



TWC

Tetherless World Constellation

# Semantic Web: “10 year update”

**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)



Original Outline (July 2000)

Nevertheless World Constellation

**Scientific American Article notes**

**[Joint starting place:]**

- I. Semantic Web Vision (TBL)
  - II. What are the enablers? (in sequence)
    - Screen Scraping (Ora and TBL)
    - Data on Web (Ora and TBL)
    - Zip code link between Data Bases (TBL)
    - Ontology Independence (JAH)
- Effect of Scale (TBL)

**“Then, a miracle occurs”**

- III. What can you do with it? (not necessarily in sequence)
  - Self-describing documents (JAH)
  - Logic to encode... (TBL)
  - Services and Advertising (Ora)
  - Devices (Ora)
  - Digital Signatures, Authentication, and Trust (TBL)



(May 21, 2001)



SEMANTIC WEB & LINKED DATA  
BUSINESS STRATEGIES



JupitermediaEvents

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, game-changing 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**



Sem Web 2010

Tetherless World Constellation

**InfoQ**  
news

446,480 Jul unique visitors

Architectures you've  
always wondered about  
NOV 1-5 2010 SAN FRANCISCO

Register  
Login  
About us  
QCon

#### Your Communities

- Java
- .NET
- Ruby
- SOA
- Agile
- Architecture
- Operations

Search

#### Featured Topics

Application  
Servers

Hadoop

RIA

Tracking change and innovation in the enterprise software development community

News

Contribute News print

## Facebook's Graph API: The Future Of Semantic Web?

Posted by Dilip Krishnan on Apr 24, 2010

Community SOA Topics REST Tags Social Networking , Semantic Web , Facebook

Share

"There are two important themes behind everything we're delivering today," says Bret Taylor, head of Facebook's platform products [in the facebook developer blog](#), about the recent announcements at the f8 conference in San Francisco. Facebook introduced Open Graph protocol, and the Graph API as the next evolution in the Facebook platform.

First, the Web is moving to a model based on the connections between people and all the things they care about. Second, this connections-based Web is well on its way to being built and providing value to both users and developers — the underlying graph of connections just needs to be mapped in a way that makes it easy to use and interoperable.

Facebook introduced three new components of Facebook Platform two of which the Open Graph protocol, and the Graph API. The API provides access to Facebook objects like people, photos, events etc. and the connections between them like friends, tags, shared content etc. via a uniform and consistent URI to access the representation. Every object can be accessed using the the URL <https://graph.facebook.com/ID>, where ID stands for the unique ID for the object in the social graph. Every connection (CONNECTION\_TYPE) that the facebook object supports can be examined using the URL [https://graph.facebook.com/ID/CONNECTION\\_TYPE](https://graph.facebook.com/ID/CONNECTION_TYPE).

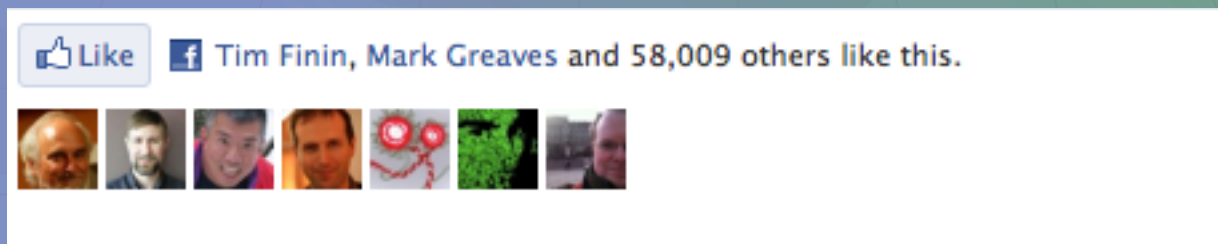
April 2010



## Facebook's Open Graph Protocol

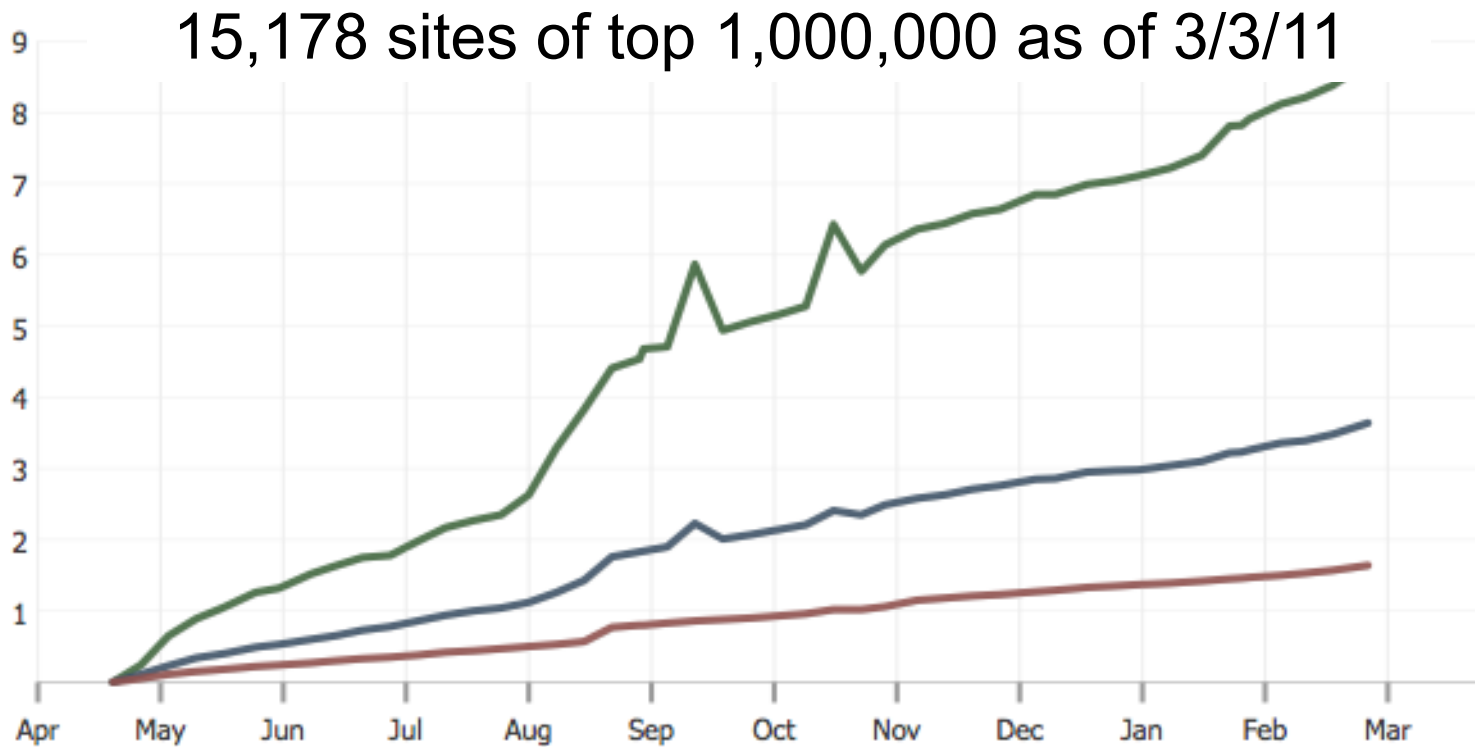
### Tetherless World Constellation

- Your Documents (XML, HTML, XHTML) contain RDFa with some FB specific vocabulary (+ links!!)
  - og:title - The title of your object as it should appear within the graph, e.g., "The Rock".
  - og:type - The type of your object, e.g., "movie". Depending on the type you specify, other properties may also be required.
  - og:image - An image URL which should represent your object within the graph.
  - og:url - The canonical URL of your object that will be used as its permanent ID in the graph
  - og:description - A one to two sentence description of your object.
  - og:site\_name - If your object is part of a larger web site, the name which should be displayed for the overall site. e.g., "IMDb".





OGP use growing quickly  
Facebook incentivizing use of RDFa like buttons  
Tetherless World Constellation



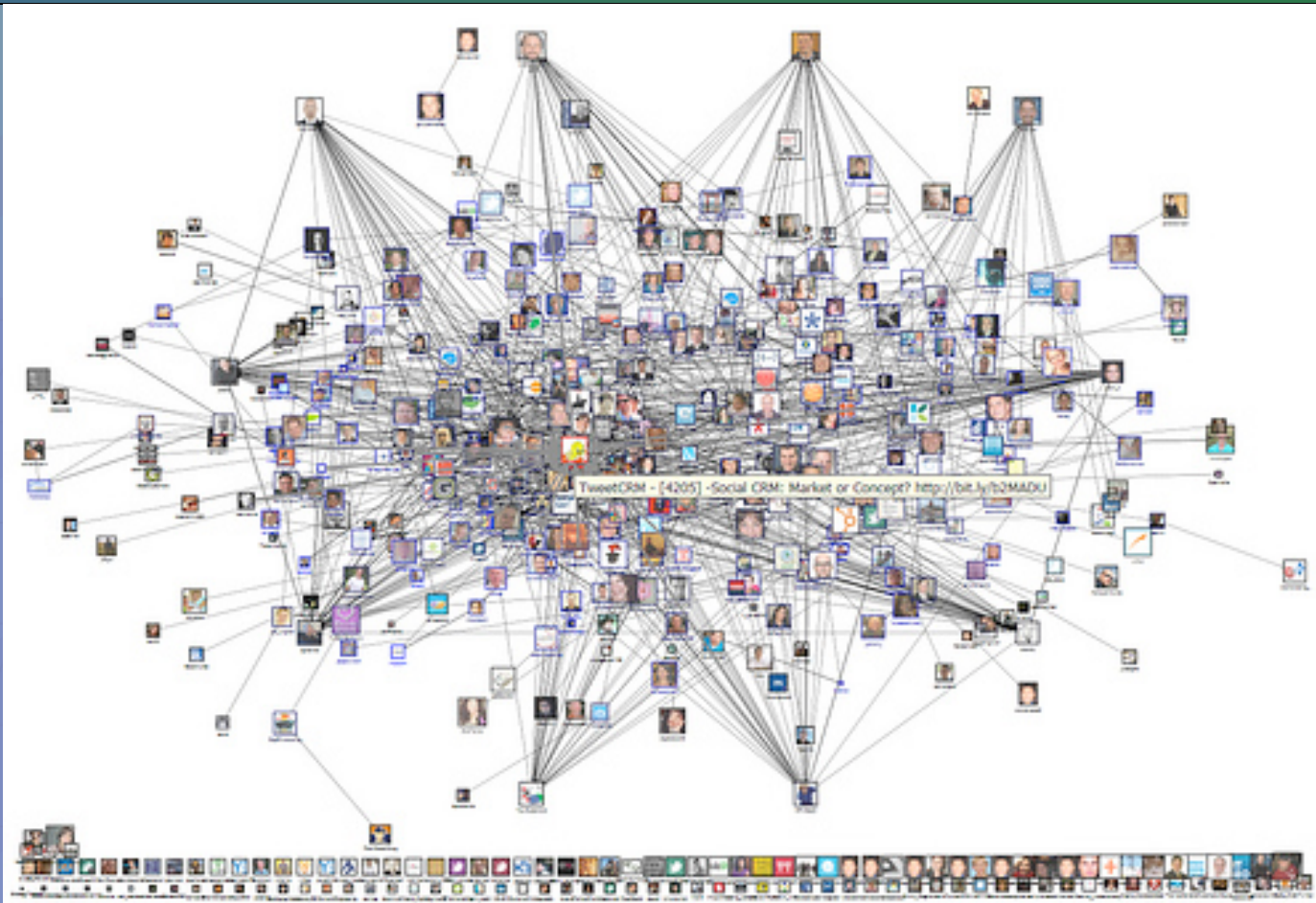
FB reports ~ 10-15% of > 3,000,000 likes per day!

Why are they pushing developers to use the RDFa version?



Because we need the links!

Tetherless World Constellation



The network of likes is where their money is made!  
(predicted >\$5B of advertising in next two years)



Creates a platform for SW-powered apps

Tetherless World Constellation

# SocialWire

facebook

Powered Recommendation Engine for E-Commerce



## PERSONALIZE

Instantly personalizes on-site shopping experience via socially engaging product recommendations.



## ENGAGE

Shoppers get birthday reminders, discover gift ideas, recommend, ask and compare products with their friends.



## RETARGET

Retargets shoppers and their friends on Facebook via personalized, product level advertising.

## Truly Plug and Play

### Step 1 - Update Header

```
<head>
...
<meta property="og:title"
<meta property="sw:price"
...
</head>
```

### Step 2 - Paste Javascript



### Step 3 - Drop Plugins



## Compatible with Facebook Open Graph



Integrating **SocialWire** is as simple as adding a Facebook Like Button to your pages.





# Semantic Web 2010

## Tetherless World Constellation


The New York Times **Technology**

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

Search Technology  Go


Inside Technology **Bits Blog** »

Internet Start-Ups Business Computing Companies


 Warren Brown  
Founder of CakeLove,  
and recipient of Dell's  
IT Makeover.

Ultra-secure Dell re  
Protect your data with the  
team of Dell, Intel® and Tr  
[TakeYourOwnPath.com](#)

### How Best Buy Is Using The Semantic Web

By RICHARD MACMANUS of  ReadWriteWeb  
Published: July 1, 2010

Yesterday we wrote about the increasing usage of [Semantic Web technologies by large commercial companies](#) like Facebook, Google and Best Buy. The [Semantic Web](#) is a Web of added meaning, which ultimately enables smarter and more personalized web apps to be built. In this post we explore how a leading U.S. retailer, [Best Buy](#), is using a Semantic Web markup language called RDFa to add semantics to its webpages.



**More News From ReadWriteWeb**

- Why We Check In: The Reasons People Use Location-Based Social Networks
- Google's Semantic Web Push: Rich Snippets Usage Growing
- How to Jailbreak iOS 4.0 for iPhone 3G
- Can Augmented Reality Help Save the Planet?
- From Your Mouth to a Thousand Ears, New Apps Make Mobile Podcasting Easy

**Technologies**

This is not just an academic exercise for Best Buy. As we will see, semantic technology has already led to increased traffic and better service to its customers. We spoke to [Jay Myers](#), Lead Web Development Engineer at BestBuy.com, to find out how.

ReadWriteWeb's Guide to The Semantic Web:

1. [It's All Semantics: Open Data, Linked Data & The Semantic Web](#)
2. [The State of Linked Data in 2010](#)
3. [Top 10 Semantic Web Products of 2009](#)
4. [ReadWriteWeb Interview With Tim Berners-Lee](#)
5. [Semantic Web Patterns: A Guide to Semantic](#)

Myers told us that the primary goal of using semantic technologies was to increase the visibility of its products and services. And with data such as store name, address, store hours and GEO data being marked up using RDFa, search engines are now able to identify each of those data components more easily and put them into context.

July 2010



# TW/C

## Sem Web 2010

### Tetherless World Constellation

**internetnews.com** Blogs

REALTIME IT NEWS Search N

Hardware Software Mobility Web Content **Search** Government Developer Business Storage

---

**Free Download: BlackBerry® Enterprise Server Express.** It works with any Internet-enabled BlackBerry data plan & enables organizations to deploy the BlackBerry solution without software or licensing costs.

[InternetNews.com](#) >> [Search](#)

### Google Snaps Up Metaweb in Semantic Web Play

Acquisition of firm offering a free, open source database of real-world information figures to inform smarter search as Google builds out semantic technology.

July 19, 2010  
By Kenneth Corbin: [More stories by this author:](#)

[Email](#) [Print](#) [Comment](#) [Share](#)

*Google's shopping spree doesn't show any signs of letting up, as the Web behemoth continues to open its wallet to acquire firms with niche technologies that can augment its developing business lines.*

*The latest target is Metaweb, a firm specializing in semantic Web technology that Google has acquired to improve its search engine. Specifically, Google thinks that Metaweb's database of real-world items can help it produce ordered and accurate results in response to specific, complex search queries. DevX has the details.*

---

In the latest in a string of acquisitions, Google has unveiled plans to purchase Metaweb, a firm that maintains a vast open source database of information about objects in the real world.

Through the acquisition, Google (NASDAQ: GOOG) aims to provide a smarter search engine that will be able to retrieve more specific answers to complex queries.

**Read the full story at DevX:**  
[Google Snags Semantic Web Firm Metaweb](#)

**TAGS:** Google, semantic Web, acquisition, search engine, Metaweb

[Email](#) [Print](#) [Comment](#) [Share](#)

0 Comments ([click to add your comment](#))

July 2010



### Semantic SEO for Google with GoodRelations and RDFa

**Breaking News:** Google has just [announced](#) that they now support the [GoodRelations vocabulary](#) developed by Prof. Hepp for Google Rich Snippets.

If you add a few extra lines to the HTML code of your Web site, you will feed Google and hundreds of evolving mobile and recommender applications in one turn.

On this page, we provide **slightly modified markup** that is **superior to the original Google pattern** for marking up your pages. Our recipe, developed by the inventor of GoodRelations for you personally, is understood by ALL RDFa-aware search engines, shopping comparison sites, and mobile services, while the Google recipe is perfect for Google only.

#### Principle and Effect

If you add a little bit of additional mark-up to your page, Google will use that information to substantially enhance the rendering of your page directly in the search results list.

There is preliminary evidence that search results with respective extensions get a **30 % higher click-through rate (CTR)**.

The flickering nature of search results in the new Google Instant technology will very likely further increase the gains of more appealing rendering in the results list. So don't put this off. **Act today!**

#### Without GoodRelations in RDFa

##### [Hepp Research Personal SCSI Controller Card](#)

The Hepp Research Personal SCSI is a 16-bit add-on card that allows attaching up to seven SCSI devices to...

[www.heppresearch.com/commercecollator](http://www.heppresearch.com/commercecollator) - [Cached](#) - [Similar](#)

#### With GoodRelations in RDFa

##### [Hepp Research Personal SCSI Controller Card](#)

★★★★★ 99 reviews - \$99.99 - in stock

The Hepp Research Personal SCSI is a 16-bit add-on card that allows attaching up to seven SCSI devices to your computer. Designed in 1991 by Martin Hepp, the maker of the GoodRelations vocabulary for e-commerce.

[www.heppresearch.com/commercecollator](http://www.heppresearch.com/commercecollator) - [Cached](#) - [Similar](#)



Nov 4, 2010



# TW/C

Sem Web 2010

Tetherless World Constellation

## Semantic Web

Semantic Web  
SUMMIT

Semantic Web  
100

### 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.

(Enterprise Sem Web)



## ORACLE DATABASE SEMANTIC TECHNOLOGIES

### SEMANTIC DATABASE FEATURES

#### NEW IN ORACLE SPATIAL 11.2

- Triple-level security with Virtual Private Database and Label Security support
- Semantic indexing for documents based on popular natural language tools
- Change management to version triples within an RDF graph
- Parallel and incremental inference, and owl:sameAs optimization
- Support for the NIH SNOMED comprehensive clinical ontology
- W3C SKOS inference support
- More OWL constructs: union, intersection, oneOf, OWL 2 property chains, disjoint

*As part of Oracle Spatial 11g, an option for Oracle Database 11g Enterprise Edition, Oracle delivers an advanced semantic data management capability not found in any other commercial or open source triple store. With native support for RDF/RDFS/OWL/SKOS standards, this semantic data store enables application developers to benefit from an open, scalable, secure, integrated, efficient platform for RDF and OWL-based applications. These semantic database features enable storing, loading, and DML access to RDF/OWL data and ontologies, inference using RDFS, OWL and SKOS semantics and user-defined rules, querying of RDF/OWL data and ontologies using SPARQL-like graph patterns embedded in SQL, and ontology-assisted querying of enterprise (relational) data.*



## The coming of "Linked Data"

Tetherless World Constellation

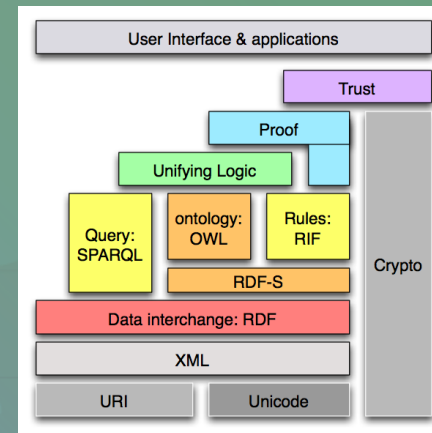
- What is different now?
  - Semantic Search
  - Advertising drives Web markets
  - "Buzz" around data on the Web
    - Esp open govt data
- Maturation of RDF technologies
  - SPARQL endpoints
  - RDFa !!!
  - Lightweight Knowledge
    - A little semantics goes a long way



# The Evolving Web (Technology View)

## Tetherless World Constellation

- Web is powered by the links between documents
  - Google worked because of the link space
- Web 2.0 is powered by "social context"
  - The network effect is in the social network
    - At scale tagging runs into usual vocabulary issues
- Web 3.0 adds data relations and vocabulary links
  - Controlled vocabularies express data relationships
    - Semantic Web standards

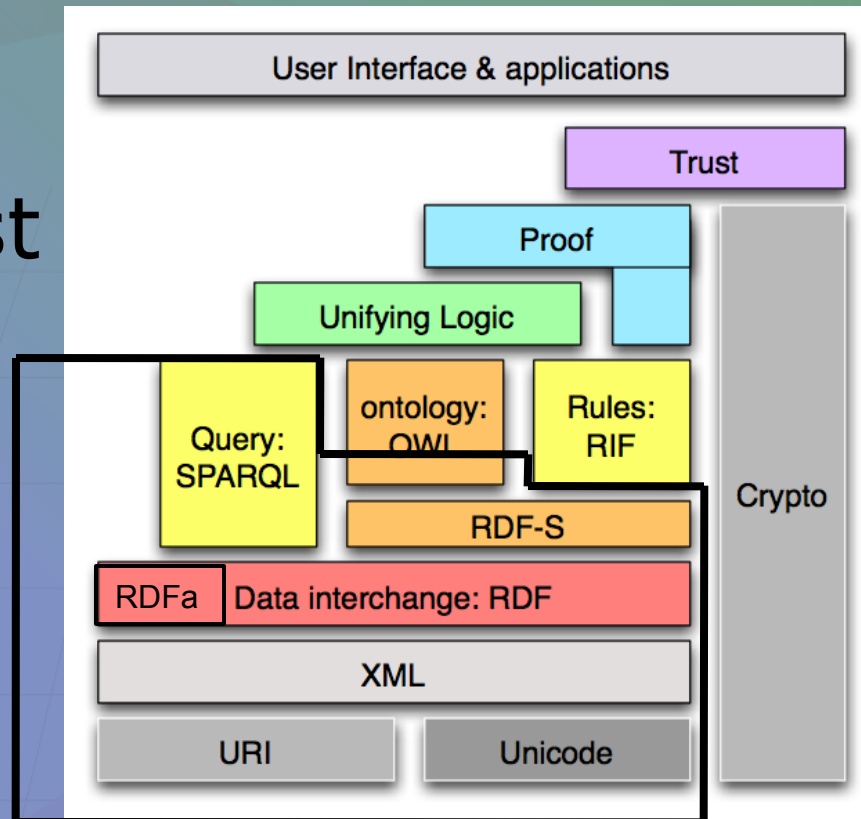




# Maturation of the "bottom" of the Semantic Web

## Tetherless World Constellation

- What is seeing the most use??

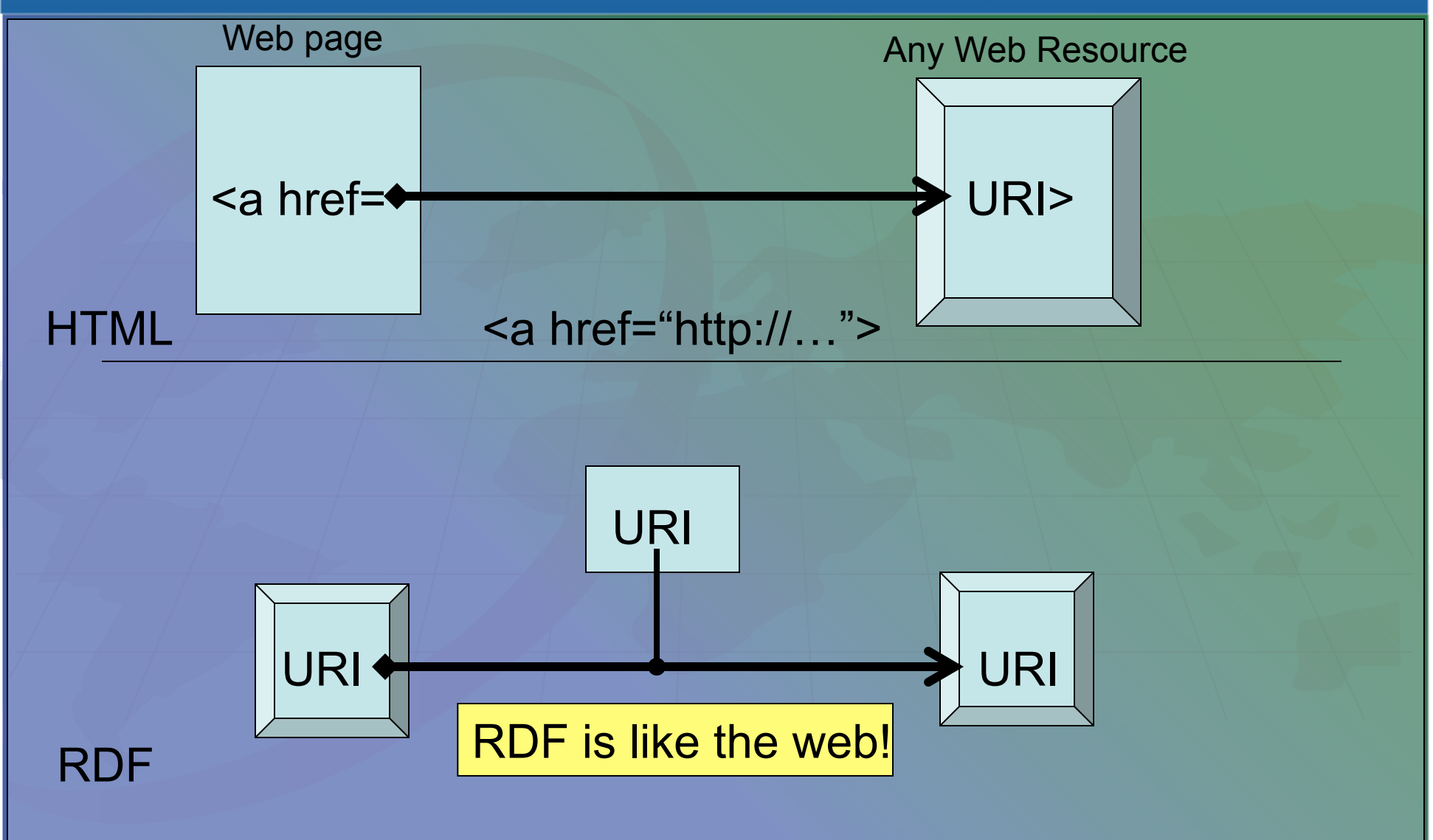






On the Web -- links are critical!

Tetherless World Constellation



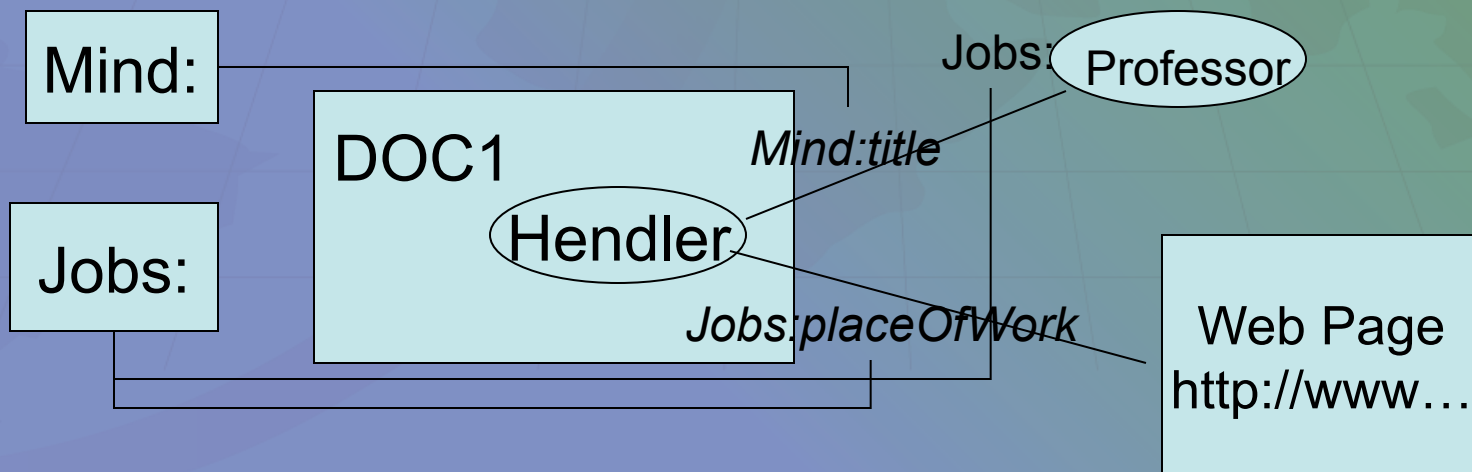


# TW/C

Links in the data

Tetherless World Constellation

```
DOC1
  <mind:Person rdf:id="Hendler">
    <mind:title jobs:Professor>
      <jobs:placeOfWork http://www.cs.rpi.edu>
  </mind:Person>
```





# Directly linking datasets

## Tetherless World Constellation

indice "James Hendler" 🔍

1. <http://southampton.rkbexplorer.com/id/person-0e287d169ba5234bf52a8c...840c1bf1cd>
2. <http://data.semanticweb.org/person/jim-hendler>
3. <http://data.semanticweb.org/person/james-hendler>
4. <http://acm.rkbexplorer.com/id/person-140392-138bb61e61c7ac2750b8a8f91708cb7e>
5. <http://acm.rkbexplorer.com/id/person-140392-4f99f5c802da7ecb9b918188a9cddb2e>
6. <http://acm.rkbexplorer.com/id/person-140392-85039bb5ec9c2d7aef47923683260c74>
7. <http://acm.rkbexplorer.com/id/person-140392-9ea8f3df3b4696b4032d090aea4e4ceff>
8. <http://dblp.rkbexplorer.com/id/people-0e287d169ba5234bf52a8c5d977aa...c7fe085558>
9. <http://dblp.rkbexplorer.com/id/people-0e287d169ba5234bf52a8c5d977aa...128e4fe020>
10. <http://dblp.rkbexplorer.com/id/people-0e287d169ba5234bf52a8c5d977aa...8d183101c0>
11. <http://dblp.rkbexplorer.com/id/people-0e287d169ba5234bf52a8c5d977aa...8ccb070262>
12. <http://dblp.rkbexplorer.com/id/people-0e287d169ba5234bf52a8c5d977aa...9ceaca902f>
13. <http://dblp.rkbexplorer.com/id/people-0e287d169ba5234bf52a8c5d977aa...a3b8c51f3c>
14. <http://dblp.rkbexplorer.com/id/people-0e287d169ba5234bf52a8c5d977aa...de0ebcebab>
15. <http://dblp.rkbexplorer.com/id/people-0e287d169ba5234bf52a8c5d977aa...6710baa554>
16. <http://citeseer.rkbexplorer.com/id/resource-CSP127877-385a279b0613b...d8b1d32fe0>
17. <http://citeseer.rkbexplorer.com/id/resource-CSP127877-f41444578a586...157ac31906>
18. <http://citeseer.rkbexplorer.com/id/resource-CSP127878-68517b365194b...da9cb531e3>
19. <http://citeseer.rkbexplorer.com/id/resource-CSP127879-25d38da0800f6...78fc33297d>
20. <http://citeseer.rkbexplorer.com/id/resource-CSP127880-1827adf5de992...7db6aed90c>
21. <http://citeseer.rkbexplorer.com/id/resource-CSP127880-321b77cdd90a4...8982b45892>
22. <http://citeseer.rkbexplorer.com/id/resource-CSP127880-496bcdd5fea5...0820253357>
23. <http://citeseer.rkbexplorer.com/id/resource-CSP127880-8a04eb6be0af0...bd2fb792ea>
24. <http://citeseer.rkbexplorer.com/id/resource-CSP127880-baad48a3d28a0...7aeaa89c7d>
25. <http://eprints.rkbexplorer.com/id/kfupm/person-0e287d169ba5234bf52a...8982b45892>
26. <http://eprints.rkbexplorer.com/id/kfupm/person-0e287d169ba5234bf52a...cf6ece8581>
27. <http://eprints.rkbexplorer.com/id/kfupm/person-0e287d169ba5234bf52a...bff2762db3>
28. <http://eprints.rkbexplorer.com/id/webscience/person-0e287d169ba5234...443a624896>
29. <http://eprints.rkbexplorer.com/id/webscience/person-0e287d169ba5234...632b8d22d8>
30. <http://eprints.rkbexplorer.com/id/edshare.soton.ac.uk/person-0e287d...ea9dff9a18>
31. <http://eprints.rkbexplorer.com/id/edshare.soton.ac.uk/person-0e287d...d860ce5080>
32. <http://eprints.rkbexplorer.com/id/edshare.soton.ac.uk/person-0e287d...245757676a>
33. <http://eprints.rkbexplorer.com/id/edshare.soton.ac.uk/person-0e287d...8c8ce31744>
34. <http://eprints.rkbexplorer.com/id/edshare.soton.ac.uk/person-0e287d...5e6cf14e15>
35. <http://dotac.rkbexplorer.com/id/www.edshare.soton.ac.uk/person-0e28...f3caea2809>
36. <http://dotac.rkbexplorer.com/id/www.edshare.soton.ac.uk/person-0e28...7053208d59>
37. <http://dotac.rkbexplorer.com/id/www.edshare.soton.ac.uk/person-0e28...9cccf907b>
38. <http://dotac.rkbexplorer.com/id/www.edshare.soton.ac.uk/person-0e28...365bbe94d>
39. <http://dotac.rkbexplorer.com/id/www.edshare.soton.ac.uk/person-0e28...e2673571da>
40. <http://eprints.ecs.soton.ac.uk/id/person/ext-b908909d811e5af11bf5735440b297e1>
41. [http://kisti.rkbexplorer.com/id/PER\\_0000000000000066693](http://kisti.rkbexplorer.com/id/PER_0000000000000066693)
42. [http://kisti.rkbexplorer.com/id/PER\\_0000000000000066695](http://kisti.rkbexplorer.com/id/PER_0000000000000066695)
43. <http://semantictweet.com/jahendler#me>
44. <http://www.cs.umd.edu/~hendler/2003/foaf.rdf#hendler>

<sameAs>

indice "http://dbpedia.org/resource/Macao" 🔍

<sameAs>

1. <http://dbpedia.org/resource/RAEM>
2. <http://dbpedia.org/resource/Aomen>
3. <http://dbpedia.org/resource/Aumen>

Show 86 more

- 90. <http://www4.wiwiss.fu-berlin.de/factbook/resource/Macao>

rdf+xml · n3 · json · text

indice "http://dbpedia.org/ontology/SoccerPlayer" 🔍

<sameAs>

1. <http://dbpedia.org/resource/Fitba>
2. <http://dbpedia.org/resource/Futbol>
3. <http://dbpedia.org/resource/Soccer>

Show 123 more

- 127. [http://www.w3.org/2006/03/wn/wn20/instances/synset-football\\_player-noun-1](http://www.w3.org/2006/03/wn/wn20/instances/synset-football_player-noun-1)

rdf+xml · n3 · json · text

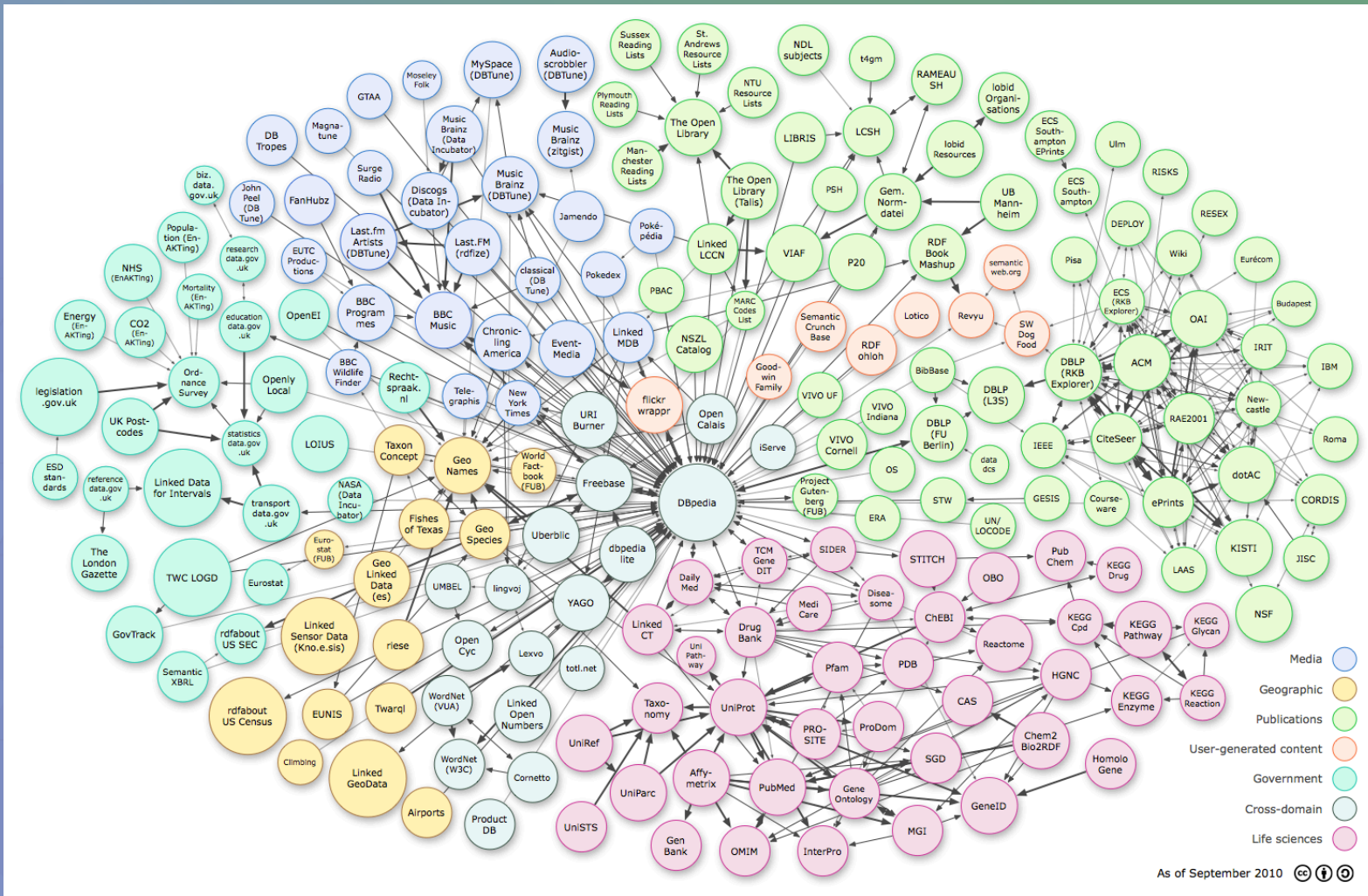


Sindice.com



# Linked Data is entering many sectors

## Tetherless World Constellation



Linkeddata.org 25 billion links



# What about ontologies?

Tetherless World Constellation

- *Consider, eg,* 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
- Compare to OGP



TWC

Widely varying use

Tetherless World Constellation

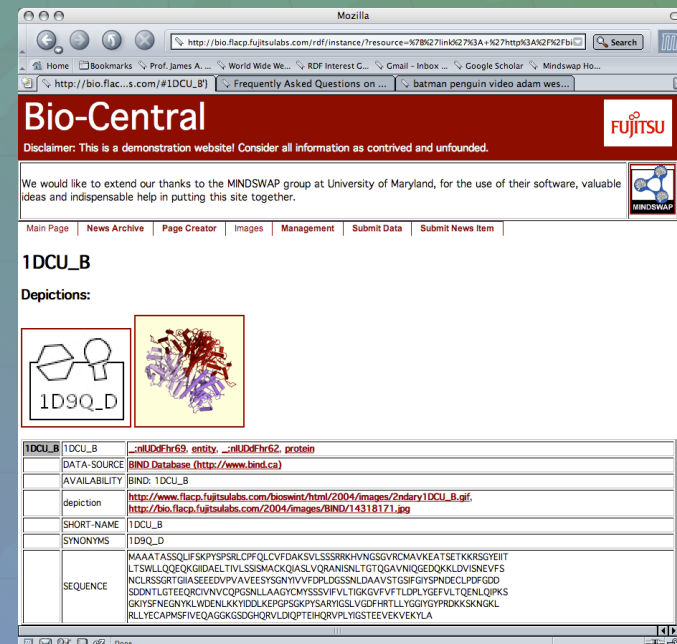
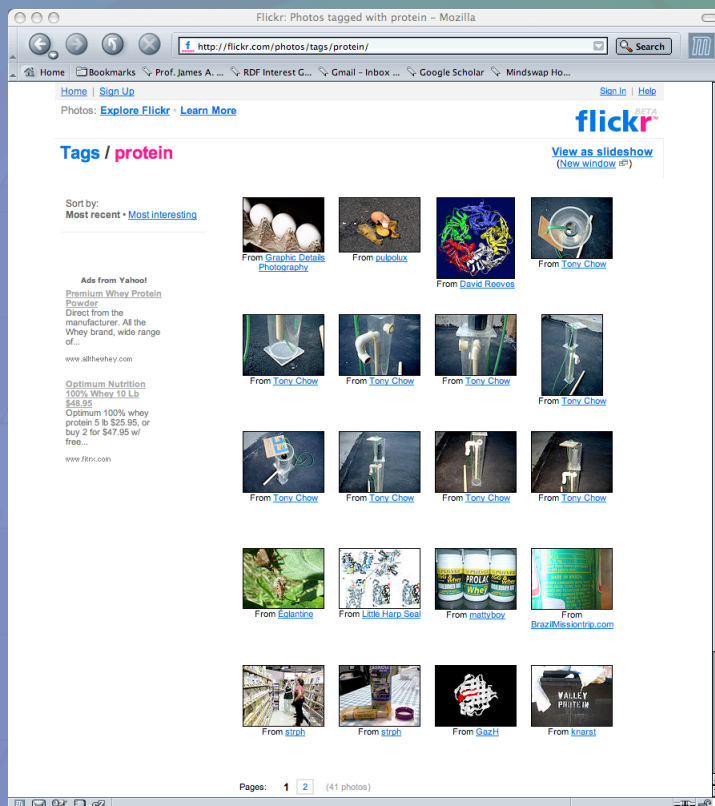
- NCBI Oncology Ontology
  - “High use” in medical community (~1200 users)
  - Very “trusted” information (provenance from NCBI)
  - Primarily terminological (relationships between cancer-related concepts), not data-oriented
- Compare to OGP
  - Hundreds of millions of users
    - Generating >1M triples/day



# 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 VERY 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
- **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





TWC

# The alternative

Tetherless World Constellation

- Linked Data approach 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

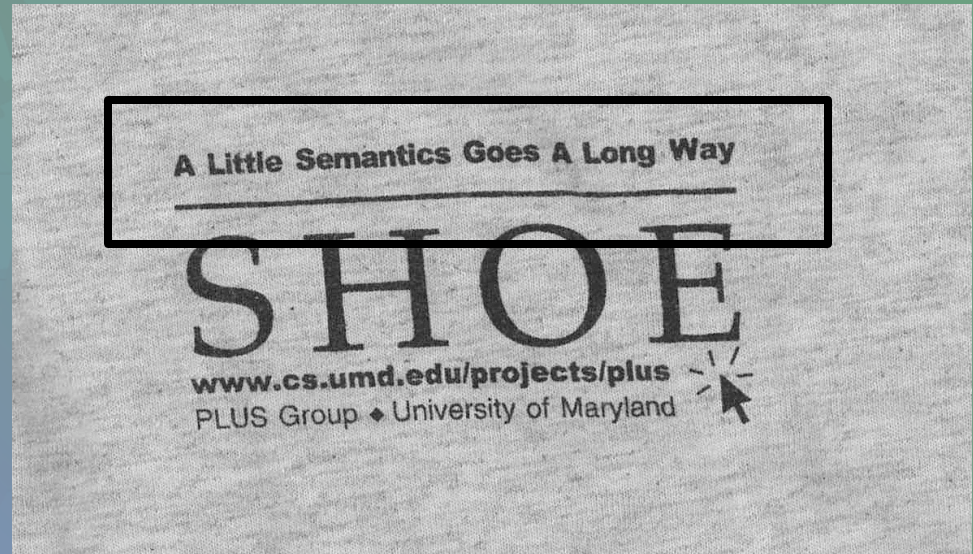
Analogy: the World Wide Web



## Linked Data + Semantics

Tetherless World Constellation

- "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!**



<http://www.cs.rpi.edu/~hendler/LittleSemanticsWeb.html>

# Example: Government Data on the Web

**Data.gov - Mozilla Firefox**  
http://www.data.gov/

FRIDAY, AUGUST 07, 2009

## DATA.GOV

HOME | CATALOGS | STATE/LOCAL | ABOUT | FAQ | CONTACT US | SUGGEST OTHER DATASETS

### DISCOVER. PARTICIPATE. ENGAGE.

Search the following Data.gov catalogs:

- CSV
- XHTML
- KML
- SHP
- "RAW" DATA CATALOG
- TOOL CATALOG
- GEODATA CATALOG

**FEATURED TOOL: U.S. GEOLOGICAL SURVEY (USGS) USGS Global Visualization Viewer for Aerial and Satellite Data**

Ten million archive images of the Earth's surface are available for immediate selection and free download via the USGS Earth Resources Observation and Science (EROS) Center's Global Visualization Viewer. Users can preview thumbnails, browse images and download full-image selections from 1.5 million aerial photos of U.S. sites and 8.5 million images captured worldwide by U.S. Earth-observing satellites.

VIEW THIS TOOL >

1 2 3 4 5

### Welcome to Data.gov

The purpose of Data.gov is to increase public access to high value, machine readable datasets generated by the Executive Branch of the Federal Government. Although the initial launch of Data.gov provides a limited portion of the rich variety of Federal datasets presently available, we invite you to actively participate in shaping the future of Data.gov by suggesting additional datasets and site enhancements to provide seamless access and use of your Federal data. Visit today with us, but come back often. With your help, Data.gov will continue to grow and change in the weeks, months, and years ahead.

### How to use Data.gov

Data.gov includes searchable **data catalogs** providing access to data in three ways: through the "raw" data catalog, the tool catalog and the geodata catalog. Please note that by accessing datasets or tools offered on Data.gov, you agree to the **Data Policy**, which should read before accessing any dataset or tool. If there are additional datasets that you would like to see included on this site, please **click here**. For more information on how to use Data.gov, **view our tutorial**.

DATA.GOV Data Policy | Accessibility | Contact Info | Privacy Policy

**Recovery.gov - Mozilla Firefox**  
http://www.recovery.gov/

RECOVERY.GOV

HOME | ABOUT | BOARD | INVESTMENTS | OPPORTUNITIES | IMPACT | NEWS | FAQ | CONTACT US

### CHAIRMAN'S CORNER

Earl E. Devaney is chairman of the Recovery Accountability and Transparency Board. The Recovery Board oversees Recovery Act spending and manages this website. His next report to the American people will focus on development of the new state-of-the-art Recovery.gov.

### FORESTRY PROJECT CREATES JOBS

The Nevada Division of Forestry has begun a \$1.3 million Recovery-act funded project in Lincoln County, creating 26 jobs and saving three. The project will reduce hazardous forest fuels through the removal of Pinyon and Juniper trees encroaching on Highway 93 and State Route 319.

LEARN MORE >

**Data and Statistics - General Reference Resources: USA.gov - Mozilla Firefox**  
http://www.usa.gov/Topics/ReferenceResources

Home | Site Index | FAQs | E-mail | Phone | Chat | Our Blog | Mobile | Español | Other Languages

1 (800) FED INFO | 1 (800) 333-4636

For Citizens | For Businesses and Nonprofits | For Government Employees | For Visitors to the U.S.

### Data and Statistics - General Reference Resources

Home > Reference Center > Data and Statistics

- 2002 Census of Governments
- Aging Statistics
- American Factfinder
- Budget of the U.S. Government
- Budget of the U.S. Government: State-by-State Tables FY 2009
- Census Data
- Child and Family Statistics

Change text size: A A A

- E-mail this page
- Print this page
- Receive updates by e-mail
- USA.gov RSS feeds
- Tutorials: Find government info

SHARE

Featured Sites

- Data.gov

More Reference Materials

**Federal IT Dashboard - Mozilla Firefox**  
http://it.usaspending.gov/

FRIDAY, AUGUST 07, 2009

## IT DASHBOARD

HOME | INVESTMENTS | DATA FEEDS | ANALYSIS | FAQ |

### your window into the federal IT portfolio

(\$) Billions

Agency: DHB HHS Treasury DOC DOT DOJ VA USDA Energy Others

### Department of Defense

Major Investments: 60  
Spending on Major Investments: \$5.6 B (FY 2009)

Overall Rating: \$

Agency Investment Spotlight



# Government Data Sharing

## Tetherless World Constellation



January 1, 2009  
 “Openness will strengthen our democracy and promote efficiency and effectiveness in Government.”  
 --- President Obama



May 21, 2009  
 data.gov online  
 57 Data Sets



December 8, 2009  
 “Open Government Directive” released  
 ~2000 Data Sets



May 21, 2010  
 data.gov relaunch with semantic web featured  
 >305,000 Data Sets

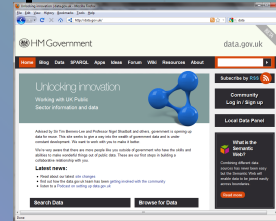
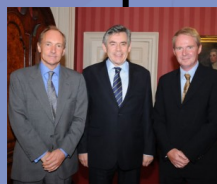
2009

2010 ...

June 30, 2009  
 Putting Govt Data online-  
 Data.gov.uk beta

January 19, 2010  
 data.gov.uk online

~6000 Data Set





# Data.gov community: International

## Tetherless World Constellation



### Examples:

US	305,000
Japan	30,184
Denmark	17,086
UK	6,000
Korea	833
Australia	700
World Health Org	400
Ireland	263
Catalonia	246

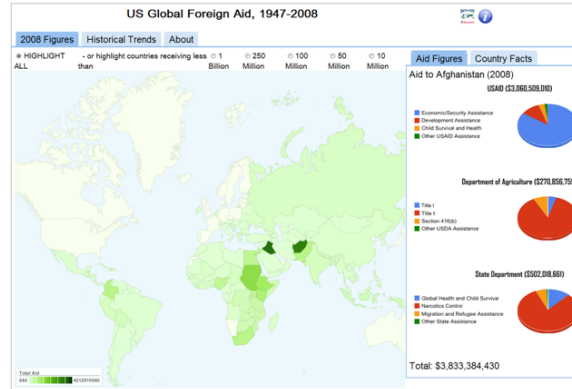


# Creating/Using Data "app" technologies

## Tetherless World Constellation



(a) White House visitor search



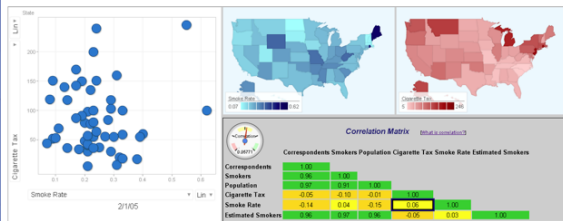
(b) US-UK Foreign Aid Comparison



(c) Agency Budget and NYTimes



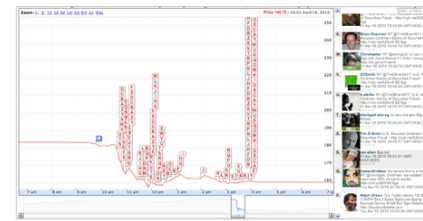
(d) Wildland fire and DBpedia



(e) [Health] Tobacco Prevalence and Correlated Factors



(f) [Policy] About Supreme Court Justices



(g) [Financial] Stock price and Twitter events



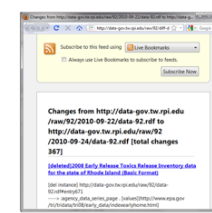
(h) [Yahoo! Pipes] World Earthquake Map



(i) [IBM ManyEyes] White House visitor network



(j) [RDFa] semantic search



(k) [RSS] data.gov updates

See more than 50 of these at <http://logd.tw.rpi.edu>



# Linking GDP of the US and China

## Tetherless World Constellation

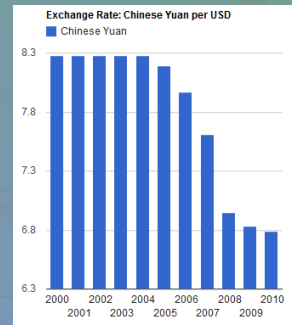
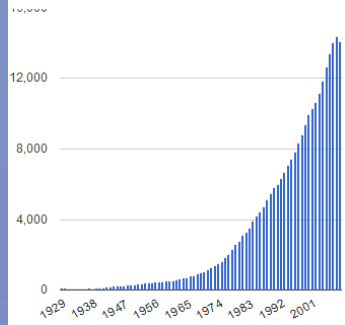


Federal Reserve Statistical Release



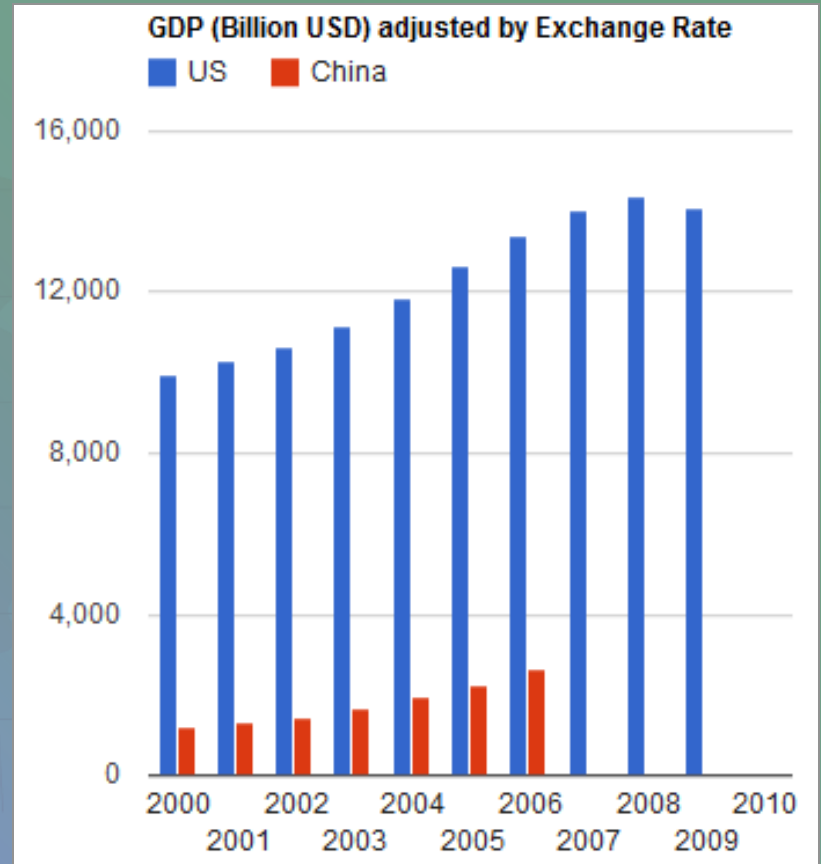
H.10  
Foreign Exchange Rates (Weekly)

GDP of the US (Billion Dollar)

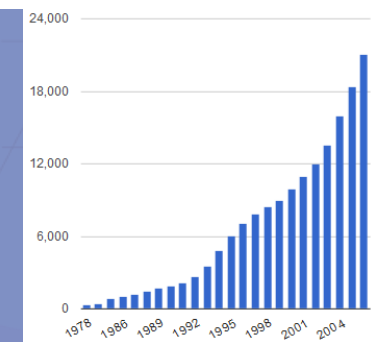


GDP (Billion USD) adjusted by Exchange Rate

■ US ■ China



GDP of China (Billion Chinese Yuan )



[Temporal Mashup] [bea.gov](http://bea.gov) + [federalreserve.gov](http://federalreserve.gov) + [stats.gov.cn](http://stats.gov.cn)



# Linking GDP of the US and China

## Tetherless World Constellation

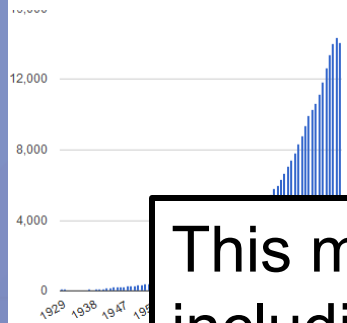


Federal Reserve Statistical Release



H.10  
Foreign Exchange Rates (Weekly)

GDP of the US (Billion Dollar)



GDP (Billion USD) adjusted by Exchange Rate

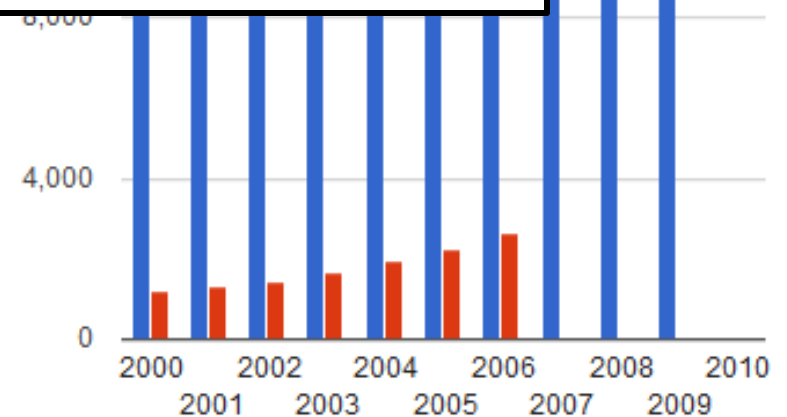
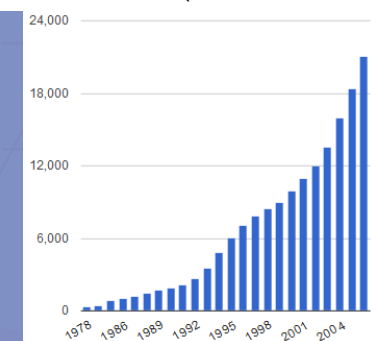
■ US ■ China

16,000

This mashup was built in less than 8 hours – including conversion of data, web interface, and visualization!



GDP of China (Billion Chinese Yuan)



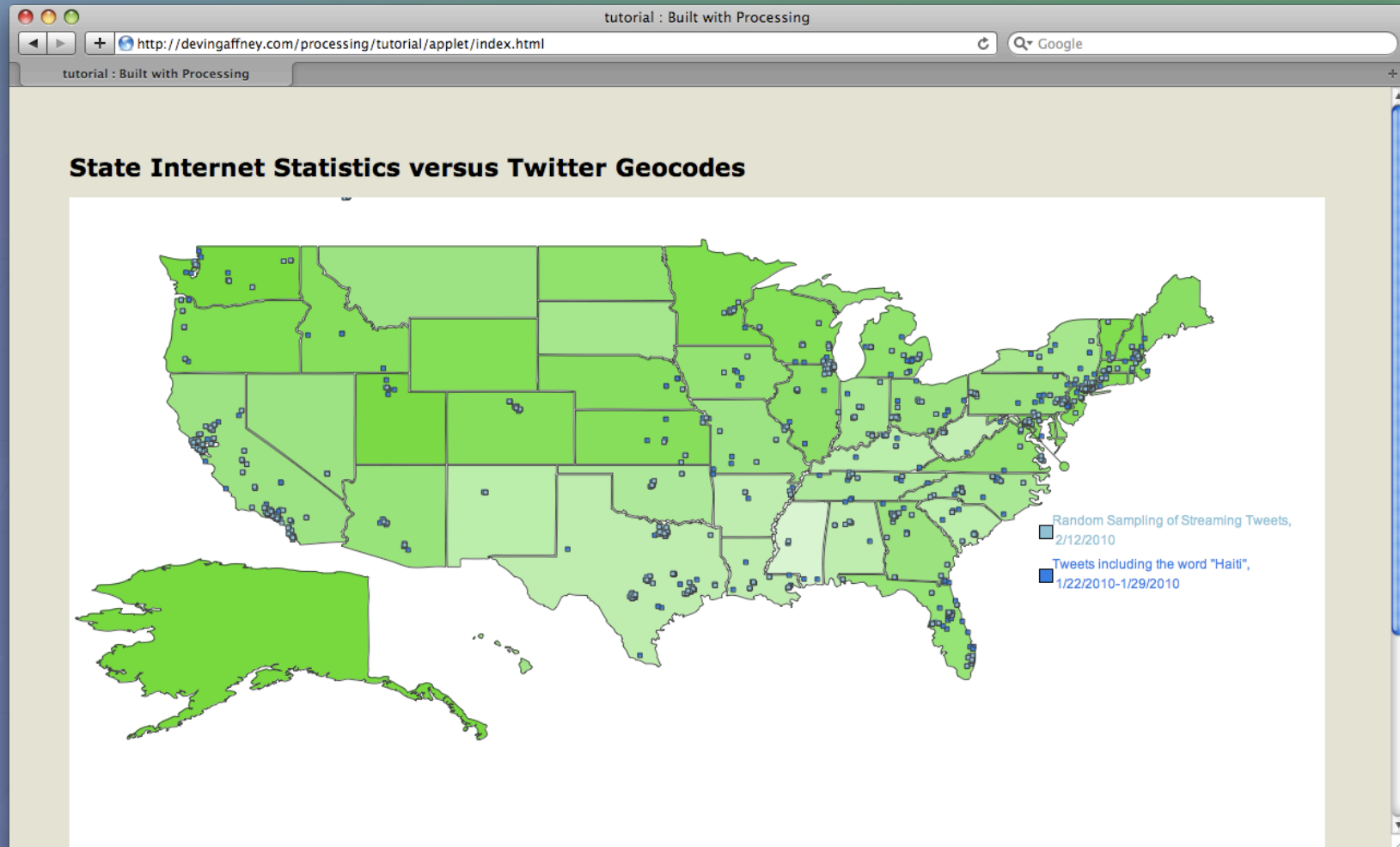
[Temporal Mashup] [bea.gov](http://bea.gov) + [federalreserve.gov](http://federalreserve.gov) + [stats.gov.cn](http://stats.gov.cn)





# Govt data linked to Social Media Metadata

## Tetherless World Constellation





There is a lot of workflow information in the mix

Tetherless World Constellation





# TW/C

## Data Search

### Tetherless World Constellation

**State-related Dataset Faceted Search**

Datasets	States	<input checked="" type="checkbox"/> Agency	Category
1 249	17 Alabama	<input type="checkbox"/> 2 Environmental Protection Agency	2 Geography and Environment
1 307	12 Alaska	<input type="checkbox"/>	
	2 American Samoa	<input checked="" type="checkbox"/>	
	6 Arizona	<input type="checkbox"/>	
	10 Arkansas	<input type="checkbox"/>	
	9 California	<input type="checkbox"/>	

2 Dataset filtered from 21 originally ([Reset All Filters](#))

**249 (link)**

label: 249

type: Dataset

URI: <http://tw2.tw.rpi.edu/data-gov/item/#249>

url: [http://data.gov.tw.rpi.edu/vocab/Dataset\\_249](http://data.gov.tw.rpi.edu/vocab/Dataset_249)

title: 2006 Toxics Release Inventory National data file of all US States and Territories

agency: Environmental Protection Agency

category: Geography and Environment

states: New Jersey, Connecticut, Wisconsin, Illinois, American Samoa, Mississippi, Ohio, Virginia, Louisiana, Nebraska, Idaho, Oklahoma, Delaware, Texas, Indiana, Missouri, Kansas, Guam, Massachusetts, Maryland, Kentucky, Florida, California, Hawaii, Utah, Michigan, Montana, Iowa

2 Dataset filtered from 21 originally ([Reset All Filters](#))

Dataset ID-	URL	Title	Agency	Category	State-related Predicates
249	<a href="http://data.gov.tw.rpi.edu/vocab/Dataset_249">http://data.gov.tw.rpi.edu/vocab/Dataset_249</a>	2006 Toxics Release Inventory National data file of all US States and Territories	Environmental Protection Agency	Geography and Environment	<ul style="list-style-type: none"> <li><a href="#">/191/stack_air_emissions_release_pounds</a></li> <li><a href="#">/191/mailling_state</a></li> <li><a href="#">/191/facility_state</a></li> <li><a href="#">/191/total_transferred_off_site_for_further_waste_management</a></li> <li><a href="#">/191/other_on_site_waste_management</a></li> <li><a href="#">/191/total_stack_air_emissions</a></li> </ul>
307	<a href="http://data.gov.tw.rpi.edu/vocab/Dataset_307">http://data.gov.tw.rpi.edu/vocab/Dataset_307</a>	2007 Toxics Release Inventory National data file of all US States and Territories	Environmental Protection Agency	Geography and Environment	<ul style="list-style-type: none"> <li><a href="#">/191/mailling_state</a></li> <li><a href="#">/191/stack_air_emissions_release_pounds</a></li> <li><a href="#">/191/facility_state</a></li> <li><a href="#">/191/total_stack_air_emissions</a></li> </ul>

How can we search for data?



# Metadata is crucial

## Tetherless World Constellation

keyword: alaska

search results (note: This demo only lists the first 200 search results)

**agency**

- 3 Department of Justice
- 1 Department of Labor
- 4 Department of the Interior
- 8 DOC/NOAA/NESDIS/NGDC > National Geophysical Data Center, NESDIS, NOAA, U.S. Department of Commerce
- 25 Environmental Protection Agency
- 3 n/a
- 1 National Oceanic and Atmospheric Administration (NOAA), National Ocean Service (NOS), Office of Coast Survey (OCS)
- 19 NOAA National Oceanographic Data Center
- 1 U.S. Geological Survey
- 5 U.S. Geological Survey, Central Energy Resources Team
- 1 U.S. Geological Survey, Central Energy Resources Team, Data Management Project
- 118 US National Oceanographic Data Center

**category**

- 31 biota
- 1 boundaries
- 19 climatology/Meteorology/Atmosphere
- 2 elevation
- 32 Geography and Environment
- 31 geoscientificInformation
- 1 Health and Nutrition
- 1 Labor Force, Employment, and Earnings
- 5 Law Enforcement, Courts, and Prisons
- 9 n/a
- 2 Natural Resources

**catalog**

- 156 geodata catalog
- 40 Raw Data Catalog
- 4 Tool Catalog

200 Items

dataset	agency	category	catalog	geographic coverage
Western Information Office	Department of Labor	Labor Force, Employment, and Earnings	Raw Data Catalog	<a href="http://logd.tw.rpi.edu/source/data-gov/dataset/92/value-of/geographic_coverage/States_Metro_Areas">http://logd.tw.rpi.edu/source/data-gov/dataset/92/value-of/geographic_coverage/States_Metro_Areas</a>
U. S. Department of Energy (DOE) Categorical Exclusion (CX) Determinations Under the National Environmental Policy Act (NEPA)	Department of Energy	Geography and Environment	Tool Catalog	<a href="http://logd.tw.rpi.edu/source/data-gov/dataset/92/value-of/geographic_coverage/U_S_and_U_S_Territories">http://logd.tw.rpi.edu/source/data-gov/dataset/92/value-of/geographic_coverage/U_S_and_U_S_Territories</a>
Total tonnage of commodities carried on commercial waterways by traffic type	Department of Defense	Transportation	Raw Data Catalog	<a href="http://logd.tw.rpi.edu/source/data-gov/dataset/92/value-of/geographic_coverage/United_States">http://logd.tw.rpi.edu/source/data-gov/dataset/92/value-of/geographic_coverage/United_States</a>
Temperature and salinity profile	NOAA National Oceanographic	oceans	Geodata Catalog	lower corner: 47.66,-158.7 upper corner: 61.07,-122.26

Accession No.: 0001332)

What kinds of metadata are: simple to create, powerful enough for search and internationalizable (esp. beyond English)



Example, integrating data and info search

Tetherless World Constellation

## Search in SciVerse Hub on Climate Change

Home | My settings

Home | ScienceDirect | Scopus | Applications

TWC RPI is logged in | Go to Scholar Suite

Home | My settings

climate change

Search

? Search Tips

About 5488464 results for ALL(**climate change**)

Previous | On to page 1 of 227458 | Next page

Sort Relevance | Date

1. **Climate Change**  
NBuAwww.greenpeace.org/usa/energy-revol... December 2009  
... Revolution Ship Tour Ship Blogs The Rainbow Memo **Climate Change** What does dangerous **climate change** mean to Asia? Philippines **Climate** Impacts Solutions Welcome to Greenpeace Philippines! **Climate Change** Port Denc **Climate change** is real and happening...

2. **IPCC Special Reports on Climate Change**  
NBuAwww.grida.no/climate/ipcc/ October 2008  
IPCC Special Reports on **Climate Change** Ordering these reports (About this web-site Emissions Scenarios Aviation and the Global Atmosphere Land Use, Land-Use **Change** and Forests Methodological and Technological issues in Technology Transfer The Regional Impacts of **Climate Change**)

3. **IPCC - Intergovernmental Panel on Climate Change**  
NBuAwww.ipcc.ch/ November 2010

Search within results

Refine Results

Limit to Exclude

Content Sources

Scopus (190842)

ScienceDirect (347245)

Digital Archives (87284)

NDLTD (48354)

RePEc (23114)

View more

Year

2011 (1467)

2010 (430557)

2009 (281737)

My Applications

Add | Manage Applications

Enlighten Enlighten...

Your query **climate change** matches these items

1. **Enlighten: Free Jobs**  
Department of Labor  
Prices

2. **Methodology: Disability**  
Department of Labor  
Labor Force, Employment, and Earnings

3. **Impact: Board: Free Jobs**  
Department of Labor  
Prices

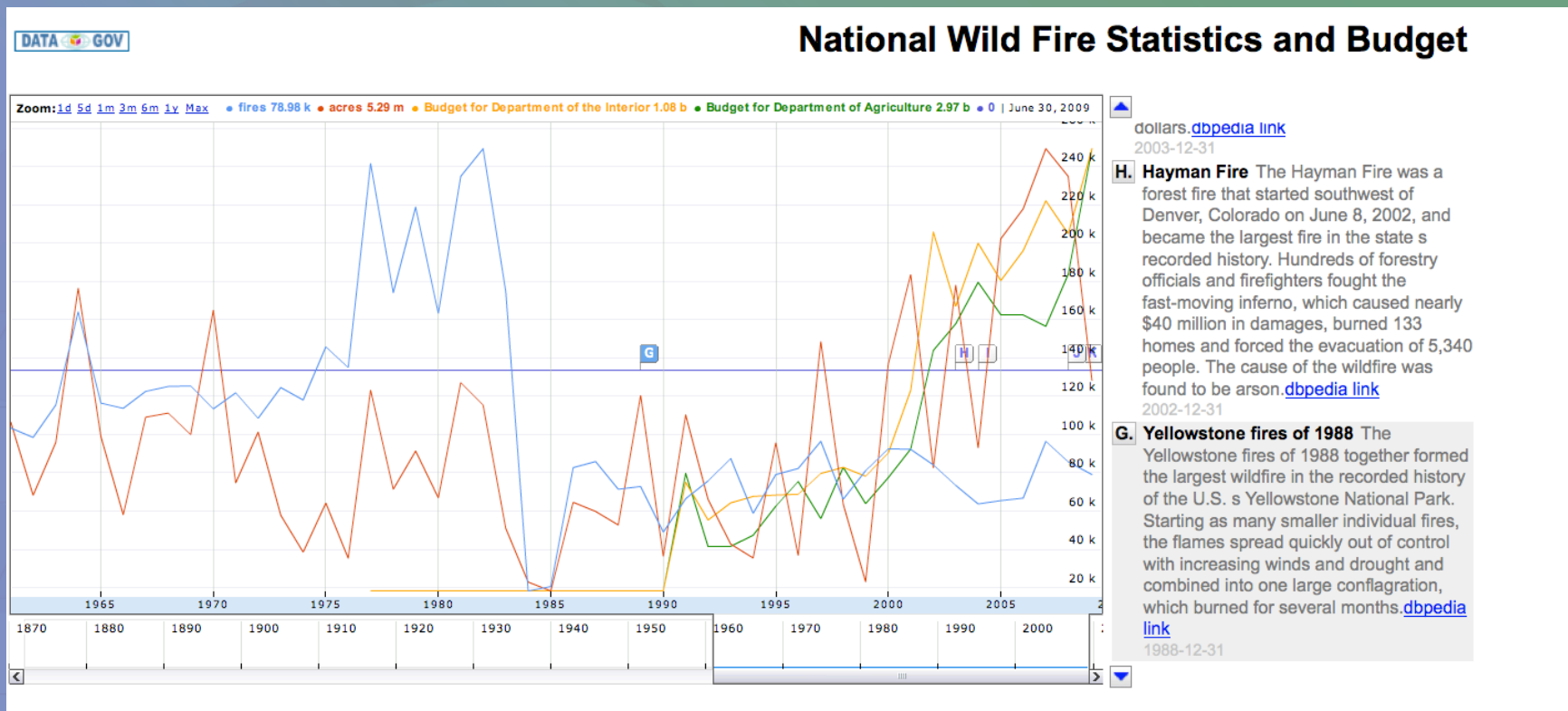
4. **Enlightenment: Cost: Taxes**  
Department of Labor  
Labor Force, Employment, and Earnings

5. **Enlightenment: Free Jobs**  
Department of Labor  
Prices



# Visualization can help identify data errors

Tetherless World Constellation



Correlates fires, acres burned, and agency budgets



# TW/C

“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, esp semantic and real-time search, 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

TipTop helps you find the best advice, opinions and people to engage and share experiences with.

Search Twitter for anything  
Web 3.0

Twitter Trends: #musicmonday #MM #aintnothinglike New Moon Thanksgiving Adam Lambert Google Wave Happy Birthday Milev Modern Warfare 2 #HollandHal

Channels: Events, Music, Sports, Food, News, Travel, Health, People, TV, Movies, Shopping

Filter Results by: web conf marketing video website technology talk family movie api app world corner personal company Web pages Types of products Types of events Companies Professions

Tip tweets: @cdomon: Thank you very much for following, hope to find/share interesting stuff. PS: I'm a fan of web 2.0, 3.0, 4.0, ... too.

pIT tweets: If web 2.0 was about social networking, web 3.0 will be about the death of anonymity.

Other tweets: 2050 web 3.0 #anticipation #futur #technology

Free whitepaper: Learn the value of Microblogging.

Data-gov RDFa Search

http://data-gov.tw.rpi.edu/ws/loclcs.php?query=environment

Semantic Search Of The Data-Gov Catalog

environment

Example: Budget, Labor, etc.

result 1 to 10 of total 65 results.

Property:92/category - Data-gov Wiki  
matched text: Geography and Environment + Dataset 1002 + Geography and Environment ... Geography and Environment + Dataset 1006 + Geography and Environment ...  
http://data-gov.tw.rpi.edu/wiki/Property:92/category

Next Generation Radar (NEXRAD) Locations  
matched text: Environmental Protection AgencyGeography and Environment11/17/20074/16/2009Data describes ... Environmental Protection AgencyGeography and Environment10/1/19936/30 ...  
http://data-gov.tw.rpi.edu/raw/92/data-92\_rdf  
types: http://xmlns.com/foaf/0.1/Document

Property:92/keywords - Data-gov Wiki  
matched text: ... program system identifier , environmental interest type, affiliation type ... program system identifier , environmental interest type, affiliation type ironmental interest type, ...  
http://data-gov.tw.rpi.edu/wiki/Property:92/keywords

Dataset 15 - Data-gov Wiki  
matched text: Geography and Environment + 92/collection mode. compilation by ... USGS, federal data download, environment, mining industry, economy, economic geology, ...  
http://data-gov.tw.rpi.edu/wiki/Dataset\_15

Dataset 1495 - Data-gov Wiki  
matched text: http://www.opic.gov/doing-business/investment/environment ... infrared radiation, global climate, environment, environmental data, environmental regulations, short tons, ...  
http://data-gov.tw.rpi.edu/wiki/Dataset\_1495

Dataset 37 - Data-gov Wiki  
matched text: Title: Lower Colorado River Daily Average Water Elevations and Releases ... environment, rivers, water, surface water, River Gauges, Lower Colorado River, ...  
http://data-gov.tw.rpi.edu/wiki/Dataset\_37  
tags: Stream, streamflow, reservoir, storage, Colorado River, federal datasets, federal data download, environment, rivers, water, surface water, River Gauges, Lower Colorado River, elevation, release  
types: http://data-gov.tw.rpi.edu/vocab/c/Dgtwc:Dataset

Dataset 962 - Data-gov Wiki

# Semantic Search Powered by RDFa



Trialx.com

Tetherless World Constellation

**TrialX** Enabling Patients to Find New Treatments. Sign Up as Patient | Sign Up as Investigator | Sign In

Home | Take a Tour | Patients | Investigators | Cures | **Help**



*Use your Personal Health Record to find matching trials*

**TrialX is a free service to enable you to find new treatments in your area >**

"TrialX matches patients to relevant clinical trials based on their health information"

**- The New York Times**

**News and Updates**

- Information resources for latest cures and treatments: [Cure Talk](#), [Cure Bytes](#) and [Curepedia](#).
- TrialX goes global. Now present in [Canada](#), [UK](#) and [India](#).
- TrialX wins the International [Semantic Web Challenge](#).

**Search Clinical Trials By Condition**

Enter Conditions:

Example: diabetes

**Find Clinical Trials**

OR

**Get Matching Trials using your Personal Health Record**

Copy FROM  OR 

Microsoft HealthVault is the way to collect, store and share information online. [Learn More](#)

TrialX is going to copy data from your HealthVault account and needs you to sign in

TrialX is a Google Health integrated partner. [Start an account](#) with Google Health to securely manage your personal health information online. Link up with TrialX to get personalized matching clinical trials.

**Browse by Clinical Trial Categories**

[Uterine Fibroids Clinical Trials](#) | [Diabetes Clinical Trials](#) | [Infertility Clinical Trials](#)  
[COPD Clinical Trials](#) | [Pancreatic Cancer Clinical Trials](#) | [Atrial Fibrillation Trials](#)  
[Breast Cancer Clinical Trials](#) | [Multiple Myeloma Clinical Trials](#) | [Depression Clinical Trials](#)  
[Asthma Clinical Trials](#) | [High Cholesterol Clinical Trials](#) | [Lung Clinical Trials](#)  
[More...](#)

**New Treatments**

[Main](#)  
[DPP-IV](#)  
[Gliptins](#)  
[Fadrozole](#)  
[Raloxifene](#)  
[More...](#)

**TrialX Facts**

- 14,000+ ongoing clinical trials approved by the US FDA
- 8000+ clinical trial investigators
- 3000+ and counting participants matched to clinical trials

**For Patients**

- [Search Treatments](#)
- [Use HealthVault](#)
- [Use Google Health](#)
- [Tour](#)
- [Join Community](#)
- [TrialX Mobile](#)

**For Investigators**

- [Submit Clinical Trials](#)
- [Tour](#)
- [Get Widget](#)
- [TrialX Mobile](#)

**Company**

- [About Us](#)
- [Press](#)
- [Blog](#)
- [Terms of Service](#)
- [Privacy](#)
- [Contact Us](#)

**As seen on**

  
 | [usatoday.com](#)  


Get FREE TrialX iPhone App



[Read More...](#)

Global Sites: [USA](#) | [Canada](#) | [United Kingdom](#) | [India](#)  
 ©2010 Applied Informatics, Inc.

Save lives



TWGC

# Lots More Tetherless World Constellation

**twine**  
jim hendler  
Albany, NY

**Freebase**  
Leicester facts - Freebase

**YAHOO! DEVELOPER NETWORK**  
MY PROJECTS | API | YDN / SearchMonkey

**SearchMonkey**  
SEARCHMONKEY IS...  
Using SearchMonkey, de...  
useful and visually appea...  
HOW DOES IT WORK...  
The SearchMonkey deve...  
you've built your app, you

**garlik**™  
The online identity experts

Home | About Garlik | Garlik products | What our customers say | News | Sign up | Log in

### DataPatrol for Individuals

Would you know...

- If your details has been stolen and were being traded online?
- If a website had published your sensitive or personal information?
- If someone had set up a postal redirect on your address to steal your mail?

As a DataPatrol customer you would know instantly and you would know exactly what to do to prevent any loss or damage before it happens.

**Financial fraud is soaring**  
Individual criminals and organised crime groups use information obtained on the internet to steal people's identities and commit financial fraud, so protecting yourself and your family is now more important than ever.

Every month DataPatrol monitors over 200,000 usernames, bank details, credit cards and email addresses being sold on the internet. By monitoring your information every day, DataPatrol provides early warning protection and alerts you if your information has been compromised and you are at risk of identity theft and financial fraud.

**Where do we look?**  
DataPatrol searches across a comprehensive variety of data sources, including:

- ✓ Public Websites
- ✓ Compromised and stolen data sources traded on the internet
- ✓ Births, Marriages and Deaths register
- ✓ Electoral Roll
- ✓ Royal Mail redirect database

**What's protected?**

<b>Personal information</b>	<ul style="list-style-type: none"> <li>✓ Name</li> <li>✓ Usernames</li> <li>✓ Aliases</li> <li>✓ Email address</li> </ul>
<b>Contact information</b>	<ul style="list-style-type: none"> <li>✓ Home address</li> <li>✓ Other address</li> <li>✓ Telephone numbers</li> </ul>

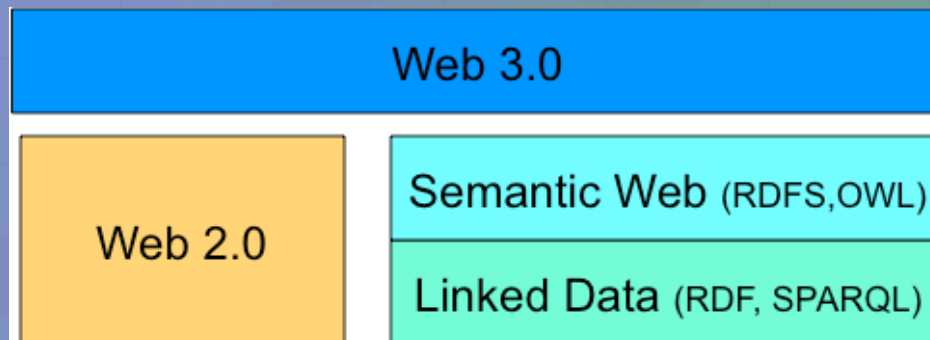
Web 3.0 Applications



## Web 3.0 excitement (hype?)

Tetherless World Constellation

- Significant and growing commercial interest...
  - Web: Google, Amazon, Travelocity...
  - Web 2.0: Facebook, Wikipedia, YouTube, Twitter...
  - Web 3.0: ??





# TWC

## Summary

### Tetherless World Constellation

- The Semantic Web is real
  - People asking “how,” not why
- So far the commercial driver has been “weak semantics”
  - Very Simple “ontologies”
  - Lots of linking
  - Metadata agreements, not ontology alignments
- Web 3.0 adds semantics as a value add to regular Web functionality
  - Data mashup
  - Semantic search
  - Semantic match
- Investor excitement: The big one is still out there