



TWC

Tetherless World Constellation

Web 3.0 emerges...

Prof. 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



Agilemedia Events

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".

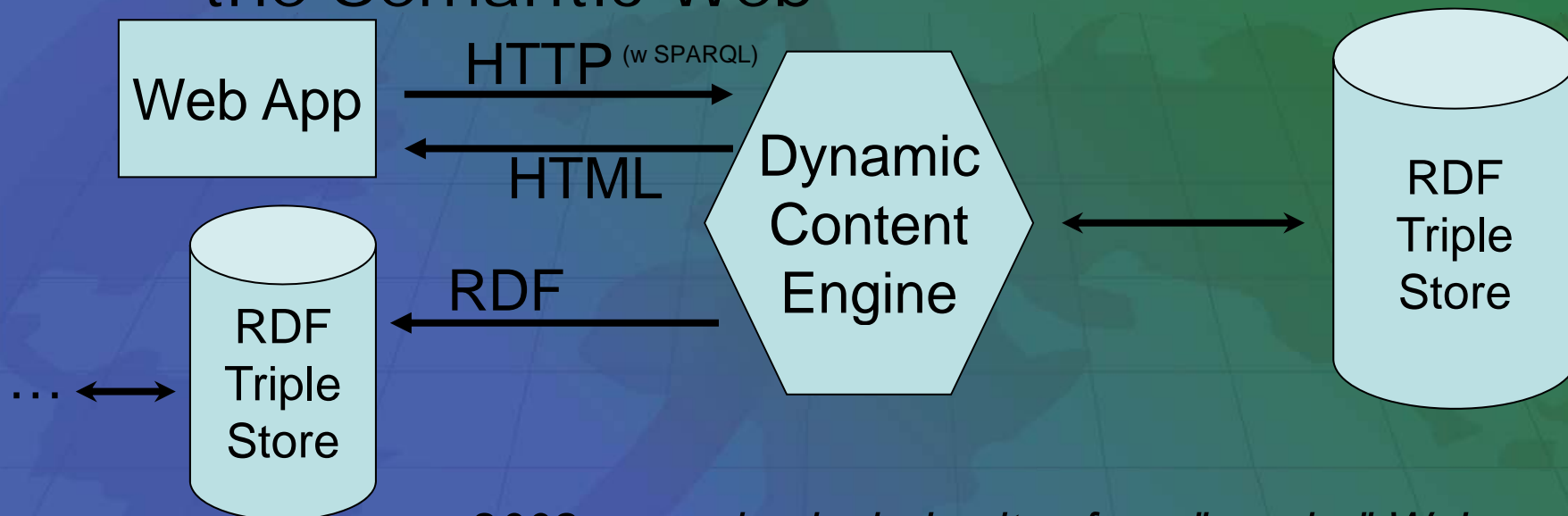
Applications, Technologies and Business Utilization



Fits Web Architecture

Tetherless World Constellation

- ~2006: Web app developers discover the Semantic Web



2008 examples include sites from "regular" Web players such as Dow Jones, Reuters and Yahoo!



Sem Web 2010

Tetherless World Constellation

InfoQ League

Tracking change and innovation in the enterprise software development community

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](#)

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 [+](#) | [f](#) [dz](#) [t](#) [g+](#) [e](#) [m](#)

"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.



Sem 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.

...





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.](#)

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



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



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.



Two very different sorts of use cases

Tetherless World Constellation

- *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

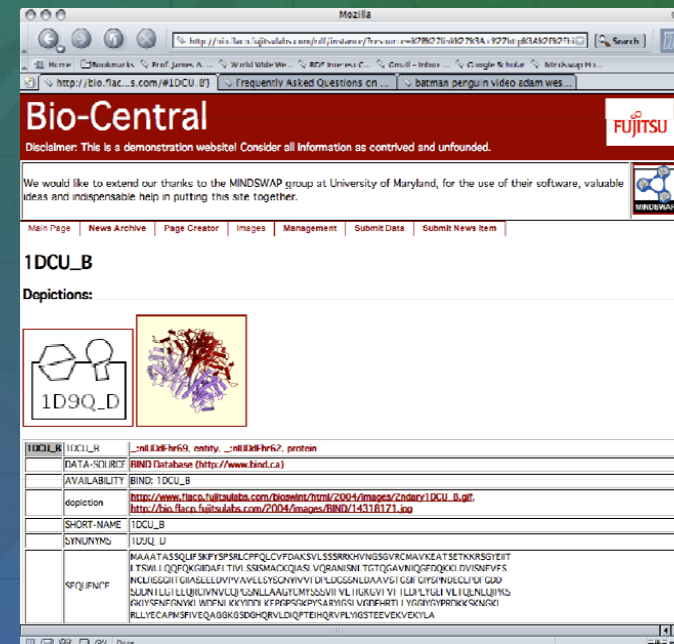
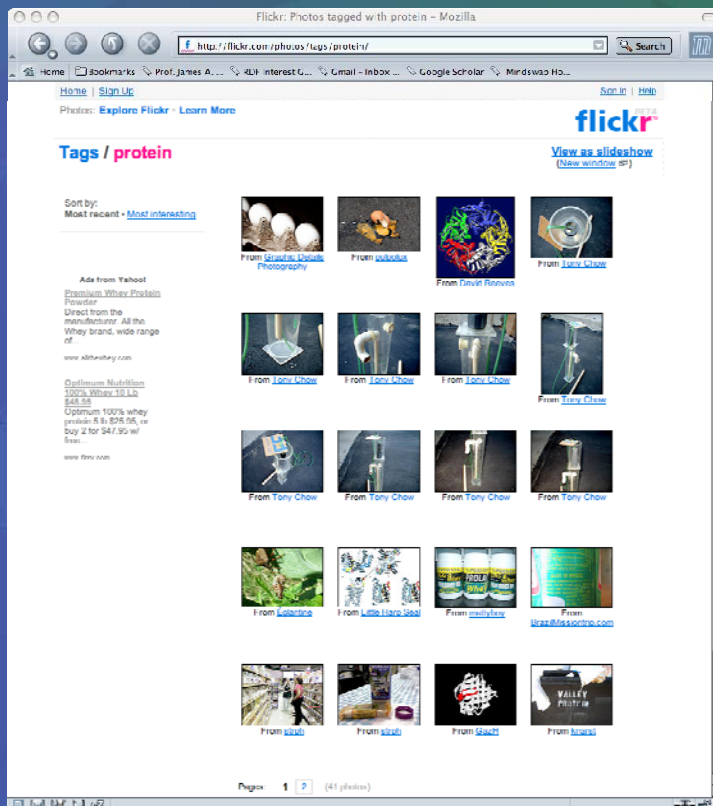
- NCBI Oncology Ontology
 - “High use” in medical community (~1200 users)
 - Very “trusted” information (provenance from NCBI)
 - Primarily terminological (relationships between cancer-related concepts), not data-oriented
- FOAF
 - >60M Foaf people (not necessarily distinct individuals) in hundreds of 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" 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!*



Raw Government Data Now 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



December 8, 2009

“Open Government Directive” released



May 21, 2010

data.gov relaunch
6.4 billion triples

June 30, 2009

Putting Government Data online



January 19, 2010

data.gov.uk online





Moving data.gov to linked data (UK)

Tetherless World Constellation

- Built around linked data with top-down push from "Number 10"

The screenshot shows the HM Government data.gov.uk website. The header includes the HM Government logo and the URL hmg.gov.uk/data. A navigation menu contains links for Home, Data, Blog, Apps, Ideas, Resources, and About. A search bar is located in the top right corner. The main content area features a banner titled "Unlocking innovation" with the subtitle "Working with UK Public Sector information and data" and a blue molecular structure graphic. Below the banner is a paragraph of text: "Advised by Sir Tim Berners-Lee and Professor Nigel Shadbolt and others, we are opening up government data for reuse. This is a preview of what a public site could look like. We want to work with you to make it better. We're very aware that there are more people like you outside of government who have the skills and abilities to make wonderful things out of public data. These are our first steps in building a collaborative relationship with you." Below this is a "Search Data" section with a search box and a "Search" button. The "Featured Data" section highlights "Education and training participation: 16-18 year olds" with a brief description and tags. The "Most Recent Apps" section lists three apps: "PlanningAlerts", "FIITthatHole", and "Where Can I Live?". The right sidebar contains a "Subscribe by RSS" button, a "What is the Semantic Web?" section, a "Digital Engagement Twitter stream" with a tweet from @innovate_uk, and a "Submit an app" form.



Moving data.gov to linked data (US)

Tetherless World Constellation

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

The screenshot shows the data.gov website interface. At the top, there's a navigation menu with links for HOME, DATA, TOOLS, COMMUNITY, METHODS, and IMAGERY. The main banner area is titled 'LINKING OPEN GOVERNMENT DATA' and includes a 'VIEW MORE' button. To the right, a 'Most Popular Datasets' section lists five items: 1. US Topo 7.5-minute map for Imperial, TX; 2. U.S. Overseas Loans and Grants (Greenbook); 3. Worldwide M+E Birthplaces, Past 7 Days; 4. Latest Volumes of Foreign Relations of the...; 5. Personal Trends by Gender/Race. Below this is a search bar labeled 'SEARCH OUR CATALOGS'. The page is divided into three columns: 'APPS' (with a 'VIEW MORE' button), 'COMMUNITY' (discussing democratizing public sector data and listing statistics like '6 Other nations establishing open data'), and 'SEMANTIC WEB' (discussing the Tetherless World Constellation and listing statistics like '8 States now offering data sites'). The footer contains the DATA.GOV logo and links for About, FAQ, Contact Info, Data Policy, Accessibility, Privacy Policy, and Sitemap.



Data.gov community

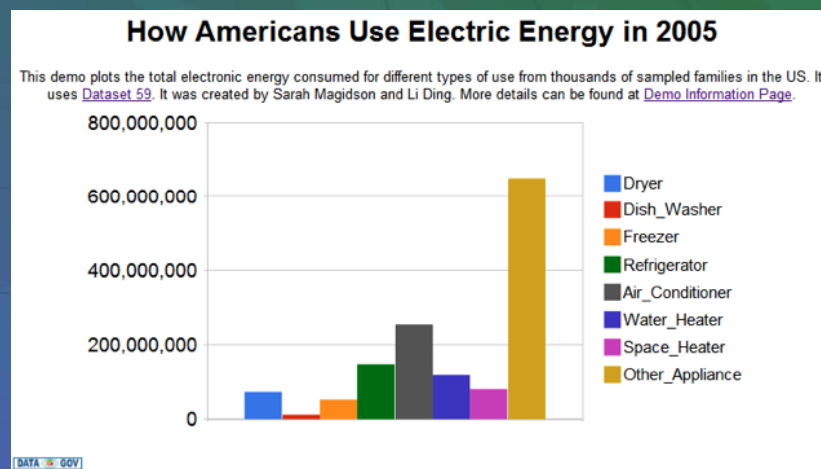
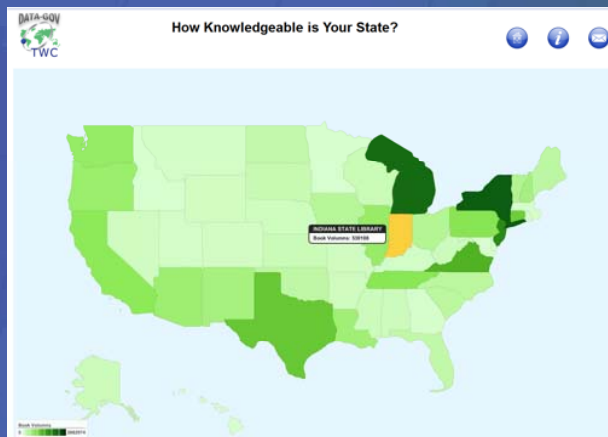
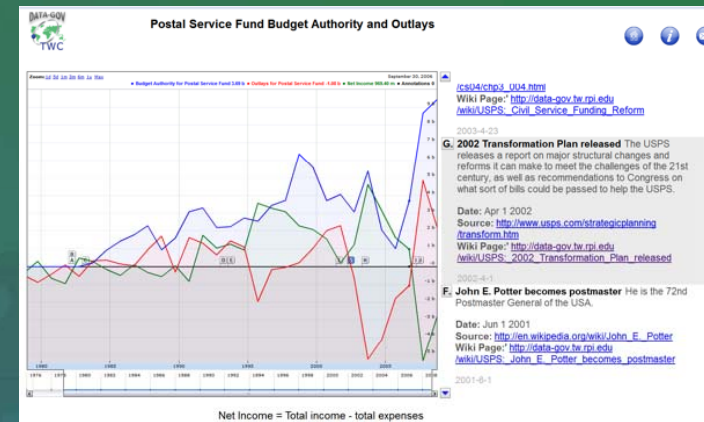
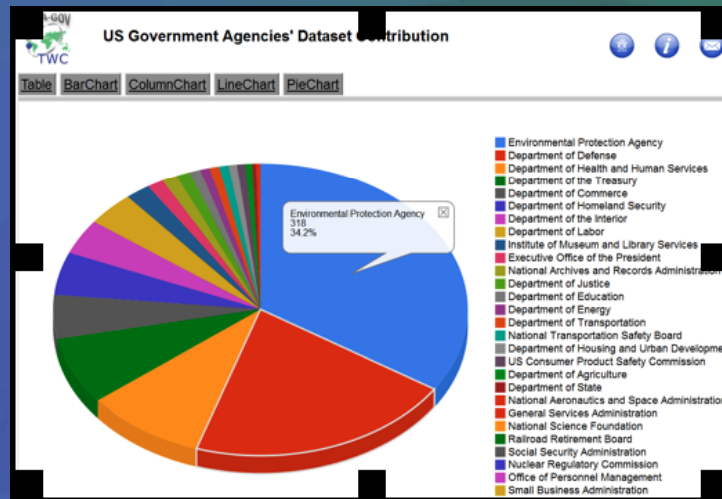
Tetherless World Constellation

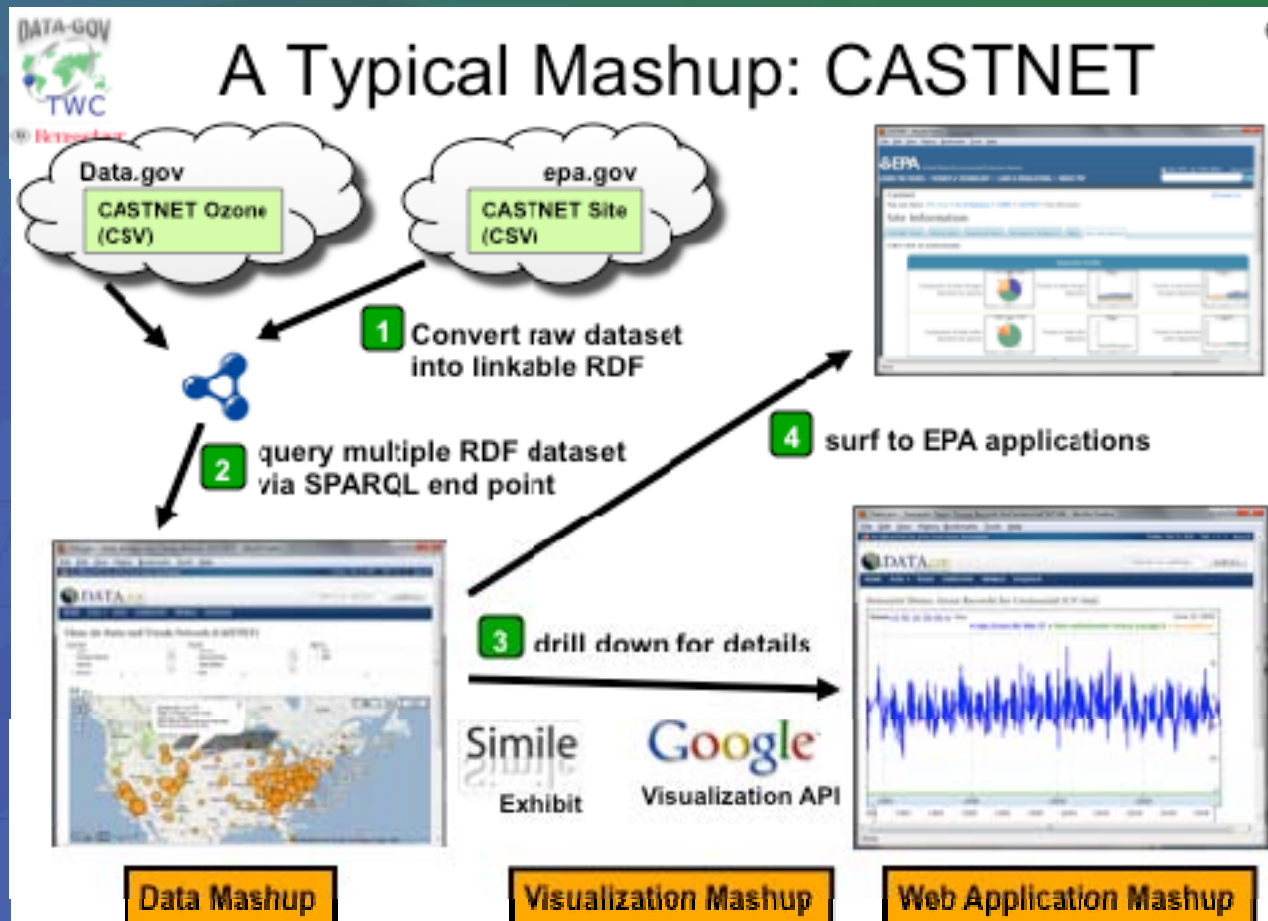
The screenshot shows a web browser window displaying the Data.gov Community page. The browser's address bar shows the URL <http://www.data.gov/community>. The page features a navigation menu with links for HOME, DATA, TOOLS, COMMUNITY, METRICS, and DIALOGUE. A search bar is located in the top right corner. Below the navigation, the word "COMMUNITY" is prominently displayed. A paragraph of text states: "Data.gov is leading the way in democratizing public sector data and driving innovation. This movement has spread throughout cities, states, and countries." Below this text are four logos: OPEN GOV, Rensselaer, Sunlight Foundation, and World Bank. A row of twelve circular icons represents various government agencies. A world map is shown with several countries highlighted in dark blue. At the bottom, a row of flags represents the United States, United Kingdom, Australia, Canada, New Zealand, Norway, and Germany.



Pump through to Google Viz for demos

Tetherless World Constellation





Created by Dominic DiFranzo, PhD student at RPI, <http://www.data.gov/semantic/Castnet/html/exhibit>



Data.gov + epa.gov

Tetherless World Constellation

DATA GOV

Castnet Ozone Map



Ozone Stations

84 Items

Land Use

- 27 Agric
- 1 Coastal/Marsh
- 3 Desert
- 40 Forest
- 1 Forest/And
- 2 Forest/Marsh

Terrain

- 37 Rolling
- 4 Mountaintop
- 1 Flat/Water
- 14 Flat
- 27 Complex
- 1 Coastal

Agency

- 55 EPA
- 25 NPS



Ozone order





Adding some Web magic

Tetherless World Constellation

ROM406 (Rocky Mtn NP)

Deposition Profile

Composition of total nitrogen deposition by species		Trends in total nitrogen deposition		Trends in wet and dry nitrogen deposition	
Composition of total sulfur deposition by species		Trends in total sulfur deposition		Trends in wet and dry sulfur deposition	

Site Profile

Site name	Rocky Mtn NP
County	Larimer
State abbreviation	CO
Latitude; decimal degrees	40.7778
Longitude; decimal degrees	-105.5453
Elevation; m	2743
Operating agency	NPS
Primary land use	Forest
Terrain surrounding site	Complex
NADP site code	CQ98 EXIT 1022
Distance to nearest NADP site; km	10.0
Does site conform to assumptions of MLM?	No

Web Analytics

NADP/NTN Monitoring Location ND00

National Atmospheric Deposition Program

Station	Thodore Roosevelt National Park-Pointed Canyon (ND00)
Location	Billings County, North Dakota
Dates of Operation	1/30/2001 - Present
Latitude	46.8947
Longitude	-103.3777
Elevation	841 meters
USGS 1:24000 Map Name	Fryburg NW
Operating Agency	Thodore Roosevelt National Park
Sponsoring Agency	NPS Air Resources Division

None Site Photographs

2005 Site Survey (13) [Large](#) [Small](#)

2001 Site Survey (10) [Large](#) [Small](#)

Miscellaneous (2) [Small](#)

[Site Survey Information](#)

NADP/NTN Wet Deposition Data Available

External links related to this site

[Trend Photo](#)

[Annual Data Summaries](#)

Annual Data

- data by calendar and water years
- wet deposition totals
- precipitation-weighted mean concentrations

Seasonal Data

- wet deposition totals
- precipitation-weighted mean concentrations

Monthly Data

- precipitation-weighted mean concentrations

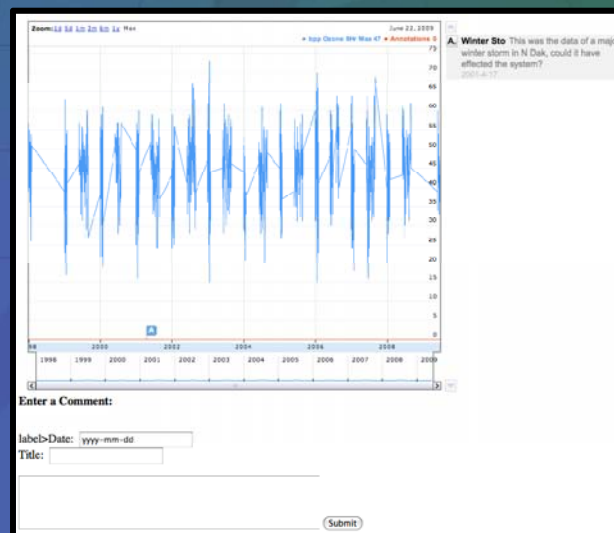
Weekly Data

- concentrations

Daily Data

- daily precipitation data

External Links



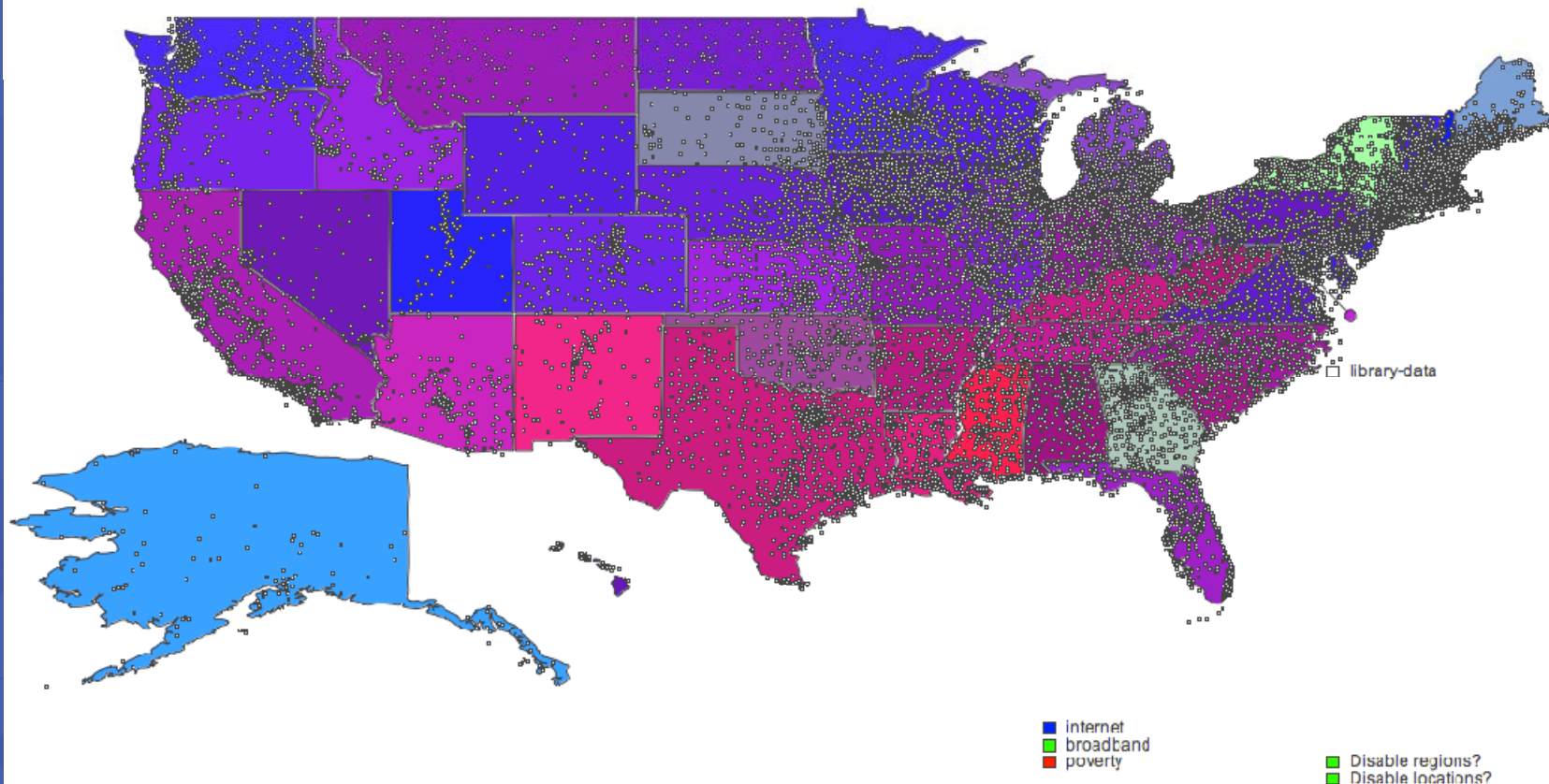
Social Data Networks



NTIA internet study vs. libraries

Tetherless World Constellation

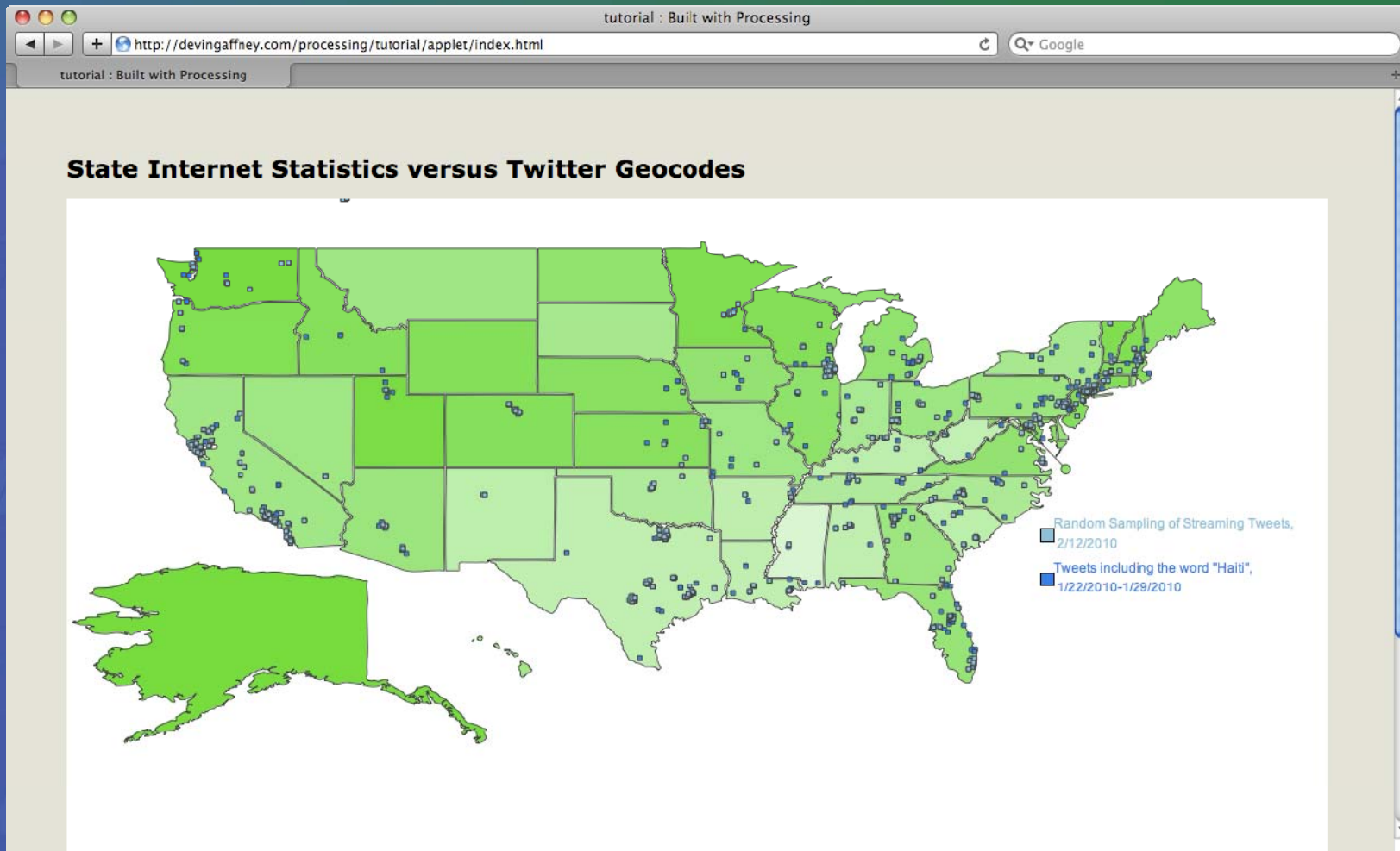
Nebraska: poverty: 14.8: broadband: null: internet: 73.6





NTIA internet funding vs. tweets about #haiti

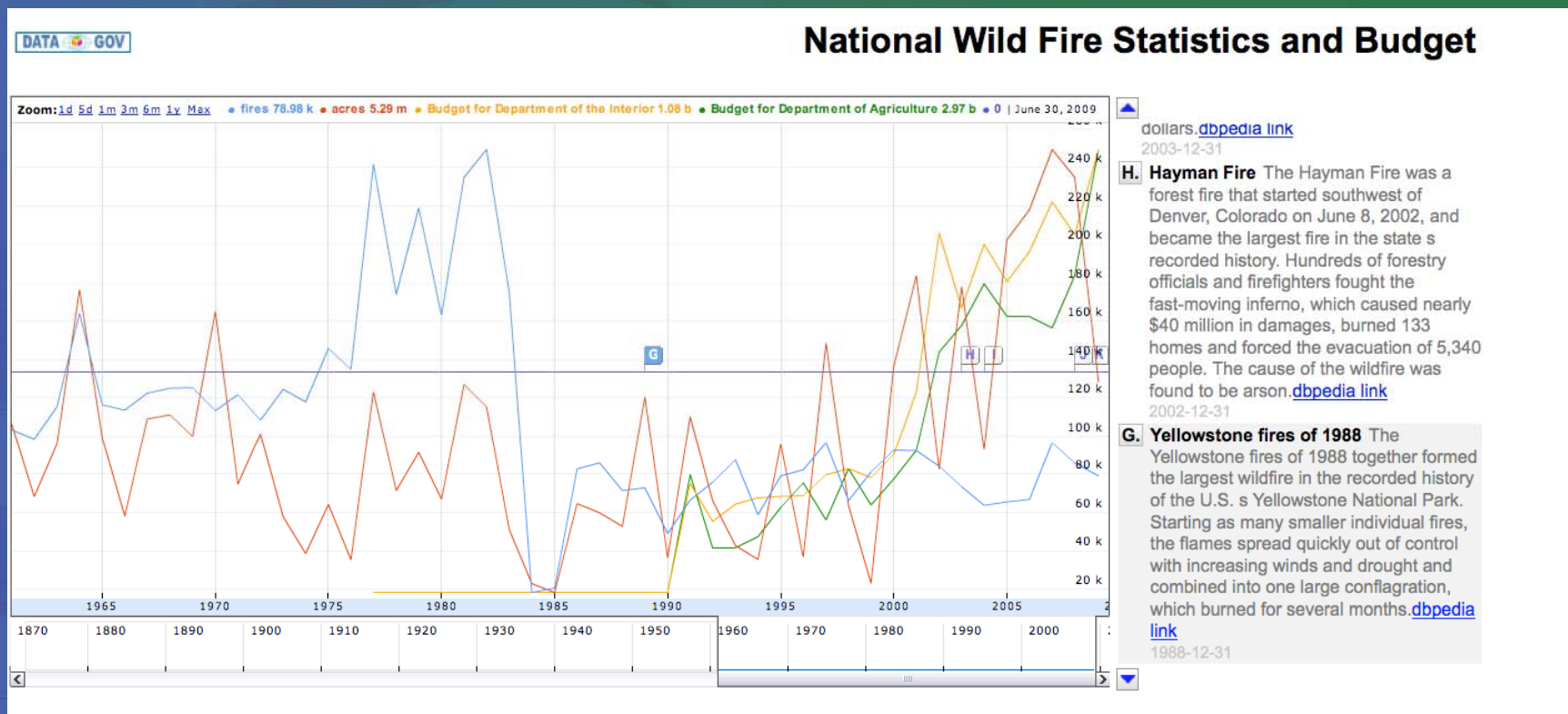
Tetherless World Constellation





Visualization can help identify data errors

Tetherless World Constellation

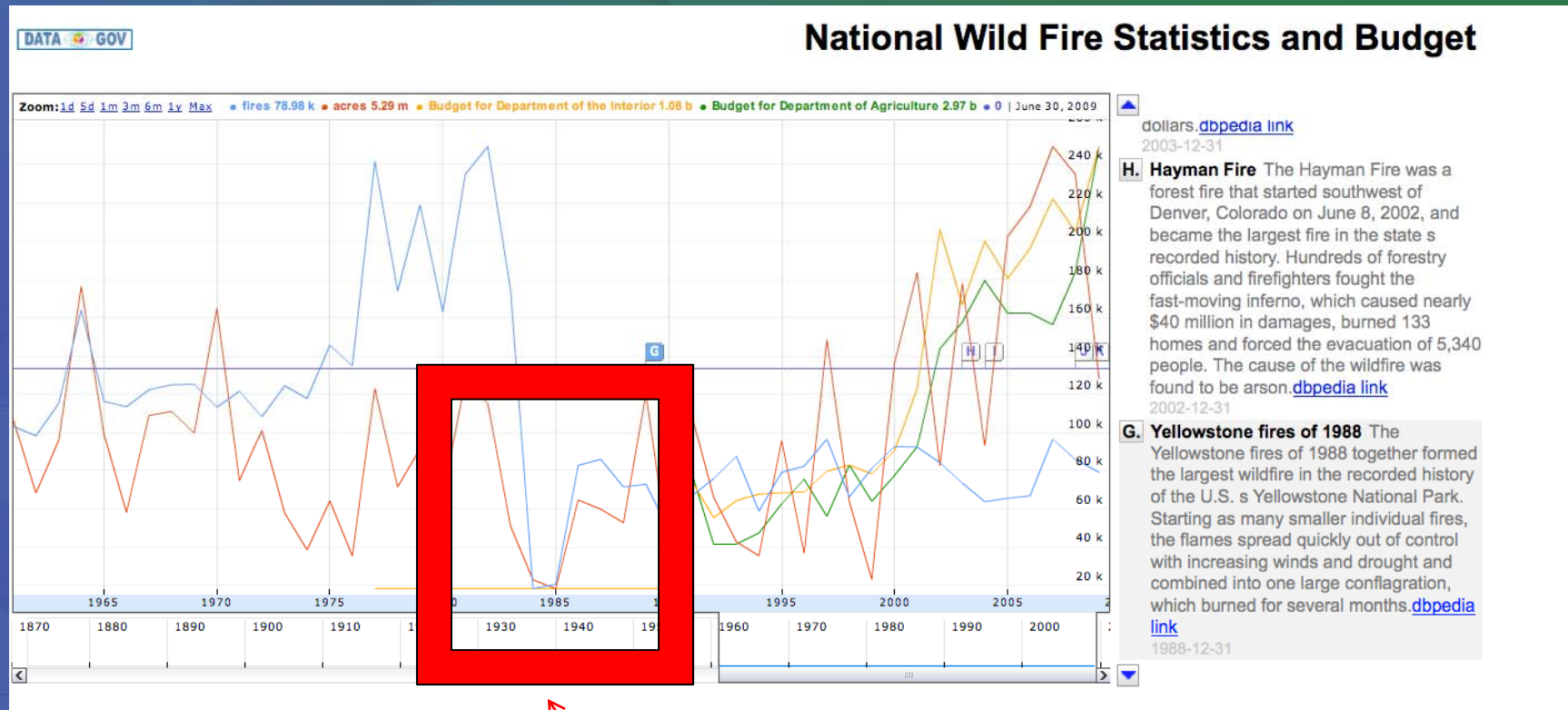


Correlates fires, acres burned, and agency budgets



Visualization can help identify data errors

Tetherless World Constellation



Were there really no fires in 1985?



Presents a challenge – different ontologies

Tetherless World Constellation

US Crime

search

Violent Crimes

- Murder and nonnegligent manslaughter
- Forcible rape
- Robbery
- Aggravated Assault

Property Crimes

- Burglary
- Larceny-theft
- Motor vehicle theft
- Arson

UK Crime

Burglary

- Criminal_damage
- Drug_offences
- Fraud_and_forgery
- Offences_against_vehicles1
- Other_offences
- Other_theft_offences
- Robbery
- Sexual_offences
- Total
- Violence_against_the_person



Presents a challenge – different ontologies

Tetherless World Constellation

US Crime

search

Violent Crimes

- Murder and nonnegligent manslaughter
- Forcible rape
- Robbery
- Aggravated Assault

Property Crimes

- Burglary
- Larceny-theft
- Motor vehicle theft
- Arson

UK Crime

Burglary

- Criminal_damage
- Drug_offences
- Fraud_and_forgery
- Offences_against_vehicles1
- Other_offences
- Other_theft_offences
- Robbery
- Sexual_offences
- Total
- Violence_against_the_person

Same or different?



Combining data from different sites

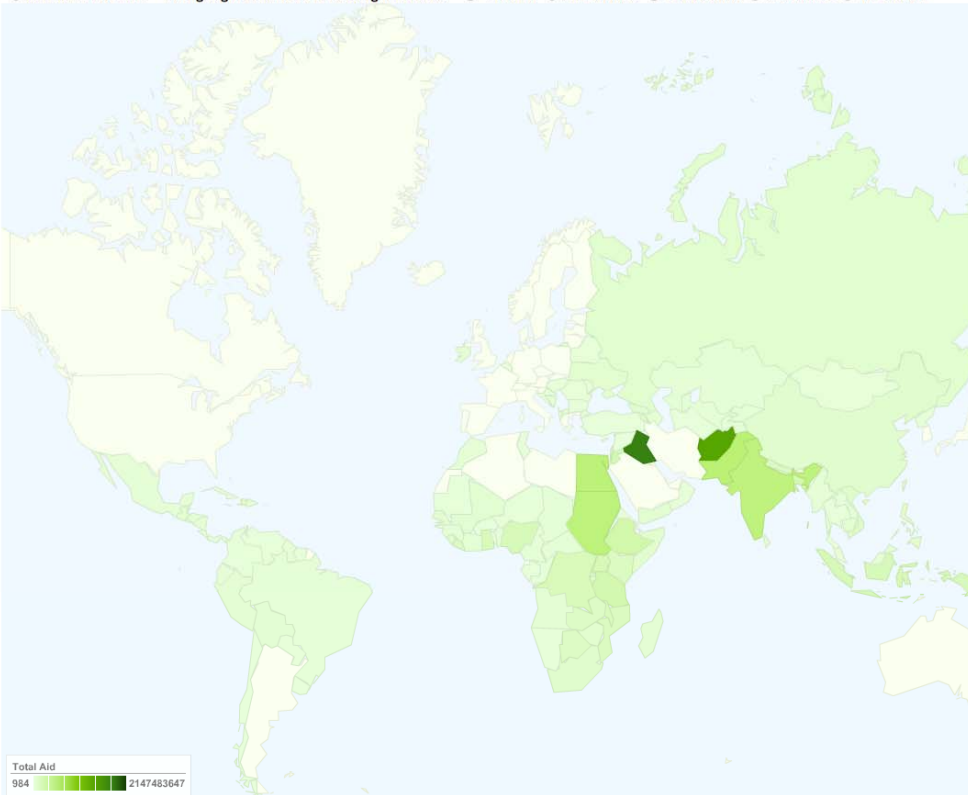
Tetherless World Constellation

Comparing US (USAID) and UK (DFID) Global Foreign Aid

Demo About

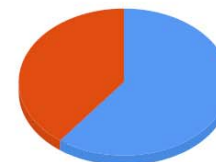
HIGHLIGHT ALL - or highlight countries receiving less than 1 Billion 250 Million 100 Million 50 Million 10 Million

Aid Figures News Country Facts



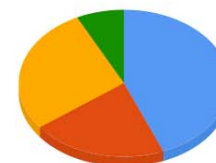
Aid to Indonesia (in USD)

USAID
DFID



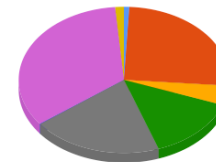
USAID Categories (in USD)

Economic/Security Assistance
Development Assistance
Child Survival and Health
Other USAID Assistance



DFID Categories (in USD)

Education
Health
Social Services
Government and Civil Society
Economic
Research
Humanitarian Assistance
Non Sector Allocable



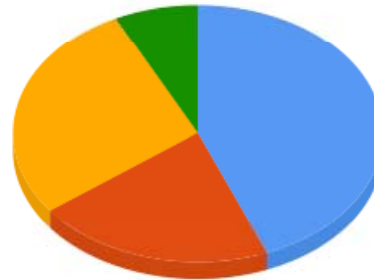


Presents a challenge – different ontologies

Tetherless World Constellation

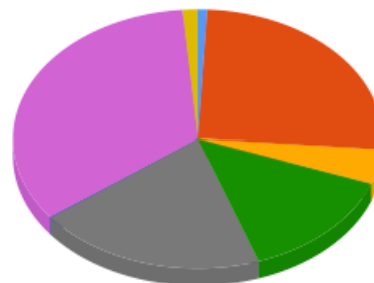
USAID Categories (in USD)

- Economic/Security Assistance
- Development Assistance
- Child Survival and Health
- Other USAID Assistance



DFID Categories (in USD)

- Education
- Health
- Social Services
- Government and Civil Society
- Economic
- Research
- Humanitarian Assistance
- Non Sector Allocable





An RPI touch – supersize it

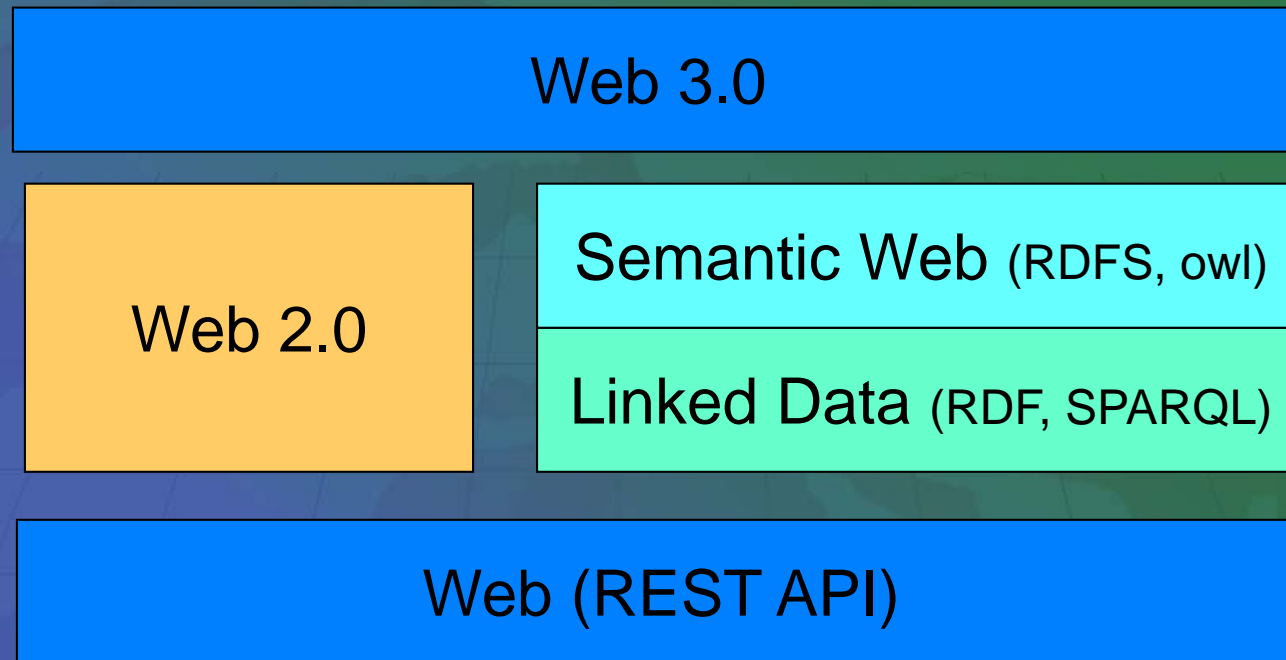
Tetherless World Constellation





This on the Web takes us into “Web 3.0”

Tetherless World Constellation



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)



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

Copy FROM **HealthVault** OR **Google Health** BETA

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 <ul style="list-style-type: none">Search TreatmentsUse HealthVaultUse Google HealthTourJoin CommunityTrialX Mobile	For Investigators <ul style="list-style-type: none">Submit Clinical TrialsTourGet WidgetTrialX Mobile	Company <ul style="list-style-type: none">About UsPressBlogTerms of ServicePrivacyContact Us	As seen on <ul style="list-style-type: none">The New York Times usatoday.com	Get FREE TrialX iPhone App Read More...
---	---	--	---	---

Global Sites: USA | Canada | United Kingdom | India
©2010 Applied Informatics, Inc.

Save lives



Lots More Tetherless World Constellation

Freebase
Leicester topic
Also known as City Of Leicester
Leicester (pronounced /ˈleɪstə/) is the traditional county town of Leicestershire, a city with a population of 350,000, the 10th largest city in England and the largest city in the East Midlands.

YAHOO! DEVELOPER NETWORK
MY PROJECTS
YDN / SearchMonkey
SEARCHMONKEY IS...
Using SearchMonkey, developers can create useful and visually appealing web applications.
HOW DOES IT WORK...
The SearchMonkey developer has built your app, you can now use it to search for data.

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

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

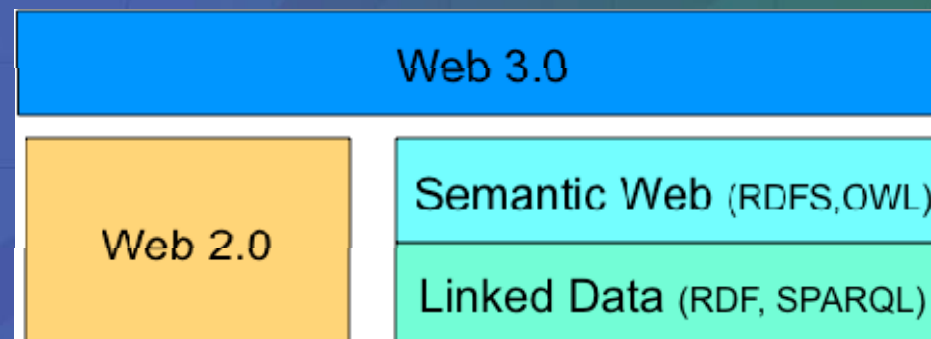
Web 3.0 Applications



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: the big ones are still out there





TWC

Summary

Tetherless World Constellation

- 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