

The Semantic Web 3.0 Mashup Universe: Coming to a Browser Near You

by Mitchell Ummel, Senior Consultant,
Cutter Consortium

The Internet is undergoing a rapid transformation from a web of hyperlinked documents to a web of semantically linked data. Recent observations lead me to believe we're seeing the emergence of what may qualify as Web 3.0 (or Semantic Web) applications.¹ These applications are consumers and providers of semantically linked data. For the purposes of this *Executive Update*, I will refer to this new generation of Internet applications as semantically aware applications (SAAs).²

WHERE IN THE WIDE WORLD IS THIS WEB OF DATA?

In a recent *Harvard Business Review* article "What You Need to Know About the Semantic Web," author Tom Ilube points out that "some 23 billion data relationships have been coded since 2000 (more than half of them in the past year alone) using a protocol known as Resource Description Framework (RDF)."³

In its simplest terms, RDF, a set of W3C specifications based on XML, can be thought of as the basic construct of the distributed semantic data model for the Internet. An RDF triple store contains a collection of statements (rows), each statement composed of a subject, predicate, and object (see Figure 1).

RDF, often implemented in consideration of a domain-specific ontology via OWL (Web Ontology Language), provides for realization of popular services such as FOAF (Friend of a Friend) and RSS/Atom (blogs, news, and Web feed syndication).⁴ At the risk of waxing anthropomorphic, we could even describe RDF, OWL, and the information so bundled and accessible among

these ontological constructs as Internet digital gray matter (IDGM).⁵

A growing proportion of IDGM is open and accessible today. Figure 2 shows some of the ontologies that the W3C's Linking Open Data community project has published and interlinked so far.⁶

Given this exponential growth in the generation and accessibility of semantic data, it doesn't take an Internet scientist to realize that we already have a robust, standards-based data foundation in place. RDF queries can now be made into and across linked triple stores. Inference engines navigate OWL-based ontologies, and through the use of natural language processing translators, we have all the means in place to pose burning questions to, and get real answers from, the Semantic Web, such as "Which other two members of the Beatles, besides Ringo Star, could play the drums?"

We can now start realizing prototypes for the next generation of "thinking" and "understanding" applications. I believe that we're at a tipping point in this evolution, with newborn SAAs, however primitive, already saturating the Web.

HOW DOES THIS SEMANTIC DATA GET CREATED AND UPDATED?

For me, the most profound result of this saturation tipping point is that semantic data, once "synthetically," and often manually, generated, is now "organically" and automatically generated by these primitive SAAs in the following three ways:

1. Based on our past and current interactions with all SAAs and the billions of digital bread crumbs of personal information we will decide to leave behind in IDGM. A case in point is an SAA such as Twine,

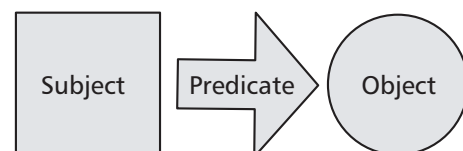


Figure 1 — An RDF triple store.

- which automatically creates and publishes RDF based on our interactions.
2. The SAAs' awareness (with our opt-in, of course) of our movement through space and time as well as the public accessibility of that geospatial and temporal data to other SAAs. Examples are the microblogging application Twitter and the hundreds of Google Maps-based mashups that have sprung up around this service.
 3. RDF-izers, which take existing Internet data, in whatever native data stores it exists, and expose it to Semantic Web linking and inference engines through RDF. DBpedia, for example, is the RDF realization of Wikipedia.

4. The SAA's ability to make the "cognitive" connections in IDGM and infer an awareness and understanding of who we are — and what we want — as individuals based on our preferences or the preferences of people who are "similar to us" in certain scenarios.

INCUBATORS FOR SEMANTIC DATA: TODAY'S SAAs

In recent months, I've been casually tracking, on a nearly daily basis, the latest generation of Web-based mashups through a community portal I configured using a mashup of Joomla! and Yahoo! Pipes called MashUpUniverse. Like a Web-based "canary in a coal mine," I've envisioned that I've set the content feed aggregator "dials" to alert me of the emergence of

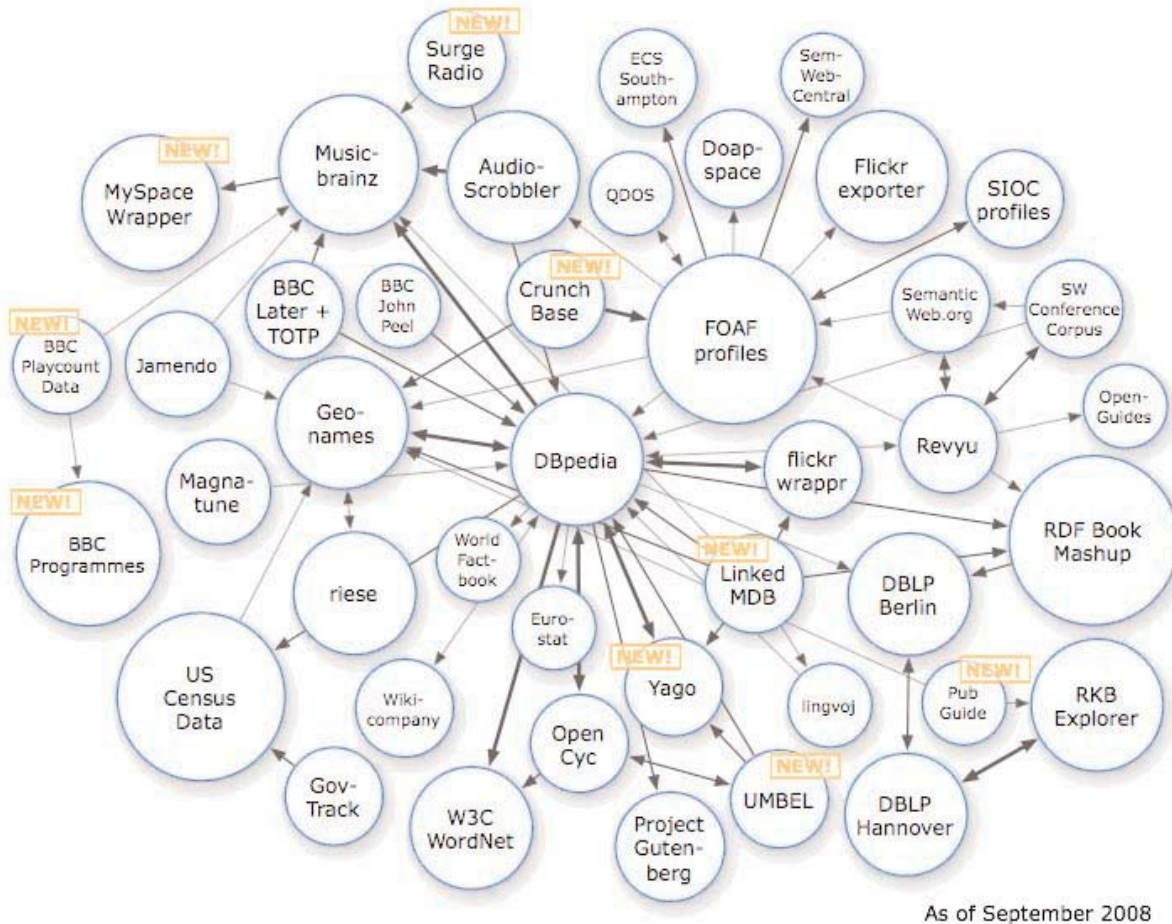


Figure 2 — W3C's Linking Open Data community project.

The *Executive Update* is a publication of the Business Intelligence Advisory Service. ©2009 by Cutter Consortium. All rights reserved. Unauthorized reproduction in any form, including photocopying, faxing, image scanning, and downloading electronic copies, is against the law. Reprints make an excellent training tool. For information about reprints and/or back issues of Cutter Consortium publications, call +1 781 648 8700 or e-mail service@cutter.com.

Web 3.0 (and the subsequent demise of Web 2.0). What I've observed during the past six months is that the majority of mashup activity has trended around and among combinations of the following types of services:

- Microblogging services such as Twitter and Facebook
- Location services such as Google Earth and Yahoo! Maps
- Entertainment-based services such as iTunes, Netflix, XM Radio, and YouTube
- Content-based services such as weather sites, news organizations, and knowledge aggregators, including the *New York Times* and Wikipedia

Calling any of the innovative, new applications that we see on a daily basis at MashUpUniverse a "true" SAA is certainly debatable. Several things we can say is that these mashups are meshing data and services from two or more sources and realizing an entirely new application (be it semantically aware or not), which, if successful, becomes much more than the sum of its parts.

More important, and through modes I described earlier, most of these new applications, or mashups of those applications, are also autogenerating new "organic" RDF content for future SAAs to build on.

MASHOUT! A NEW SPECIES OF MASHUP?

On a side note, we're also seeing something new in the wild that looks very much like a mashup and uses the same Web 2.0 technologies, but it doesn't meet the traditional definition of a mashup, as it uses data/service from only one source.

Perhaps we need a new term for these single-source Web application hybrids. I propose the term "mashout," which takes a popular service — let's say Wikipedia or Netflix — and attempts to improve on the presentation, navigation, or perhaps add innovative features that are not part of the original source service. A couple examples are:

- A better, faster Wikipedia? Try it for yourself at www.lexisum.com.
- A better, faster way to select Netflix Watch Now selections? Try it for yourself at <http://instantwatcher.com>.

Suppose that these mashouts, through our interactions with them (and after asking for user opt-in permission), also generated and published semantic data through their RDF-izer that other SAAs could use ... one hopes,

on our behalf. This is one trend we're already seeing with mashouts. Imagine the possibilities. Also imagine the security and privacy implications.

REAL WORLD WIDE SEMANTIC WEB USAGE SCENARIOS

Use cases for future SAAs are fun to visualize. Here's a couple examples.

Use Case 1

My guest and I are on a business trip in a rental car driving through Chicago. We have four hours to kill before our business meeting. I'm not familiar with the area (my Semantic Web-attached GPS "knows" this — why shouldn't it?). So it suggests an itinerary for us based on a number of factors, including (1) our interests (latest computer gadgetry and classical art), (2) our recent experiences (I haven't been to a computer superstore in three months, and my guest has never seen Van Gogh's "Starry Night"), and (3) our location (these services are close by and there's a traffic-friendly route between them). As a bonus (I didn't ask, you must realize), my GPS suggests, through its interface with IDGM, that our itinerary leaves just enough time for lunch at a local seafood restaurant. The GPS knows this due to our location, our planned routes, time of day (11 am), our general/shared food preferences, and the fact that neither of us has recently dined on seafood in the past week.

Use Case 2

Using my new Semantic Web "browser,"⁷⁷ I type in (or speak or the Web just "knows" without my asking), my simple request: "Please give me three options by 5 pm today, with pricing and full itinerary, for an enjoyable four-day spring break trip for my family to someplace warm." We can all imagine how an automated Web "agent," even today, could pull together options and costs for airfare, cruise, rental car, itinerary, and lodging across the smorgasbord of travel services already in place on the Internet. That's not enough. In order to fully "grok" this request, the semantic agent of the future must know a little more about who I am, what relationships I have (in this case, understanding the term "family" in the context of who I am), and what my family's travel preferences may be (in this case, interpreting the term "enjoyable" in the context of who my family is). The agent may be able to infer additional and deeper understanding of my request, through navigation of the Semantic Web (at least our family's slice of IDGM that I've made available to share), and infer additional meaning from facts as diverse as "families

like us enjoyed vacations to warm places such as....” or “our family had a bad experience on a holiday cruise to the Bahamas three years ago” or “our family generally takes spring break vacations in the price range of....”

Using today’s mashup technologies and secure access to today’s existing semantic data stores, I suggest that each of these scenarios could be realized as a SAA today.

WHAT DOES ALL THIS MEAN TO BUSINESS AND ENTERPRISE DATA/APPLICATION SYSTEMS OF THE FUTURE?

The opportunities and impact to our way of thinking about enterprise systems development in a Semantic Web-based world is yet to be explored and is well beyond the scope of this *Update*.⁸

We can say only one thing for certain: through the years, Internet-borne technologies have a perfect track record of being conceived, realized, and proven on the Internet, and then they eventually find utility and application within the inner operations of the enterprise. The list is long, so we won’t rehash it here — just start with TCP/IP and end with whatever Semantic Web evolves to be.

As for applicability within the enterprise, one might suggest that traditional enterprise systems are from Mars and Semantic Web systems are from Venus — in reference to the fundamental differences in principles and design assumptions underlying each.

Traditional Enterprise Systems Are from Mars

Within the confines of an enterprise, traditional applications generally lead us to operate under what are known as closed-world assumptions (CWA). This implies that all decisions and business rules can be applied to data or facts that are known to, and managed by, the enterprise as a closed world. In a CWA environment, we operate enterprise resource planning (ERP), accounting, and payroll systems. Rest assured that the relational data on which decisions are made is complete, accurate, and under our control. As such, we can confidently infer meaning from facts or, alternatively, from the absence or converse of facts.

For example, in an application based on CWA, if an order for a customer is not found in the order processing system, our application logic will assume that no order was placed. The answer to the question “Does the customer have an order?” is no.

Semantic Web Systems Are from Venus

The technologies, applications, and data of the Semantic Web, including RDF and OWL, live in what is known as open-world assumptions (OWAs). This means simply that data or facts are assumed to be incomplete and will generally never be fully known. In OWA, we can’t infer meaning from an absence of a fact or from the converse of a known fact. In these cases, we must simply infer “unknown.”

In an application based on OWA, if an order for a customer is not found in the order processing system, we can’t say for certain whether an order exists. The answer to the question “Does the customer have an order?” is unknown.

The point is that today’s realization of SAA technologies are likely not directly applicable to such traditional applications as ERP, accounting, and payroll systems; however, there are certain classes of enterprise systems where we may find applicability with Semantic Web technologies. They are:

- **Advanced data warehouse reporting and visualization.** Crosscutting both structured and unstructured content and strongly qualitative (rather than quantitative) in the nature of the questions, we ask about the millions or billions of “facts” comprising the problem space.
- **Knowledge-based business rules, requiring fuzzy logic and use of inference engines.** These are used where institutional knowledge (managed as facts in massive RDF triple-store repositories) drive enterprise process optimization. This is the business case for what we once referred to as knowledge management (KM), now finally realized.
- **Enterprise systems that, due to their mission, must interface and collaborate with OWA applications living in other private enterprise fact stores or in the public Semantic Web.** Imagine your company’s R&D business area inferring new product strategies through visibility of all of your (prospective) customer’s KM systems.

CONCLUSION

Let your imagination run wild. The timelines are not clear, but the possibilities in a Semantic Web 3.0 mashup universe are endless.

ENDNOTES

¹In “Is the World Ready for the Semantic Web?” Mark Choate also discusses whether the Semantic Web is on the verge of becoming mainstream (Cutter Consortium Business Technology Trends & Impacts *Executive Update*, Vol. 9, No. 8, 2008).

²Regarding the term “SAAs,” this is proposed at the risk of trampling on IBM’s “Systems Application Architecture” (SAA) of the late 1980s.

³Ilube, Tom. “What You Need to Know About the Semantic Web.” *Harvard Business Review*, February 2009.

⁴In “Semantics, Ontologies, and Data Modeling,” Cutter Senior Consultant David C. Hay takes us on a much more in-depth tour of the technologies of the Semantic Web, including but not limited to OWL and RDF, and in the process also contrasts traditional data modeling to ontological engineering (Cutter Consortium Business Intelligence *Executive Report*, Vol. 6, No. 7, 2006).

⁵While often used in the past to describe specific, advanced, software- or hardware-based technologies, the phrase “digital gray matter,” to my knowledge, has not been used in anthropomorphic reference to the Internet, Semantic Web, or Web 3.0.

⁶Linking Open Data (<http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>).

⁷What do we call it when we interact with Web 3.0, aka the Semantic Web? We need to invent a new word. In the “old school” Web 1.0-Web 2.0, we described our interactions as using a Web browser, browsing, Googling, looking it up on the Web, and/or surfing the Internet. These remain good descriptions for navigating billions of pages of static HTML “documents” in a not-so-intelligent Internet world. In the Web 3.0/Semantic Web of the future, humans will interact with the entire World Wide Web of semantically linked data, and “it” will know us, respond to us (through artificial intelligence, inference engines, etc.), and eventually, interact on its own with us, much as a human interacts with another intelligent human. In reverence to one of my favorite writers, Robert Heinlein, author of the sci-fi classic *Stranger in a Strange Land*, I propose resurrection of the term “grok” or “grokking” — this is what I believe to be the perfect word to describe how we will interact with Web 3.0. For more about “grok,” see Wikipedia at <http://en.wikipedia.org/wiki/Grok>.

⁸In “Ontology-Supported BI,” Paola Di Maio supplies further evidence and provides additional examples for the potential for Semantic Web technologies to be used for BI applications within the enterprise (Cutter Consortium Business Intelligence *Executive Update*, Vol. 8, No. 22, 2008).

ABOUT THE AUTHOR

Mitchell Ummel is a Senior Consultant with Cutter Consortium’s Enterprise Architecture practice and President of UmmelGroup International, Inc., a US-based business and technology management consulting firm. Mr. Ummel is a visionary who is most well known for championing practical application of innovative but lightweight IT process improvement methodologies into today’s culture within large organizations. In recent years, he has enjoyed advising state governments and private-sector enterprises in their planning and architecture for large, multimillion-dollar business and technology transformations. Mr. Ummel’s IT experience spans 25 years and includes service in a variety of CIO, executive management, training, coaching, mentoring, and consulting roles for government, telecommunications, electric/gas utilities, monitored home security, health insurance, regulatory/licensure, law enforcement, justice, and a variety of Internet-based product or services companies. Earlier in his career, he served as a state of Kansas agency CIO, where he contributed to the inception of the Kansas statewide technical architecture, helped define next-generation project and portfolio governance processes, and led technology initiatives enabling electronic health surveillance records management, vital records security/privacy, national bioterrorism response, GIS integration, and public health and environmental informatics among local, state, and federal stakeholders. Mr. Ummel holds advanced degrees in mathematics and computer science from Fort Hays State University. He is a graduate of, and contributor to, Kauffman Foundation’s Entrepreneurial Leadership and New Tech Venture Development Programs, is a frequent speaker at national business and technology conferences, and is a member of the Project Management Institute. He can be reached at mummel@cutter.com.

Workshop Developers/ Presenters

Every workshop is led by one of Cutter Consortium's expert Senior Consultants — experienced IT professionals who have honed their skills and developed their methodologies over years in the field, at companies like yours.

Verna Allee
 Sanjiv Augustine
 Rob Austin
 Christopher Avery
 Sam Bayer
 Kent Beck
 E.M. Bennatan
 Bob Benson
 Tom Bugnitz
 David J. Caruso
 Ken Collier
 Rachel Davies
 Tom DeMarco
 Sid Henkin
 Jim Highsmith
 Wendell Jones
 Jeff Kaplan
 Joshua Kerievsky
 Bartoz Kiepuszewski
 Tim Lister
 Michael Mah
 Terry Merriman
 Larissa Moss
 Ken Orr
 Carl Pritchard
 Ken Rau
 Thomas Redman
 Suzanne Robertson
 Mike Rosen
 Michael Schmitz
 David Spann
 Rob Thomsett
 John Tibbetts
 Jim Watson
 Karl Wiig
 Bob Wysocki

Cutter Consortium
 37 Broadway, Suite 1
 Arlington, MA 02474-5552, USA

Tel: +1 781 648 8700
 Fax: +1 781 648 1950
 Web: www.cutter.com
 E-mail: sales@cutter.com



Workshops

In these times of intense pressure to make every development dollar and every development minute count, the maxim *you are only as strong as your weakest link* has never rung truer.

Moving your development organization up the productivity curve will improve the ROI of every one of your projects. Just trace this back and you'll discover the ROI in training is immense. And with training and workshops designed and delivered by Cutter Consortium's Senior Consultants, you can add to that equation the peace of mind you get from being trained by the best of the best.

Cutter Consortium offers inhouse training solutions from IT project management techniques to software development methodologies, improving data quality, architecting Web services applications, aligning business and IT objectives, and more.

●●● Workshop Topics

- Agile Development Methodologies
- Agile Project Management
- Business-IT Alignment
- CIO Dashboards
- Data Quality
- Data Warehousing
- Enterprise Architecture
- Estimation Techniques
- Extreme Programming
- IT Strategic Planning
- Knowledge Management
- Metrics/Benchmarking
- Outsourcing
- Requirements Management
- Risk Management
- Software Development Practices
- Teamwork and Leadership
- Web 2.0 Technologies

For details about the courses offered in each of these areas, contact Jack Wainwright at +1 781 641 5122 or jwainwright@cutter.com, or visit www.cutter.com.