A Semantic Web Business Case

Jeff Pollock, Oracle Corporation

on behalf of the W3C Semantic Web Education and Outreach interest group

Terms like "Semantic Web," "Web 3.0," and the "Data Web" have been interchangeably used to describe the underlying vision behind recently approved technology standards created by the World Wide Web Consortium (W3C). However catchy, none of these buzz words supply any hint about this new technology's ability to transform the foundation of enterprise software, empower radical new business capabilities, and throttle back IT spending in the notoriously expensive areas of data integration, master data management, and enterprise information management.

The Semantic Web is a fundamentally unique way of specifying data and data relationships. It is more declarative, more expressive, and more consistently repeatable than Java/C++, Relational Database Management Systems (RDBMS), and XML documents. It builds upon and preserves the conventional data models' respective strengths.

This Business Case will articulate why the Semantic Web will:

- Empower, directly and indirectly, new business capabilities
- Throttle back IT expenditures within medium and large businesses
- ...By transforming the foundation of enterprise software, and data integration in particular

We encourage the reader to take the following actions:

- invest in training and skills development now
- prototype a solution and explore the new tools now
- probe your software vendors about their semantic technology roadmap now
- compel your enterprise architects to formulate a multi-year metadata strategy now

By the end of this short paper, the reader should understand the overall superiority of Semantic Web technologies and be able to describe why it is very likely that they will be embedded in the fabric of nearly all data-intensive software within several years.

BACKGROUND

Enterprise software buyers and technology approvers make decisions that impact about $150 Billion [1] worth of mainstream software spending each year, primarily coming from medium and large businesses that typically have revenues greater than $50 million per year. Although the Semantic Web itself may have an impact on other software markets, like Consumer Software or embedded software for Electronics on other Devices, we are not exploring those market areas for this paper. For the purposes of this Business Case, we specifically look at the kinds of software procured within corporations for industrial use that may include:

- Infrastructure Software, including: RDBMS, Data Warehousing and Appliances, Enterprise Integration Technologies (SOA, Data Integration, Master Data Management, etc), Security
- Packaged Applications, including: Enterprise Resource Planning and Vertical Specialty Applications
- Information Management, including: Business Intelligence, Performance Management, Content Management, Master Data Management

These enterprise technologies are typically purchased in order to achieve specific program objectives that are driven by core business owners, including:

- Competitive Pressures — the cost of keeping up with overall industry improvements
- Executive Mandates — to fulfill new business initiatives mandated by sponsoring executives
- Cost Controls — to streamline outdated processes and generate new efficiencies
- Regulatory Demands — the requirement to meet corporate, local, state and federal governance
- Strategic Advantages — gaining business advantage through use of information; for example, in collaboration

But what factors impact a buying decision for enterprise software? Contrary to popular belief, the relative goodness or technical superiority of the software is rarely a decision-making advantage for the vendors. Likewise, the long-term strategic fit of the technology is usually not enough for a substantial enterprise software buy. Instead, the following selection criteria most frequently drive how large sums of money for enterprise software changes hands:

- Lowest Risk Option — where risk is calculated on the basis of overall fit and vendor reputation
- Tactical Fit — where the short-term requirements trump any long-term disadvantages
- Partner Choices — where the important ties between customer and vendor matter

Thus, although it may be a straightforward task to promote the Semantic Web technology stack on the basis of its technical and strategic

superiority for enterprise software, we must first start with building a business case that speaks towards the real buying pressures in the market: Low Risk, Tactical Fit, and Vendor Relations.

Only then can we explain why Semantic Web is a superior technical choice for many hard data problems.

LOW RISK SOFTWARE CHOICE

Let's be very direct: the Semantic Web technology seems like a riskier alternative to conventional data-centric technologies like RDBMS, basic XML, and programming based software approaches using UML and Java. For starters, the Semantic Web technology stack currently faces:

- Minimal large-vendor support for development tools
- Expensive hard-to-find skill sets to hire
- Few proven reference implementations in the public domain
- A very real paradigm shift in modeling, design, and declarative programming

But if we make a very small change of mindset, and shift the risk horizon to five years from now — the status quo technologies begin to look like the more risky option. For instance, software professionals know that when new software is developed on the basis of purely tactical decisions, the resulting chaotic architecture produces incomprehensible data silos that are much more costly and risky to handle in the long run. [2] Status quo is risky:

- Data Proliferation — in incompatible formats, multiple naming conventions, different applications, using different metadata
- Sensor (Instrument) Proliferation — creating more and more silos of data, faster, and with more expectations
- Complexity Explosion — data models, transformation rules, business rules, XML, UML, Java, etc.
- Executive Mandates — expectations on IT are higher than ever and becoming more demanding

In fact, new thinking about innovation shows how the Discounted Cash Flow (DCF) Trap [3] can distort conventional business risk assessments by incorrectly favoring do-nothing strategies. By shifting our time horizons, we begin to see that the limits of plain old Java, XML and RDBMS simply can't adapt quickly enough to the new world of enterprise software. Under the strain of 1000's of systems, try asking a Fortune 500 CIO to change a core business data definition; or ask them to produce a report that shows which enterprise software systems handle Purchase Order data. Seemingly simple tasks for a computer become unsolvable situations when the data is disconnected, inconsistently formatted and invisible to any sort of cohesive view.

Old technology itself is not the problem. The uncoordinated proliferation of old technology is the actual problem. And the "uncoordinated" part is a non-negotiable reality of 21st century big business.

Thus, while choosing only the traditional technology that seems low-risk today seems like a smart idea — it only takes a little foresight to stretch your risk horizon, avoid the DCF Trap, and agree that the data-management status quo is unsustainable at current rates of data proliferation and complexity.

Semantic Web technologies are lower-risk in the medium-term timescale, and in the short-term they are also most likely to become the roadmap for traditional data technologies — based on technical merits alone.

In terms of absolute risk, accounting for a long horizon and the DCF Trap, the low risk choice to make right now for an information-centric organization is to begin investing in Semantic Web technology as a foundation control point for disparate information.

TACTICAL SOFTWARE CHOICE

By definition, most enterprise software projects will have a simple tactical software solution available to them — the enterprise software industry itself has evolved to a sufficiently mature state that most software problems in most industries will have a specific vendor with a specific solution as at least one option. But now more than ever, these tactical solutions are seen for what they are — often a stop-gap series of temporary fixes that usually create new silos of disconnected data and rarely fit within an organization's strategic direction. Nonetheless, there is usually compelling business and financial motivation to choose a strong tactical enterprise software solution, where a top-priority pain-point can be temporarily fixed, even if the bigger technical problem remains unresolved.

In contrast, a Semantic Web based solution almost never looks like a tactical fit from the surface. But dig a little deeper, and more narrowly define the meaning of a "tactical fit" — Semantic Web technologies will look a lot more down-to-earth.

For instance, many enterprise software projects revolve around the notion of *information-centric operations*. When the tactical-fit of enterprise software depends on information-centricity, it is hard to beat the power of Semantic Web data specifications.

Tactical projects for a large information-centric organization might include:

- Data Warehousing and Business Intelligence
- Service-Oriented Architecture (SOA) Data Services
- Portal Applications and Data Mashups

- Data Integration, Replication, and Migration

Each of these tactical areas has both large and small vendor solutions servicing demand by using Semantic Web technologies. Although still a minority, the vendors using Semantic Web technology to supply tactical software solutions in these project areas would certainly expect to be measured against all the typical tactical metrics the industry has adopted.

In essence, when the buying organization is committed to information-centric operations and defines tactical success as measure of data flexibility, audit-ability, and reuse, the Semantic Web-based products will often be best-of-breed for those specific tactical needs. In particular cases, a vendor may or may not specify which technologies are being used, but may rather choose to market the benefits the technology provides. For example, Oracle, IBM and Software AG all leverage Semantic Web technologies in their SOA products, but you won't see them advertize the technology itself — only the features they provide.

So, contrary to popular belief, some Semantic Web technology can be very tactical in nature. And as is frequently observed, a little semantics goes a long way. [4]

PARTNERING CHOICE

Strong business relationships will trump other buying factors in most cases. Usually it is a function of risk: if you've been successful with a partner previously, you inherently trust them more, and view their suggestions as less risky than a new, unknown, vendor. In most cases, this is just simple, smart business.

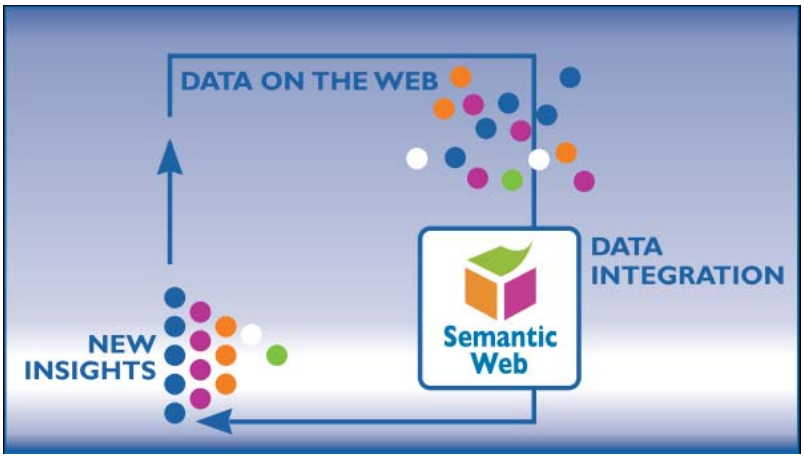
The Semantic Web cannot explicitly bolster any particular partnership choices, nor can the technology itself help buyers overcome any personal doubts about a particular vendor's employees.

However, it is worth noting that most large enterprise software vendors, and many small ones, have already begun to adopt Semantic Web technologies and embed them into their mainstream products. In fact, leading enterprise software vendors like HP, IBM, Microsoft, Oracle, SAP, and Software AG all currently provide Applications and Tools that support Semantic Web specifications.[5][6][7]

Ask your partners about their plans to adopt Semantic Web standards for metadata and data.

If your mainstream partners are unwilling or unable to articulate clear guidance about their roadmap for data and metadata management, there are many mid-size vendors who would appreciate your time and can give you details about the future of Semantic Web technology for enterprise software. [5][6]

Trusted relationships will usually lead to good business decisions, but in the realm of technology and data management your trusted advisors must be innovative as well as safe.



TECHNICAL SUPERIORITY

Data is different than information. In the context of software, information is data that references or is referenced by a computational model. That Information Model is a necessary, logically consistent interface for accessing data. These Information Models are always accompanied by metadata about the model itself. The Semantic Web specifications (RDF, OWL, SPARQL, GRDDL, SAWSDL) define consistent computational interfaces for enterprise software to declaratively interact with data.

Besides Semantic Web specifications, other computational metadata specifications for information models typically include:

- Entity Relationship Model and DDL Scripting — all RDBMS
- Meta Object Facility Models and Model Transformers — all UML compatible languages
- XML Infoset Model and Custom Program Implementations — all XML interchange

Databases, UML, and XML technologies constitute how the vast majority of enterprise software applications store and manage data today. But the Semantic Web presents a newer, more computationally powerful metadata specification [8] that can be as reliable as a database [9], as portable as XML [10], and as powerful as native programming logic [11].

The Semantic Web specifications, in particular RDF and OWL, are the only technology specifications that were purpose-built for use as a metadata language, entirely dedicated to describing and linking data of all sorts and at Web scale. Over 30 years ago, relational databases were conceived for the storage and consistently fast retrieval of data records. Over 15 years ago, UML was conceived as a unified approach for visually modeling structure software programs. Almost 10 years ago, XML was wrought from SGML as a way to give structure to documents and messages. Yet today, software developers routinely misuse XML, UML, and relational databases for purposes that they were not intended for.

Areas where RDBMS, UML, and XML technologies are misused, and where Semantic Web technologies excel, include the following:

- Specification of computationally sound business information models
- Specification of linking and relationship (meta)data across physical data locations
- Specification of dynamic structural logic and rules that are part of the data realm
- Specification of a federation approach for geographically separate data records

It is not that RDBMS, UML, and XML technology cannot be made to solve some of these technical challenges, it is that they must be contorted out of their sweet-spots to do so — and also that attempts to make them work have led to non-standard, one-off, vendor-implemented, heuristics-based solutions that have absolutely zero portability and therefore no chance at solving Enterprise scale information problems. Semantic Web specifications are the only purpose-built solution for large-scale metadata intensive data problems in enterprise software.

Being purpose-built for change is a particularly striking difference between Semantic Web technology and conventional data languages. Conventional approaches rely on static data models and complex query logic, which causes a type of software development lifecycle that favors the up-front specification of system behavior. But software developers can rarely envision how a given system will be used in practice many years from the point at which requirements were developed. In fact, application data will always be used in unanticipated ways.

The Semantic Web specifications are different because they provide for continually changing data models, inferred classification of data and taxonomy, and all the richness and power of a declarative query language.

Key Semantic Web specifications themselves were commissioned by US and European government agencies in the early 2000's [12][13] precisely because their defense research scientists knew that RDBMS, UML, and XML technology could not, by themselves, solve our information challenges of the next century. Even the standards bodies that control conventional data standards are selecting Semantic Web standards as a foundation for their own next-generation specifications. For instance:

- Object Management Group — controls UML and CWM specifications, is adopting RDF and OWL as the centerpiece specification for their core Definitional Metamodels [14]
- International Standards Organization — controls various EDI and Metadata specifications, is adopting RDF and OWL within several ISO specification families [15][16]
- World Wide Web Consortium — controls XML and SOA specifications, is adopting RDF and OWL as extensions to existing XML and Web Service specifications [17]
- OASIS — controls many vertically-oriented business data specifications, is adopting RDF and OWL as a core feature in standards for Documents, Data Centers, Security, and Business Process Management [18][19]

But global conglomorates and federal agencies are not idly waiting for the enterprise software vendors and standards bodies to supply the Semantic Web on a silver platter. Industrial use cases [7] are emerging from these end-user organizations that demonstrate both the necessity and power of the Semantic Web technical approach. Organizations are investing in this technology, in most cases, because there isn't a viable alternative that can address the size, scope or complexity of their legacy data problems.

It should be plain to see that the Semantic Web specifications provide a superior technical capability for information-intensive enterprise software problems, which have a high degree of dependence on metadata for operational reliability, portability, and dynamic behavior.

ENABLING FOUNDATION

Now that the technical case has been established for the Semantic Web's superiority in the realm of metadata and data management, and it is clear that it can be a safe low-risk and tactically-oriented solution that is well supported by traditional partners, we can turn now to the business benefits of the technology.

While the Semantic Web by itself cannot supply any magic revenue boost to an enterprise, it can indeed provide a means to rationalize incredibly complex information ecosystems. Without Semantic Web technologies, businesses and federal agencies must use conventional RDBMS, UML, and XML technology combined with liberal amounts of expensive manpower, in order to rein in and achieve a modicum of visibility into their enterprise information.

The Semantic Web can be tactically applied to the following projects:

- Data Integration, at the XML, RDBMS, and Object software tiers
- Data Warehousing and Business Intelligence
- Service-Oriented Architecture (SOA) Data Services
- IT Maintenance and IT Infrastructure Management
- Portal Applications and Data Mashups
- Data Replication, Migration and Transformation

And it can be strategically applied to the following business initiatives:

- Enterprise Information Management
- Enterprise Governance and Risk (including Policy Compliance)

Far from being a silver-bullet, the Semantic Web is itself an enabler. It enables smoother running systems as a result of better metadata; it enables less expensive manual efforts to keep disparate information linked up; and it enables much stronger capabilities for auditing, tracking, and defining actionable rules on top of shared enterprise data. The Semantic Web specifications will:

- Empower, directly and indirectly, new business capabilities because they enable stronger and more consistent metadata linking, automatic inference for dynamic data structures, and a more declarative foundation model for shared business information
- Throttle back IT expenditures within medium and large businesses via reduced head-count requirements for the management of enterprise information assets, decrease the long-term costs of integration, and simplify decentralized data architectures
- Transform the foundation of enterprise software as all major software vendors adopt Semantic Web specifications within the context of their own mainstream tools.

In short, the Semantic Web can help smash the silos of data that currently cost the enterprise time and money to make interoperable. Start training and planning for it now. Talk to your vendors about it now.

ACKNOWLEDGEMENTS

This paper is a product of the W3C Semantic Web Education & Outreach (SWEO) interest group. SWEO includes the following members:

- Melliyal Annamalai, Oracle Corporation
- Sami Bhiri, DERI Galway
- Dan Brickley, Asemantics S.R.L.
- John Davies, BT
- Martin Dzbor, The Open University
- Dunja Ewinger, REWERSE
- Steve Harris, Garlik
- Kingsley Idehen, OpenLink Software Inc.
- Jacek Kopecký, University of Innsbruck
- Alberto Mijares, Fundación CTIC
- Karen Myers, W3C
- Benjamin Nowack, semsol web semantics
- Paula-Lavinia Patranjan, REWERSE
- Pasquale Popolizio, International Webmasters Association / HTML Writers Guild (IWA-HWG)
- David Rooks, Segala
- Leo Sauermann, German Research Center for Artificial Intelligence (DFKI) Gmbh
- Elena Simperl, University of Innsbruck
- Susie Stephens, Eli Lilly and Company
- Michael Wilson, STFC (Science & Technology Facilities Council)
- Danny Ayers, Talis Information Limited
- Uldis Bojars, DERI Galway
- Frank Chum, Chevron
- Ian Davis, Talis Information Limited
- Orri Erling, OpenLink Software Inc.
- Lee Feigenbaum, Cambridge Semantics Inc.
- Ivan Herman, W3C
- Kjetil Kjernsmo, Computas AS
- Andrea Kulas, REWERSE
- Paul Miller, Talis Information Limited
- Pantelis Nasikas, Institute of Informatics & Telecommunications (IIT), NCSR
- Yue Pan, IBM
- Jeff Pollock, Oracle Corporation
- David Provost, Metatomix
- Charles Rudgard, Garlik
- Jeffrey Schiffel, The Boeing Company
- Anastasios Skarlatidis, Institute of Informatics & Telecommunications (IIT), NCSR
- Paul Walsh, Segala

REFERENCES

1. Ovum Analysts, http://store.ovum.com/Product.asp?pid=36856
2. NIST, Interoperability Cost Analysis of the US Automotive Supply Chain. 1999, http://www.nist.gov/director/prog-ofc/report99-1.pdf
3. Christensen et al; Innovation Killers, Harvard Business Review, 2008 http://harvardbusinessonline.hbsp.harvard.edu/hbsp/hbr/articles/article.jsp?articleID=R0801F&productId=R0801F&ml_action=ge

4. Jim Hendler & Ora Lassila: "Semantic Web @ 5 (years)" Semantic Technologies 2006, http://www.cs.umd.edu/~hendler/presentations/SemTech2006-keynote.pdf
5. Jeff Pollock & Susie Stephens: "Enterprise Semantic Web," Semantic Technologies 2007, http://me.jtpollock.us/pubs/2007.05-Pollock.STC.2007.pdf
6. Semantic Web Tools, http://esw.w3.org/topic/SemanticWebTools
7. Semantic Web Use Cases: http://www.w3.org/2001/sw/sweo/public/UseCases/
8. Deb McGuinness: "Question Answering on the Semantic Web," IEEE Intelligent Systems, Volume 19, No. 1, January/February, 2004
9. Ora Lassila: "The Resource Description Framework," IEEE Intelligent Systems, Volume 15(6): 67-69 November/December 2000
10. Berners-Lee, Testimony to the United States Congress, Hearing on the "Digital Future of the United States: Part I -- The Future of the World Wide Web", http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web.html
11. Horrocks at al; "From SHIQ and RDF to OWL: The Making of a Web Ontology Language." Journal of Web Semantics, 1(1):7-26, 2003, http://web.comlab.ox.ac.uk/oucl/work/ian.horrocks/Publications/download/2003/HoPH03a.pdf
12. OIL EC/IKT Commission, http://www.ontoknowledge.org/oil/misc.shtml#part
13. DAML DARPA Commission, http://www.daml.org/about.html
14. Object Management Group, "Ontology Definition Metamodel, OMG Adopted Specification" http://www.omg.org/docs/ptc/07-09-09.pdf
15. International Standards Organization, ISO/IEC 24752-5 "User Interface Resource Specification" http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=42313
16. International Standards Organization, "Industrial automation systems and integration—Integration of life-cycle data for process plants including oil and gas production facilities", http://en.wikipedia.org/wiki/ISO_15926
17. W3C GRDDL/SAWSDL citation, http://www.w3.org/2001/sw/grddl-wg/ and http://www.w3.org/2002/ws/sawsdl/
18. OASIS DCML/SAML citation, http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security
19. OASIS Open Document Format: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office

---