Integration's Dirty Little Secret: It's a Matter of Semantics



JEFFREY T. POLLOCK
FEBRUARY 2002

MODULANT™
The Interoperability Company

WHAT'S GOOD NEED NOT BE SECRET, AND WHAT'S SECRET IS NOT GOOD

- Chinese Proverb

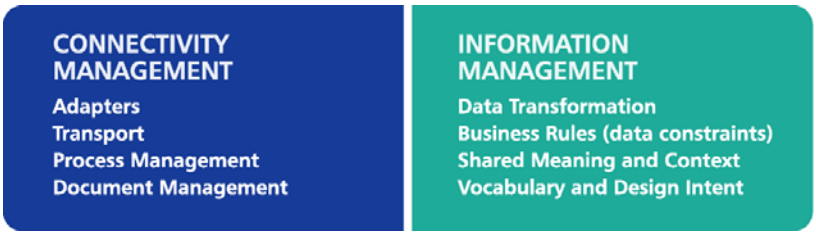
Standing at the counter of a major U.S. airline in Frankfurt, Germany trying to get my e-ticket transferred to one of its European partner airlines, I was hugely disappointed. Not only had the U.S. carrier cancelled my flight, but they were also instructing me on the finer points of booking another flight on a different airline. I had erroneously assumed that since both airlines are members of a large global alliance, it would to be relatively straightforward for them to transfer my ticket to a partner airline and have it show up in their system. Unfortunately, this was not the case. I had to physically take my printed ticket and receipt from the U.S. carrier's counter, walk over to the partner carrier's counter, and stand in yet another line to wait for the appropriate re-booking to be made. The irony hit me as I was walking to the other counter. What irony you ask? I was on my way, via Frankfurt, to an international conference on airline customer satisfaction!

I was set to meet up with some movers and shakers in the airline industry to discuss emerging solutions in the area of IT systems interoperability, so this situation did indeed provide timely fuel for the fire. You see, the ostensibly simple process of transferring e-tickets among partner airlines involves more than the mere transfer of seat reservations from one carrier to another. E-ticket transfers impact not only the passenger reservation systems of both airlines, but also their finance, membership rewards, and customer relationship software systems as well.

INTEGRATION CHALLENGES

The issues that arise while attempting to enable disparate computer systems to exchange information in a more efficient and effective manner continue to plague not only large, highly-visible enterprise alliances, but also most businesses across all industries.

First, there are the challenges relating to the actual physical connectivity and networking of the systems involved in a given exchange – the concerns of data gateways over ODBC, message-oriented-middleware, composite integration technology and other technical issues concerning software adapters, message transport services, and routing of the right messages to the right systems at the right times as dictated by business process rules. Second, there are the challenges of enabling each system to appropriately understand the information that is being shared – this is the logical, and non-technical, aspect of using and sharing data based on its intended meaning.



For example, the physical concerns of connecting software systems in the global airline alliance primarily involve the software protocols and tools that link together their various applications and enable them to exchange messages. These protocols and tools are the COBOL adapters and transport mechanisms that hook legacy software up to the data warehouses that contain information about how many miles each of us fly every year.

In contrast, the logical concerns involve understanding what the data in one reservation system actually means in that system, and how it relates to each and every one of the other alliance passenger systems. Addressing these **semantic** concerns involves discovering how information is used differently by each of the members, and how that information maps to the normative alliance view. Currently, most semantic interoperability issues are handled by either: (a) a common vocabulary of terms that each partner must adhere to when communicating to the group, and/or (b) custom-coded, point-to-point bridges that translate how one particular vocabulary is supposed to relate to the group's vocabulary or another carrier's vocabulary.

Integration vendors have generally solved the first challenge – namely, issues regarding physical connectivity and process control. The second challenge – ensuring that shared data is understood by receiving systems as intended – has not yet been efficiently addressed by most commercially available integration products, which still require cumbersome custom coding or rigid adoption of standard vocabularies to achieve information exchange. This lack of effective alternatives to custom coding or vocabulary standards has led to anemic success rates among integration projects today. Up to 88% of the integration projects among all industries end in failures involving overrun budgets, under-delivered functionality, or – in some cases – outright cancellation of entire projects. The issues surrounding these integration failures are fairly well understood by practitioners out in the field. Typically, it is a myriad of contributing factors unique to a given project, but some anti-patterns have certainly emerged throughout the years.

THE HUMAN ELEMENT

It may come as a bit of a surprise to learn that most problems contributing to the high failure rates of integration projects are not technical in nature. By this I mean that the ineffective human processes involved in capturing the

right kinds of expertise about the information, design intent, and usage of the data that needs to become interoperable have been a core contributor to many of these failures. In a nutshell, the issues really surround how, and to what degree, proper analysis occurs prior to 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.



"I don't know what you mean by 'glory,'" Alice said.

Humpty Dumpty smiled contemptuously. "Of course you don't—till I tell you. I meant 'there's a nice knock-down argument for you!'"

"But 'glory' doesn't mean 'a nice knock-down argument,'" objected Alice.

"When I use a word," Humpty Dumpty said, in a rather scornful tone, "it means just what I chose it to mean—neither more nor less."

Lewis Carroll, Through The Looking-Glass

The logical issues involved in managing information – how it is used, how it is communicated, and what it means – are the human elements of the equation, where databases, EDI messages and XML documents fail to deliver information about what things mean and how they should be used.

DATA MEANINGS AND INFORMATION

The issues surrounding data meanings and information are not just textbook material for computer science graduate students. Consider some of the following quotes:

"...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 Magazine, Jan 15, 2002

"...today's enterprises cannot readily exchange the necessary data on a consistent, reliable, normal course of business basis."

- Automotive Industry Action Group

"Trying to engage with too many partners too fast is one of the main reasons that so many on-line 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 eBusiness exchange for the automotive industry, has confronted this problem directly. Bruce Parker, of Modulant, as cited in a recent Forrester report, explains, "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 are not in the movement of data from one system to another, they exist in the logical domain of **information interoperability**. The information interoperability issues of data meaning, terminology, and shared vocabularies are pervasive, little understood, and full of popular misconceptions in the marketplace.

MISCONCEPTIONS ABOUT INFORMATION INTEROPERABILITY

1. EAI SOLUTIONS SOLVE THIS PROBLEM – If they did, I wouldn't be writing this article. When you get right down to the nuts and bolts of 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 study done by the Defense Logistics Agency shows how the total cost of ownership (TCO) for a particular implementation of a top-tier EAI tool provider's solution was dominated not by acquisition or implementation costs but by ongoing operational and maintenance costs. The vendor in this example, whom you've certainly heard about, expounds at length on the positive economics of the "information bus technology" – but still requires point-to-point translation scripts to link messages together. In this particular study, looking back over three years of operation, more than **95%** of the total cost of ownership was centered on maintenance and responding to process changes. This is not at all atypical for any given EAI vendor – and we haven't even explored the 88% of integration projects that fail before launch!

2. WEB SERVICES SOLVE THIS PROBLEM – Perhaps someday, but not unless we all agree to use the same vocabulary…fat chance! 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 does not addresses the problem of disparate terminology and vocabularies in a given exchange environment in any way whatsoever. Today, solutions such as Grand Central and Bowstreet (leading innovators in Web Services delivery) require either (a) agreement on semantic standards or (b) custom developed code to translate formats and data from one message to another. Most of the Web services community is aligning themselves around ebXML for both the advanced business process capabilities as well as the standard vocabularies embedded in the specification. But will a single standard or vocabulary ever unify all industries?

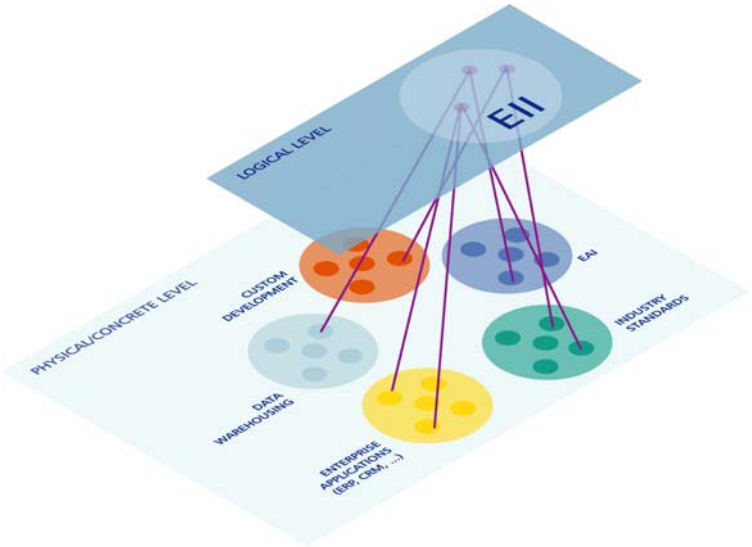
3. STANDARD VOCABULARIES SOLVE THIS PROBLEM – Yeah, and we all love EDI because we all want – and are able – to speak about everything in the same way. If thirty years of standards development in EDI, ASC x12, EDIFACT, CORBA, and XML have taught us anything, it's that (a) standards are a painfully slow way to develop common vocabularies, (b) they're not easy to change, and (c) once certified, they are certainly misused and abused. These issues are not likely to go away anytime soon. Recently, many great strides have been made by creating business process standards for certain vertical markets (ebXML and RossettaNet) and by establishing meta process models that can be used as an overall process framework (OMG), but nothing points to significant unity on the vocabularies that real businesses need to use on a day-to-day operational basis.

4. INTEROPERABILITY AND INTEGRATION ARE THE SAME THING – Have you ever tried to integrate with your computer? It wouldn't be pretty. The words integration and interoperability seem to be used interchangeably today, but they represent fundamentally different concepts. A simple example: when you use your computer to edit a letter, you are interoperating with it. Integrating with a computer conjures up images of Star Trek-like Borg who are "assimilated by the collective." From an IT perspective, 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.

A REAL SOLUTION FOR INFORMATION INTEROPERABILITY

Semantics-based information interoperability solutions are an emerging category of methodology and tools focused on solving the problem of disjointed vocabularies, data definitions, terminologies and world-views in enterprise IT systems. Information interoperability is the missing link in integration. It is the alternative to custom coded point-to-point rules and strictly enforced monolithic standardized data vocabularies. Information

interoperability is concerned with the transference of meaning and intent – the necessary ingredients for truly rich collaboration.



The benefits of a semantics-based information interoperability solution are direct corollaries to some of the problems we've already discussed. With an information interoperability solution in place you can expect to be able to non-invasively connect multiple information systems to each other while maintaining the complete logical autonomy of the information resources that you need to share. Use of an information interoperability approach also means that participants will not have to agree on standard vocabularies to exchange complex information about products, design materials, passenger data, or anything else. A semantics-based information interoperability solution creates highly flexible information frameworks that increase overall operational agility while driving down operational costs.

Information interoperability solutions have practical application in the most serious and important of IT systems. On that fateful day last September, several individuals identified as security threats in multiple local, state, federal, and international databases boarded airplanes departing from Boston International Airport, with tragic results. The prospect of sharing information among all levels of security databases worldwide, and also among all airline systems, is a daunting task – especially if solutions are limited to traditional integration technologies. A semantics-based information interoperability solution will allow multitudes of autonomous and secure IT systems to share relevant data in a loosely coupled and non-invasive manner, while preserving data meaning and field security levels. This type of information-centric solution means that no single integration vendor, no confusing spaghetti code, and no monolithic standardized vocabulary will stand in the way of a complete solution. Homeland security

in the United States, perhaps more than any other opportunity, stands to benefit from this emerging technology with deep roots.

The roots of semantics-based information interoperability technologies are completely unrelated to EAI technologies. Semantic technologies derive in part from work done in the artificial intelligence (AI) community dating back to the early 1970's. At that time, computer scientists were grappling with issues of knowledge representation in digital systems, attempting to solve the problem of how to represent human knowledge digitally. Today's typical semantics-based interoperability solutions have three key characteristics that are reminiscent of old school AI technologies, and share many common aspects with Tim Berners-Lee's efforts embodied in the Semantic Web proposals of the W3C.

1. SEMANTIC MEDIATION – Information interoperability solutions use an ontology – a model that makes concepts explicit – as a mediation layer in order to abstract particular data terms, vocabularies, and information into a sharable and 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.

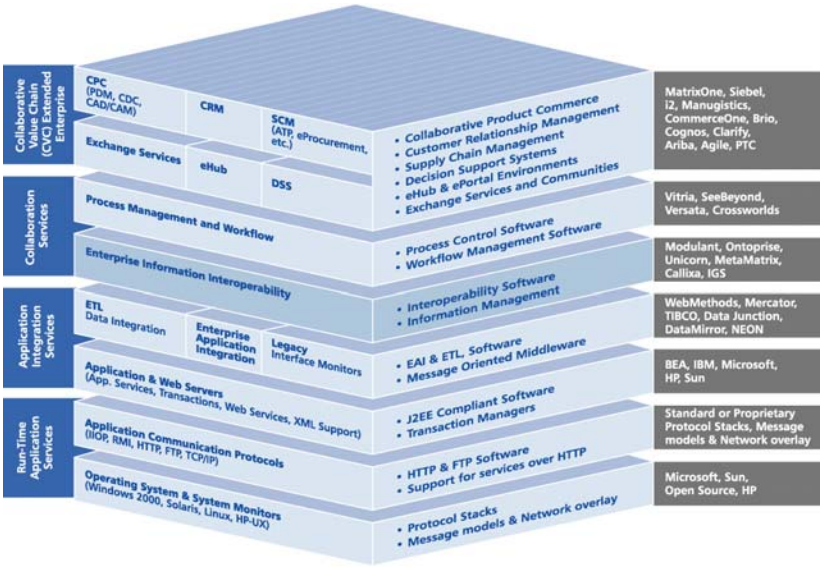
2. SEMANTIC MAPPING – An ontology is only as good as the quality of the map that associates the enterprise data to it. Mapping to an ontology preserves the native semantics of the data and eliminates the need for custom code. In information interoperability solutions, mapping accounts for much more than simple many-to-many data formatting rules or data syntax arbitrations – it is how the semantics are captured, aligned, and structured in relation to the data itself, thereby creating information out of "dumb" data repositories.

3. CONTEXT SENSITIVITY – As linguists know very well, 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 runtime server process.

Information interoperability software tools used in real-world environments vary by vendor, but typically they include at least a server-based process that can mediate real-time and batch-oriented messages and data exchanges. In addition, the design-time GUI tools will usually include a mapping tool that can build the semantic and contextual structures that link together ERP, PDM, EDI, XML and other data formats to the ontology. By focusing on the elements of interoperability that pertain to vocabularies, taxonomy, and data meanings, information interoperability solutions cross established boundaries of B2Bi and EAI solutions – working equally well for both internal and external information exchanges.

UNDERSTANDING THE BIG PICTURE

It is important to keep in mind that information interoperability technologies are a component of an overall solution, so don't look to them to address every EAI, XML, or other integration shortcoming. The business needs of interoperability still dictate that software adaptors, message transport services, and business process management techniques must be accounted for. This is exactly why the information interoperability space dovetails so well with Web services technology: whereas Web services creates a loosely-coupled framework for business process, connectivity, and discovery, information interoperability solutions create a loosely-coupled framework for the core information exchange that is required during the overall process. In addition, there are significant advantages to using an information interoperability solution as a complement to more traditional integration technologies such as EAI, ETL, and composite technologies (CORBA, EJB, and COM).



We also shouldn't look to information interoperability solutions to shorten implementation times of complex integration projects. The vast majority of time and resources spent on a typical integration implementation involve analysis – understanding business rules, understanding the data, and engaging the appropriate subject matter experts – and no current technology solution can speed up this fundamental human activity. Nevertheless, information interoperability solutions do offer a distinct advantage in that once the analysis has been done: rather than embedding all that knowledge in custom code or proprietary scripts, information interoperability solutions

allow the project to capture that knowledge in the semantic and context maps, thereby making the work reusable for all those who follow. The essential human expertise, rules, and data understandings will have been formalized in a way that is long-lived. Each additional system that joins an information interoperability community is incrementally cheaper to bring online – in stark contrast to the exponentially higher costs for adding or removing systems in more traditional integration solutions.

Partner airlines in the same alliance should be capable of easily exchanging all the necessary financial, passenger, loyalty and security information as members of their industry alliance. If this particular alliance asked all of its members to join an information interoperability community, the incremental costs of new participants joining would be significantly lower and the overall degree of collaboration would rise. In the information interoperability approach, a neutral ontology would be created that represents the conceptual ideas of all the information that may eventually be shared among participants. Then, each member would be able to non-invasively map their own data vocabularies to the shared abstract vocabulary – enabling the loose coupling of information sources so that no universal standards would be required and each member could maintain its complete autonomy.

USHERING IN THE AGE OF INTEROPERABILITY

Integration's dirty little secret is that current EAI and B2Bi technologies still require point-to-point custom code or standardized vocabularies for information exchange, leading to skyrocketing failure rates and long-term maintenance costs. Tomorrow's technologies, like Web services and enterprise information interoperability, may well lead to the holy grail of true interoperability, in a manner that is affordable and maintainable. By combining the strengths of loosely-coupled physical connectivity with loosely-coupled information interoperability, Web services and information interoperability technology will enable real, legitimate business transformation on a massive scale.

Technology visionaries all over the world have reached the same conclusion: ontology-based semantic mediation of disparate information resources is the most viable alternative to today's inefficient and inflexible solutions involving custom code and monolithic standards. Semantics-based technologies will be an essential part of all interoperability solutions in the very near future.

Modulant sums up the vision of semantics-based information interoperability with a simple phrase: the semantically connected world. The interoperability of business information assets, internally and externally, is a prerequisite for genuine IT-led business transformation. Historically, IT professionals have made tremendous progress clearing the hurdles of building network protocols, software adaptors, and business process tools; the next great leap will happen when the communication of information,

independent of disparate vocabulary and data meanings, becomes ubiquitous.

As information interoperability becomes more prevalent in the future, I will not have to wait in different lines to transfer e-tickets with partner airlines; I will be able to buy cars that have been designed and assembled from suppliers worldwide who used collaborative product commerce to speed design cycles and reduce costs; my tax dollars will not be wasted on outrageously expensive and recurring integration projects within the government; and I will know that the person sitting next me on my flight has had their profile checked against local, state, federal, and international security databases. The future looks bright indeed!

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. Jeff has successfully architected, designed, and built application server and middleware solutions for dozens of Fortune 500 companies. Prior to Modulant, Jeff managed IT projects as a Principal Engineer with Modem Media and worked on leading edge technology for several years at Ernst and Young's Center for Technology Enablement. Jeff has lectured at JavaOne and San Francisco State University, and he also instructs courses at the University of California at Berkeley Extension on the topics of software engineering, object-oriented analysis and development, and large-scale software architecture.

This document is protected by copyright and may not, in whole or in part, be copied, reproduced, translated, or reduced to any electronic medium or machine-readable form without prior written authorization of Modulant.

Modulant™, the Modulant logo, and Curtana™ are trademarks of Modulant Solutions, Inc. Other trademarks, service marks, trade names, and company logos referenced are the property of their respective owners.

RESTRICTED RIGHTS: Use, duplication, or disclosure by the U.S. Government is subject to restrictions of FAR 52.227-14(g) (6/87) and FAR 52.227-19 (6/87), or DFAR 252.227 7015(b) (6/95) and DFAR 227.7202-3(a).

Information in this document is subject to change without notice and does not represent a commitment on the part of Modulant. This document is provided “as is” and all express or implied conditions, representations, and warranties, including any implied warranty of merchantability, fitness for a particular purpose, or non-infringement, are disclaimed, except to the extent that such disclaimers are held to be legally invalid.



425 Brannan Street, Suite 200
San Francisco, CA 94107
Tel: 415.281.4500
Fax: 415.537.3550
info@modulant.com
www.modulant.com