

News: Data-Driven Enterprise--Slouching toward the Semantic Web

Greg Goth

As the networked world extends into ever-more-remote digital crannies, technology experts and enterprise leaders agree a data accessibility crisis is on the horizon. What they can't agree on is whether the Semantic Web (<http://www.w3.org/2001/sw/>)—the principle of linking machine-readable data contained in documents—is the answer to that problem.

The concept's potential is certainly attractive. "Suppose you could link Environmental Protection Agency data about the chemicals found in some location to a database of chemical effects on various genes to PubMed articles on those genes and the diseases that affect them," says James Hendler, director of the University of Maryland's Joint Institute for Knowledge Discovery (<http://zaphod.mindlab.umd.edu:16080/JIKD/>), and one of the Semantic Web's pioneers. "You might be able to realize that the rise in, say, Burkitt's Lymphoma was likely to be linked to the dumping of some chemical. In short, the ability to link texts to databases, to data sets, et cetera, could lead to information linking well beyond current capabilities."

Getting to that point won't be easy, even though the Semantic Web's most visible standard bearer is the knighted Tim Berners-Lee (<http://www.w3.org/People/Berners-Lee/>), the man who invented the World Wide Web. So the media still pays attention, even though industry analysts, corporate network architects, and frontline administrators can't decide whether the Semantic Web will turn out to be the networking equivalent of Ford's iconic Mustang or tragicomic Edsel.

Getting outside corporate boundaries

Many organizations are still trying to figure out how to build a mostly object-oriented Web services system, says Dave McComb, president and chairman of Semantic Arts, an enterprise software architecture consultancy. In explaining the Semantic Web's enterprise possibilities, McComb emphasizes the benefits of direct access to the data within applications rather than integrating the applications themselves.

"The Big Picture message is, whether you've acknowledged it or not yet, you're drowned with data, and you're putting more and more data online per year," he says. "And, as there are more and more expectations that enterprises must get outside their own boundaries, you can no longer say, 'I'm going to just do a query within my own system.'"

Outside their own boundaries, however, enterprises face two big problems, according to McComb. One is the absence of technology that supports deep data queries. "You can't run SQL and have it hit stuff all over the Web," he says. The second problem is the sheer volume of metadata to keep up with. "For a long while, we were just able to keep very large schemas in a few people's heads. We've been able to rely on that person who knew 'To do this query, you have to exclude this particular value because there's an asterisk in this field.' We're at that limit now. There just aren't enough smart-enough go-to guys."

R. Todd Stephens, director of the metadata services unit for the BellSouth regional telecommunications company, says corporate technologists have addressed much of the data deluge over the past several years through network application architectures and increased transport capacities. However, he says the next era of intercorporate communications will likely present problems like those McComb describes.

Stephens says technologists can design viable search mechanisms for a corporate architecture of 100,000 to 200,000 objects, but the next generation's networks will have to combine not only far-flung intranets but also external connections. "Then you go to this collaborative environment, 200,000 objects goes to 1 million," he says, "and all of a sudden your pipe doesn't work anymore."

So, if savvy technologists can see problems the Semantic Web might successfully address, why is it still in the gee-whiz phase? Just what might it take for IT executives to start investing big money in it?

Immature integration capabilities

Some industry observers still see the Semantic Web as a theoretician's dream. Enterprises are still trying to figure out how to integrate applications via Web services and service-oriented architectures, so asking them to begin drawing up Semantic Web-capable taxonomies anytime soon isn't realistic.

"From my perspective, this is deep research," says Anne Thomas Manes, vice president and research director at the Burton Group technology analysis firm. Manes says even the most established Semantic Web standard, the Resource Description Framework (<http://www.w3.org/TR/rdf-primer/>) for a metadata model, isn't yet ready for enterprise problems. "I have yet to find a real practical application for RDF," she says.

Hawaii-based software developer Seth Ladd, whose SemErgence blog (<http://www.picklematrix.net/>) follows Semantic Web technology developments, says RDF-based integration capabilities haven't matured enough to make enterprise technologists demand semantic solutions. The RDF metadata model describes resources in the form of an RDF *triple*, consisting of a subject-object-predicate expression that describes the resource, a trait or aspect about it, and a value. "There isn't a good large-scale deductive database capable of integrating with standard RDBMS packages like Oracle and able to reason across millions and millions of triples," Ladd says.

To enter the enterprise, Semantic Web technologies must integrate with the large relational databases that represent enormous company investments in tools, training, and technologies. They must also extend RDBMS functionality in ways that are easier and more efficient than traditional means.

According to Ladd, the knowledge sharing that would entice enterprises would allow relational databases to directly import ontologies and rules just as easily as they use standard SQL insert, update, and delete statements. "Ontologies can act as the glue between multiple information sources, providing a consistent view of all the information, augmenting it and filling in the gaps," he says. "To efficiently do this, however, requires direct integration with the data source because it is impossible to perform enterprise-wide reasoning in memory only. Much like why a relational database doesn't load up all tables into memory before doing a search, performing ontology-driven reasoning efficiently requires intelligent integration with the data on disc as well as in memory."

Despite this obstacle to widespread enterprise adoption, however, one major vendor, Oracle, has begun supporting RDF in its 10g database release. "It even supports rules,

including built-in RDF Schema rules, although only in a read-only, data-warehouse type of usage," Ladd says. "While not sufficient for OLTP [online transaction processing] type applications, Oracle's RDF support is an important first step at putting Semantic Web technologies into the enterprise."

BellSouth's Stephens likens the current state of Semantic Web adoption with HTML's early days, before tools for non-expert users became available. "The HTML environment didn't take off when it was stuck with a bunch of professors and technologists talking about it," he says. "It took off when Front Page made it easier for the average user to drag and drop and create their own documents. So RDF and OWL [the W3C's OWL Web Ontology language] and all these things are great, but until we have a tool as useful as FrontPage or OneNote, it isn't going to matter. The average IT person isn't going to do it."

Ladd suggests the best route for success will be to stop emphasizing what the Semantic Web might do in the future and work instead on practical steps to improve the architecture. For starters, he calls the current RDF XML syntax "awful"—hard to understand and incompatible with existing XML tools. "Creating a new XML serialization of an RDF graph, one that can slip seamlessly into existing XML processing pipelines, would do wonders to help RDF integrate with the rest of the document processing systems in use today," Ladd says.

Another practical issue is RDF's lack of a formal or standardized integration point with XHTML, the XML-derived markup language that allows processing to be done automatically via a standard XML library. While this integration point is not directly related to the goals of enterprise-level knowledge, Ladd says ignoring it limits RDF visibility. "It should be very easy to insert an RDF triple into an XHTML document," he says, "and until then, RDF adoption will be stalled. If more documents on the Web had RDF triples embedded in them, then awareness for RDF would increase, thus helping overall adoption."

Folksonomies: From bottom-up to button-down?

Corporate decision-makers will likely bide their time until semantic tools are more mature. The first widespread Semantic Web advances are likely to emerge in the bottom-up taxonomies that ad hoc communities create. However, whether any of the "folksonomy" movement's principles translate to the enterprise's more rigorous needs is still debatable.

"I think folksonomy and the stuff coming out of RDFS [RDF-Schema] and OWL have some very interesting potential relations," says Hendler, "but they're certainly not the same thing." At the corporate level, he says, the ontology side makes much more sense than the folksonomy side. "If I'm marking up pictures of my family to share with people who know us well, then folksonomy is going to work great. If I'm a biologist trying to find some scanning microscope image of a particular protein, I don't want to hear about protein powders. All folksonomy really is, is a new way of putting keywords on things and then eventually using Google again."

The corporate world includes spaces that go beyond Google, he says. "When you're looking for all the policies that affect the shipping of a particular part, the last thing you want is something that's guessing."

Conclusion

BellSouth's Stephens says the adoption of Semantic Web concepts might not proceed as planned at World Wide Web Consortium (W3C) workshops, but the business requirements of data proliferation will demand a solution related to them in some way.

"It's just an enormous problem with major corporations, but I think executives are becoming aware the collaborative solution is just becoming a dumping ground," Stephens says. "We're starting to ask whether there could there be a better solution, and whether that would be an ontology or taxonomy-based architecture."

Stephens says the nirvana of a quickly deployed Semantic Web is a "giant leap" that probably won't happen, but that shouldn't detract from concerted efforts to build "incremental things that work."

Ladd, too, hopes to see more ground-level work being done to bring the masses to semantic-type technologies; maybe not the heretofore pie-in-the-sky vision of a grand semantic architecture, but smaller pieces of data integration that will lead to gradual adoption of semantic building blocks. "Through combined efforts of enterprise software vendors delivering real products, and new marketing campaigns that stress what is possible now, the Semantic Web can deliver on some of its promises."

Cite this article: Greg Goth, "Data-Driven Enterprise: Slouching toward the Semantic Web," *IEEE Distributed Systems Online*, vol. 7, no. 3, 2006, art. no. 0306-o3006.