

Web 3.0 emerges...

Jim Hendler

Tetherless World Professor of Computer and Cognitive Science

Assistant Dean of Information Technology and Web Science

Rensselaer Polytechnic Institute http://www.cs.rpi.edu/~hendler @jahendler (twitter)



SEMANTIC WEB & LINKED DATA BUSINESS STRATEGIES



October 16-17, 2008 • Santa Clara, CA

ellation

The Next Generation Web



Web 3.0



Semantic Web



Linked Data

Web 3.0 Conference & Expo is focused on bringing together the key proponents and components delivering the promise of next generation web applications, technologies and business utilization.

Web 3.0 showcases in case-study format, the explosive, gamechanging promise and disruptive opportunities as they develop in order to help the entire community realize the promise of Web 3.0.

Semantic Web takes that hyper-data and enriches it with meaning through semantic web standards, technologies and strategies, and ultimately leads to the next generation Web.

Linked Data brings traditional hyperlinks and today's "web of documents" into the era of an interconnected, standards-based "web of linked data".

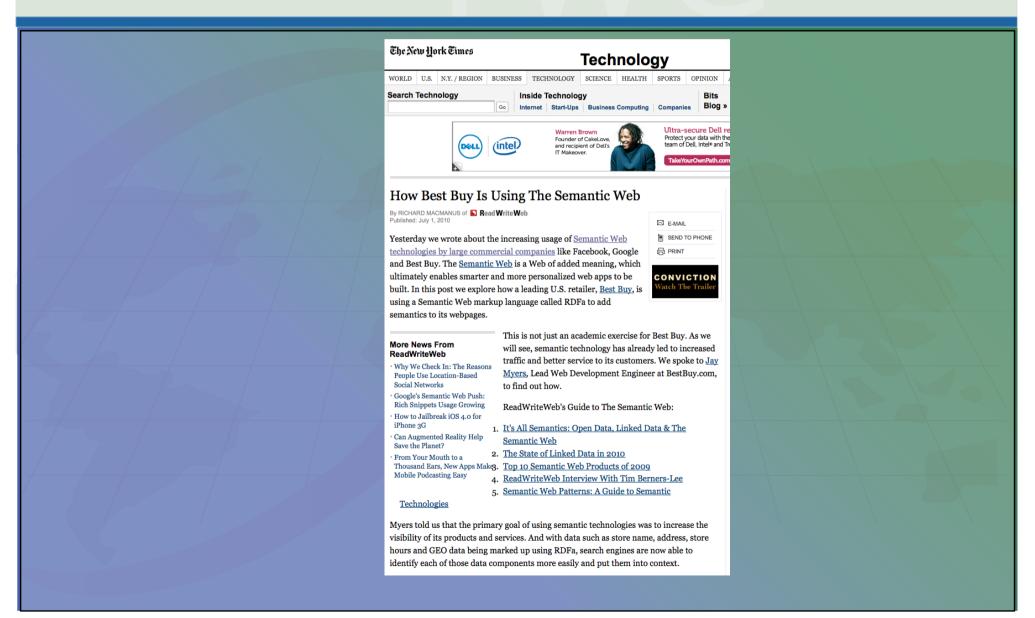
Applications, Technologies and Business Utilization







Semantic Web 2010





Tetherless World Constellation

Google, Twitter and Facebook build the semantic web

-) 02 August 2010 by Jim Giles
- Magazine issue 2771. Subscribe and save

A TRULY meaningful way of interacting with the web may finally be here, and it is called the semantic web. The idea was proposed over a decade ago by Tim Berners-Lee, among others. Now a triumvirate of internet heavyweights - Google, Twitter and Facebook - are making it real.

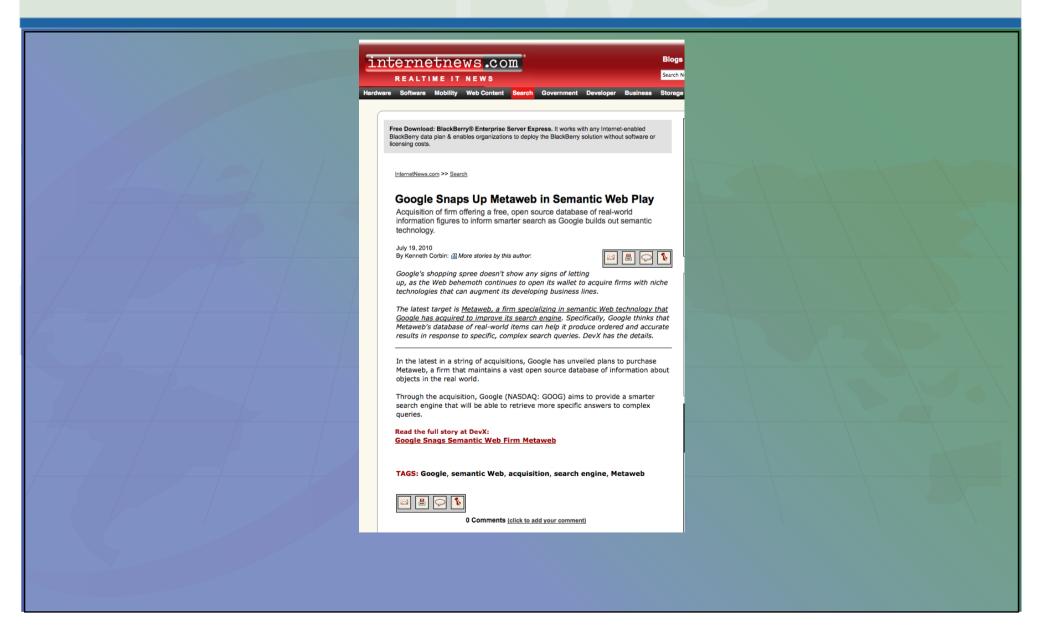
The defining characteristic of the semantic web is that information should be stored in a machine-readable format. Crucially, that would allow computers to handle information in ways we would find more useful, because they would be processing the concepts within documents rather than just the documents themselves.

Imagine bookmarking a story about Barack Obama: your computer will store the URL, but it has no way of knowing whether the content relates to politics or, say, cookery. If, however, each web page were to be tagged with information about its content, we can ask the web questions and expect sensible answers.



...







Tetherless World Constellation

Semantic Web

Semantic Web Semantic Web

Semantic Enterprise: What Are The Gorillas Doing? (Oracle, IBM, HP, Cisco, Microsoft and SAP)

By Bernard Lunn on Aug 03, 2010 07:00 AM



In Crossing The Chasm terminology, "gorillas" are the dominant vendors. Simple message for start-ups - don't mess with them!

In this post, we want to understand what the gorillas are doing to apply semantic web technology to the enterprise. The gorillas in this market are: Oracle, IBM, HP, Cisco, Microsoft and SAP.

Oracle: Embrace & Extend

Oracle is active in the semantic web. This matters to them. They cannot afford another database management system based on RDF to replace Oracle and MySQL. Oracle look at RDF as just another thing to store - like objects. The saw off the threat from object management systems and they aim to see off any threat from RDF triples.



- cf. US National Center for Biotechnology Information, "Oncology Metathesaurus"
 - 50,000+ classes, ~8 people supporting full time, monthly updates, mandated for use by NIH-funded cancer researchers
 - OWL DL rigorously followed
 - Provably consistent
- cf. Friend of a Friend (Foaf)
 - 30+ classes, Dan Brickley and Libby Miller made it, maintained by consensus in a small community of developers
 - Violates DL rules (undecidable)
 - Used inconsistently



Widely varying use

Tetherless World Constellation

- NCBI Oncology Ontology
 - "High use" in medical community (~1200 users)
 - Very "trusted" information (provenance from NCBI)
 - Primarily terminological (relationships between cancer-related concepts), not data-oriented

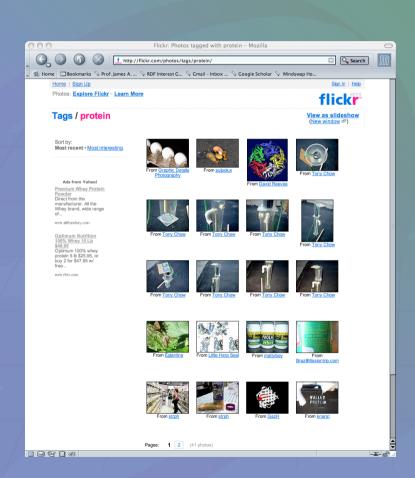
FOAF

- >60M Foaf people (not necessarily distinct individuals) in hundreds of applications touched by a large community (~1,000,000,000 users)
- Used by a number of large providers
 - If you use LiveJournal, you have a FOAF file
 - Also flickr, ecademy, tribe, joost, ...
 - And you can export Foaf from Facebook and many other social networking sites
- Becoming de facto standard for open social networking

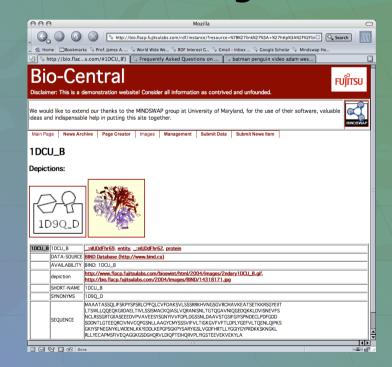


The argument for NCBI seems compelling

Tetherless World Constellation



When "folksonomy" isn't enough...



Which one do you want your doctor to use?



But the cost is too high

Tetherless World Constellation

- Formal modeling finds its use cases in verticals and enterprises
 - Where the vocabulary can be controlled
 - Where finding things in the data is important
- Example
 - Drug discovery from data
 - Model the molecule (site, chemical properties, etc) as faithfully and expressively as possible
 - Use "Realization" to categorize data assets against the ontology
 - Bad or missed answers are money down the drain
- But the modeling is very expensive and the return on investment must be very high!
 - Which is part of why the "expert systems revolution" wasn't one
 - Became part of the technology tool kit, a useful niche in the programming pantheon, but didn't change the world

Analogy: the pre-web hypertext world



The alternative

Tetherless World Constellation

- OWL is based on RDF, a language designed for the (Semantic) Web
 - Built with Web architecture in mind
 - Exploits Web infrastructure, respects W3C TAG recommendations
 - Internationalization, accessibility, extensibility
 - Fits the Web culture
 - Open and extensible, supports communities of interest
 - If you don't like my ontology, extend it, change it, or build your own
 - Fits the Web application development paradigm
 - Scales like "databases"
 - With some new ways of linking to formal models
 - Heavy use of a small amount of RDFS and a tiny bit of OWL
 - Generally used "like it sounds" not like the formal model
 - Example "owl:sameAs" debate

"linked data" often used to describe this low semantics Semantic Web

Analogy: the World Wide Web

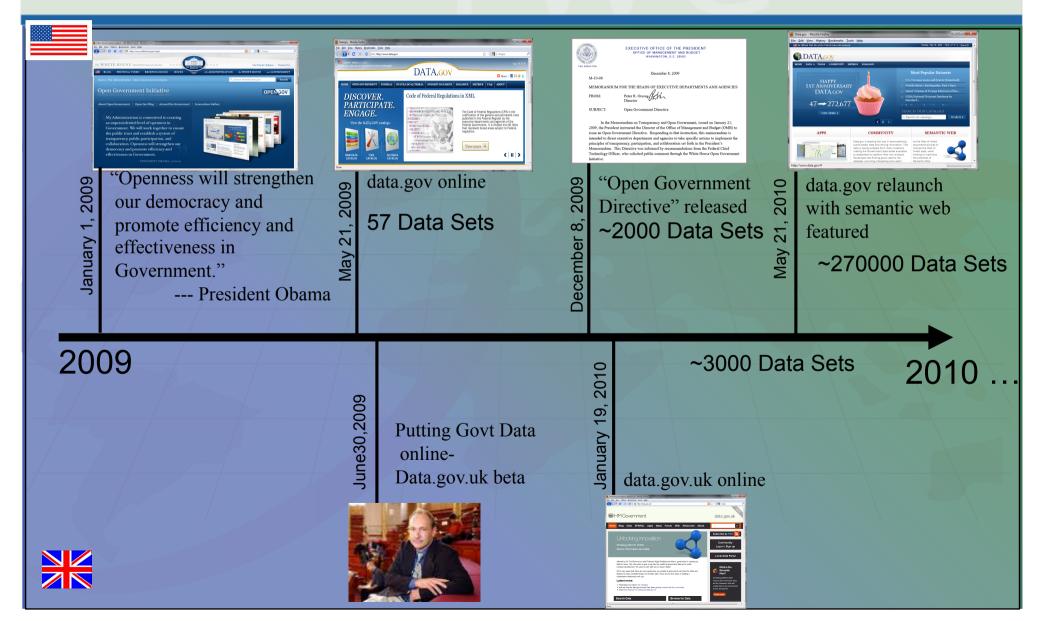


Linked Data + Semantics

- "Linked Data" approach finds its use cases in Web Applications (at Web scales)
 - A lot of data, a little semantics
 - Finding anything in the mess can be a win!
- Example
 - Declare simple inferable relationships and apply, at scale, to large, heterogeneous data collections
 - eg. Use InverseFunctional triangulation to find the entities that can be inferred to be the same
 - These are "heuristics" not every answer must be right (qua Google)
 - But remember *time* = *money*!



Government Data Sharing

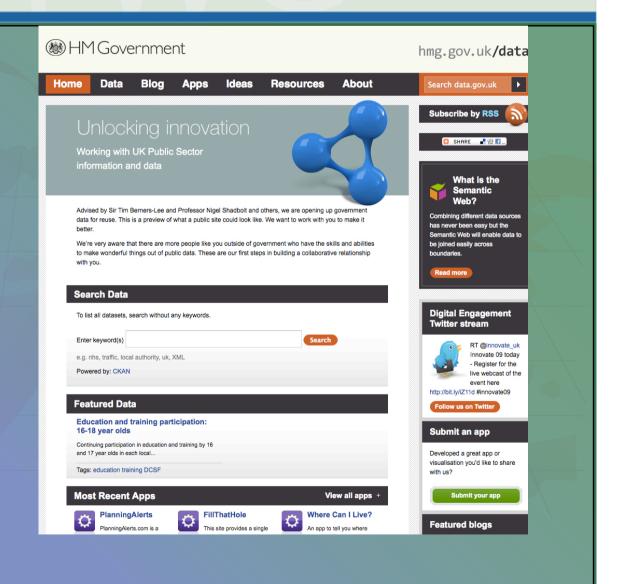




Moving data.gov to linked data (UK)

Tetherless World Constellation

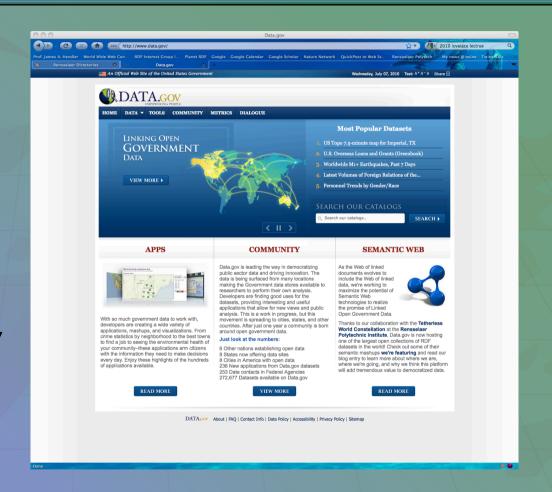
 Built around linked data with topdown push from "Number 10"





Moving data.gov to linked data (US)

- Originated with 3rd parties doing Sem Web bottom up
- May 21 2010, embraced by data.gov



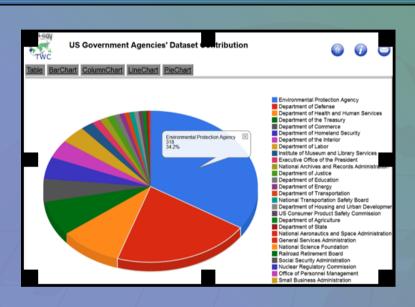


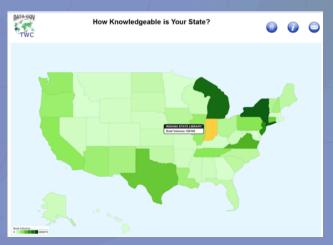
Data.gov community

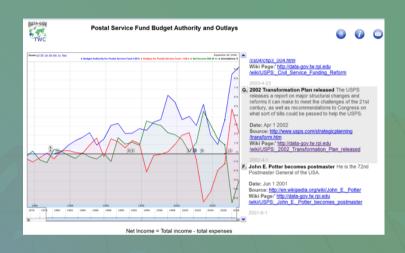


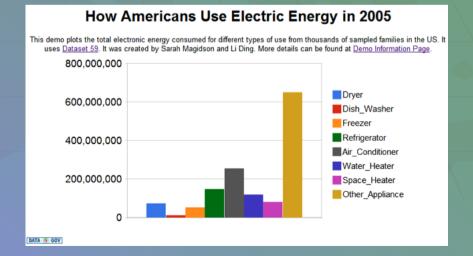


Pump through to Google Viz for demos

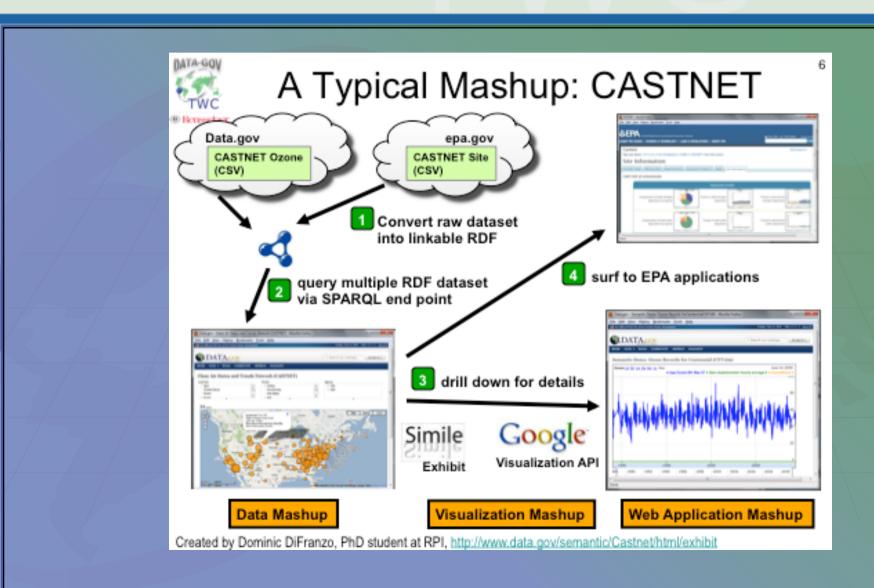














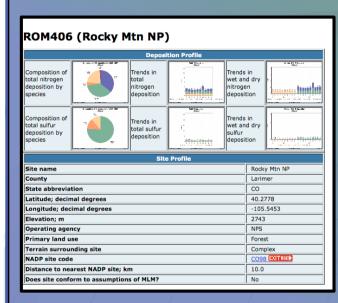
Data.gov + epa.gov



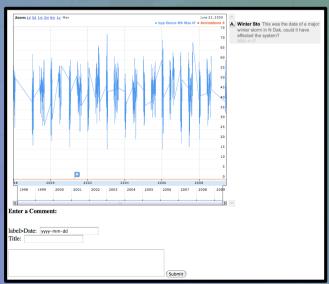


Adding some Web magic

Tetherless World Constellation



Web Analytics

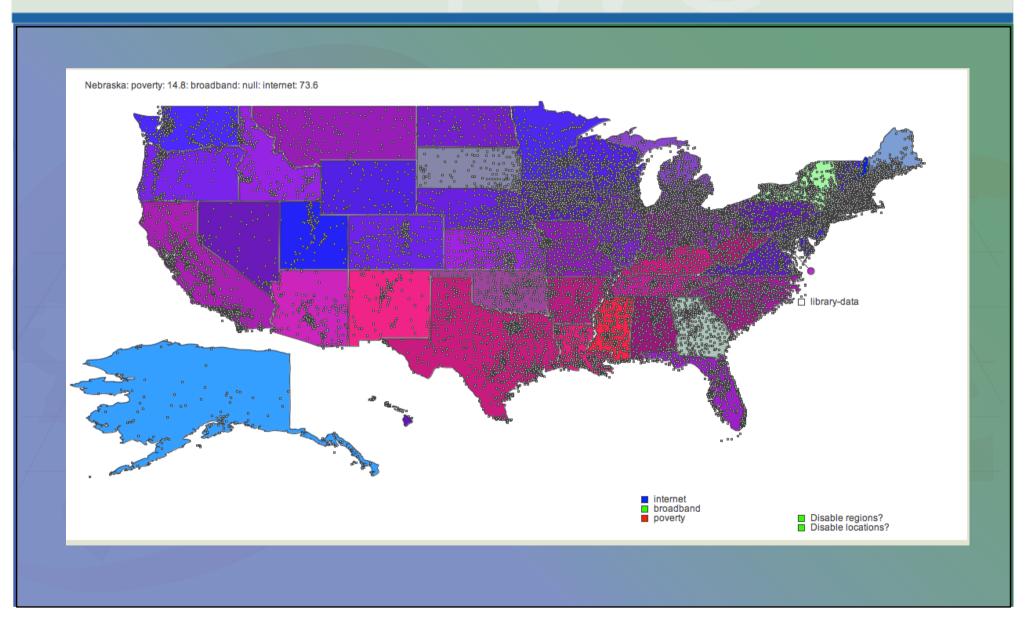


External Links

Social Data Networks

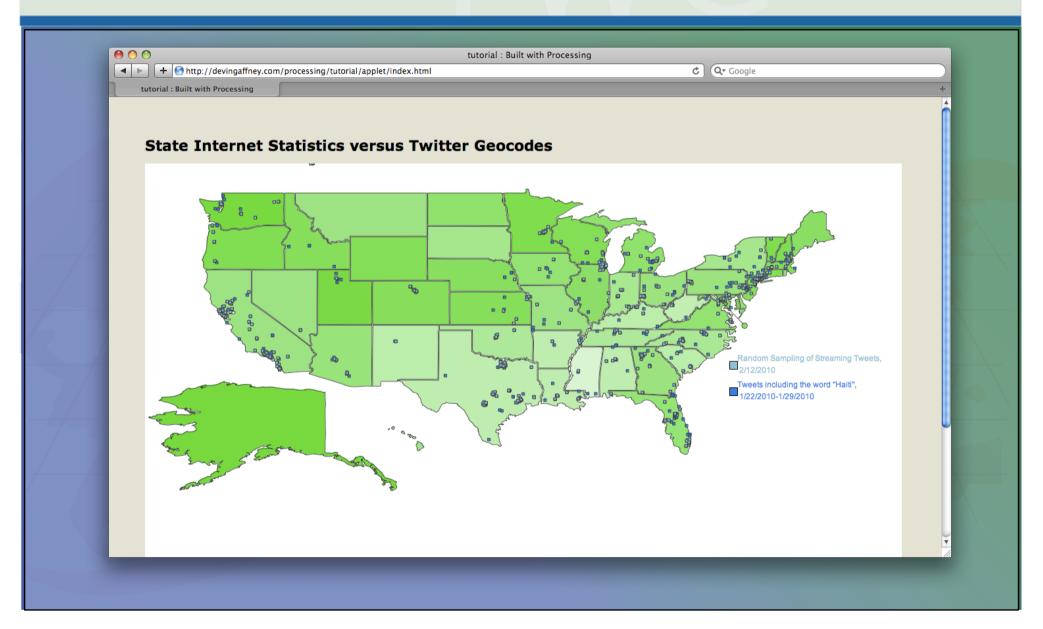


NTIA internet study vs. libraries





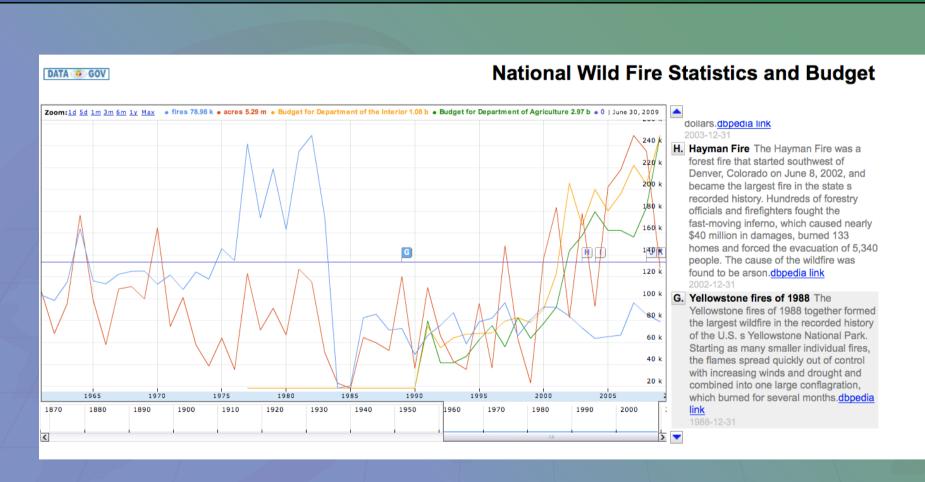
NTIA internet funding vs. tweets about #haiti





Visualization can help identify data errors

Tetherless World Constellation



Correlates fires, acres burned, and agency budgets

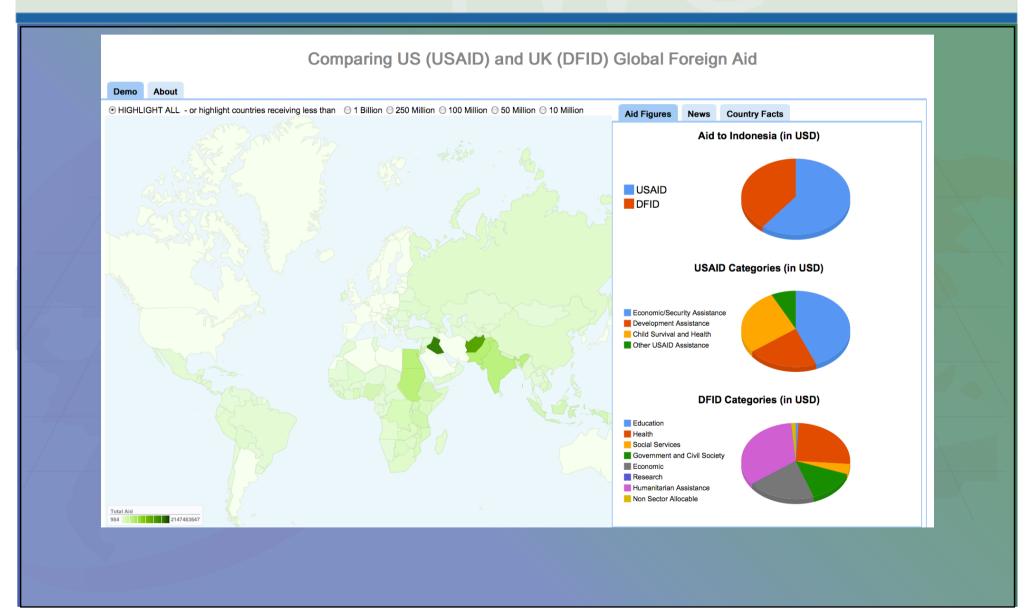


Visualization can help identify data errors



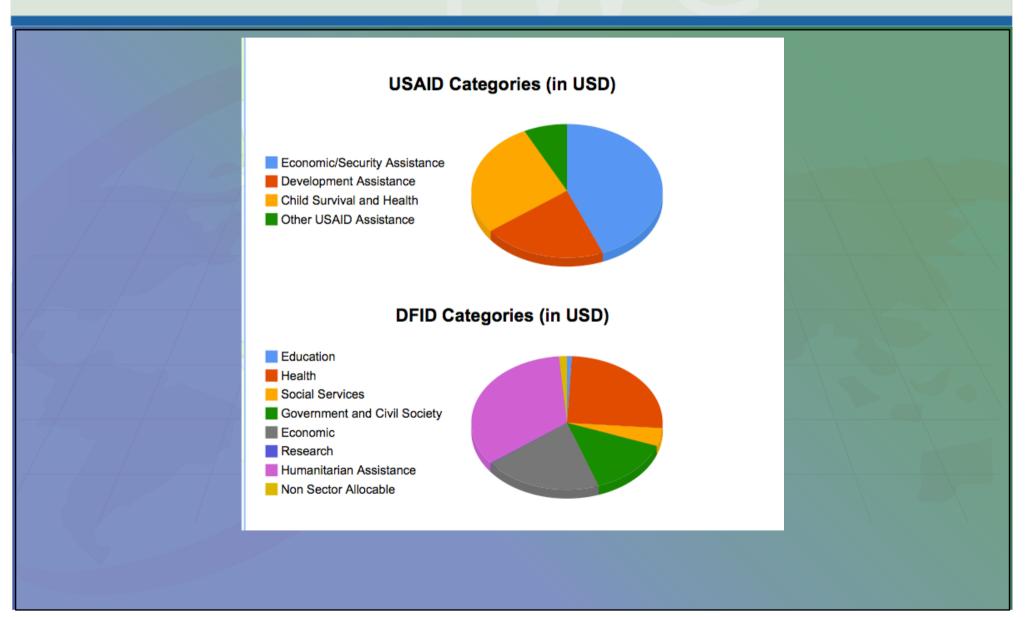


Combining data from different sites





Different ontologies





This on the Web takes us into "Web 3.0"

Tetherless World Constellation

Web 3.0

Web 2.0

Semantic Web (RDFS, owl)

Linked Data (RDF, SPARQL)

Web (REST API)

Web 3.0 extends current Web applications using Semantic Web technologies and graph-based, open data.



Semantic Search

Tetherless World Constellation

WEB TECHNOLOGIES

Web 3.0: The Dawn of Semantic Search

James Hendler, Rensselaer Polytechnic Institute



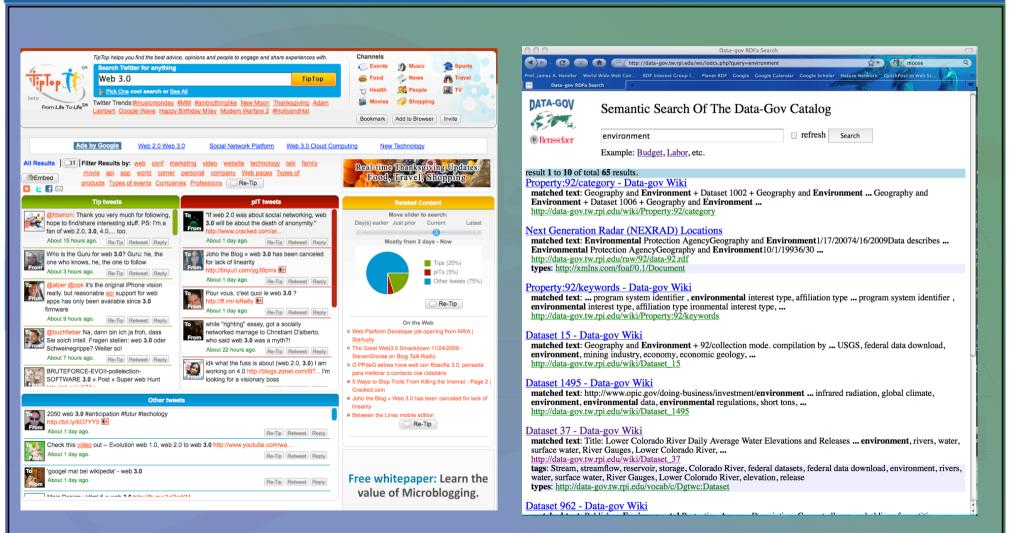
Emerging Web 3.0 applications use semantic technologies to augment the underlying Web system's functionalities.

IEEE Computer, Jan 2010; IEEE Computing Now, Feb 2010 (free)



Semantic Search

Tetherless World Constellation

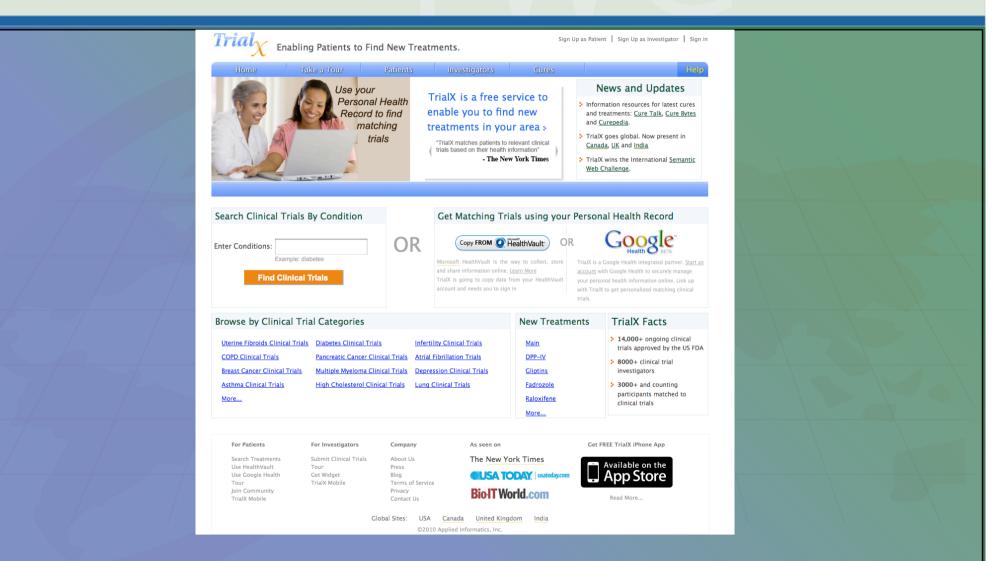


Semantic Search Powered by RDFa



Trialx.com

Tetherless World Constellation

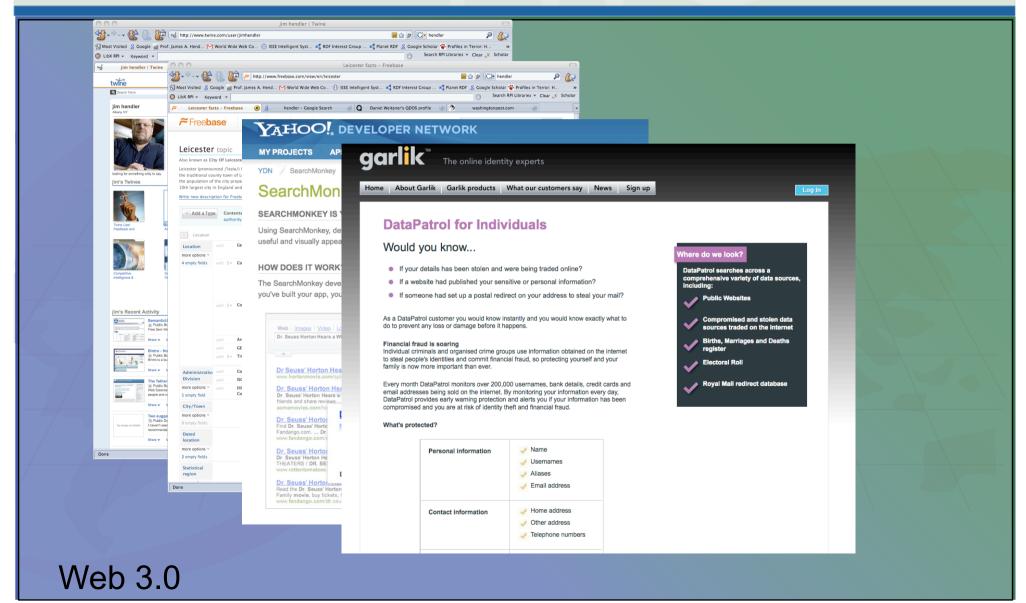


Save lives



Lots More

Tetherless World Constellation

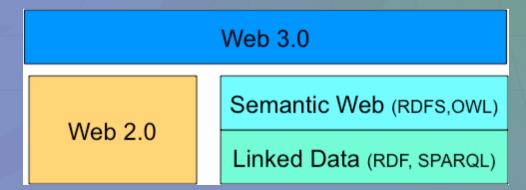


Applications



Web 3.0 excitement (hype?)

- Significant and growing commercial interest...
 - Web: Google, Amazon, Travelocity...
 - Web 2.0: Facebook, Wikipedia, YouTube, Twitter...
 - Web 3.0: the big ones are still out there





Summary

- The Semantic Web is going just fine thank you
 - People asking "how," not why
- So far the commercial driver has been "weak semantics"
 - In the enterprise
- Web 3.0 adds semantics as a value add to regular Web functionality
 - Semantic search
 - Semantic match
 - Semantic etc
- The big one is still out there