a multidimensional perspective (see Figure 8). This integrated repository supports querying across process definitions, entities, application code, and data. Some of the potential benefits are improved productivity and reduced cost of quality due to better visibility into the dependencies. The repository also delivers views at varying levels of abstraction for multiple stakeholders from the basic data by using declarative rules to precisely define the mapping between high-level concepts and detail data (see Figure 9).

CHALLENGES IN SEMANTIC TECHNOLOGY DEPLOYMENT

While SWT offers compelling benefits, there are a number of challenges in deploying semantic technology into the enterprise:

- **There is a mismatch between semantic technology and existing data technology.** The rules language within the semantic technology stack and SQL are both based on the abstract logic-based language Datalog. However, the complete integration of rules into the semantic stack is still a work in progress due to differences in primary assumptions between semantic technology and Datalog. SWTs, especially OWL and RDF, are primarily designed to operate at the Web scale without any central control. They presuppose incomplete distributed data that is not centrally controlled and therefore assumes no unique name and relies on "open-world" reasoning. Existing database technology operates in a complete and controlled environment, where it is safe to assume unique names for individuals and "closed-world" reasoning. Both paradigms provide the same results when facts are known and expressed in a knowledge base. However, when dealing with negated conditions, results differ. For example, consider the definition of a "childless couple" as a husband and wife who have no children. It is possible that the facts about children for some couples are not captured in the knowledge base. In this scenario, closed-world reasoning simply assumes that what is not mentioned in the knowledge base is not true, thus implying that all such couples are in fact childless. Presented with the same knowledge base, open-world reasoning concludes only that these couples *may* be childless.
- **The effectiveness of solutions depends largely on the quality of ontologies.** Semantic models involve some social agreement about the words used to describe concepts and relations in any domain. The translation of existing expertise and knowledge into machine-processable semantic models is an error-prone manual activity. Instead of humans devising semantic models, it is also possible to apply machine learning methods to "learn" concepts from existing data and documents. In either case, the quality of the semantic models will ultimately drive the value derived from information integration and search applications.

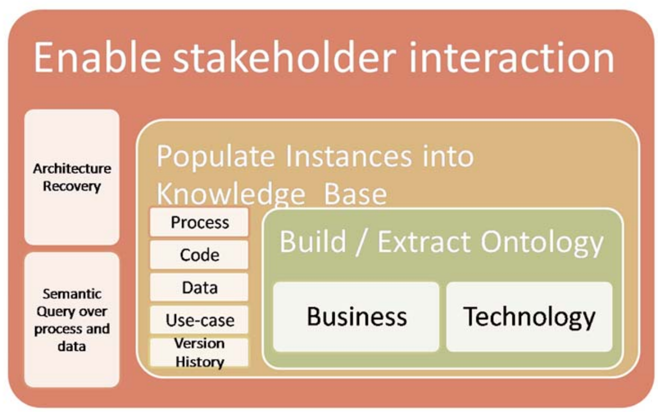


Figure 8 — Integrated knowledge base for the IT portfolio.

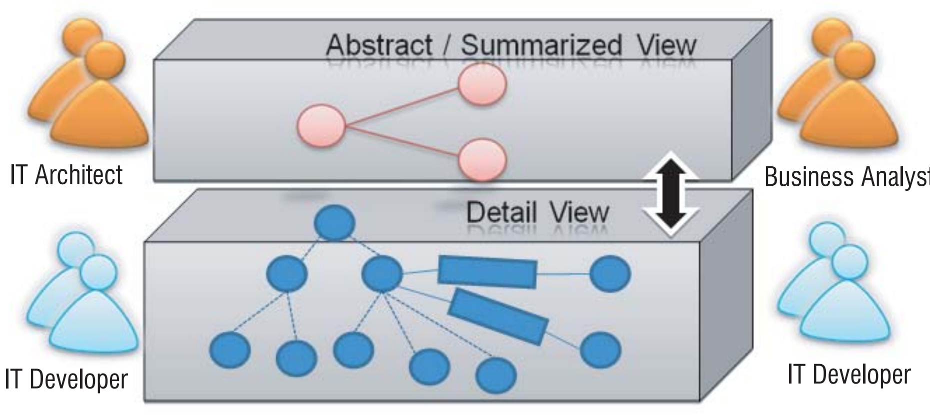


Figure 9 — The integrated knowledge repository uses declarative rules to define the mapping between high-level concepts and detail data.

- **Ontology creation in OWL presents a high barrier to entry due to the learning involved.** Semantic modeling involves understanding the various constructs of knowledge representation. The foundations of the modeling language constructs are borrowed from a subset of first-order predicate logic known as description logic.[11] This can be intimidating for modelers, so emerging semantic modeling tools hide the complexity by providing a graphical environment to model and maintain logical consistency. However, the task still demands some skill and initial learning about constructs of knowledge representation.
- **Integrating and aligning multiple ontologies is a nontrivial task.** An enterprise will require many ontologies to cover the various products, geographies, operating functions, units, and information categories. There will be some overlap between these ontologies, so discovering and expressing the precise overlap and mapping between various ontologies or semantic models within an enterprise is absolutely essential for deriving value from the technology.
- **User interaction with a knowledge base requires better visualization technology.** Enabling human-computer interaction through SWT-based knowledge repositories requires new techniques for visualizing data relationships. The ontology and the triple-based data store represent massive information graphs. The challenge here is to enable users to see what is of interest to them in terms of concepts and relations that express information at the right level of granularity.

CONCLUSION

SWT is ready for enterprise deployment. It has clearly moved out of academic and research contexts into actual industrial use. There is some literature[12, 13] that offers use cases and guidance for IT managers and architects. There are also reports from early adopters of this technology across all industry segments, ranging from pharmaceuticals to healthcare, banking, insurance, telecommunications, and retail.

I recommend starting small with a clear problem definition and set of use cases rather than attempting to go "big bang" with an enterprise-wide ontology modeling activity. This may require partnering with vendors to assess the applicability criteria, define the architecture and phased technology induction and training plan, provide tool support, and, finally, supply the implementation and maintenance services associated with SWT.

ENDNOTES

[1]Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web." *Scientific American*, May 2001.

[2]Brunnermeier, Smita B., and Sheila A. Martin. *Interoperability Cost Analysis of the US Automotive Supply Chain*. Prepared by Research Triangle Institute for National Institute for Standards and Technology (NIST), March 1999.

[3]Dublin Core Metadata Initiative (http://dublincore.org).

[4]"ISO/IEC 11179-1:2004: Information technology — Metadata Registries (MDR) — Part 1: Framework." International Organization for Standardization (ISO), 2004.

[5]Baader, Franz, Diego Calvanese, Deborah. L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider (eds). *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[6]Brachman, Ronald, and Hector Levesque. *Knowledge Representation and Reasoning*. Morgan Kaufmann, 2004.

[7]Sowa, John F. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Brooks/Cole Publishing, 1999.

[8]McGuinness, Deborah L. "Ontologies Come of Age." In *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, edited by Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolfgang Wahlster. MIT Press, 2003.

[9]McGuinness, Deborah L., and Frank van Harmelen (eds). "The Web Ontology Language Overview: W3C Recommendation 10 February 2004." W3C, 2004 (www.w3.org/TR/owl-features).

[10]McBride, Brian. "The Resource Description Framework (RDF) and Its Vocabulary Description Language RDFS." In *Handbook on Ontologies: International Handbook on Information Systems*, edited by Steffen Staab and Rudi Studer. Springer, 2004.

[11]Baader et al. See 5.

[12]"Semantic Web Case Studies and Use Cases." W3C, 1994-2009 (www.w3.org/2001/sw/sweo/public/UseCases).

[13]Pollock, Jeffrey T. "Enterprise Semantic Web in Practice." Oracle, 2007 (http://me.jtpollock.us/pubs/2007.05-Pollock.STC.2007.pdf).

John Kuriakose is a Principal Architect with the Center for Knowledge-Driven Information Systems (CKDIS) labs at Infosys Technologies Ltd., India. The CKDIS labs is part of SETLabs, the applied research and development unit at Infosys, and it aims to create the basic reusable primitives that will help Infosys clients transform themselves from being data-centric enterprises to being knowledge-centric enterprises. Mr. Kuriakose has 17 years' experience in the IT industry, and his current research focuses on the application of semantic technology in creating next-generation, knowledge-driven IT solutions. He can be reached at john_kuriakose@infosys.com.

Leveraging the Semantic Web for Data Integration

by Shamod Lacoul

Data integration is a key issue for any enterprise. If data from multiple sources can't be mixed and matched, it gets hard to gather the right information at the right time. As a result, it becomes a major hurdle to make informed decisions within or across an enterprise and for employees to reach set goals. How can one overcome this obstruction of the lifeblood of a company (i.e., enterprise data)? One promising set of technologies is based on the Semantic Web. Semantic Web technologies (SWTs) promise to enrich enterprise data by adding a layer to ease interoperability. Easier interoperability of data is critical in an enterprise, where information handling is always a top priority to complete tasks on time and within budget. Let's peek into the capabilities of the Semantic Web and explore what it offers for data integration.

CURRENT STATE-OF-THE-ART DATA INTEGRATION SOLUTIONS

Enterprises use such technologies as extract, transform, load (ETL), enterprise application integration (EAI), and enterprise information integration (EII) for data integration. Lately, IT is gradually moving from these technologies to service-oriented architecture (SOA), where service contracts are exposed to minimize the interoperability maze. SOA has been a buzzword in enterprise data management for a while now, and we hear about it among many developers trying to resolve data integration by exposing services.

Business process management (BPM), as a complementary discipline, has been the de facto solution to EAI problems for several decades. BPM allows packets of data to be transmitted within a business process workflow among applications in the enterprise. In BPM, various attributes of disparate data formats are mediated to synchronize applications residing in silos.

SOA and BPM are state-of-the-art data integration technologies in the current enterprise IT infrastructure. Unfortunately, both technologies have been criticized in the enterprise for being rigid, brittle, and time-consuming tools for integrating data among applications. SOA creates an ecosystem of integration where interoperability becomes a maze, and BPM requires a multitude of person-months to map attributes from source data to its destination format. Both solutions are cost-inefficient and time-consuming as far as maintenance and management are concerned. So people are starting to look around for alternatives to provide more nimble solutions. Is there anything out there for those looking to minimize the pain of traditional integration? The answer is maybe — maybe there is something in the Semantic Web that can help us minimize the pain. However, the fact is that Semantic Web technologies cannot provide a complete solution to data integration — in fact, no technology can — but they certainly show great potential by providing dynamic data and a flexible architecture. Let's find out how.

SEMANTIC DATA INTEGRATION

Ever since the second business application was written, the problem of integration has been ingrained in the data management space. Data integration is a common problem, and any solution that can diminish this challenge is welcome in the enterprise. Can the Semantic Web help reduce the pain of integrating data? If there were only one application to control the entire set of enterprise data, then we would not have seen this issue at all. Unfortunately, just the contrary holds true in the real world, where a variety of applications are implemented to meet specific business requirements. As a result, numerous applications proliferate into an unmanageable set of assets at organizations everywhere.

Within the current enterprise boundary, the data model technologies underlying these unmanageable assets are mainly:

- Relational database management systems (RDBMSs)
- XML exposed through Web services
- Various types of unstructured data and text

An RDBMS manages relational data in a tabular format with relations between tables in primary/foreign key relationships. XML uses a tree-based XSD schema to define structures that contain instances. Unstructured text, on the other hand, loosely defines the structure of the document. All these known formats make it hard for machines to understand entities and their properties. Hence, whenever we reconcile the formats of data in any of these silos, it becomes difficult to map attributes residing in one format of the schema to attributes residing in another format. One is essentially passing data values to a destination schema from their source and patching the links with syntactic glue. This type of integration is syntactic integration. You may notice that none of these formats explicitly supports the transfer of semantics. The overarching problem is the lack of a shared semantic model.

So it seems as though a flexible means of expressing semantics is missing from the data models available today. What can be done to add meaning to these data structures? The term "semantic" in the Semantic Web indicates an intention to add this missing meaning to current data models. The most prevalent semantic data models use a branch of first-order logic called description logic to define relations between concepts in the data by adding structural metadata. Every class (concept) in a document can be related to other concepts in the same or a remotely located object by adding properties that define hierarchies, equivalencies, differences, and so on. By augmenting these metadata tags, the model becomes enriched with structures that were previously unavailable.

Furthermore, unlike relational databases, in which the underlying mathematics is based on relational or set theory, the semantic data model is based on graph theory. Any unit of information in the semantic model is always represented as a graph, which is easier to evolve. The graph model, by default, offers a lightweight data integration framework where two or more models can be merged simply by making a union of the graphs, as depicted in Figure 1.

Every unit of information in the semantic data model is represented by a Uniform Resource Identifier (URI), by virtue of which you can explicitly define relationships between two or more disparately located entities. For example, you can add a metadata tag to a unit of information and specify it to be the same as another unit of information within the same or another document. Likewise, there are multitudes of other tags that are available as explicit syntax to express whether several concepts are equivalent, different, inverse of, subclass of, and so forth. One key advantage that a distributed semantic data model brings to the table along with these metadata tags is its ability to adapt to the ever-changing requirements of a business.

A semantic model is often called an ontology. Many ontologies are written using the Resource Description Framework (RDF) and defined in a W3C-recommended standard language called the Web Ontology Language (OWL). OWL was created to support the definition of abstractions of real-world things, including concepts from the enterprise world. For example, you can easily define employees, purchase orders, or even expense reports in an OWL ontology and use it to create applications that are semantically aware.

In essence, an ontology can define the metadata of enterprise artifacts where the mapping of domain-level sets becomes interconnected over URI namespaces. This interconnection of disparate data sets creates a richer, more unified base of metadata for machine processing and interoperability. Compare this to the interconnection in mapping relational or XML data, as mentioned above. In non-ontological data models, we simply try to connect the values from source attributes to their destination counterparts. Each link is based on syntax, and every connection is on a node-by-node basis. However, when we align ontologies, we align them on a conceptual level. For example, if we need to map two purchase orders — one in System A and one in System B — all we do is figure out the relationships between these two nodes and explicitly define the metadata hierarchy. For instance, if it is a subsumption hierarchy (i.e., if it includes things in a child-parent relationship), we can say node A is a subclass of node B. If these two nodes are opposites, we can say node A is an inverse of node B.

As we can see, ontologies give us an option for defining a link between different parts of the data space in a precise manner. In this way, semantic models can save a significant amount of time when integrating disparate data sets by mixing and merging concepts instead of

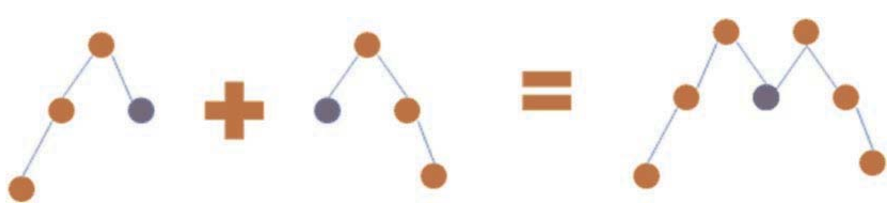


Figure 1 — Union of two graphs creates a merged graph.

mapping one-to-one as in traditional methods. Table 1 compares the two methods.

POWER OF SPARQL

SPARQL is a W3C-recommended SQL-like query language for accessing RDF data. It is gaining quick adoption among Web developers, and the real power behind it is its ability to query any models that are federated via graphs. Figure 2 gives an example of a SPARQL query to gather data from two different (fictional) URL locations.

The SELECT statement fetches two values (?firstname and ?lastname) for each person from two disparate graphs located in fictional URLs http://www.example1.xyz and http://www.example2.xyz. The "WHERE" clause in the query is an option for filtering the output. When an application executes this query, the output is displayed as it is shown on the right side of Figure 2. SPARQL is easier than traditional query languages like SQL. It eliminates cumbersome joins when querying graph models. In the query above, no join logic is executed, contrary to how we implement a query in a relational database.

A system based on the graph data principles provides a layer of abstraction that SPARQL rides on top of. As long as the data that SPARQL reads is in the form of URIs within RDF graphs, tapping into many data sources becomes feasible and less painful.[1] Legacy data stored in silos can easily be exposed as a "queryable" link in the form of a SPARQL endpoint. Some available tools for exposing legacy data in RDF are D2RQ[2] and Triplify.[3]

So what can SPARQL do to assist in enterprise integration? Imagine a platform where the existing data, be it internal or external, is mapped to RDF and exposed via URIs just as you would expose Web services. Now, instead of calling the API, you can write queries against your federated data explicitly and merge the output using the same query. This provides a nimble way to embed business logic in the query by which you can fetch data from distributed data models that are rich,

Table 1 — Traditional vs. Semantic Data Integration (Source: PricewaterhouseCoopers.)

	Traditional Data Integration	**Semantic Data Integration**
Data structure	Predominantly relational: focuses on sets of similar data	More flexible: focuses on relationships between things regardless of similarity
Data integration method	Extract from original source, transform to local data definitions, load on own servers	Link to source of data using data definitions in shared ontology
Data integration scalability	Each new data source expands costs exponentially	New data sources accessible at minimal cost; business domains share the federation cost
Contextual richness	Constrained by costs and central staff workloads	Benefits from the network effect: context added with new data and linkages
Information source bias	Internal	Internal and external
Business unit involvement	Report requesters	Managers of their own ontology and external data-linking activities
Standardization method	One standard, no exceptions, loss of valuable information context	Explicitly allows both standard data and contextual information

Query

```
SELECT ?firstname ?lastname
FROM <http://www.example1.xyz>
FROM <http://www.example2.xyz>
WHERE { ?person <http://www.example1.xyz#first> ?firstname.
      ? person <http://www.example2.xyz#last> ?lastname. }
```

Output

?firstname	**?lastname**
John	Doe
Mike	Smith
Shamod	Lacoul

Figure 2 — Sample SPARQL query.

dynamic, and evolving. In essence, the broader your data set, the more specific and relevant your query results can be. SPARQL provides the ability to filter results in an efficient manner and thus helps extract more relevant insights.[4] For example, say your company has its customer data spread out among an SAP ERP server, an Oracle database, and SalesForce CRM. Previously, you would either have integrated all three systems to synchronize customer data among applications or used some type of master data management (MDM) hub to normalize data to a single version of the truth. Now, given that all customer data from the three systems can be exposed as RDF for access, you can simply SPARQL these disparate customer data sets and retrieve whatever customer information you need without much hassle.

When such a gigantic amount of data becomes linked and exposed, it generates a network effect and adds tremendous value to any enterprise information supplied with additional context.

LINKED OPEN DATA CLOUD

Linked Open Data (LOD) is a movement on the Web in which structured RDF data is linked to other forms of RDF data to formulate a cloud of data where one can traverse federated models. It is one of the fastest-growing phenomena on the Web, and there are new data sets joining the LOD cloud almost daily. The latest one to join is the *New York Times* with its thesaurus data, a development that was announced at the *2009 Semantic Technology Conference*. The LOD cloud is revolutionary and offers great value to the World Wide Web as another layer to augment the Web's structured knowledge. Figure 3 shows the traction it has already gained.

Now, the striking feature we get out of this is an ability to query the web of data to fetch answers from the LOD cloud. When such a gigantic amount of data becomes linked and exposed, it generates a network effect and adds tremendous value to any enterprise information supplied with additional context. Think of mashups today. A mashup can mix and match data from various sources if the format complies with the access layer. In the LOD world, however, we do not need to worry about the access layer; we simply utilize the power of SPARQL to point to any URI and execute the query. On top of the instant access, it also provides publicly available contextual data and helps create a "mashup on steroids," merging internal as well as external information.

Some of the popular public ontologies enterprises can use to link to the LOD cloud are GoodRelations (for defining your company and products)[5] and those being developed by the (US) National Center for Biomedical Ontology (NCBO)[6] and oeGov (for eGovernment).[7]

SEMANTIC WEB FOR CROSS-ENTERPRISE INTEGRATION

Cross-enterprise, or B2B, integration is a major data integration challenge faced by many companies today that need to collaborate on a data level with one or more of their partners, customers, and sometimes even competitors. Standards such as EDI, RosettaNet, UBL, ebXML, and so on have been used to exchange documents between enterprises since the 1970s, but the sad truth is that "standards help make more standards, and no one follows any." In other words, everyone tends to create his or her own idiosyncratic standards for business exchange documents. As a result, it becomes hard even to comply with the internal standards within the same business ecosystem.

Ontologies are a structured approach to exposing the choices companies must make between operational standards and operational flexibility. They become a platform for creating a shared understanding of the formal business language where flexibility at a local level within the enterprise is encouraged.[8] Similarly, this formal business language can be extended beyond the enterprise firewall, exposing the data to an extended ecosystem of suppliers and partners in order to add value for customers. By giving explicit access to the ontologies within the ecosystem, participants can contribute their distinctive views and evolve the inter-organizational relationships in an organic manner.

At the same time, when users query data over the Web via SPARQL, they receive direct access to cross-enterprise data stores exposed as ontologies. If companies expose their data as is done in the LOD cloud, then any partner with proper security access can query the business data quite easily using SPARQL. Instead of mapping data point-to-point into other standards, they can simply query the data and then map it to whatever standard they need internally. This helps, on the one hand, to leave data at its source and, on the other, to pull only the subset of the data that the users need. For example, consider a scenario in which SupplierA needs to share his daily report with BuyerB to reconcile orders

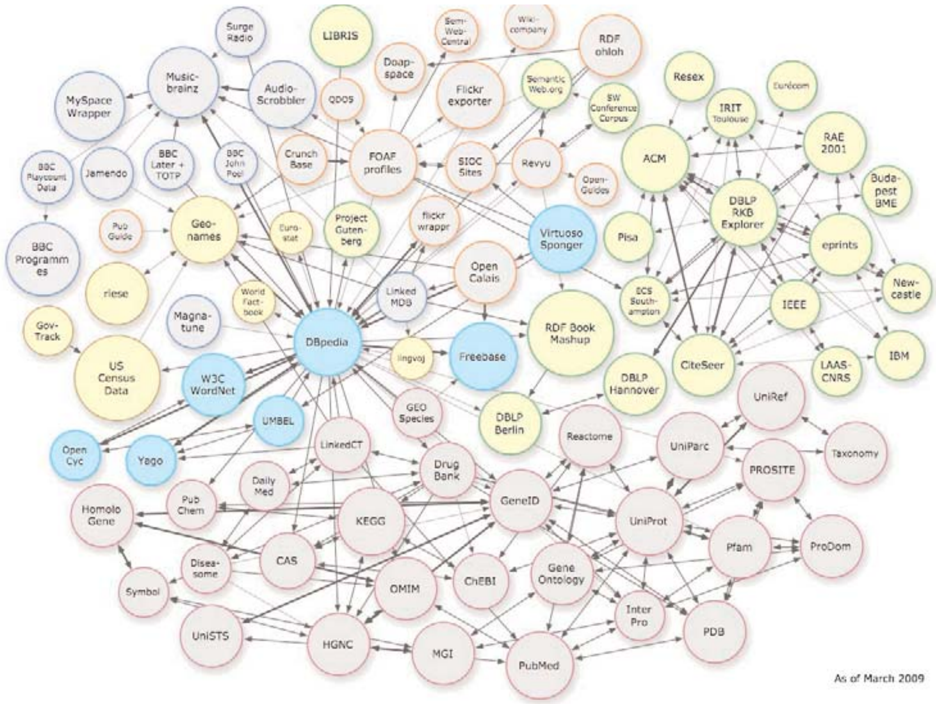


Figure 3 — The growing Linked Open Data cloud.
(Source: www4.wiwiss.fu-berlin.de/bizer/pub/lod-datasets_2009-03-05_colored.png.)

at the end of every business day. Say SupplierA exposes his reporting data to BuyerB in an RDF format with proper access and hands BuyerB a SPARQL query to execute. When BuyerB executes the query, she gets the data directly from SupplierA's source and extracts only the subset that she needs. Better still, if the requirements change tomorrow, all the partners need to do is change the logic in the query.

CONCLUSION

We know that data integration is a common problem that enterprises face daily. System integrators waste hundreds of hours trying to connect the dots, and information workers find it laborious to extract sensible information out of disparate datasets. Semantic technologies are emerging with a promise to revolutionize this troublesome issue. Let us explore this stack of technology that shows capabilities for evolving and quickly integrating discordant data within and outside our enterprises.

ENDNOTES

1 "Spinning a Data Web." *Technology Forecast*, PricewaterhouseCoopers, Spring 2009.

2 "The D2RQ Platform — Treating Non-RDF Databases as Virtual RDF Graphs." D2RQ Freie Universität Berlin, 2009 (www4.wiwiss.fu-berlin.de/bizer/D2RQ).

3 Triplify (http://triplify.org/Overview).

4 PricewaterhouseCoopers. See 1.

5 GoodRelations is a lightweight ontology for annotating offerings on the Web (www.heppnetz.de/projects/goodrelations).

6 The National Center for Biomedical Ontology (NCBO) is a consortium of leading biologists, clinicians, informaticians, and ontologists (http://bioontology.org).

7 oeGOV is making and publishing W3C OWL ontologies for eGovernment (www.oegov.org).

8 "Making Semantic Web Connections." *Technology Forecast*, PricewaterhouseCoopers, Spring 2009.

Shamod Lacoul has been working in the enterprise data integration space for past six years and has been instrumental in leveraging Semantic Web technologies for data integration. Mr. Lacoul has a BS in computer science and an MS in software engineering from San Jose State University. He is also a co-organizer of the Silicon Valley Semantic Technology Meetup group. Mr. Lacoul can be reached at shamod@gmail.com.

Accruing Business Value Through the Adoption of Semantic Web Technologies

by Bhuvan Unhelkar and San Murugesan

The Internet and Web have enabled easy and nearly instantaneous dissemination and exchange of different types of information almost free of cost. Most information currently on the Web is document (page) based and presented in textual natural language, because it is primarily meant for use by humans. It's hard for computers to understand and make use of this information content, as it may lack structure, precise meaning, or context, or because a given word or phrase could refer to more than one thing. The Semantic Web is an evolving extension of the World Wide Web in which Web content can be expressed not only in natural language, but also in a format that can be read and used by computers, thus permitting a more automated and effective way to find, share, and integrate information. Semantic Web technologies (SWTs) facilitate exchange of information among various applications in a *meaningful* way by providing precise and unique meaning and context to the content and easing interaction between the human users and a computer system, or between two computer applications.

A semantic enterprise (SE) is an organization that uses SWTs in its communications, business processes, software applications, structured information (data) architecture, and unstructured information (e.g., multimedia content) delivery, resulting in deployment of new kinds of applications that were otherwise infeasible — or previously unimaginable. By integrating and using information from disparate, incompatible sources in a cohesive and meaningful way, an SE is better able to enhance and personalize the experience it provides to its users, create and effectively use new bodies of knowledge from the information available within and outside the organization, and become a more agile business by dynamically adapting its business processes depending on the situation or user context. By exploiting the technologies of the Semantic Web, an SE can create a people-machine continuum that enhances business agility.

UNDERSTANDING THE SEMANTIC WEB

As we noted above, the Semantic Web is a *meaningful* Web. This means that SWTs offer substantially more value than mere sophistication in exchange and shared use of information. They not only make it easier to connect content that may be residing in different information silos, but also enable applications to learn through the process of execution. Consider, for example, a user who wants to buy a music video. Instead of making the user specify the title and related details of the music video, a semantic application can utilize the user's "hints" and the overall context or environment in which he is operating to help narrow the search and identify the music video he is looking for, thereby personalizing and enhancing the user experience. The system could also learn the user's preferences and perceived interests from his previous purchases and queries and so offer smarter and more tailored responses when he tries to purchase another music video or, perhaps, a totally different item. Applications based on semantic technologies open up opportunities for the organization to dynamically gain new knowledge pertaining to its users, knowledge that can then be used to positively affect the organization's internal and external processes and its overall operating environment.

The Semantic Web is a combination of people-centric technologies with machine-understandable semantics that results in an ontology of data, information, and knowledge. Such an "ontology" can be open and shared and/or owned and maintained by an organization that provides it, along with the standards for exchange of information and the rules for interpreting it. The Semantic Web also provides the basis for the provisioning and consumption of dynamically created bodies of knowledge for organizations that collaborate to supply a product or service to the user.

Ontological approaches are used to create a knowledge base of users, which is built based on the initial understanding of the users together with their interests, regional preferences, and seasonal choices. Such knowledge bases facilitate socialization of users on the Web through formation of social networks. For example, blogs, wikis, RSS feeds, and "tweets" are all part of the people-to-people interaction that began with the mere exchange of information and eventually led to the formation of dynamic groups and networks.

Ontologies used in social networking technologies have made it easy for service providers to customize their searches and outputs to provide context and relevance to users *as they use the Web*. The Semantic Web combines people and processes together in order to not only offer the users what they are looking for, but also to develop their understanding of what they may actually want or need. Extending the people-to-people interaction with machine-facilitated understanding of information exchange on the Web opens the door for many semantically aware applications (SAAs) to collect, transmit, receive, interpret, comprehend, and manipulate information that is continuously and dynamically updated based on the user's behavior.

Say a student is undertaking a school science project. Her first instinct might well be to fire off a number of Google or Amazon searches, which will doubtless yield some useful and some not-so-useful results. One might suggest instead a more semantic manner of responding to the user's queries. Semantic search results can be based on several factors, including:

- **Input the user has provided.** These are the user's actual query terms via text or voice. The user may not necessarily have a precise query in mind; hence, this information might have to be collected through more than one interaction between the person and the machine. For example, instead of searching for information on "potentially hazardous asteroids (PHAs)," the student may query on "falling objects" *and* "hazard."
- **Information the user provided about herself in the past.** This is based on preferences and interests of the user as recorded from earlier interactions. This information can be used with the current query in order to give a response that has better value to the student. For example, if the student has interacted earlier with the Web on meteorites, the query can further point to the latest stellar events.
- **Additional information the user inadvertently provided.** This information includes the user's physical location, IP addresses, and the like. While the use of such information may raise privacy and security concerns, there is nevertheless the opportunity, within legal bounds, for an SAA to understand the context and/or background of the query and offer relevant responses. For example, a student enquiring about PHAs in the US or UK (northern hemisphere) may have a different context than a student in Australia or New Zealand (southern hemisphere).
- **Information the user may not be willing to provide.** This category includes such information as age-based behavior patterns and related shopping and consumption patterns, which may be sourced from third-party services. Here is the Semantic Web at its best — supplying information from externally available and internally built sources to put together a schema for the user that will enable the applications to offer far more related information than a query called for. For example, students with an interest in science may want to know about science forums, related books and conferences, and prizes for which they might compete.

SAAs can use all available information that is directly and indirectly provided by the user, that surrounds the user, and that can be gleaned from previous providers of service to the user to offer an intuitive and productive user experience. Realizing such applications will require communications networks to furnish the technology basis for rapid information exchange as well as Web services to enable applications to understand each other irrespective of the underlying technical environments.

> **Ontologies used in social networking technologies have made it easy for service providers to customize their searches and outputs to provide context and relevance to users *as they use the Web.***

THE BUSINESS VALUE OF THE SEMANTIC WEB

A business can gain significant value from the Semantic Web by drawing on its capability to combine and interoperate with several technologies and services, encompassing data warehouses, disparate operating systems, and myriad types of messaging. The resultant "cohesive" technological platform allows in-depth user participation and collaboration that also reveals new and meaningful relationships among information silos and applications that may not be obvious otherwise to the business. By deploying a Semantic Web–based enterprise information platform, a business can launch new systems and applications that enhance enterprise agility and create synergy among networks, services, open technologies, and security measures, resulting in interoperability among data, applications, business processes, and services.

However, as authors T. Jeffrey Pollock and Ralph Hodgson[1] note, semantic interoperability among an organization's various applications is only a subset of the Semantic Web. While this interoperability enables the exchange of information in a meaningful way within organizational boundaries, the Semantic Web also enables exchange of information and execution of applications for for multiple organizations over the Internet cloud.

A semantic enterprise can glean value by judiciously and innovatively embracing SWTs in four key business areas (see Figure 1):

1. Internal and external communication and interaction
2. Business processes
3. Software applications
4. Multimedia information provisioning

An SE can capitalize on this cohesive Semantic Web platform to adapt and augment its internal operational aspects as well as the external customer-centric and business partner–centric aspects. In this way, the SE forges better, tighter collaboration with its business partners and individual users both within and outside the enterprise.[2] The end result is value-added services to its customers, employees, and business partners that deliver an enhanced end-user experience.

As Nova Spivack, CEO and founder of Radar Networks, has outlined, the Semantic Web presents the business world many new possibilities, including:[3]

- **Creation of new multimedia content**. Offering a new product or service on the Web requires sourcing of varied content, which may be generated by the organization, its partners, and even its users. The Semantic Web enables sourcing of this content in its myriad formats. Furthermore, this uniquely sourced content can be easily updated by authorized people and/or computer systems, thereby ensuring that the information stays current.
- **Creation of new business offerings**. Businesses are able to become and remain more agile than before by effectively exploiting a technical platform that can be used for launching new applications that can then support new products and services.
- **Creation of better-connected consumer networks.** Enterprises can create unique social networks and communities based on common interests of customers derived from an understanding of the customers' behavior. An SE can thereby provide opportunities for its customers to relate to each other. This interaction allows customers to understand the organization better, and the same interaction on the Web can be used to glean feedback to help the organization decide on the strategies for its newer products and services.

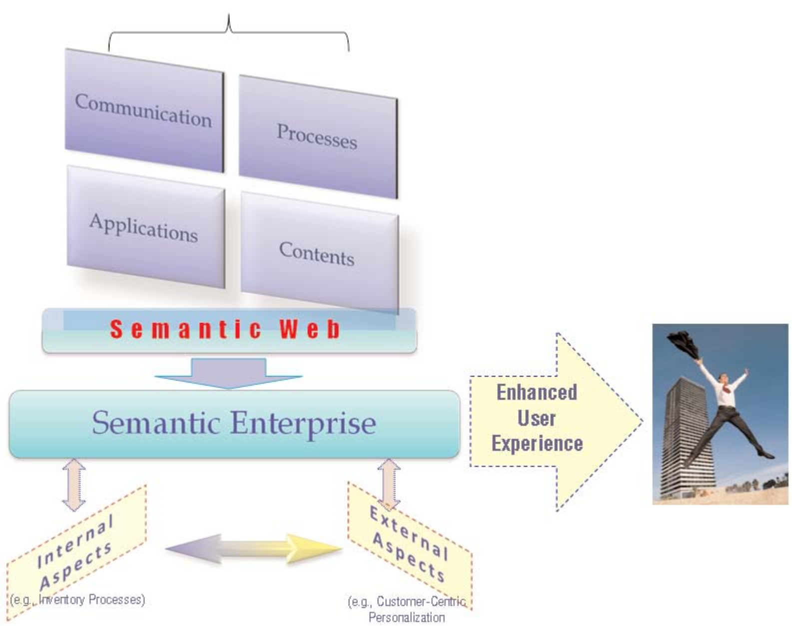


Figure 1 — By embracing SWTs in four key business areas, a semantic enterprise can forge better collaboration with its business partners and enhance the user experience.

- **Creation of new collaborative opportunities**. SAAs are able to seek and consume services outside the organizational boundary in order to create new products and services. The collaborative nature of the Semantic Web is thus of immense benefit to global business and one of the best sources of value the Semantic Web offers an enterprise.

HOW SAAs WORK

In a semantic enterprise, a Semantic Web–based technical platform, shown in Figure 2, facilitates direct information exchange amongst siloed applications, Internet-based exchange and sharing of information amongst those applications, and, finally, ontology-based collaboration among multiple applications and databases. Furthermore, as highlighted in Figure 2, an SE not only aims to use the Semantic Web, but also has a specific goal of influencing the people and processes associated with it. Hence, an SE focuses on enabling people to make gainful use of applications that take them beyond the specific transaction they are engaging in with the organization.

SAAs cover a wide array of business areas. These include marketing/advertising, semantic enterprise search, business intelligence, smarter business process management, and customer relationship management (CRM), to name but a few.[4, 5] SAAs are in a "continuous learning" mode, so that every interaction a user has with these applications supplies them with more knowledge about what the user is looking for. In addition, as mentioned above, these applications make use of the context, past habits, and demographic information about the user to provide the solutions.

For example, a person looking for a certain book may wish to consult the past history of similar purchases from people with similar demographics. To make this possible, new and existing applications need to "talk" with each other in order to understand both the needs of a specific user *and* the potential needs of an entire cross-section of people with similar demographics. This results in business processes that are not restricted to a single requirement, but are collaborative in nature. Collaborative business processes are continually transcending technological as well as organizational boundaries. These collaborative "global" processes require precise modeling for multiple users in their many different contexts.

Modeling and implementation of these business processes assumes greater importance as these processes use semantically enriched and widely linked data and information. The era of using rigidly defined components to put together software applications has passed. Most modern-day applications are built dynamically by modeling the business processes they are meant to service and making use of a large number of smaller-sized software components that are available as services over the Internet. Therefore, the

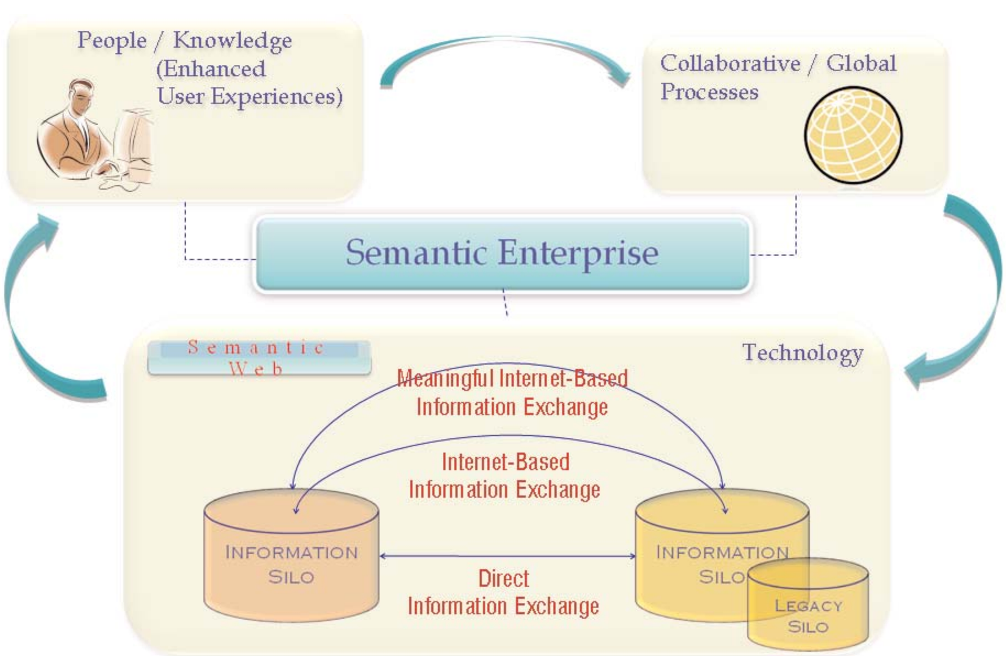


Figure 2 — Increasingly meaningful exchange of data and information leads to collaborative processes and knowledge.

use of modeling standards, such as the Business Process Modeling Notation (BPMN) and Unified Modeling Language (UML), becomes even more important in ensuring that the interoperability between applications is translated into corresponding business value. It is the combination of interoperable and meaningful applications, an understanding of the context and needs of the users, and the ability to transcend organizational boundaries and engage in collaborative processes that together make a semantic enterprise.

SAA ADVANTAGES AND CHALLENGES

With their use of context, multimedia content, and location awareness, SAAs offer many advantages:

- **Personalizing the user experience.** This requires semantic applications to understand the context of the user and his personalization needs. It is worth noting that the user's context can change dynamically as he uses the system. While mobile-enabled applications make personalization easier, they do pose security and privacy challenges.[6]
- **Enhancing the customer relationship strategy.** In addition to incorporating enhanced customer experiences, an SE's customer relationship strategy needs to deal with the privacy and security issues that invariably pop up with SAA usage.[7] This strategy should address the use of existing CRM software packages and the changes that must be made to these packages in order to enable them to use the Semantic Web. The enterprise will need to model its corporate customers separately from its individual customers, as corporate customers will most likely have their own SAAs with which the application under consideration may need to be integrated.
- **Creating new bodies of knowledge.** This requires the data and information silos within and outside the organization to be connected to each other in a meaningful way. For example, if a user belonging to a certain demographic (e.g., a junior high school student) is looking for certain information for a science project, then supplying her with additional material of potential interest may be relevant. On the other hand, offering the same user a listing of auto mechanics in the area will probably not be of value (different demographics). Therefore, establishing correlations between information silos requires multitiered ontologies. The logic used to relate a set of information with another set of information itself requires further correlation with another group of logic — leading to the concept of ontologies within ontologies.[8]
- **Helping businesses become more agile.** To better serve its customers and clients, a business can become more agile by dynamically changing its processes depending on the situation. This requires the applications to understand the commercial environment, the behavior of business partners, regulatory requirements, and services in order to enable changes to the business processes. We consider location independence in the architecture and design of business processes to be crucial for agile businesses.

SAAs also present challenges that need to be carefully managed in order to provide business value. Since the multimedia data in SAAs is sourced from multiple entities, one of the key issues with these applications is the ownership of this data. Moreover, collaborative global processes,[9] which form a core part of SAA, are successfully produced only through requirements that are modeled keeping the interoperability among multiple applications in mind. Semantic Arts President Dave McComb, in a discussion of semantic enterprise architecture (SEA), has also highlighted the need for a reference architecture to facilitate application interoperability.[10]

As noted above, semantic applications are made up of a large — and at times even unknown — number of software components that are diversely spread out over the Internet cloud. Therefore, implementation of an SAA is not a simple case of developing and testing code. SAAs are made up of executable software components/services, content sourced dynamically from content providers, personalized content created and made available by users, business rules/ontologies, and business processes that change according to context. Consequently, from a software development viewpoint, implementing these applications is less a matter of development and more one of integration. Table 1 summarizes the impact of SAAs' dynamically changing requirements and need for continuous integration on the traditional software development lifecycle (SDLC) phases.

RECOMMENDATIONS

To successfully embrace the Semantic Web and begin to transform your organization into a semantic enterprise, we recommend that you:

- Assess the ability of the enterprise to benefit from semantic applications. Identify the key users (customers and employees), model their behavior, and incorporate a suitable implementation approach. Make provision for users to create and manage their own personalized experiences.

Table 1 — SAA Development Considerations by SDLC Phase

SDLC Phase	Impact on Application Development	Relevance to a Semantic Enterprise
Requirements (which also incorporates data modeling, known as ontological engineering)	Data may need to be exchanged and shared across multiple applications and organizations. Model the applications using BPMN and/or UML.	Develop the ability to cater to unique and dynamically changing user requirements. Ease of translating requirements to a solution or an application is a key advantage. Combine actual requirements with "expected" requirements in the future.
Design	Use reusable components that are encapsulated and interoperable.	Reusable designs integrate components from different sources (services), which results in minimal inhouse development. Applications can be cheaper, but security and business challenges need to be addressed appropriately.
Implementation	The development environment can be varied/independent for various components or services.	SAAs will require continuous implementation, with a high focus on end-user testing rather than testing within given technical environments. Agile approaches to development are appropriate for these applications.
Deployment	Deploy generic applications that may be tailored to the user's preferences.	Employee (user) training is vital. However, customer training hardly ever happens. Incorporate help and guidance for end users who may not have training.

- Understand the current "inventory" of software applications, networks, content, and processes that the business uses.
- Map your new enterprise information (data) architecture to Semantic Web ontologies that exist within Open Linked Data or specific industry domains. Identify where gaps exist and/or where ontological engineering may be required to extend those ontologies.
- Model the new business processes that will be collaborative — and perhaps global — in nature using known process modeling techniques, such as BPMN and UML. We reiterate that the major challenge in implementing these business processes within a semantic architecture will be integration.
- Identify collaborative opportunities between the organization and external parties — this is where the machine-centric nature of the Semantic Web can provide input.
- Map the business processes to the SAAs available to the organization. The crucial difference here in the use of the Semantic Web is that the applications being mapped can be both internal and external to the organization.
- Identify, address, and discuss the risks associated with the use of SAAs (e.g., privacy and security). Choose and apply appropriate risk metrics to the creation of dynamic business processes.
- Deploy applications. This may happen through downloadable client modules or directly as a service on the Web.
- Organize for training and support. This could happen formally through organized face-to-face training, on the job, or as the user tries to use the system.
- Enhance the user experience beyond the direct "customer and employee" interactions of the organization and move toward a more personalized and location-independent relationship with the customer.
- Enable users to create and manage their own personalized experiences with the organization.
- Source products and services from collaborating/partner organizations.
- Enable incorporation of future technologies; use open architecture where feasible.

- Where required,[11] take an SAA-based approach to implementing enterprise-wide carbon emissions management systems.

The emerging Semantic Web and SWTs have great potential for providing new business value to an enterprise. Enterprises should start to embrace SWTs for both intra- and interorganizational applications, focusing first on applications that are of significance and offer higher value to them. The time is ripe for beginning the transformation to a semantic enterprise!

ENDNOTES

[1]Pollock, T. Jeffrey, and Ralph Hodgson. *Adaptive Information: Improving Business Through Semantic Interoperability, Grid Computing, and Enterprise Integrations*. John Wiley & Sons, 2004.

[2]Unhelkar, Bhuvan, Abbass Ghanbary, and Houman Younessi. *Collaborative Business Process Engineering and Global Organizations: Frameworks for Service Integration*. Business Science Reference, 2009.

[3]Spivack, Nova. "Minding the Planet: The Meaning and Future of the Semantic Web." Lifeboat Foundation, 6 November 2006 (http://lifeboat.com/ex/minding.the.planet).

[4]Murugesan, San. "Get Ready to Embrace Web 3.0." Cutter Consortium Business Intelligence *Executive Report*, Vol. 7, No. 8, 2007.

[5]Murugesan, San. *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications*. Information Science Publishing, 2009.

[6]Unhelkar, Bhuvan. "Mobile Enterprise Architecture: Model and Application." Cutter Consortium Enterprise Architecture *Executive Report*, Vol. 11, No. 3, 2008. The discussion in this report pertains to mobile business applications, but it is also applicable to the current discussion of semantic applications.

[7]Murugesan. See 4.

[8]Murugesan. See 4.

[9]Unhelkar et al. See 2.

[10]McComb, Dave. "A Semantic Enterprise Architecture (SEA)." Semantic Arts, September 2004 (www.semanticarts.com/DesktopModules/ViewArticle.aspx?ArticleID=905&mid=3475).

[11]Large organizations and businesses in Australia and other countries will soon be required to comply with mandatory carbon emissions reporting (www.climatechange.gov.au/reporting/index.html). An SAA-based approach is ideal for the implementation of such systems.

Bhuvan Unhelkar (BE, MDBA, MSc, PhD; FACS) is a Senior Consultant with Cutter Consortium's Business-IT Strategies and Enterprise Architecture practices. He has more than 27 years of strategic as well as hands-on professional experience in information and communication technologies (ICTs) and their application to business and management. He is the founder of MethodScience.com and has notable practical consulting and training expertise in software engineering, enterprise architecture, project management, collaborative Web services, and mobile business. This hands-on experience led to the creation of two- to five-day industrial courses in the aforementioned areas, which have now been delivered during the last decade in Australia, New Zealand, US, UK, India, China, and Singapore. Dr. Unhelkar earned his PhD in the area of object orientation from the University of Technology, Sydney, in 1997. Subsequently, he designed and delivered course units, such as global information systems, object-oriented analysis and design, business process reengineering, and IT project management in various universities in Australia, China, and India. He leads the Mobile Internet Research and Applications Group (MIRAG) at the University of Western Sydney, where he is also an adjunct associate professor. He has written/edited 14 books in the areas of globalization, mobile business, software quality, business analysis, business processes, and UML and has extensively presented and published research papers and case studies. Dr. Unhelkar is a sought-after speaker, fellow of the Australian Computer Society, life member of the Computer Society of India, Rotarian at St. Ives, discovery volunteer at NSW National Parks and Wildlife Service, and a previous TiE Mentor. He can be reached at bunhelkar@cutter.com.

San Murugesan is a Senior Consultant with Cutter Consortium's Business-IT Strategies advisory service; a professor at Multimedia University, Malaysia; an Adjunct Professor in the School of Computing and Mathematics at the University of Western Sydney, Australia; and an independent business, IT, and education consultant. Dr. Murugesan's expertise and interests include green computing, Web 2.0 and 3.0, mobile computing, Web engineering, information retrieval, e-business, innovation, IT project management, and offshoring. In 2006, the Council of the Australian Computer Society (ACS) elected Dr. Murugesan to the grade of ACS Fellow, an honor that is awarded to those who have made a distinguished contribution to the field of information and communication technology. Dr. Murugesan is a distinguished visitor of the IEEE Computer Society and a sought-after speaker. He is the editor of the Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications. *He was General Chair of the* 5th International Conference on Web Engineering, *and he has served as cochair and organizer of other international workshops and conferences. He serves as Associate Editor of the* Journal of Web Engineering, International Journal of E-Business Research, *and* International Journal of Health Information Systems and Informatics. *Dr. Murugesan held various senior positions at Southern Cross University and the University of Western Sydney, both in Australia, and the Indian Space Research Organization, Bangalore, India. He also served as Senior Research Fellow of the US National Research Council at the NASA Ames Research Center. He can be reached at smurugesan@cutter.com.*

Intranet Architecture for the Semantic Enterprise

by Hyoung-Gon (Ken) Lee and Edmund W. Schuster

ERP STRENGTHS AND WEAKNESSES

As a mature technology, enterprise resource planning (ERP) has a strong history of increasing productivity through organizational integration and the application of the packaged software concept. ERP provides the majority of computer applications associated with day-to-day business operations for medium-sized and large firms, including accounting, human resources management, and the coordination of manufacturing operations.

ERP vendors typically offer production scheduling systems, and other models, as part of large, standardized software packages. To meet the needs of a diverse user base, ERP vendors limit the amount of functionality to the lowest common level. Firms generally use business process reengineering as the means of adapting to the limited functionality. They seldom rewrite individual modules, because this reduces the possibility of an upgrade if the ERP vendor provides an improved version sometime in the future.

In practice, it becomes clear that there is no single ERP solution for all industries. This is especially true for global corporations with operations in many countries and a diverse set of products and manufacturing processes. In these situations, a standardized ERP module, such as a single production-scheduling model, has little chance of working equally well across all divisions and manufacturing environments.

As global markets and the needs of consumers become more complex, corporations can benefit from recent innovations in computing architecture that are capable of overcoming the emerging weaknesses of ERP (i.e., the lack of flexibility and appropriateness of software and mathematical models for specific problems particular to a firm). We believe that Semantic Web technologies (SWTs) combined with existing standards form the critical elements for improving the flexibility of delivering various mathematical models to users, a common function of ERP systems. This will enhance business decision making for many corporations during the next decade.

A SEMANTIC TECHNOLOGY ROADSHOW

In 2008, Ed made a visit to Taiwan that brought to light a significant shortcoming common to all ERP systems. The purpose of Ed's visit was a demonstration of a new SWT. For the past five years, the technology had undergone extensive research and development at the Massachusetts Institute of Technology (MIT), and it was ready for commercial application. It involved connecting data and decision-making models across the Internet.

The core of the system was an online dictionary with a unique way of building ontologies for use with XML. This contrasts with W3C standards, which have few robust means of handling semantics for OWL beyond the use of triples. Current practice is to treat words and text as symbols connected via an ontology. Reasoning occurs through the tree of relationships and assumes that no semantic ambiguity exists. Yet we know this is hardly true in the real world. For example, the word "apple" could mean a fruit or a brand of computers, depending on the context. Since computing within ERP does not include machine-understandable semantics, there is no way to know for sure the exact meaning of "apple" or how the word fits into existing ontologies for fruit or computers.

Our experience is that few ontologies are absolute and all-encompassing, because about 10% of English words commonly used in business have multiple meanings. W3C standards include no provision for an online dictionary of any type. Within the W3C framework, various artificial intelligence (AI) approaches are the primary means employed to achieve machine-understandable semantics. However, AI has a mixed record of success in practice.

The firm Ed visited was Taiwan Semiconductor Company (TSMC), and the ERP application involved production scheduling. While the operations managers at TSMC liked the new heuristic model for production scheduling well enough, they expressed even greater interest in the semantic enterprise architecture (SEA).

There was a consensus at TSMC that one of the biggest problems with the company's current ERP system was the increasing difficulty and cost of installing new software.

Enabling Experimentation

TSMC managers stated several times that the ability to implement new models quickly would create a level of flexibility not currently available in commercial ERP systems. This relative lack of flexibility also characterizes the end of the lifecycle, making the removal of outdated models equally difficult. It is frequently the case that models are implemented and then phased out as the needs of the business change or modeling technology improves.

The rapid implementation of new models for operations or supply chain management provides an opportunity for experimentation, something very much needed in standard ERP systems. These systems usually consist of modules designed to do specialized tasks. This is at odds with the realities of modern corporations, because the technology available for modeling is constantly changing, as are the business problems that models analyze. This increases the role of rapid adaptation as the basic element for management success. Current ERP systems fall short in this area.

For TSMC, it is important to schedule the sequence of manufacturing (using a mathematical model), as this determines end-item inventory levels and ultimately customer service. It is critical to meet demand for semiconductors in a timely manner because these are the essential components for electronics such as digital cameras, cell phones, and computers. A flexible approach for the delivery of a model matched to a particular operational situation would therefore be very valuable to TSMC — and to many firms worldwide.

Introducing LSSEA

With these thoughts in mind, we will examine the structure and advantages of the Lee-Schuster Semantic Enterprise Architecture (LSSEA), the technology Ed presented to TSMC. LSSEA allows access to mathematical models via an intranet connection and provides interoperability for data and models. The architecture exposes the interface to a mathematical model via standard protocols. Although the prototype of LSSEA developed at MIT involves production planning, the architecture can deliver any mathematical model written in computer code.[1]

The components of LSSEA are threefold:

1. A computer model for calculating a production schedule, called the **Modified Dixon-Silver Heuristic (MODS)**,[2] which resides on a server
2. The **M Dictionary**,[3-5] an Internet-based way of achieving machine-understandable semantics for describing data fields that are inputs, outputs, and attributes of a mathematical model
3. A **spreadsheet interface**

The intent of LSSEA is to standardize and speed the process of modeling across the enterprise, thus eliminating a significant barrier to practical use.[6] LSSEA essentially creates a supply network for mathematical models within the enterprise. The approach uses established Web standards combined with several innovations and provides real-time documentation to users. The implementation requirement to run a model is minimal, creating the flexibility needed for experimentation.

LSSEA originally was an Internet-only application; however, the architecture will find quicker application and wider use at the enterprise level. Its characteristics and capabilities fit with the emerging use of Web 3.0 as a tool to organize intrafirm information and to make connections between data and decision-making models.

The primary difference between LSSEA and OWL is the use of a high-tech, Web services–accessible dictionary that contains unambiguous ontologies, among other things. This offers a great advantage in consistent definitions for words and noun phrases, along with reuseable ontologies.

HOW LSSEA WORKS

The core concept of LSSEA is similar to software as a service (SaaS), which envisions a demand-led software market in which businesses assemble and provide various services as needed to address a particular industrial requirement.[7] The SaaS approach has gained worldwide attention with the introduction of Apple's iTunes and Google Apps.[8] Almost immediately, both of these services were extremely successful among younger Internet users.

We adapt the SaaS approach to the internal delivery of models using several innovations, including the M Dictionary and a spreadsheet interface. Most SaaS involves an Internet-based browser for manual data entry; however, we believe there are advantages in using spreadsheets. Long term, the majority of spreadsheet applications will be Web-enabled and open, offering another reason to adopt this technology for the user interaction.

Creating an SEA for mathematical models involves two elements:

1. A simple user interface
2. A means to connect the interface to computer code located on servers

The central idea is to host a model, written in a structured computer language, such as Java or C++, on a single server with a simple interface that can be loaded onto any computer. The interface then connects to the central server when running the model. Such a system allows users located anywhere in the enterprise to connect to a particular model via personal computer quickly and cheaply.

For LSSEA, a Microsoft Excel spreadsheet serves as the end-user interface. Spreadsheets are easy to understand, and many firms already do many model calculations in Excel using custom approaches developed internally. Enhancements in Excel 2003 and 2007 allow for direct interaction with a server, where computer code for a specific model can reside. In this type of application, the spreadsheet is just an interface, although it is still possible to customize and do other calculations.

Creating an SEA using spreadsheets also requires a robust way to treat semantics. LSSEA uses the M Dictionary to provide consistent, machine-understandable semantics for words and noun phrases contained in the spreadsheet interface. In this way, there is no ambiguity regarding the definition of terms used to describe the data fields. Figure 1 shows the high-level diagram of how the Excel spreadsheet interacts with the server and the M Dictionary.

In summary, LSSEA provides a way to connect models hosted on a server to users via a spreadsheet interface. The M Dictionary supplies common semantics to describe the inputs, outputs, and attributes of a mathematical model. This provides consistency across the enterprise, a valuable thing for large multinational corporations. Ultimately, the goal of LSSEA is to have many models loaded on a server, accessible though spreadsheets distributed to users. This provides central control over versions of the code while giving users the flexibility to experiment with different models. Implementation is rapid and low cost.

In the next section, we describe the important aspects of the M Dictionary in relation to LSSEA.

THE M DICTIONARY

The fundamental problem with employing words as descriptors in enterprise systems is that a single word can have several definitions, while multiple words can have the same definition. This paradox means that natural language often does not have the internal consistency required for straightforward application as an identifier or a unit of meaning within computer systems. Take the word "cell," for example. It might mean "cellular phone," "biological cell," "fuel cell," or "terrorist cell." Without some idea of the context, it is impossible to know the meaning of the word.

Achieving the goal of word descriptions that are machine understandable requires a deeper appreciation of the role of semantics in enterprise computing systems. In M, every word has only one definition. This is an extremely important characteristic because computers that communicate using M do not need to understand the context or usage of a word to know its meaning. To overcome the ambiguity issue, the M Dictionary includes a numeric extension to denote individual words. For example, a specific meaning of "cell" becomes `cell.1`.

To account for several definitions of the same word, the M Language allows multiple numeric extensions, one for each definition. Thus, `cell.1` is a word in M and `cell.2` is a different word. With this method, every word has only one meaning.

Figure 2 shows a screen shot of the M Dictionary (mlanguage.mit.edu) for the word "forecast." In this example, there are five separate definitions. Depending on usage, some compound words are also part of the

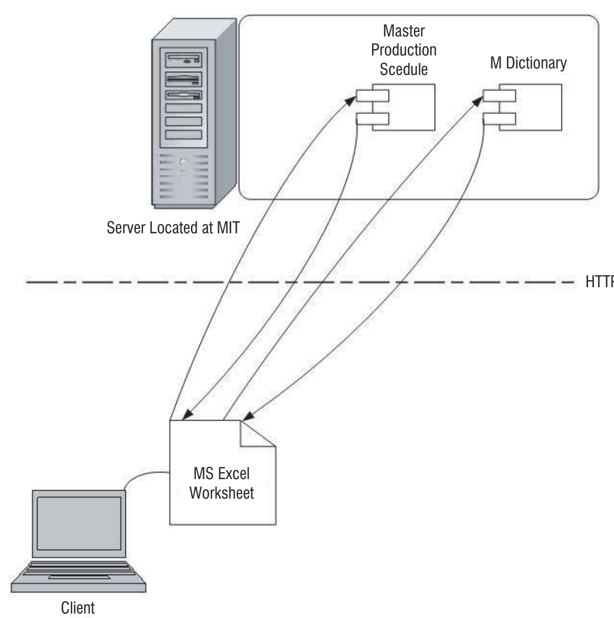


Figure 1 — The Overall Architecture of the Open System for Master Production Scheduling (OSMPS), an earlier version of LSSEA.

M Dictionary. An example is "operations research," represented as `operations_research.1`.

In addition to the definition, the dictionary entry also contains word relations (sometimes called an ontology). These are simply the connections between words. Typical relationships include synonyms, antonyms, types, and parts. Types refer to word generalizations. For example, `automobile.1` is a *type of* `motor_vehicle.1`. Parts are words that are components of another word. This is often the case when thinking about physical objects (e.g., a `wing.4` is a *part of* `airplane.1`), although it could also hold true for abstractions.

Both types and parts establish a hierarchy within the dictionary through making connections between entries. These word connections are valuable in a number of different ways, including improved search.

The M Dictionary uses the wiki approach with several important modifications, including improved security through user registration, maintenance of the integrity of word relations, a monitoring function to reduce the chances of near-identical definitions, and administrative controls to ensure accuracy. The M Dictionary also has various statistical features that measure usage. Currently, there are 210,000 words in the M Dictionary and more than 700,000 ontological relationships.

Perhaps the most important aspect of the M Dictionary is its ability to create a precise semantic for enterprise search that is machine understandable. Since words in M have only one meaning, a keyword search in M yields matches based on the word definition, not on the character string. This allows for search based on meaning rather than keyword.

LSSEA IN ACTION

In this final section of the article, we will highlight an example of LSSEA called the Open System for Master Production Scheduling (OSMPS). Initially an Internet-based architecture, the OSMPS also works as an enterprise application via an intranet. In general terms, the OSMPS:

- Can be implemented using any ERP system along with an available intranet
- Decouples various production planning models from the ERP system, allowing a firm to avoid vendor/solution "lock-in" and providing standardization across multiple ERP frameworks that might exist in a global enterprise.

As an initial step, the Excel spreadsheet partially shown in Figure 3 must be loaded on a personal computer. Immediately after opening, the spreadsheet will automatically connect to the server where the code for the model resides. In this case, the model is the MODS, a sophisticated production-scheduling program for make-to-stock manufacturing.

First, a dialog box will appear that asks the user whether he or she wishes to update the M words contained in the spreadsheet. If the answer is yes, then the spreadsheet connects to the M Dictionary, and all word definitions are updated. This process takes about 30 seconds.

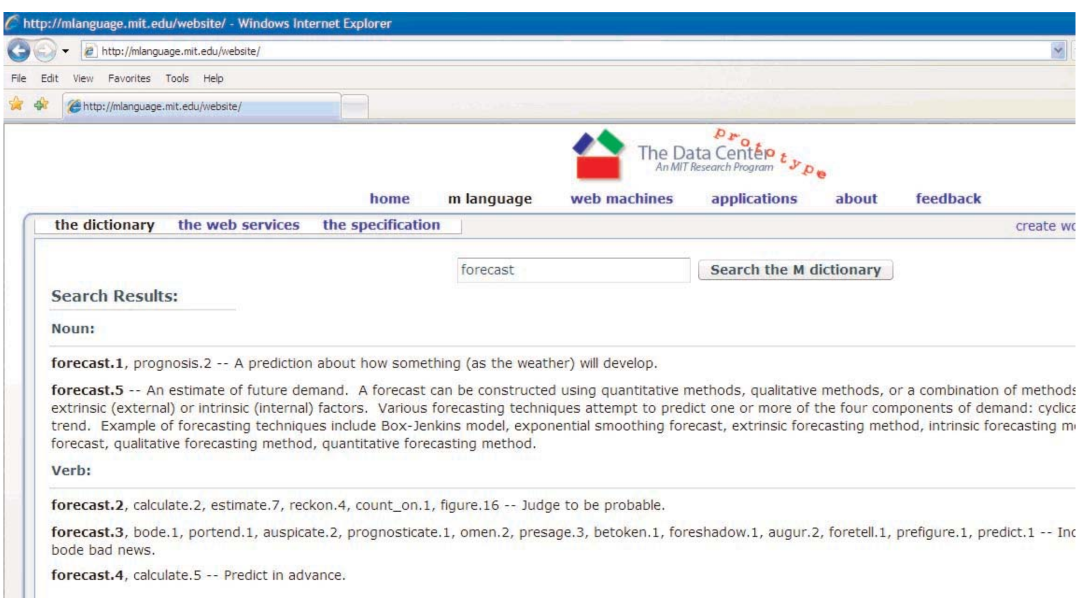


Figure 2 — An example from the M Dictionary.

Next, a dialog box appears, asking the user whether it is OK to clear the model outputs in preparation for a new run of MODS. Clicking yes clears these cell values so that the user can see new values appear as MODS runs. The spreadsheet is now open for entry of new data. The final step is to run the model by clicking RUN OSMPS.

Figure 3 shows part of the Excel model for 32 items spanning 52 weeks. In addition, words from the M Dictionary label various data fields within the spreadsheet. Clicking on an M word opens a yellow box that provides a detailed definition of the word.

For example, in Figure 3, the definition of `production_capacity.1` appears. For the OSMPS, nearly all of the M words take the form of phrases with an underscore separating each word and the numeric extension appearing at the end of the phrase. The ability to use phrases that have exact semantic meaning is the biggest advantage of using the M Dictionary. In many ways, this approach replaces paper and online-based dictionaries operated by many professional groups such as the American Production and Inventory Control Society (APICS). Unlike M Dictionary, none of the online dictionaries has the capability for machine-understandable semantics.

The relationship, or ontology, for the various M words used to describe the data inputs, outputs, and attributes of MODS appear in Figure 4. For instance, the `total_holding_cost.1` is a *type of* `holding_cost.1` (inventory carrying cost per unit, per time), which is also a *part of* `total_cost.1`, one of the outputs from a MODS run. Though this ontology is specific, it offers the flexibility to add words describing other, non-OSMPS approaches linked to `master_production_schedule.1`.

The M words used to describe data inputs, outputs, and attributes of MODS offer a great advantage in conducting enterprise searches. By placing M words into the "properties" of an Excel workbook, they are exposed to enterprise search tools. This becomes a metatag for the workbook with machine-understandable semantics.

LSSEA AND THE FUTURE OF THE INTERNET

LSSEA is an effective way to deliver mathematical models to users. Prototype industrial applications are in operation. Implementation costs are low, and the M Dictionary gives users a precise definition for data inputs, outputs, and attributes. Since most practitioners are familiar with spreadsheets, the interface is simple and allows for flexibility in customization. LSSEA has the potential to offer users a variety of models. Utilizing a semantically precise ontology located in the M Dictionary, LSSEA connects words and phrases in an exact way. This eliminates inconsistencies sometimes associated with OWL. The architecture allows for effective enterprise search and the capability to reason through ontological trees of relationships, forming a new base for business intelligence.

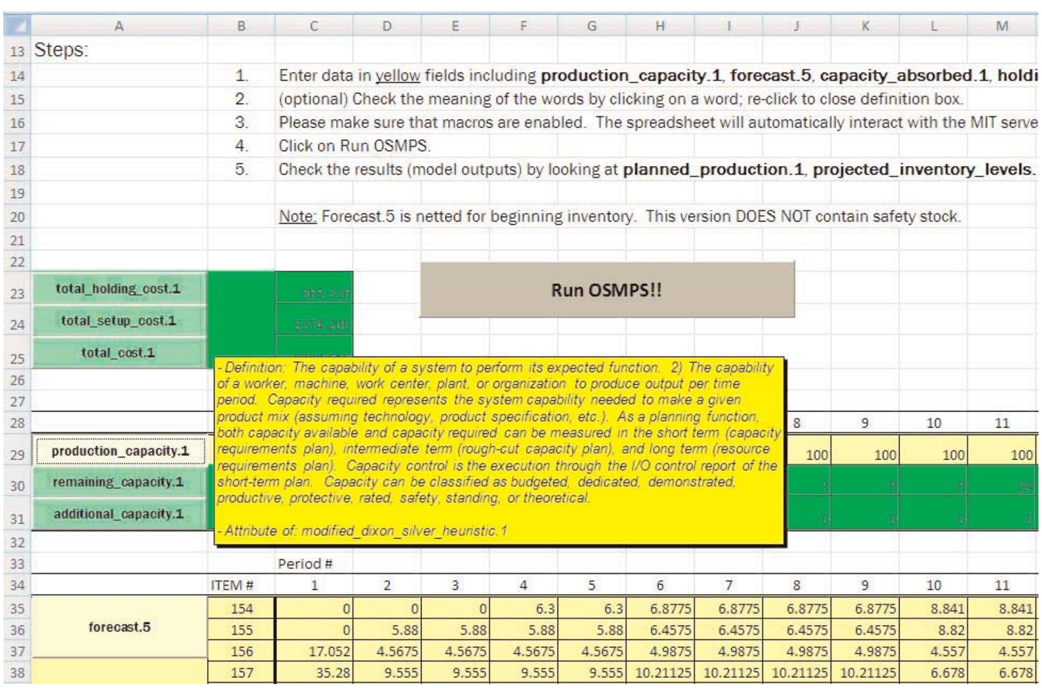


Figure 3 — The OSMPS Excel spreadsheet interface.

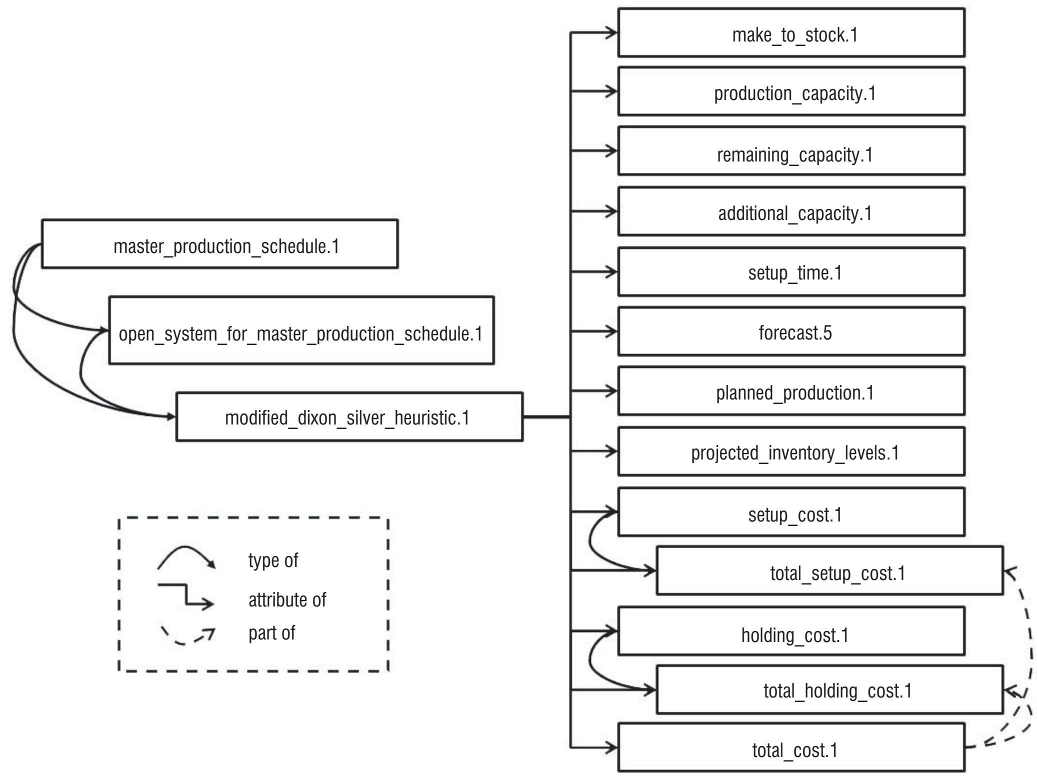


Figure 4 — The OSMPS-related ontology.

The rise of the semantic enterprise reflects powerful trends in Internet computing, which are perhaps only measured in decades. We envision three phases of Internet development, starting with the current state of affairs, the *Internet of Information*. Characterized as a "static repository of unstructured data," the present-day Internet requires substantial human interaction through Web browsers and search engines to overcome the lack of structure.[9]

With this technology as a base, the second phase began in 2003 and is best described as the *Internet of Things*. An attempt to connect physical objects to the Internet using RFID tags, the Internet of Things is essential to improving supply chain efficiency through unique identification. In addition, this technology is also the base for automation.

The final phase that we envision is the *Internet of Abstractions*. Yet to be realized, this phase involves the connection of data and mathematical models within the enterprise and across the Internet. The application of SWTs is a critical part of the evolution of the Internet of Abstractions along with the technology to connect mathematical models for decision making in a fluid way across the Internet. From our research, we very much believe this is the direction of the Internet; however, there is a great deal of work yet to do in achieving the beginning stages of phase three, the Internet of Abstractions.

Though LSSEA was originally an Internet-only application, it is likely to be a more effective tool at the enterprise level. The prospect of using M words in billions of Internet Web pages is not practical. However, it is possible to focus on specific industry applications involving multiple enterprises. For example, potential applications in agriculture are currently under consideration. Eventually, the M Dictionary will evolve from an Internet application to enterprise software. The application of SWTs to software approaches such as ERP systems represents the next frontier in information infrastructure needed for improved business productivity.

As a final note, we have developed other applications of the M Dictionary.[10] These involve employing sophisticated database technologies to make XML more effective in practice. The practical outcome is the rapid merging of weather data and surface observations of plant disease, creating a system to do spatial modeling in precision agriculture. Several commercial applications are moving forward in this area.

ENDNOTES

[1]Earlier versions of LSSEA were called the Open System for Master Production Scheduling (OSMPS).

[2]Allen, Stuart J., Jack L. Martin, and Edmund W. Schuster. "A Simple Method for the Multi-item, Single-level, Capacitated Scheduling Problem with Setup Times and Costs." *Production and Inventory Management Journal*, Vol. 38, No. 4, 1997, pp. 39-47.

[3]Brock, David L., Edmund W. Schuster, Stuart J. Allen, and Pinaki Kar. "An Introduction to Semantic Modeling for Logistical Systems." *Journal of Business Logistics,* Vol. 26, No. 2, 2005, pp. 97-117.

[4]Brock, David L., Edmund W. Schuster, and Timothy J. Kutz, Sr. *An Overview of the M Language* (MIT-DATACENTER-WH-09). MIT Data Center, 2006.

[5]Schuster, Edmund W., Stuart J. Allen, and David L. Brock. *Global RFID: The Value of the EPCglobal Network for Supply Chain Management*. Springer, 2007.

[6]Chakravorty, Satya S., Douglas. N. Hales, and James I. Herbert. "How Problem-Solving Really Works." *International Journal of Data Analysis Techniques and Strategies*, Vol. 1, No. 1, 21 August 2008, pp. 44-59.

[7]Turner, Mark, David Budgen, and Pearl Brereton. "Turning Software into a Service." *IEEE Computer*, Vol. 36, No. 10, October 2003, pp. 38-44.

[8]Lee, Hyoung-Gon, and Edmund W. Schuster. "MRP-as-a-Service: An Alternative Way to Execute Models for Business Processes." Cutter Consortium Business Technology Trends & Impacts *E-Mail Advisor*, 29 November 2007.

[9]Fensel, Dieter, James Hendler, Henry Lieberman, and Wolfgang Wahlster (eds.). *Spinning the Semantic Web*. MIT Press, 2003.

[10]Three commercial software licenses are available through the MIT Technology Licensing Office: Lee-Schuster Semantic Enterprise Architecture (LSSEA), Modified Dixon-Silver Heuristic (MODS), and Kratulos (semantics and syntax for XML expression).

Hyoung-Gon Lee is a postdoctoral associate at the MIT Laboratory for Manufacturing and Productivity, cosponsored by uCIC (u-Computing Innovation Center), Seoul National University, Korea. Dr. Lee holds a bachelor's degree and PhD from Seoul National University in industrial engineering, specializing in bringing computational advances to manufacturing such systems as MRP and APS. At MIT, he is currently working on next-generation applications of M Language, an outcome of Auto-ID Technology. Dr. Lee holds several software licenses through MIT. He can be reached at hg_lee@mit.edu.

Edmund W. Schuster is a research engineer at the MIT Laboratory for Manufacturing and Productivity. He is coauthor of Global RFID: The Value of the EPCglobal Network for Supply Chain Management*, a popular book about RFID. Mr. Schuster has also published more than 100 articles on the application of management science. Before coming to MIT, he worked as Corporate Manager of Operations Planning at Welch Foods, and he has been an active member in professional societies, including the American Production and Inventory Control Society (APICS). Mr. Schuster can be reached at edmund_w@mit.edu.*

CUTTER CONSORTIUM

ACCESS TO THE EXPERTS

Cutter’s Inquiry Program

Get Guidance and Answers from the Best Minds in the Business

Cutter’s Access to the Experts Inquiry Program, available to Cutter Consortium Enterprise Suite clients, provides you with astute analysis and advice directly from Cutter’s renowned experts. Every inquiry is fielded by a Cutter Senior Consultant or Fellow with hands-on expertise addressing IT challenges at organizations worldwide. You’ll get insightful answers to your questions from the pros, based on real-world experience, which is markedly different from what other analyst firms provide.

Answers You Can Put Into Action

Clients can leverage the hours with Cutter’s experts in many ways. Allot some of the inquiry time to the CIO, pairing him or her with a top Cutter business-IT strategist for monthly brainstorming sessions. Or set up monthly calls for your Director of Enterprise Architecture with one of Cutter’s EA experts. In either case, the arrangement allows the two of them to build a rapport and gives the Cutter expert a growing understanding of the issues your enterprise is facing. The Cutter expert quickly becomes a valuable advisor. You can also submit e-mail inquiries to Cutter. A Cutter Senior Consultant or Fellow with relevant hands-on expertise will answer your question directly, either in writing, by phone, or both ways if that better meets your needs.

From advice on your governance strategy from Cutter’s Enterprise Risk Management & Governance expert Bob Charette, to counsel on integrating agile approaches into your existing project management methodology from agile project management expert Jim Highsmith; counsel on how to spread the culture of innovation throughout your enterprise from innovation expert Rob Austin; input on creating and deploying your enterprise architecture from EA expert Mike Rosen; advice about how to protect the personally identifiable information your organization collects from privacy guru Rebecca Herold; to advice on how to uncover IT cost reduction opportunities from IT strategy expert Bob Benson, Cutter’s team of experts is uniquely positioned to provide you with solutions.

While some inquiries may be answered informally by our experts, others will require research and further conversation in order to provide the best advice. But in each and every case, you’ll get one of the best minds in the industry working on your problem right away.

Your Client Research Manager

Every Access to the Experts client organization is assigned a client research manager who will field your inquiries, matching the right expert to your inquiry. Your client research manager ensures that your questions are answered promptly and he or she handles the logistics for teleconferences. We dedicate the time to coordinating the communication, so you can concentrate on getting the answers you need.

Rob Austin | Ron Blitstein | Christine Davis | Tom DeMarco | Lynne Ellyn | Tim Lister | Lou Mazzucchelli | Ken Orr | Mark Seiden | Ed Yourdon

Cutter Business Technology Council

Here's a sampling of experts you can call upon with Cutter Consortium's Access to the Experts inquiry program

- Stephen Andriole
- Sanjiv Augustine
- Rob Austin
- Sam Bayer
- Bob Benson
- David J. Caruso
- Jens Coldewey
- Ken Collier
- Sara Cullen
- Tom DeMarco
- Khaled El Eman
- Michael Enright
- Clive Finkelstein
- Israel Gat
- Curt Hall
- David Hay
- Rebecca Herold
- Jim Highsmith
- Jeff Kaplan
- Vince Kellen
- Josh Kerievsky
- Tim Lister
- Michael Mah
- Jason Matthews
- Lou Mazzucchelli
- James Odell
- Ken Orr
- Bart Perkins
- David Rasmussen
- Ken Rau
- Suzanne Robertson
- Mike Rosen
- Oliver Sims
- William Ulrich
- Jim Watson
- Karl Wiig
- William Zucker

... and others

Access to the Experts: Case Studies

Strategic Advice on Opportunities Opened with Mobile and Wireless Technology

Senior Consultant Michael Enright worked with a *Fortune* 500 automotive company to identify future trends in mobile and wireless technology. The company was exploring ways to create new and innovative embedded systems as well as ways to make its workforce more mobile and, therefore, more agile. Mr. Enright served as the expert voice on an internal mobility R&D department created to help the company use mobile technology as a strategic differentiator.

CIO Mentoring

Cutter Fellow Rob Austin and the CIO of a *Fortune* 500 and *Financial Times* Global 500 firm meet monthly by phone to discuss the latest challenges IT confronts due to the organization's rapidly changing and mission-critical industry.

Business Requirements Engineering Assessment

Cutter Fellow Steve Andriole worked with a *Fortune* 500 energy company to analyze and revamp its business requirements engineering process. The company struggled with determining the right requirements and how to communicate them properly. Dr. Andriole helped develop a roadmap to a more robust and highly developed requirements process.

Protecting Personally Identifiable Information

Senior Consultant Rebecca Herold helped a *Fortune* 500 financial institution assess its capability to protect personally identifiable information and its current incident-monitoring system to evaluate how breaches were monitored, reported, and resolved. Ms. Herold provided valuable insight and advice on how to make the system more accurate and responsive in order to cut down on data breaches and information vulnerabilities.

Understanding the Drivers for PC Replacement

Cutter Fellow Lou Mazzucchelli helped a biomedical firm understand industry best practices and the business case for proactive PC refresh. He described the typical lifespan of laptop and desktop PCs and helped the organization understand the impact that hardware, operating system, and software upgrades have on replacement cycles.

For More Information

For details on Cutter Consortium's Access to the Experts inquiry program, contact your account manager directly, or call +1 781 648 8700 or send e-mail to sales@cutter.com.

North American Sales:

Vice President of Sales:	Jack Wainwright	+1 781 641 5122	jwainwright@cutter.com
New England:	Chris Getz	+1 781 641 5121	cgetz@cutter.com
Mid Atlantic/Southeast:	Jack Wainwright	+1 781 641 5122	jwainwright@cutter.com
Mid West/Canada:	Victoria Kyle	+1 905 852 4449	vkyle@cutter.com
Southwest:	Rob Curtis	+1 781 641 5124	rcurtis@cutter.com
West Coast:	Dennis Crowley	+1 781 641 5125	dcrowley@cutter.com
Client Research Manager:	Mark Rosania	+1 781 641 5120	mrosania@cutter.com

Regional Offices:

Europe, Mideast, Africa:	Jonathan Durrant, VP	+44(0)1252 40 4016	jdurrant@cutter.com
Eastern Europe:	Marta Jakubiuk	+22 534 74 00-01	mjakubiuk@ivmx.pl
Latin America:	Cuitláhuac Osorio	+52(55)554 48253	cosorio@cutter.com
Israel:	Jimmy Schwarzkopf	+972(0)974 44474	jschwarzkopf@cutter.com
Rest of the World:	Jack Wainwright	+1 781 641 5122	jwainwright@cutter.com

CUTTER CONSORTIUM
Access to the Experts

About Cutter Consortium

Cutter Consortium is a truly unique IT advisory firm, comprising a group of more than 100 internationally recognized experts who have come together to offer content, consulting, and training to our clients. These experts are committed to delivering top-level, critical, and objective advice. They have done, and are doing, groundbreaking work in organizations worldwide, helping companies deal with issues in the core areas of software development and agile project management, enterprise architecture, business technology trends and strategies, enterprise risk management, metrics, and sourcing.

Cutter offers a different value proposition than other IT research firms: We give you Access to the Experts. You get practitioners' points of view, derived from hands-on experience with the same critical issues you are facing, not the perspective of a desk-bound analyst who can only make predictions and observations on what's happening in the marketplace. With Cutter Consortium, you get the best practices and lessons learned from the world's leading experts, experts who are implementing these techniques at companies like yours right now.

Cutter's clients are able to tap into its expertise in a variety of formats, including content via online advisory services and journals, mentoring, workshops, training, and consulting. And by customizing our information products and training/consulting services, you get the solutions you need, while staying within your budget.

Cutter Consortium's philosophy is that there is no single right solution for all enterprises, or all departments within one enterprise, or even all projects within a department. Cutter believes that the complexity of the business technology issues confronting corporations today demands multiple detailed perspectives from which a company can view its opportunities and risks in order to make the right strategic and tactical decisions. The simplistic pronouncements other analyst firms make do not take into account the unique situation of each organization. This is another reason to present the several sides to each issue: to enable clients to determine the course of action that best fits their unique situation.

For more information, contact Cutter Consortium at +1 781 648 8700 or sales@cutter.com.

The Cutter Business Technology Council

The Cutter Business Technology Council was established by Cutter Consortium to help spot emerging trends in IT, digital technology, and the marketplace. Its members are IT specialists whose ideas have become important building blocks of today's wide-band, digitally connected, global economy. This brain trust includes:

- Rob Austin
- Ron Blitstein
- Christine Davis
- Tom DeMarco
- Lynne Ellyn
- Tim Lister
- Lou Mazzucchelli
- Ken Orr
- Mark Seiden
- Ed Yourdon