Cover Story

eAI JOURNAL
www.eaijournal.com

The Web Services Scandal

How Data Semantics Have Been Overlooked in Integration Solutions

By Jeffrey T. Pollock

If ever there were a technology that could generate trillions of dollars in savings worldwide while simultaneously creating new wealth through market expansion, it would be the technology that makes business information interoperable. Ask any technology-savvy professional. They'll agree that today, there's a significant gap between the vast potential of enterprise integration software solutions and what's actually achieved.

Let's call a spade a spade: Current enterprise integration technologies are failing. *Fortune 500* IT organizations list integration as their number-one challenge. Venture capitalists are pouring money into the sector at astounding rates. Integration failures can be traced to nearly every aborted industry exchange, value chain integration project, and government modernization program. Integration technology solutions have been too difficult, too expensive, too complex, and too rigid to accommodate the constant, inevitable changes in business environments.

The industry acknowledges the strong marketplace demand for improvement in enterprise integration technologies. Web services represent a great leap forward in capabilities, yet they're only incremental steps. The scandal about Web services is really part of a much larger integration scandal — the ongoing reliance on custom code at the core of systems interoperability solutions.

Web services technology, despite its potential benefits, is limited in its ability to work with randomly formatted, non-standard data or data not based on XML. It's not yet a ubiquitous solution. The Simple Object Access Protocol (SOAP) envelope is as far as the Application Program Interface (API) and protocol standards will get you, but it's what's inside the envelope that counts. Most of the world does not use ebXML or RossettaNet as their data dictionary.

No Business Semantics in Web Services

With Web services, there's an assumption that either every system using them will be speaking the exact same language and dialect or a translation service will have to be coded to enable systems to communicate well. Where's the value in that? If we're still forced to bind our messages together with code or agree to speak about everything in the same way, then we're still stuck with most of the same problems that have plagued the Enterprise Application Integration (EAI) market. Those include failed projects, misused and abused standards, and misunderstood vocabularies.

Steve Vinoski, chief architect at IONA, says it best, "While many claim that supporting dynamic discovery through systems such as UDDI is trivial, it's actually a difficult problem. It requires clients and services to share common definitions and understandings of the concepts (the ontology) involved in any Web service through which they hope to interact. Without a shared understanding of the concepts, metadata, and semantics associated with a particular Web service, a client cannot hope to dynamically interact with it in the correct manner."

Today, Web services promise to get us out of the vendor lock-in scenario and free us from the "integrating the integrators" problem. Ultimately, Web services are relying heavily on the standard vocabularies solution to the problem of disparate semantics. These vocabulary standards will no doubt be significant for some industries. But the notion that a single set of standards will enable businesses everywhere to speak precisely the same vocabulary is as misplaced as the idea that EDI solved the integration challenge 30 years ago. At best, the lessons we've learned from industry's work with EDI have shown us that:

- Defining standard vocabularies is difficult and time-consuming.
- Once defined, standards don't adapt well.
- People don't implement standards correctly anyway.

There's no effective way to accommodate variant business semantics within the Web services framework. That's not surprising because Web services were designed for a different purpose. In fact, these limitations are really a legacy of traditional business systems integration approaches.

No Business Semantics in Integration

Whether one considers business-to-business integration (B2Bi), EDI, Common Object Request Broker Architecture (CORBA), EAI tools, or data integration as the true integration technology, none of these solves the problem of incompatible business semantics. Anybody who's ever implemented any of these technologies knows that the incompatible business rules and data vocabularies get resolved via custom code. Custom-coded solutions can come disguised in many ways: feature-rich integrated development environments, Graphical User Interface (GUI)-driven drag-and-drop tools, or more traditional text editors with shortcut keys for compiling the code. But these coded and often point-to-point transformation routines always result in brittle structures that are expensive to implement and maintain.

These issues aren't just textbook material for computer science graduate students. Consider these quotes:

- "Lack of technologies and products to dynamically mediate discrepancies in business semantics will limit the adoption of advanced Web services for large public communities whose participants have disparate business processes."
  — Gartner Research
  February 28, 2002

- "Data definitions usually vary from company to company — or even within a single company. That's the bigger problem than being technically able to pass data from one system to the other."
  — *CIO*
  January 15, 2002.

- "XML is only the first step to ensuring that computers can communicate freely. XML is an alphabet for computers and, as everyone who travels in Europe knows, knowing the alphabet doesn't mean you can speak Italian or French."
  — *BusinessWeek*
  March 18, 2002

- "Trying to engage with too many partners too fast is one of the main reasons that so many online market makers have foundered. The transactions they had viewed as simple and routine actually involved many subtle distinctions in terminology and meaning."
  — *Harvard Business Review*
  October 2001

Covisint, the world's largest e-business exchange for the automotive industry, has confronted this problem directly. Bruce Parker, of Modulant, explained in a recent Forrester report that, "The reason for Covisint [to exist] is to solve the integration problems the OEMs created by spinning out Delphi and Visteon in the first place. It has to solve the data interchange issue."

Further, many of the ongoing problems with scaling Covisint's business center around the current need for "swivel chair integration" solutions, where highly paid engineers key in data from one system to another.

Today's technology challenges aren't in the movement or extraction of data from one system to another; they exist in the logical domain of semantic interoperability. The information interoperability issues of data meaning, terminology, and shared



vocabularies are pervasive, little understood, and full of popular misconceptions.

Myths About Semantics

- ***B2B or EAI solutions solve this problem***. Within any established EAI vendor tool suite, you'll find some type of GUI interface that generates code, usually in a point-to-point manner, that provides translation routines for how to move data from one message format to another. A recent Defense Logistics Agency study shows how the Total Cost of Ownership (TCO) for a particular implementation of a leading EAI tool provider's solution was dominated by ongoing operational and maintenance costs, not acquisition or implementation costs. This particular vendor cites the positive economics of the "information bus technology," but still requires point-to-point translation scripts to link messages together. In this study, covering more than three years of operation, more than 95 percent of the TCO was centered on maintenance and responding to process changes. This is not uncommon for any given EAI vendor — and we haven't even explored the 88 percent of integration projects that fail before launch!
- ***Web services solve this problem***. Not unless we all agree to use the same vocabulary, which is unlikely! Web services are a breakthrough way of designing and building the mechanisms to discover, bind to, and transport messages to collaborative systems in a loosely coupled manner. However, Web services don't resolve the problem of disparate terminology and vocabularies in a given exchange environment. Today, solutions such as Grand Central and Bowstreet (leading innovators in Web services delivery) require either agreement on semantic standards or custom code to translate formats and data. Most of the Web services community is aligning around ebXML for the advanced business process capabilities and standard vocabularies embedded in the specification. But will a single standard or vocabulary ever unify all industries?
- ***Standard vocabularies solve this problem***. If 30 years of standards development in EDI, CORBA, and XML have taught us anything, it's that standards are a painfully slow way to develop common vocabularies, they're not easy to change, and they're certainly misused and abused. These issues aren't likely to go away. The ebXML and RossettaNet standards were helpful in creating business process standards for certain vertical markets and establishing meta process models that can be used as an overall framework (OMG). However, nothing points to near-term, significant unity on the vocabularies that real businesses need to use.
- ***Interoperability and integration are the same thing***. The words integration and interoperability seem to be used interchangeably, but they're different concepts. When you use your computer to edit a letter, you're interoperating with it. Interoperability is about maintaining the autonomy of the participating members of an exchange community, allowing them to share and use information intelligently while maintaining their own vocabularies, computing environments, and general perspective on the data. The capabilities for true interoperability are only now becoming a reality.

Semantics Are Messy!

Most problems contributing to the high failure rates of integration projects aren't technical in nature. The human activities involved in capturing the right kinds of expertise about the information, design intent, business rules, and usage of the data that needs to be interoperable have been core contributors to many of these failures. The issues surround how, and to what degree, proper analysis occurs before attempting to link multiple information systems together. Too many project managers, architects, and business people assume that with enough technology, enough business experts, and enough money, the problem can be solved. The assumption is that, somehow, the work of tying together all this disparate information across businesses, computer systems, company cultures, and international boundaries can be accomplished by putting enough smart people in a room together. Wrong.

The logical issues involved in managing information — how it's used, how it's communicated, and what it means — are the human elements of the equation. This is where databases, EDI messages, and XML documents fail to deliver information about what things mean and how they should be used. This is the messy part. If the vast majority of enterprise software maintenance costs, failed projects, and budget overruns are related to undefined human jargon, terminology, and localized rules on a per-project basis, how could a technology-based solution ever solve these issues? Actually, it turns out that lots of smart people have been thinking about these problems for quite some time now.

Emerging Solutions

Tim Berners-Lee and Marvin Minsky are just two of the many visionaries who have confronted the issues of data semantics in their standards groups and academic efforts. Today, the semantic Web activity within the W3C is viewed as some of the most interesting and visionary work to emerge in years. Additionally, dozens of leading-edge companies worldwide are building semantics-based tools for solving enterprise middleware challenges. It's this momentum, coupled with the fact that these solutions are becoming available today, which makes the space so exciting.

Semantics-based enterprise information interoperability solutions are an emerging category of tools focused on solving the problems of disjointed vocabularies, data definitions, terminology, and world-views of enterprise IT systems. Semantic interoperability solutions let engineers and business people build loosely coupled information webs that can dynamically enable rich, semantically precise collaboration and understanding — without the need for custom code or standardized vocabularies. Information interoperability solves a completely different set of problems than either Web services or B2Bi and EAI products. It focuses on the logical information infrastructure of an enterprise rather than the physical connectivity and routing infrastructure.

The roots of semantics-based interoperability technologies are also completely unrelated to EAI and Web services technologies. Semantic technologies derive from parts of the work done in the Artificial Intelligence (AI) community in the early 1970s. At that time, computer scientists were grappling with issues of knowledge representation in digital systems, attempting to solve the problem of how to digitally represent human knowledge. Today's typical semantic interoperability solutions represent a practical application of some of Berners-Lee's ideas embodied in the Semantic Web proposals of the W3C. They have three key characteristics reminiscent of old school AI technologies:

- **Semantic mediation** — Information interoperability solutions use an ontology — a model that makes concepts explicit — as a mediation layer to abstract particular data terms, vocabularies, and information into a shareable, distributable model. This is akin to some of the old notions about creating a "model-driven enterprise," using core information models as a lens to refract enterprise data in whatever form required.
- **Semantic mapping** — Ontology is only as good as the quality of the map that associates the enterprise data to it. Mapping to ontology preserves the native semantics of the data and eliminates the need for custom code. In information interoperability solutions, mapping accounts for much more than simple many-to-many data formatting rules or data syntax arbitrations — it's how the semantics are captured, aligned, and structured in relation to the data itself, thereby creating information out of "dumb" data repositories.
- **Context sensitivity** — As linguists know, the meaning of any data is always bound to a particular perspective or context. Thus, any information interoperability solution set must accommodate the fact that the same data can mean many different things from different viewpoints. Typically, the business rules, context definitions, and environmental metadata are captured and stored during the mapping process, making them reusable in any run-time server process.

Technology leaders have reached the same conclusion: Ontology-based semantic mediation of disparate information resources is the most viable alternative to today's inefficient, inflexible solutions involving custom code and monolithic standard vocabularies. Although data semantics have been overlooked for far too long, people are now mobilizing worldwide to confront the semantics issue.

Exposing the Scandal

Integration technologies stand to benefit greatly from applying the semantic Web approach to enterprise software. Imagine if a Web services exchange could provide not only Universal Description, Discovery, and Integration (UDDI) discovery of location, but also of vocabulary, concepts, and of context maps that provide rules for how data can move between vocabularies. Imagine if Web Services Description Language (WSDL) provided descriptions of those ontologies and maps, so that any requesting system could check to see if it shared a point of reference or could dynamically discover maps that enabled rich communication. Finally, as with all Web services, XML, SOAP, and HyperText Transfer Protocol (HTTP) could be the standard envelope and exchange protocol but the message contents could be in any of a variety of formats. You would still be able to discover, evaluate, and begin to use these services dynamically.

This is the promise of Web services and their value, but today's Web services scandal is that:

- The hype around dynamic discovery is overblown.
- Vendors aren't focused on the semantic element of the problem.
- One wonders if the technology's evolution is not misdirected.

The controversial steps that the W3C is taking with the semantic Web proposals are visionary and addressing a real problem in the enterprise integration software space. A complete middleware solution that's both loosely coupled and inexpensive to deploy and maintain must have at least three core elements:

- A sophisticated, loosely coupled physical infrastructure — Web services
- A sophisticated, loosely coupled information infrastructure — semantic interoperability
- Quality analysis and planning — the human element.



All serious, difficult problems — including interoperability and integration — are human problems. Ultimately, the disjointed vocabularies of disparate computer systems are the results of differing cultures, ideas, business processes, folklore, and viewpoints. A semantic interoperability solution focuses on the needs of users and organizations that are at the core of any sophisticated information exchange.

So even in jest, as we discuss the scandal and its unwitting participants, it should be clear that a call to action about the large-scale problems involved with integration is needed. As waves of new integration technologies such as Web services are hoisted upon IT professionals and positioned as the next great thing, we should remain level-headed about how far connectivity and message-routing software can get us. Trillions of dollars are spent annually on the human power that imbues meaning into brittle technical algorithms that encode enterprise information. The power to formalize that effort into reusable, strong building blocks is well within reach. eAI

About the Author



Jeffrey T. Pollock is the vice president of technology strategy at San Francisco-based Modulant Solutions. He is responsible for the technology vision and direction of Modulant's semantics-based middleware platform. e-Mail: jtpollock@Modulant.com; Website: www.modulant.com.