

# Enterprise-scale Semantic Systems

by Mike Bergman - Tuesday, January 08, 2013

<http://www.mkbergman.com/1037/enterprise-scale-semantic-systems/>



## Introduction: Part 1 of a New Series

For about the past two years, [Fred Giasson](#) and I have had the good fortune to work with some cutting-edge, reference enterprise deployments of semantic technologies. Our work at [Structured Dynamics](#) is about to see the light of day from these initiatives, which our sponsors will be unveiling shortly. These efforts in enterprise-scale systems have been eye opening. One eye has been opened with respect to how semantic technologies need to integrate and adapt to existing enterprise practices and deployments. The other eye has been opened with respect to how semantic technologies need to be presented and sold to internal enterprise stakeholders.

Our emerging series on enterprise-scale systems, which this first article introduces, attempts to package and share the lessons we have gained from these enterprise-scale deployments. This series marks a subtle -- but substantive -- shift from many of my prior writings. Those earlier writings over the past five to six years represent an attempt to introduce and describe some of the key underpinnings of semantic technologies and the mindsets behind them. Probably the best summary of these earlier messages resides in my article on the [Seven Pillars of the Open Semantic Enterprise](#), published nearly three years ago today.

But logic, theory and foundational descriptions can only go so far. Ultimately, if the constructs at the core of these semantic technologies are to be realized, the conceptual needs to be brought down to the practical. Those are the efforts that Fred and I have been pursuing for the past two years, and those are the efforts that this new series on enterprise-scale semantic systems attempts to capture.

## A Bit of History

The foundational underpinnings to the semantic Web -- upon which semantic enterprise technologies are based [\[1\]](#) -- extends back now nearly 12 to 15 years. Predecessor data models to RDF were being described in the late 1990s (actually, even earlier, but not within a Web framework), with the

foundational [Resource Description Framework](#) approach first promulgated as a standard by the World Wide Web Consortium ([W3C](#)) in 1999. The first social semantic Web vocabulary, [FOAF](#), was started in 2000. Schema extensions to RDF and then the Web Ontology Language ([OWL](#)) for formalizing semantic Web schema were first published in 2004. Many related efforts in supporting formats followed soon thereafter, including some of the baseline semantic Web vocabularies such as [SKOS](#). This decade or more of effort has now resulted in a rich set of vocabularies, standards and best practices.

Many in the community look to the techniques associated with [linked data](#) and the complementary project to make Wikipedia content accessible as structured data via [DBpedia](#) as the essential turning point for the semantic Web. The so-called linked open data (LOD) voice has become a prominent one within the community. While we embrace linked data techniques and believe them often to be best practices, our own experience indicates that linked data alone is not a key driver to enterprise adoption of semantic technologies. In our experience, linked data advocacy may be best characterized as neutral to negative in helping to foster enterprise semantic technology adoption.

At a consumer level, efforts from DBpedia to [Siri](#) and semantic search and various structured data initiatives are validating the use of semantic technologies as foundational elements to many current information architectures. The use of semantic knowledge graphs and graph-oriented databases (DBs) and structures is, for example, pervasive to many standard Web offerings from Google to Facebook. These adoptions tend to be incremental and subtle; looking at the functional and relational capabilities of major Web properties today in comparison with their same offerings of even a few years ago verifies these transitions.

But, generally, these semantic transitions are subtle and incremental. Semantic technologies are not expressing themselves as revolutionary new "killer apps" or in-your-face differences. Rather, they are background improvements that act to better inform how we find stuff and what relevant information gets presented to us.

Semantic technology advocates often appear to have been caught in a vise of their own making. On the one hand, "revolutionary" improvements in data access and management were promised, the idea that data would have an equivalent impact to the initial adoption of the Web. On the other hand, since such huge data management changes have not been glaringly evident, it is clear that semantic technologies have not achieved that vaunted potential. The advocate's strawman has not apparently appeared, and can not either be knocked down nor recognized.

The seeming "failure" of semantic technologies to achieve their advocated potential leads both to (sometimes) expressions of despair for why that "failure" has occurred as well as strident arguments for certain community-focused advocacies, such as linked data. From an enterprise perspective, most all of this seems quite parochial and beside the point. From the customer perspective, data integration and usefulness is the driving motivation, not linked data.

The "drivers" for semantic technologies in the enterprise remain the same as they have been for decades: better integration of existing and desired information assets at lower cost and with better insight for business purposes. These are the metrics of interest to enterprise decisionmakers, not the internal advocacy positions of linked data academics. We thus have the strange confluence of the market embracing and accepting semantic relationships within data while advocates perceive a lack of adoption.

All of this is occurring within the backdrop of software development shifting from the past few years of consumer prominence to the [re-emergence of enterprise uses](#). Though stupid and overstated, some have expressed this enterprise shift as a [trillion dollar opportunity](#). For sure, the opportunity is big, but no one believes the incumbent enterprise providers will be overcome easily and much enterprise stuff will be shared from the consumer side of things. But, in any event, the enterprise market opportunity remains compelling.

### Some Good News

Like all good market opportunities, there is much positive to discover once one scratches below the initial semantic surface. The same compelling needs for data integration and interoperability that have been a commonplace of the enterprise market for more than three decades remain today. Information use of all kinds within the enterprise sucks, and has now for many decades. The litany of information management failures in the enterprise extends from stovepiped data silos to lack of use of internal unstructured data (documents) and the virtual lack of integration with external information sources. It has taken the massive success of the Web and its distributed model of resources to make clear just how dysfunctional most enterprise information systems truly are.

Though certainly not yet evident to all or even most, it is clear that the promises in the logic, theory and foundational bases of semantic technologies are real and are relevant. The RDF data model works, as does the use of ontologies as governing schema. Natural language processing (NLP) techniques married to RDF have made unstructured documents equal first class citizens with structured data. Open world approaches are showing how schema and integration development can be incremental and cost effective, while overcoming past brittleness in how to organize and manage information. Web-oriented architectures are proving the same benefits to the enterprise as is shown on the broader public Web. These architectures and rather simple connectors or RDFizers are showing how legacy systems and assets can be leveraged in place to transition to a semantic future without undue cost or disruption to existing practices.

Daily we see success of semantic technologies in multiple locations, and the market is coming to understand the uses and potential benefits. The benefits of graph-based knowledge structures in search and recommendation systems are becoming accepted. We see how basic search is being enhanced with entity recognition and characterization, as well as richer links between entities. The ability of the RDF data model and ontologies to act as integration frameworks is no longer an assertion, but a fact. Despite the despair of some semantic advocates, the market place is increasingly understanding the concepts and potential of semantic technologies.

There is real and good money to be made in this marketplace. Fred and I turn away work and our company, Structured Dynamics, has been self-financed and profitable since its inception. Our revenues increase substantially each year and we have significant monies banked to protect us in a downturn or to fund our own initiatives. No one owns any portion of our company and we are not in debt or obligated to follow the directives of any venture firm. Even with our profitable operations, we still offer cheaper and faster ways for enterprises to achieve their information objectives than through conventional means.

### Some Realistic News

Semantic technologies have not yet reached the point of fulfilling their own prophecy nor of being sufficiently buzz-worthy to fuel their own demand. Enterprise customers are intrigued with the idea of semantic solutions, but still need to be convinced. Better search is often the telling leverage point in the sale. Enterprises do not appear to be interested in linked data alone (if at all), though some like the idea of possibly contributing linked data back to others. In any event, linked data (at least in our experience) is not a material factor to the sale.

The material factors to a sale have been data integration and interoperability, fulfilled through a distributed Web architecture that is now apparent all around us. Yet, even despite a general positive predilection to semantic technologies, the conceptual and technology transfer barriers to overcome are quite daunting. We (I) pride myself on being able to communicate complicated ideas and concepts relatively simply. But, despite hundreds of pages of documentation and many polished write-ups (see, for example, our [TechWiki](#) and the [chronology](#) of this blog), semantic concepts are not (generally) intuitive to content editors, information architects, project managers or fellow developers or project vendors. It is absolutely imperative to engage in continuous training and knowledge transfer during a semantic deployment.

These imperatives increase when multiple parties and components are being brought to bear for a large-scale enterprise deployment. Each part of the puzzle -- from portal and content management system to middleware to security to information repository -- has its own lingo and concepts for quite similar things. Because of the central role of semantics to these integration problems, it is critical that concepts in all legacy areas be properly "mapped" to the terminology and concepts of the semantic solution. One component's '*entity*' is another component's '*instance*'; one component's '*schema*' is another component's '*ontology*'.

Inter-team communications must be grounded in shared vocabulary and concepts. Yet, even then, it is still necessary to continuously describe and explicate the benefits due to semantic approaches over conventional ones. Because of its general foundational nature, semantic approaches are often hidden or at the core of the information solution. It is not always self-evident what the advantages of semantic approaches are, because their results can be mimicked via conventional approaches (though at greater cost with greater brittleness).

In no instance are we aware of enterprises having much interest in public data, except as judicious supplements. Most all information challenges are based on private, internal data, with much concern over security and access. Where public data enters the equation, it is from very limited sources of excellent quality and provenance. Thus, information solutions geared to the enterprise must have security and differential access baked into the cake, and not be an afterthought. In this regard, the semantic enterprise is quite unlike the semantic Web. Interoperability, data quality and data reliability take huge precedence over such ideas as serendipity or [follow your nose](#), advantages often put forward in a public Web context.

Unlike just a few years back, we no longer see resistance to open source solutions. In fact, for early semantic adopters, open source is a positive feature. But with open source in a complicated enterprise environment comes its own challenges. Support is often poor and integrating the pieces becomes one of the key project responsibilities and risks. Simple assertions of open APIs and a commitment to Web service endpoints still can lead to significant integration challenges. Encoding mismatches or how error messages get generated or treated, as two examples, point to some of the challenges in creating an

integrated enterprise environment from multiple open source pieces.

Though enterprise funding sure beats the funding behind most consumer-oriented projects, enterprise IT budgets have also come under their own pressures. The justification for many projects resides in being to offset annual licensing and maintenance fees, which can impose delivery constraints based on renewal dates. Existing enterprise IT budgets have also been made more incremental, with milestone achievements often required for moving forward. These trends are putting a premium on [agile development](#) and the need for enterprise-scale deployment and testing tools. Repeatable build processes and scripts are an essential component now of complicated stack deployments.

Many of the issues that emerge in enterprise deployments are ancillary to or independent of specific semantic components. Logging, testing, security, access, service buses and deployment builds are an umbrella over entire deployments. In these regards, incorporation of semantic technologies means that these contributions, too, must adhere to enterprise build practices and standards. This works to put a premium on repeatable build and testing scripts and improved deployment documentation and practices.

What all this means is that semantic technologies and practices need to grow up to adhere to standard enterprise practices, which themselves are undergoing rapid change as incremental, agile development becomes more prevalent. Much of what SD has learned in the past two years relates to the development and deployment environments that both aid and govern modern enterprise IT projects. In these regards, semantic technologies are merely another set of components in a broader, enterprise-wide stack.

Lastly, another reality of semantic technologies in the enterprise is that there are precious few champions and advocates within any given enterprise. Means must be found to communicate to semantic newbies and to enlist the aid of these champions in carrying the message forward within their organizations. In multi-vendor deployment environments it is important to find single points of contact that can also help communicate with their colleagues.

## What is to Come

These general points set the context for some of the specifics in the series to come. Attention will be given in the series to a number of topics, not necessarily in this order nor scope:

- [Three leading for semantic technologies](#)
- [Architecting semantic technologies for the enterprise](#)
- [Making text a first-class citizen](#)
- [The primacy of search](#)
- [Access control using datasets](#)
- Harvesting and [ETL](#) considerations
- Authoring
- Workflow integration and differences posed by semantic technologies
- Security considerations with semantic technologies
- Adopting and use of comprehensive deployment and testing environments (*e.g.*, [JIRA](#), [Confluence](#), [Bamboo](#))
- Automated testing
- [Continuous integration](#)

- Working with a Web-oriented architecture and endpoints
- Multiple, modular ontologies to capture enterprise schema
- Tech transfer and tools for ontology growth and development
- Integrating semantic technologies into publishing platforms
- Version control for semantic data and ontologies
- Gaps in enterprise-readiness of semantic technologies (bespoke tools, etc.)
- Challenges in communicating the benefits of semantic technologies
- Broader challenges in adoption of semantic technologies.

As appropriate, these topics will be addressed in forthcoming installments in this series. We will be culminating this series with overviews of two enterprise initiatives with high visibility for which Structured Dynamics has been the lead semantics contractor.

**NOTE:** This is the first part in an ongoing series on [enterprise-scale semantic systems](#) (ESSS), which has its own category on this blog. Simply click on that [category link](#) to see other articles in this series.

---

[1] With the growing popularity of semantic technologies, many entities claim the "semantic" mantle based solely on ways to negotiate the meaning ("semantics") of various concepts. This effort is legitimate, but tends to undercut a more strict interpretation. As we use herein, and have for some time, semantic technologies means the explicit use of the semantic Web languages and specifications adopted by the W3C along with the various software applications to use them.