IDG.net

GENERAL MOTORS ■ AMERICAN AIRLINES ■ THE VANGUARD GROUP ■ AMAZON.COM ■ DOW CHEMICAL ■ FEDEX

AUGUST 15, 2002 · CIO.COM

# CIO

The Magazine for Information Executives

THE 15TH ANNUAL CIO-100 AWARDS




# THE INTEGRATION IMPERATIVE

The money's been spent. On CRM, ERP, supply chain.... But to unlock the value, you must connect all those systems, people and processes. These 100 enterprises have succeeded. Learn from them.

UNICEF ■ O'REILLY AUTOMOTIVE ■ WASTE MANAGEMENT ■ RAYTHEON ■ BEST BUY ■ NORTHROP GRUMMAN ■ SOLECTRON ■ EQUIFAX ■ NESTLÉ USA

ROADWAY EXPRESS ■ BOISE CASCADE OFFICE PRODUCTS ■ J.P. MORGAN CHASE ■ CHARLES SCHWAB ■ WAL-MART ■ WELLS FARGO ■ BARTON MALOW

UPS ■ KRISPY KREME DOUGHNUT ■ STATE STREET ■ METLIFE ■ HARRAH'S ■ STAPLES ■ PROCTER & GAMBLE ■ PIRELLI




ILLUSTRATION BY JOHN WEBER

# Doing It with Meaning

*Semantics tools promise a world of universally compatible data*

**BY JOHN EDWARDS**

**EARLY LAST YEAR,** Michael Dreiling faced a stomach-churning problem. The vice president of technology for Quadrem U.S., a Dallas-based global electronic marketplace serving the mining, minerals and metals industries, needed to find a way to seamlessly integrate data from more than 1,000 companies.

Traditional middleware products could take care of the nuts-and-bolts job of converting files spewed out in EDI, legacy data formats and various flavors of XML. What they couldn't do was discern the meanings contained within the files. To cure his data integration indigestion, Dreiling looked into a new type of middleware: semantics-based integration tools.

Like Dreiling, an increasing number of CIOs and their staffs are being asked to integrate data from multiple, dissimilar sources into an electronic marketplace hub, EAI platform or database. The need may arise because of a merger, an acquisition, a CRM effort, the building of a data warehouse or a company's participation in an e-marketplace. And getting all the various fields and rules to mesh seamlessly into a single location is no easy job.

**Semantics integration tools are considered by analysts and industry players to be a kind of middleware that thinks.**

To accurately map source fields to targets, it's important for the middleware to completely understand the full semantic meaning of each data source element and how it behaves over the entire scope of source data. "Essentially, you're looking for semantical equivalents," says Jess Thompson, a research director for the application integration and middleware strategies group with Gartner Research in Stamford, Conn.

Unfortunately, while conventional middleware and hard-wired solutions are generally good at connecting non-compatible systems—converting protocols and formats—they often fall flat when it comes to interpreting the meaning of specific information and then applying it to a new environment. While a standard middleware product

may know, for example, that Nov. 14, 2002, 11/14/2002 and 11-14-02 are all the same shipping dates, it may not be able to understand whether the "shipping date" means "available for shipping," "on the dock" or "released to carrier." That's where semantics-based integration tools step in. These products, while not perfect, aim to make sense out of ambiguous and potentially conflicting information, reducing the labor-intensive need to manually link and synchronize data. "Semantics-based middleware could eliminate the need for application-specific APIs, traditionally used in integration today," says Eric Austvold, a research director at AMR Research in Boston.

## Putting It Together

Semantics-based integration technology is being pioneered by a variety of software companies, including Contivo, Modulant Solutions, Network Inference and Unicorn Solutions. Semantics-based integration tools are considered by many analysts and industry players to be a kind of middleware that thinks. That isn't surprising, since many of the tools utilize a combination of natural language analysis, pattern recognition, artificial intelligence and other leading-edge cognitive technologies.

At the heart of most semantics-based integration products is a powerful engine that mediates conflicting meanings among disparate data sources. The technology is designed to eliminate the need to manually analyze and map each source's various meanings, and then to remap those meanings each time a new data format arrives.

Modulant's Contextia, for example, provides automated tools that capture the meaning, relationships and context of data elements, and then maps them all into reusable models. A "transformation engine" performs run-time data conversions between the source and target application. Proprietary technology ensures semantic preservation when data is transformed between formats. "This modeling and mapping approach requires no custom coding yet provides the capabilities to capture implicit information in the data," says Jeffrey T. Pollock, CTO for San Francisco-based Modulant.

**Modulant's Contextia provides automated tools that capture the meaning, relationships and context of data elements, and then maps them all into reusable models.**

Contivo, on the other hand, supplies a "thesaurus" that contains databases that allow any two interfaces to map to each other with little or no human intervention. Each time a map is created, the databases store the synonyms and rules associated with both source and target interfaces. Contivo then outputs a transformation code, which is used as a road map for information flows in a run-time environment.

Dave Hollander, CTO of Mountain View, Calif.-based Contivo, says that as semantics developers get a better understanding of how to mediate conflicting meanings among various data sources, semantics-based integration tools are becoming easier to use and highly intuitive. He notes that Contivo can even build bridges between human languages, an important consideration in today's increasingly globalized business world. "It knows that 'street' and 'strasse' are the same thing," he says.

Quadrem's Dreiling, who selected Contivo for use with his e-marketplace hub, says the technology has worked well, slickly translating between various data formats and deftly removing ambiguities. He notes that the technology has been a significant time-saver. "Otherwise, we would be manually mapping from scratch, and we would have a longer time frame to actually getting trading partners integrated," he says.

Although they have no direct involvement with Contivo, Quadrem's trading partners also benefit from the software. "They are typically surprised to find that we're able to facilitate the integration and make that communication mediation at a much easier and faster rate than they anticipate," says Dreiling.

While basking in its benefits, Dreiling remains realistic about the technology's limitations. He notes that semantics-based integration software isn't a magic bullet for an enterprise's data integration troubles. "[CIOs] need to understand and make the commitment for the training of their personnel and then have the discipline to deploy it."

Training and deployment issues remain important because, while semantics-based integration software can go a long way toward easing data integration woes, they're not quite perfect. That's because the current crop of semantics-based integration tools still require some degree of manual input or configuration. Humans also have to double-check the software's handiwork to see that it hasn't made any faulty assumptions. Fully automatic semantics-based integration would require no external fiddling. "It is definitely the right idea, but it's years ahead of any implementation on a broad scale,"

says AMR Research's Austvold. Michael Lees, founder and COO of Manchester, England-based Network Inference, is even more succinct: "Proper semantic integration is still in its infancy."

## Weaving the Semantic Web

Concurrent with the creation of semantics-based integration tools is an even more ambitious initiative: the development of a Semantic Web. Advocates, including the World Wide Web Consortium (W3C), are planning to string together a Web that not only links documents to each other but that also recognizes the meaning of data contained within those documents. While semantics-based integration tools are designed to work with various types of structured data, the Semantic Web aims to unify the unstructured information scattered across the wild Web. The ultimate goal is to transform the Web from a display-oriented publishing medium into an environment where information can be interpreted, exchanged and processed.

Developing the necessary tools, such as highly descriptive Web content meta-tags and techniques that will allow different programs to relate and share meta-data from various websites, is shaping up to be an immense task. Yet the potential payoff—a Web that acts like a single giant database—would be worth all the effort. "The buzz is starting to feel like the buzz at the start of the original World Wide Web," says Lees. "There are still a lot of unanswered questions and a lot of unproven ideas, but at the core is a technology that will change the Web."

**Combined with Web services applications, semantics-based integration technology could doom middleware as it's currently known.**

By making data understandable to all systems, a Semantic Web would dovetail nicely with Web services, the budding technology that allows incompatible applications to talk to each other. Adding Semantic Web support to Web services would open the door to new worlds of potentially meaningful data. "The whole idea of the Semantic Web is to make all information machine-processable," says Ramana Venkata, CTO and cofounder of Stratify, a Mountain View, Calif., company that produces software for organizing and managing unstructured data. "Think of Web services as a sort of down payment today for the Semantic Web of five years hence."

## The End of Middleware?

While everyone waits for the Semantic Web to arrive, current semantics-based integration tools are destined to become increasingly powerful and capable. Combined with Web services applications, the technology could doom middleware as it's currently known. "If the concept of semantics-based integration takes off, Web services, as well as standards like RosettaNet, will evolve and adapt the concept into their approaches to integration," says AMR Research's Austvold. Adds Venkata: "The limitations that exist today on Web services-based interactions will go away."

Such a development would lead to one of technology's Holy Grails: universally compatible data. "It would mean the elimination of application-specific knowledge needed to integrate enterprise applications together," says Austvold.

That day isn't here yet. But if semantics-based integration tools live up to their promise, it soon will be. ■

---

John Edwards is a freelance writer based in Gilbert, Ariz. He can be reached at *john@john-edwards.com*.

MODULANT™
The Interoperability Company

425 Brannan Street, Suite 200
San Francisco, CA 94107
415.281.4551
www.modulant.com

Reprinted by Reprint Services with permission from CXO Media, Inc.
For reorders call Reprint Services 651-582-3800. For subscription information call 818-734-1520.
225620