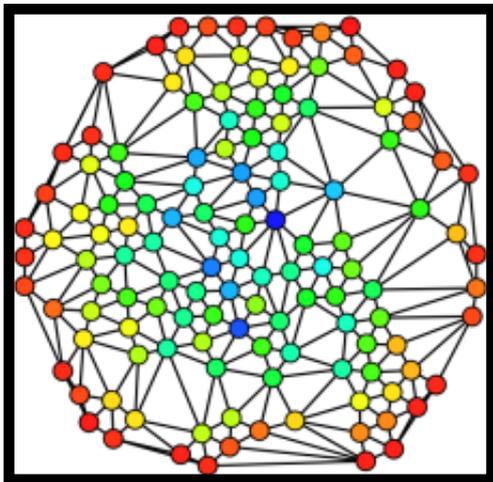


The Rationale for Semantic Technologies

by Mike Bergman - Monday, July 02, 2012

<http://www.mkbergman.com/1015/the-rationale-for-semantic-technologies/>



Conventional IT Systems are Poorly Suited to Knowledge Applications

Frequently customers ask me why semantic technologies should be used instead of conventional information technologies. In the areas of knowledge representation (KR) and knowledge management (KM), there are compelling reasons and benefits for selecting semantic technologies over conventional approaches. This article attempts to summarize these rationales from a layperson perspective.

It is important to recognize that semantic technologies are orthogonal to the buzz around some other current technologies, including cloud computing and big data. Semantic technologies are also not limited to open data: they are equivalently useful to private or proprietary data. It is also important to note that semantic technologies do not imply some grand, shared schema for organizing all information. Semantic technologies are not “one ring to rule them all,” but rather a way to capture the world views of particular domains and groups of stakeholders. Lastly, semantic technologies done properly are not a replacement for existing information technologies, but rather an added layer that can leverage those assets for interoperability and to overcome the semantic barriers between existing information silos.

Nature of the World

The world is a messy place. Not only is it complicated and richly diverse, but our ways of describing and understanding it are made more complex by differences in language and culture.

We also know the world to be interconnected and interdependent. Effects of one change can propagate into subtle and unforeseen effects. And, not only is the world constantly changing, but so is our understanding of what exists in the world and how it affects and is affected by everything else.

This means we are always uncertain to a degree about how the world works and the dynamics of its working. Through education and research we continually strive to learn more about the world, but often

in that process find what we thought was true is no longer so and even our own human existence is modifying our world in manifest ways.

Knowledge is very similar to this nature of the world. We find that knowledge is never complete and it can be found anywhere and everywhere. We capture and codify knowledge in structured, semi-structured and unstructured forms, ranging from “soft” to “hard” information. We find that the structure of knowledge evolves with the incorporation of more information.

We often see that knowledge is not absolute, but contextual. That does not mean that there is no such thing as truth, but that knowledge should be coherent, to reflect a logical consistency and structure that comports with our observations about the physical world. Knowledge, like the world, is constantly changing; we thus must constantly adapt to what we observe and learn.

Knowledge Representation, Not Transactions

These observations about the world and knowledge are not platitudes but important guideposts for how we should organize and manage information, the field known as “information technology.” For IT to truly serve the knowledge function, its logical bases should be consistent with the inherent nature of the world and knowledge.

By knowledge functions we mean those areas of various computer applications that come under the rubrics of search, business intelligence, competitive intelligence, planning, forecasting, data federation, data warehousing, knowledge management, enterprise information integration, master data management, knowledge representation, and so forth. These applications are distinctly different than the earliest and traditional concerns of IT systems: accounting and transactions.

A transaction system -- such as calculating revenue based on seats on a plane, the plane’s occupancy, and various rate classes -- is a closed system. We can count the seats, we know the number of customers on board, and we know their rate classes and payments. Much can be done with this information, including yield and profitability analysis and other conventional ways of accounting for costs or revenues or optimizations.

But, as noted, neither the world nor knowledge is a closed system. Trying to apply legacy IT approaches to knowledge problems is fraught with difficulties. That is the reason that for more than four decades enterprises have seen massive cost overruns and failed projects in applying conventional IT approaches to knowledge problems: traditional IT is fundamentally mismatched to the nature of the problems at hand.

What works efficiently for transactions and accounting is a miserable failure applied to knowledge problems. Traditional relational databases work best with structured data; are inflexible and fragile when the nature (schema) of the world changes; and thus require constant (and expensive) re-architecting in the face of new knowledge or new relationships.

Of course, often knowledge problems do consider fixed entities with fixed attributes to describe them. In these cases, relational data systems can continue to act as valuable contributors and data managers of entities and their attributes. But, in the role of organizing across schema or dealing with semantics and differences of definition and scope – that is, the common types of knowledge questions – a much different

integration layer with a much different logic basis is demanded.

The New Open World Paradigm

The first change that is demanded is to shift the logic paradigm of how knowledge and the world are modeled. In contrast to the closed-world approach of transaction systems, IT systems based on the logical premise of the open world assumption (OWA) mean:

- Lack of a given assertion does not imply whether it is true or false; it simply is not known
- A lack of knowledge does not imply falsity
- Everything is permitted until it is prohibited
- Schema can be incremental without re-architecting prior schema (“extensible”), and
- Information at various levels of incompleteness can be combined.

Much more can be said about OWA, including formal definitions of the logics underlying it [\[1\]](#), but even from the statements above, we can see that the right logic for most knowledge representation (KR) problems is the open world approach.

This logic mismatch is perhaps the most fundamental cause of failures, cost overruns, and disappointing deliverables for KM and KR projects over the years. But, like the fingertip between the eyes that cannot be seen because it is too close at hand, the importance of this logic mismatch strangely continues to be overlooked.

Integrating All Forms of Information

Data exists in many forms and of many natures. As one classification scheme, there are:

- *Structured data* — information presented according to a defined data model, often found in relational databases or other forms of tabular data
- *Semi-structured data* — does not conform to the formal structure of data models, but contains tags or other markers to denote fields within the content. Markup languages embedded in text are a common form of such sources
- *Unstructured data* — information content, generally oriented to text, that lacks an explicit data model or schema; structured information can be obtained from it via data mining or information extraction.

Further, these types of data may be “soft”, such as social information or opinion, or “hard”, more akin to measurable facts or quantities.

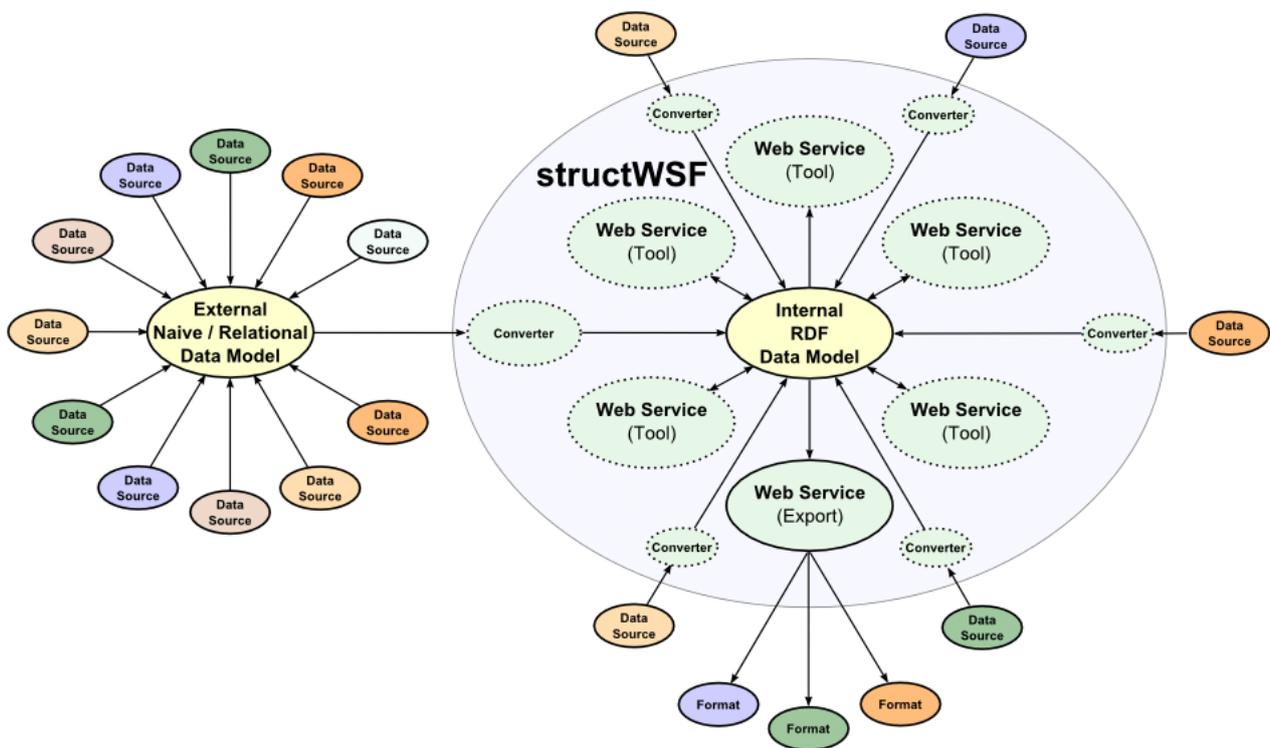
These various forms may also be serialized in a variety of data formats or data transfer protocols, some using straight text with a myriad of syntax or markup vocabularies, ranging to scripts or forms encoded or binary.

Still further, any of these data forms may be organized according to a separate schema that describes the semantics and relationships within the data.

These variations further complicate the inherently diverse nature of the world and knowledge of it. A suitable data model for knowledge representation must therefore have the power to be able to capture the form, format, serialization or schema of any existing data within the diversity of these options.

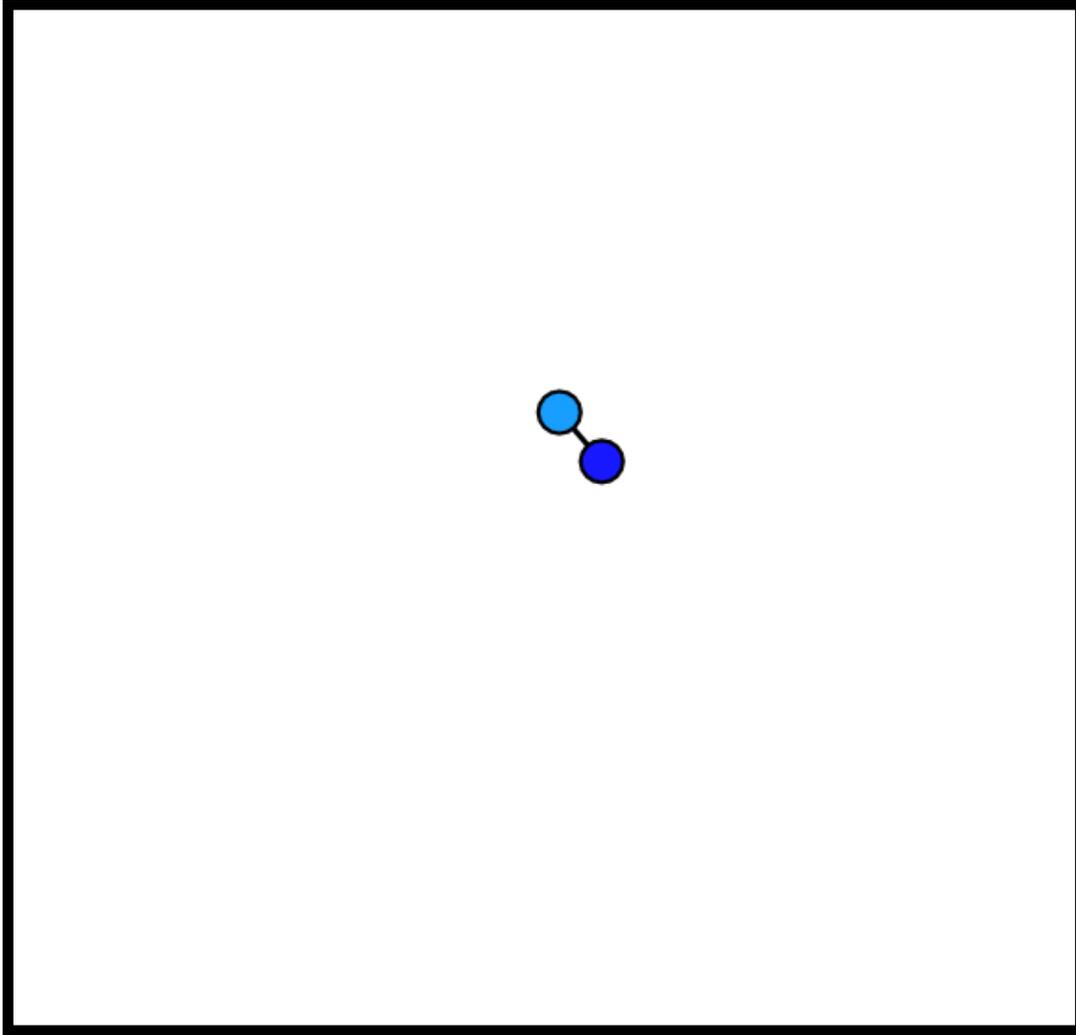
The Resource Description Framework (RDF) data model has such capabilities [2]. Any extant data form or schema (from the simple to the complex) can be converted to the RDF data model. This capability enables RDF to act as a “universal solvent” for all information.

Once converted to this “canonical” form, RDF can then act as a single representation around which to design applications and other converters (for “round-tripping” to legacy systems, for example), as illustrated by this diagram:



Generic tools can then be driven by the RDF data model, which leads to fewer applications required and lower overall development costs.

Lastly, RDF can represent simple assertions (“Jane runs fast”) to complex vocabularies and languages. It is in this latter role that RDF can begin to represent the complexity of an entire domain via what is called an “ontology” or “knowledge graph.”



Connections Create Graphs

When representing knowledge, more things and concepts get drawn into consideration. In turn, the relationships of these things lead to connections between them to capture the inherent interdependence and linkages of the world. As still more things get considered, more connections are made and proliferate.

This process naturally leads to a graph structure, with the things in the graphs represented as nodes and the relationships between them represented as connecting edges. More things and more connections lead to more structure. Insofar as this structure and its connections are coherent, the natural structure of the knowledge graph itself can help lead to more knowledge and understanding.

How one such graph may emerge is shown by this portion of the recently announced Google Knowledge Graph [\[3\]](#), showing female Nobel prize winners:



Unlike traditional data tables, graphs have a number of inherent benefits, particularly for knowledge representations. They provide:

- A coherent way to navigate the knowledge space
- Flexible entry points for each user to access that knowledge (since every node is a potential

starting point)

- Inferencing and reasoning structures about the space
- Connections to related information
- Ability to connect to any form of information
- Concept mapping, and thus the ability to integrate external content
- A framework to disambiguate concepts based on relations and context, and
- A common vocabulary to drive content “tagging”.

Graphs are the natural structures for knowledge domains.

Network Analysis is the New Algebra

Once built, graphs offer some analytical capabilities not available through traditional means of information structure. Graph analysis is a rapidly emerging field, but already some unique measures of knowledge domains are now possible to gauge:

- Influence
- Relatedness
- Proximity
- Centrality
- Inference
- Clustering
- Shortest paths
- Diffusion.

As science is coming to appreciate, graphs can represent any extant structure or schema. This gives graphs a universal character in terms of analytic tools. Further, many structures can only be represented by graphs.

Information and Interaction is Distributed

The nature of knowledge is such that relevant information is everywhere. Further, because of the interconnectedness of things, we can also appreciate that external information needs to be integrated with internal information. Meanwhile, the nature of the world is such that users and stakeholders may be anywhere.

These observations suggest a knowledge representation architecture that needs to be truly distributed. Both sources and users may be found in multiple locations.

In order to preserve existing information assets as much as possible (see further below) and to codify the earlier observation regarding the broad diversity of data formats, the resulting knowledge architecture should also attempt to put in place a thin layer or protocol that provides uniform access to any source or target node on the physical network. A thin, uniform abstraction layer – with appropriate access rights and security considerations – means knowledge networks may grow and expand at will at acceptable costs with minimal central coordination or overhead.

Properly designed, then, such architectures are not only necessary to represent the distributed nature of users and knowledge, but can also facilitate and contribute to knowledge development and exchange.

The Web is the Perfect Medium

The items above suggest the Web as an appropriate protocol for distributed access and information exchange. When combined with the following considerations, it becomes clear that the Web is *the* perfect medium for knowledge networks:

- Potentially, all information may be accessed via the Web
- All information may be given unique Web identifiers (URIs)
- All Web tools are available for use and integration
- All Web information may be integrated
- Web-oriented architectures (WOA) have proven:
 - Scalability
 - Robustness
 - Substitutability
- Most Web technologies are open source.

It is not surprising that the largest extant knowledge networks on the globe – such as Google, Wikipedia, Amazon and Facebook – are Web-based. These pioneers have demonstrated the wisdom of WOA for cost-effective scalability and universal access.

Also, the combination of RDF with Web identifiers also means that any and all information from a given knowledge repository may be exposed and made available to others as linked data. This approach makes the Web a global, universal database. And it is in keeping with the general benefits of integrating external information sources.

Leveraging – *Not Replacing* – Existing IT Assets

Existing IT assets represent massive sunk costs, legacy knowledge and expertise, and (often) stakeholder consensus. Yet, these systems are still largely stovepiped.

Strategies that counsel replacement of existing IT systems risk wasting existing assets and are therefore unlikely to be adopted. Ways must be found to leverage the value already embodied in these systems, while promoting interoperability and integration.

The beauty of semantic technologies – properly designed and deployed in a Web-oriented architecture – is that a thin interoperability layer may be placed over existing IT assets to achieve these aims. The knowledge graph structure may be used to provide the semantic mappings between schema, while the Web service framework that is part of the WOA provides the source conversion to the canonical RDF data model.

Via these approaches, prior investments in knowledge, information and IT assets may be preserved while enabling interoperability. The existing systems can continue to provide the functionality for which they

were originally designed and deployed. Meanwhile, the KR-related aspects may be exposed and integrated with other knowledge assets on the physical network.

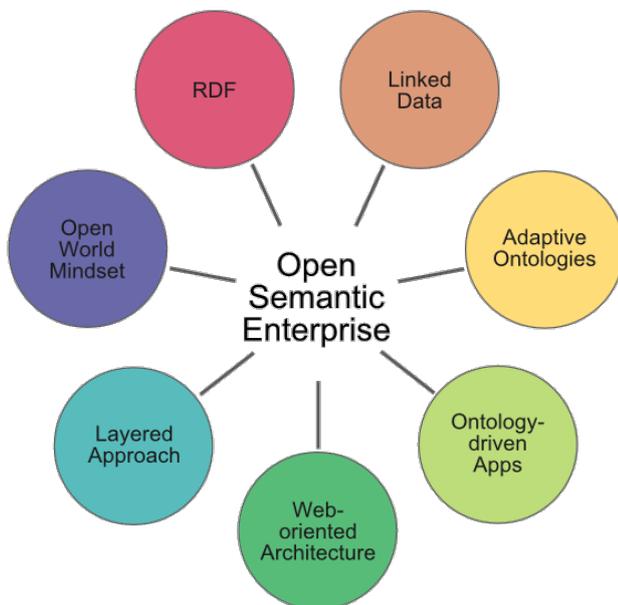
Democratizing the Knowledge Function

These kinds of approaches represent a fundamental shift in power and roles with respect to IT in the enterprise. IT departments and their bottlenecks in writing queries and bespoke application development can now be bypassed; the departments may be relegated to more appropriate support roles. Developers and consultants can now devote more of their time to developing generic applications driven by graph structures [4].

In turn, the consumers of knowledge applications – namely subject matter experts, employees, partners and stakeholders – now become the active contributors to the graphs themselves, focusing on reconciling terminology and ensuring adequate entity and concept coverage. Knowledge graphs are relatively straightforward structures to build and maintain. Those that rely on them can also be those that have the lead role in building and maintaining them.

Thus, graph-driven applications can be made generic by function with broader and more diverse information visualization capabilities. Simple instructions in the graphs can indicate what types of information can be displayed with what kind of widget. Graph-driven applications also mean that those closest to the knowledge problems will also be those directly augmenting the graphs. These changes act to democratize the knowledge function, and lower overall IT costs and risks.

Seven Pillars of the Semantic Enterprise



Elsewhere we have discussed the specific components that go into enabling the development of a semantic enterprise, what we have termed the seven pillars [5]. Most of these points have been covered to

one degree or another in the discussion above.

There are off-the-shelf starter kits for enterprises to embrace to begin this process. The major starting requirements are to develop appropriate knowledge graphs (ontologies) for the given domain and to convert existing information assets into appropriate interoperable RDF form.

Beyond that, enterprise staff may be readily trained in the use and growth of the graphs, and in the staging and conversion of data. With an appropriate technology transfer component, these semantic technology systems can be maintained solely by the enterprise itself without further outside assistance.

Summary of Semantic Technology Benefits

Unlike conventional IT systems with their closed-world approach, semantic technologies that adhere to these guidelines can be deployed incrementally at lower cost and with lower risk. Further, we have seen that semantic technologies offer an excellent integration approach, with no need to re-do schema because of changed circumstances. The approach further leverages existing information assets and brings the responsibility for the knowledge function more directly to its users and consumers.

Semantic technologies are thus well-suited for knowledge applications. With their graph structures and the ability to capture semantic differences and meanings, these technologies can also accommodate multiple viewpoints and stakeholders. There are also excellent capabilities to relate all available information – from documents and images and metadata to tables and databases – into a common footing.

These advantages will immediately accrue through better integration and interoperability of diverse information assets. But, for early adopters, perhaps the most immediate benefit will come from visible leadership in embracing these enabling technologies in advance of what will surely become the preferred approach to knowledge problems.

Note: there is a version of this article on Slideshare:

[The Rationale for Semantic Technologies](#)

View more presentations from [Mike Bergman](#).

[1] For more on the open world assumption (OWA), see the various entries on this topic on Michael Bergman's *AI3:::Adaptive Information* blog. This link is a [good search string](#) to discover more.

[2] M.K. Bergman, 2009. *Advantages and Myths of RDF*, white paper from Structured Dynamics LLC, April 22, 2009, 13 pp. See http://www.mkbergman.com/wp-content/themes/ai3v2/files/2009Posts/Advantages_Myths_RDF_090422.pdf.

[3] Google Knowledge Graph; see <http://www.google.com/insidesearch/features/search/knowledge.html>.

[4] For the most comprehensive discussion of graph-driven apps, see M. K. Bergman, 2011. " [Ontology-Driven Apps Using Generic Applications](#)," posted on the *AI3:::Adaptive Information* blog, March 7, 2011. You may also search on that blog for '[ODapps](#)' to see related content.

[5] M.K. Bergman, 2010. “Seven Pillars of the Open Semantic Enterprise,” in *AI3::Adaptive Information* blog, January 12, 2010; see <http://www.mkbergman.com/859/seven-pillars-of-the-open-semantic-enterprise/>.

PDF generated by *AI3::Adaptive Information* blog