

MAKING THE ARGUMENT FOR SEMANTIC TECHNOLOGIES

Michael K. Bergman¹, Coralville, Iowa USA

September 12, 2011

AI3::Adaptive Information blog

There have been some notable attempts of late to make [elevator pitches \[1\]](#) for semantic technologies, as well as Lee Feigenbaum's recent series on *Are We Asking the Wrong Question?* about semantic technologies [\[2\]](#). Some have attempted to downplay semantic Web connotations entirely and to replace the pitch with Linked Data (capitalized). These are part of a history of various ways to try to make a business case around semantic approaches [\[3\]](#).

What all of these attempts have in common is a view — an *angst*, if you will — that somehow semantic approaches have not fulfilled their promise. Marketing has failed semantic approaches. Killer apps have not appeared. The public has not embraced the semantic Web consonant with its destiny. Academics and researchers can not make the semantic argument like entrepreneurs can.

Such hand wringing, I believe, is misplaced on two grounds. First, if one looks to end user apps that solely distinguish themselves by the sizzle they offer, semantic technologies are clearly not essential. There are very effective mash-up and data-intensive sites such as many of the investment sites ([Fidelity](#), [TDAmeritrade](#), [Morningstar](#), among many), real estate sites ([Trulia](#), [Zillow](#), among many), community data sites ([American FactFinder](#), [CensusScope](#), [City-Data.com](#), among many), shopping sites ([Amazon](#), [Kayak](#), among many), data visualization sites ([Tableau](#), [Factual](#), among many), etc., etc., that work well, are intuitive and integrate much disparate information. For the most part, these sites rely on conventional relational database backends and have little semantic grounding. Effective data-intensive sites do not require semantics *per se* [\[4\]](#).

Second, despite common perceptions, semantics are in fact becoming pervasive components of many common and conventional Web sites. We see natural language processing (NLP) and extraction technologies becoming common for most search services. Google and Bing sprinkle semantic results and characterizations across their standard search results. Recommendation engines and targeted ad technologies now routinely use semantic approaches. Ontologies are creeping into the commercial spaces once occupied by taxonomies and controlled vocabularies. Semantics-based suggestion systems are now the common technology used. A surprising number of smartphone apps have semantics at their core.

So, I agree with Lee Feigenbaum that we are asking the wrong question. But I would also add that we are not even looking in the right places when we try to understand the role and place of semantic technologies.

The unwise attempt to supplant the idea of semantic technologies with linked data is only furthering this confusion. Linked data is merely a means for publishing and exposing structured data. While linked data can lead to easier automatic consumption of data, it is not necessary to effective semantic approaches and is actually a burden on data publishers [\[5\]](#). While that burden may be willingly taken by publishers because of its consumption advantages, linked data is by no means an essential precursor to semantic approaches. None of the unique advantages for semantic technologies noted below rely on or need to be preceded by linked data. In semantic speak, linked data is not the same as semantic technologies.

¹Email: mike@mkbergman.com

The essential thing to know about semantic technologies is that they are a conceptual and logical foundation to how information is modeled and interrelated. In these senses, semantic technologies are infrastructural and groundings, not applications *per se*. There is a mindset and worldview associated with the use of semantic technologies that is far more essential to understand than linked data techniques and is certainly more fundamental than elevator pitches or “killer apps.”

Five Unique Advantages

Thus, the argument for semantic technologies needs to be grounded in their foundations. It is within the five unique advantages of semantic technologies described below that the benefits to enterprises ultimately reside.

#1: Modern, Back-end Data Federation

The RDF data model — and its ability to represent the simplest of data up through complicated domain schema and vocabularies via the OWL ontology language — means that any existing schema or structure can be represented. Because of this expressiveness and flexibility, any extant data source or schema can be represented via RDF and its extensions. This breadth means that a common representation for any existing schema may be expressed. That expressiveness, in turn, means that any and all data representations can be described in a canonical way.

A shared, canonical representation of all existing schema and data types means that all of that information can now be federated and interrelated. The canonical means of federating information via the RDF data model is the foundational benefit of semantic technologies. Further, the practice of giving URIs as unique identifiers to all of the constituent items in this approach makes it perfectly suitable to today’s reality of distributed data accessible via the Web [6].

#2: Universal Solvent for Structure

I have stated many times that I have not met a form of structured data I did not like [7]. Any extant data structure or format can be represented as RDF. RDF can readily express information contained within structured (conventional databases), semi-structured (Web page or XML data streams), or unstructured (documents and images) information sources. Indeed, the use of ontologies and entity instance records in RDF is a powerful basis for driving the extraction systems now common for tagging unstructured sources.

(One of the disservices perpetuated by an insistence on linked data is to undercut this representational flexibility of RDF. Since most linked data is merely communicating value-attribute pairs for instance data, virtually any common data format can be used as the transmittal form.)

The ease of representing any existing data format or structure and the ability to extract meaningful structure from unstructured sources makes RDF a “universal solvent” for any and all information. Thus, with only minor conversion or extraction penalties, all information in its extant form can be staged and related together via RDF.

#3: Adaptive, Resilient Schema

A singular difference between semantic technologies (as [we](#) practice them) and conventional relational data systems is the use of an [open world approach](#) [8]. The relational model is a paradigm where the information must be complete and it must be described by a schema defined in advance. The relational model assumes that the only objects and relationships that exist in the domain are those that are explicitly represented in the database. This makes the closed world of relational systems a very poor choice when attempting to combine information from multiple sources, to deal with uncertainty or incompleteness in the world, or to try to integrate internal, proprietary information with external data.

Semantic technologies, on the other hand, allow domains to be captured and modeled in an incremental manner. As new knowledge is gained or new integrations occur, the underlying schema can be added to and modified without affecting the information that already exists in the system. This adaptability is generally the biggest source of economic benefits to the enterprise from semantic technologies. It is also a benefit that enables experimentation and lowers risk.

#4: Unmatched Productivity

Having all information in a canonical form means that generic tools and applications can be designed to work against that form. That, in turn, leads to user productivity and developer productivity. New datasets, structure and relationships can be added at any time to the system, but how the tools that manipulate that information behave remains unchanged.

User productivity arises from only needing to learn and master a limited number of toolsets. The relationships in the constituent datasets are modeled at the schema (that is, ontology) level. Since manipulation of the information at the user interface level consists of generic paradigms regarding the selection, view or modification of the simple constructs of datasets, types and instances, adding or changing out new data does not change the interface behavior whatsoever. The same bases for manipulating information can be applied no matter the datasets, the types of things within them, or the relationships between things. The behavior of semantic technology applications is very much akin to having generic mashups.

Developer productivity results from leveraging generic interfaces and APIs and not bespoke ones that change every time a new dataset is added to the system. In this regard, ontology-driven applications [9] arising from a properly designed semantic technology framework also work on the simple constructs of datasets, types and instances. The resulting generalization enables the developer to focus on creating logical “packages” of functionality (mapping, viewing, editing, filtering, etc.) designed to operate at the construct level, and not the level of the atomic data.

#5: Natural, Connected Knowledge Systems

All of these factors combine to enable more and disparate information to be assembled and related to one another. That, in turn, supports the idea of capturing entire knowledge domains, which can then be expanded and shifted in direction and emphasis at will. These combinations begin to finally achieve knowledge capture and representation in its desired form.

Any kind of information, any relationship between information, and any perspective on that information can be captured and modeled. When done, the information remains amenable to inspection and manipulation through a set of generic tools. Rather simple and direct converters can move that canonical information to other external forms for use by existing external tools. Similarly, external information in its various forms can be readily converted to the internal canonical form.

These capabilities are the direct opposite to today’s information silos. From its very foundations, semantic technologies are perfectly suited to capture the natural connections and nature of relevant knowledge systems.

A Summary of Advantages Greater than the Parts

There are no other IT approaches available to the enterprise that can come close to matching these unique advantages. The ideal of total information integration, both public and private, with the potential for incremental changes to how that information is captured, manipulated and combined, is exciting. And, it is achievable today.

With semantic technologies, more can be done with less and done faster. It can be done with less risk. And, it can be implemented on a pay-as-you-benefit basis [10] responsive to the current economic climate.

But awareness of this reality is not yet widespread. This lack of awareness is the result of a couple of factors. One factor is that semantic technologies are relatively new and embody a different mindset. Enterprises are only beginning to get acquainted with these potentials. Semantic technologies require both new concepts to be learned, and old prejudices and practices to be questioned.

A second factor is the semantic community itself. The early idea of autonomic agents and the heavy AI emphasis of the initial semantic Web advocacy now feels dated and premature at best. Then, the community hardly improved matters with its shift in emphasis to linked data, which is merely a technique and which completely overlooks the advantages noted above.

However, none of this likely matters. The five unique advantages for enterprises from semantic technologies are real and demonstrable today. While my crystal ball is cloudy as to how fast these realities will become understood and widely embraced, I have no question they will be. The foundational benefits of semantic technologies are compelling.

I think I'll take this to the bank while others ride the elevator.

Acknowledgements

This article was originally posted on the *AI3::Adaptive Information* Web site at <http://www.mkbergman.com/974/making-the-argument-for-semantic-technologies/>. This version has been edited and reformatted slightly for PDF distribution. We thank Cognonto Corporation for making this content freely available.

-
- [1] This series was [called for](#) by Eric Franzon of [SemanticWeb.com](#). Contributions to date have been provided by [Sandro Hawke](#), [David Wood](#), and [Mark Montgomery](#).
 - [2] See Lee Feigenbaum, 2011. "Why Semantic Web Technologies: Are We Asking the Wrong Question?," [TechnicaLee Speaking blog](#), August 22, 2011; see http://www.thefigtrees.net/lee/blog/2011/08/why_semantic_web_technologies.html, and its follow up on "The Magic Crank," August 29, 2011; see http://www.thefigtrees.net/lee/blog/2011/08/the_magic_crank.html. For a further perspective on this issue from Lee's firm, Cambridge Semantics, see Sean Martin, 2010. "Taking the Tech Out of SemTech," presentation at the *2010 Semantic Technology Conference*, June 23, 2010. See <http://www.slideshare.net/LeeFeigenbaum/taking-the-tech-out-of-semtech>.
 - [3] See, for example, Jeff Pollock, 2008. "A Semantic Web Business Case," Oracle Corporation; see <http://www.w3.org/2001/sw/sweo/public/BusinessCase/BusinessCase.pdf>.
 - [4] Indeed, many semantics-based sites are disappointingly ugly with data and triples and URIs shoved in the user's face rather than sizzle.
 - [5] Linked data and its linking predicates are also all too often misused or misapplied, leading to poor quality of integrations. See, for example, M.K. Bergman and F. Giasson, 2009. "When Linked Data Rules Fail," [AI3::Adaptive Innovation](#) blog, November 16, 2009. See <http://www.mkbergman.com/846/when-linked-data-rules-fail/>.
 - [6] Greater elaboration on all of these advantages is provided in M. K. Bergman, 2009. "Advantages and Myths of RDF," [AI3::Adaptive Innovation](#) blog, April 8, 2009. See <http://www.mkbergman.com/483/advantages-and-myths-of-rdf/>.
 - [7] See M.K. Bergman, 2009. "'Structs': Naïve Data Formats and the ABox," [AI3::Adaptive Innovation](#) blog, January 22, 2009. See <http://www.mkbergman.com/471/structs-naive-data-formats-and-the-abox/>.
 - [8] A considerable expansion on this theme is provided in M.K. Bergman, 2009. "'The Open World Assumption: Elephant in the Room,'" [AI3::Adaptive Innovation](#) blog, December 21, 2009. See <http://www.mkbergman.com/852/the-open-world-assumption-elephant-in-the-room/>.
 - [9] For a full expansion on this topic, see M.K. Bergman, 2011. "Ontology-driven Apps Using Generic Applications," [AI3::Adaptive Innovation](#) blog, March 7, 2011. See <http://www.mkbergman.com/948/ontology-driven-apps-using-generic-applications/>.

[10] See M.K. Bergman, 2010. “‘Pay as You Benefit’: A New Enterprise IT Strategy,” *AIS::Adaptive Innovation* blog, July 12, 2010. See <http://www.mkbergman.com/896/pay-as-you-benefit-a-new-enterprise-it-strategy/>.