Semantic Web Science

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)

AND NOW FOR
SOMETHING
COMPLETELY
DIFFERENT

- Semantic Web is real
- Growing at a fast pace
- Producing lots of interesting networks
- That no one is really analyzing from a network science perspective
- Which could hugely help those of us trying to use this for some really hard real world problems
    - For example, open govt data

Sem Web 2010

Tetherless World Constellation

InfoQ

446,480 Jul unique visitors

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.

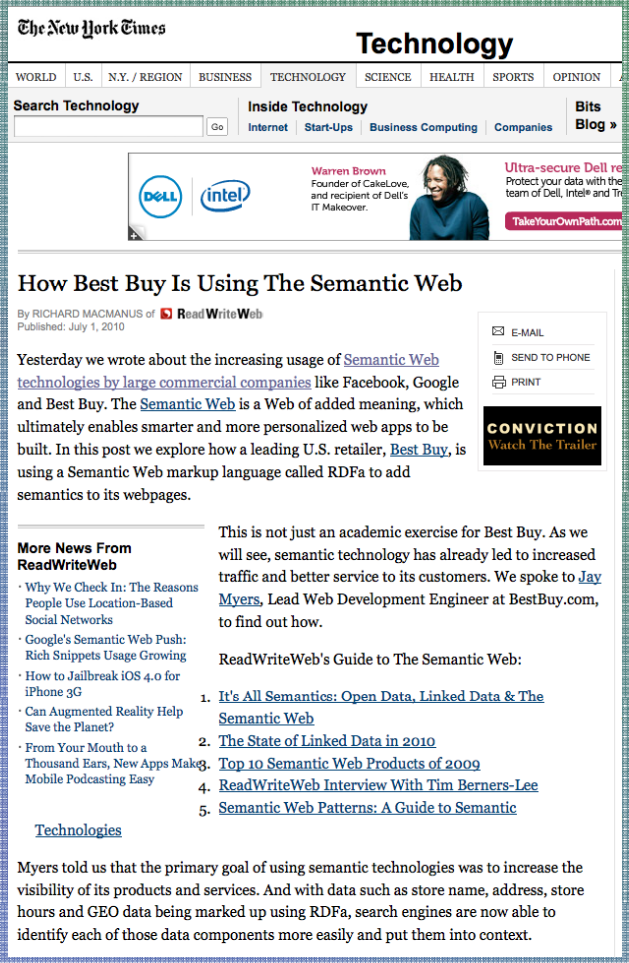
> First, the Web is moving to a model based on the connections between people and all the things they care about. Second, this connections-based Web is well on its way to being built and providing value to both users and developers — the underlying graph of connections just needs to be mapped in a way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

4/2010

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 | Inside Technology: Internet | Start-Ups | Business Computing | Companies | Bits Blog »

DELL intel
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

E-MAIL
SEND TO PHONE
PRINT

CONVICTION
Watch The Trailer

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.

7/2010

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)**.

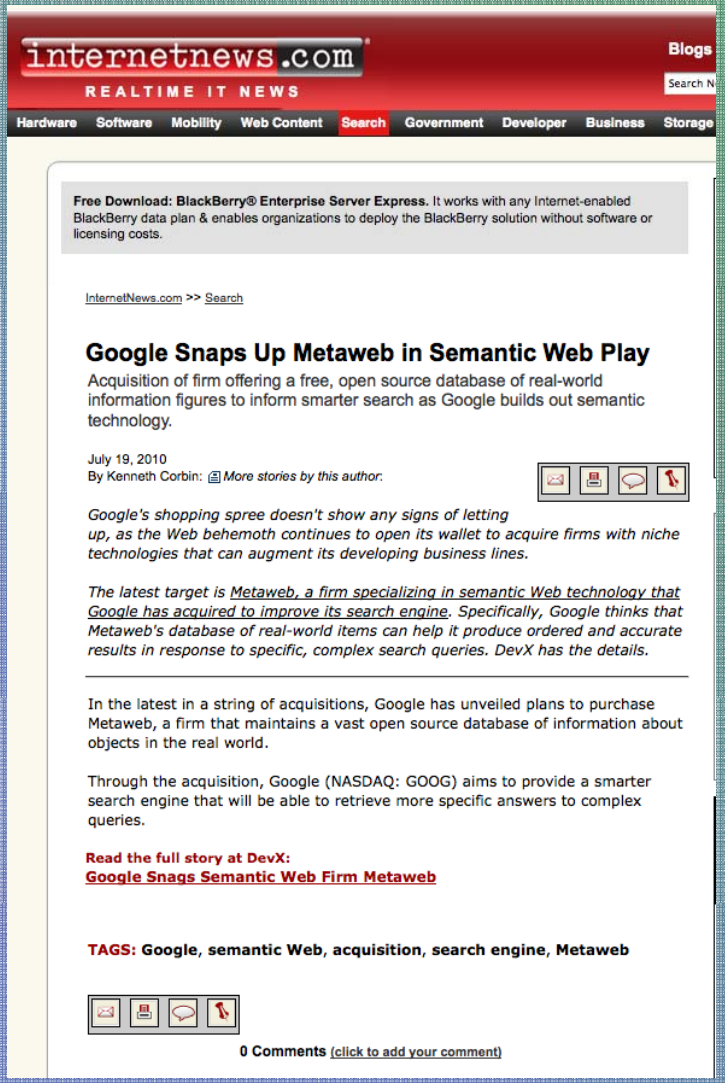
The flickering nature of search results in the new Google Instant technology will very likely further increase the gains of more appealing rendering in the results list. So don't put this off. **Act today!**

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 - 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 - Cached - Similar



internetnews.com
REALTIME IT NEWS

Blogs
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)

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.

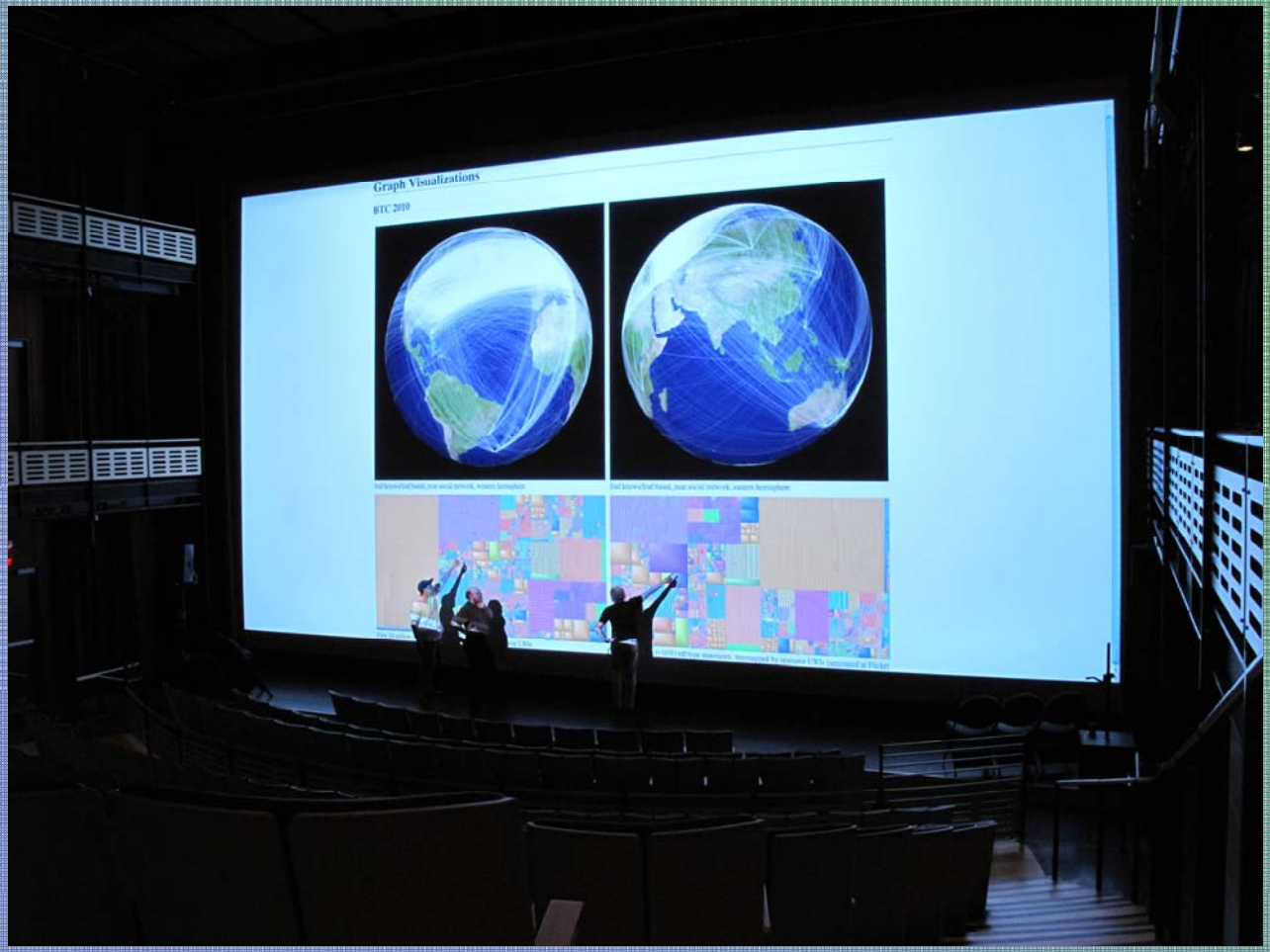
Imagine bookmarking a story about Barack Obama: your computer will store the URL, but it has no way of knowing whether the content relates to politics or, say, cookery. If, however, each web page were to be tagged with information about its content, we can ask the web questions and expect sensible answers.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

- What is different now?
  - Semantics in Search
  - Advertising drives Web markets
  - "Buzz" around data on the Web
    - Facebook OGP, Open Govt Data, ...
- Maturation of RDF technologies
  - SPARQL endpoints
  - RDFa
  - Lightweight Knowledge
    - A little semantics goes a long way

- FOAF
  - >60M Foaf people (not necessarily distinct individuals) in hundreds of applications touched by a large community (> 100,000,000 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

FOAF Network has been explored as Social network *per se*

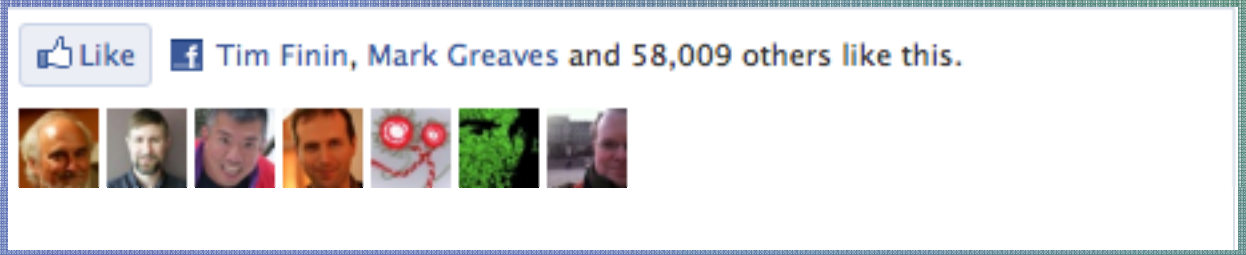
Tetherless World Constellation



Foaf complicated compared to OGP (Facebook's Open Graph Protocol)

Tetherless World Constellation

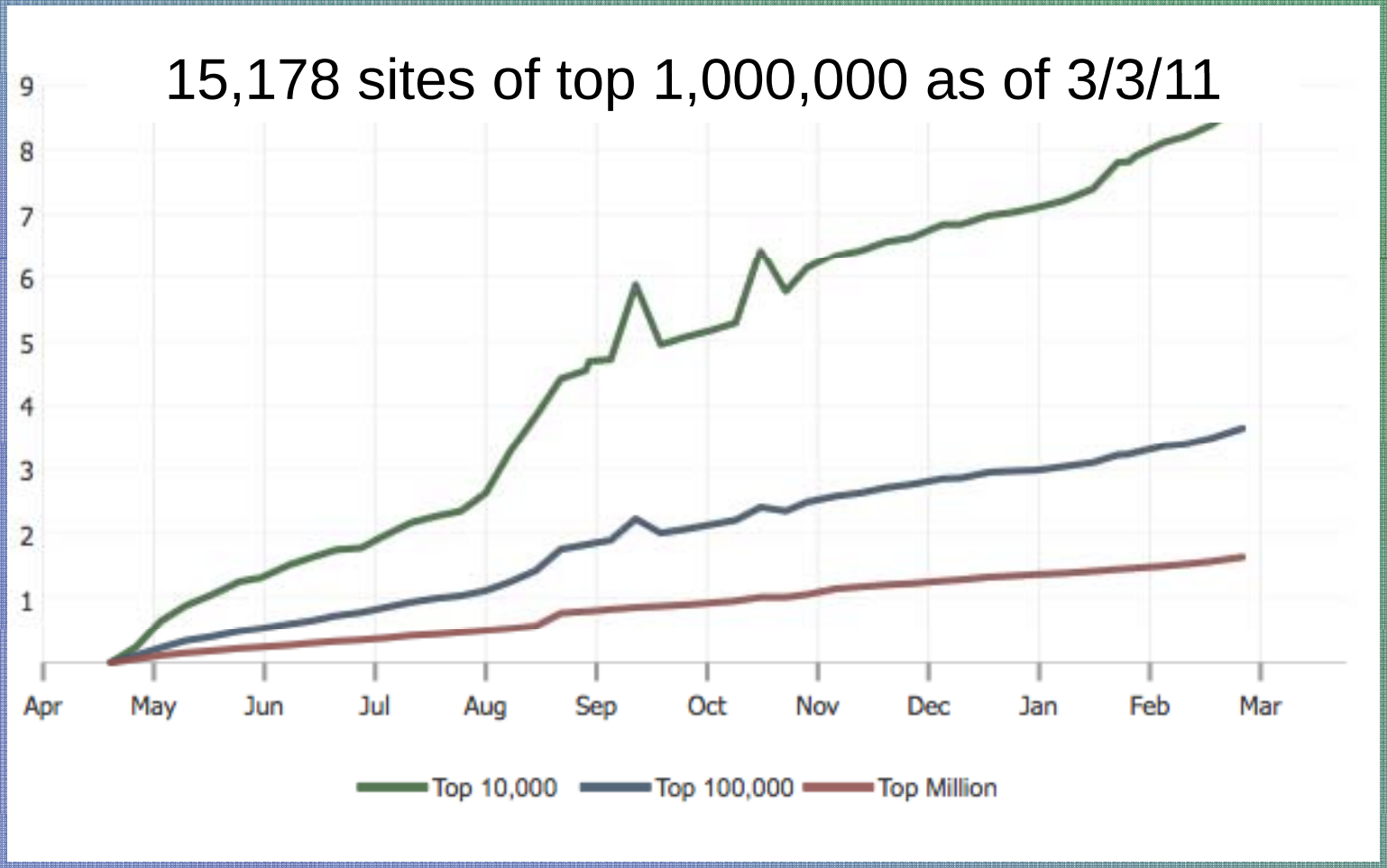
- og:title - The title of your object as it should appear within the graph, e.g., "The Rock".
- og:type - The type of your object, e.g., "movie". Depending on the type you specify, other properties may also be required.
- og:image - An image URL which should represent your object within the graph.
- 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

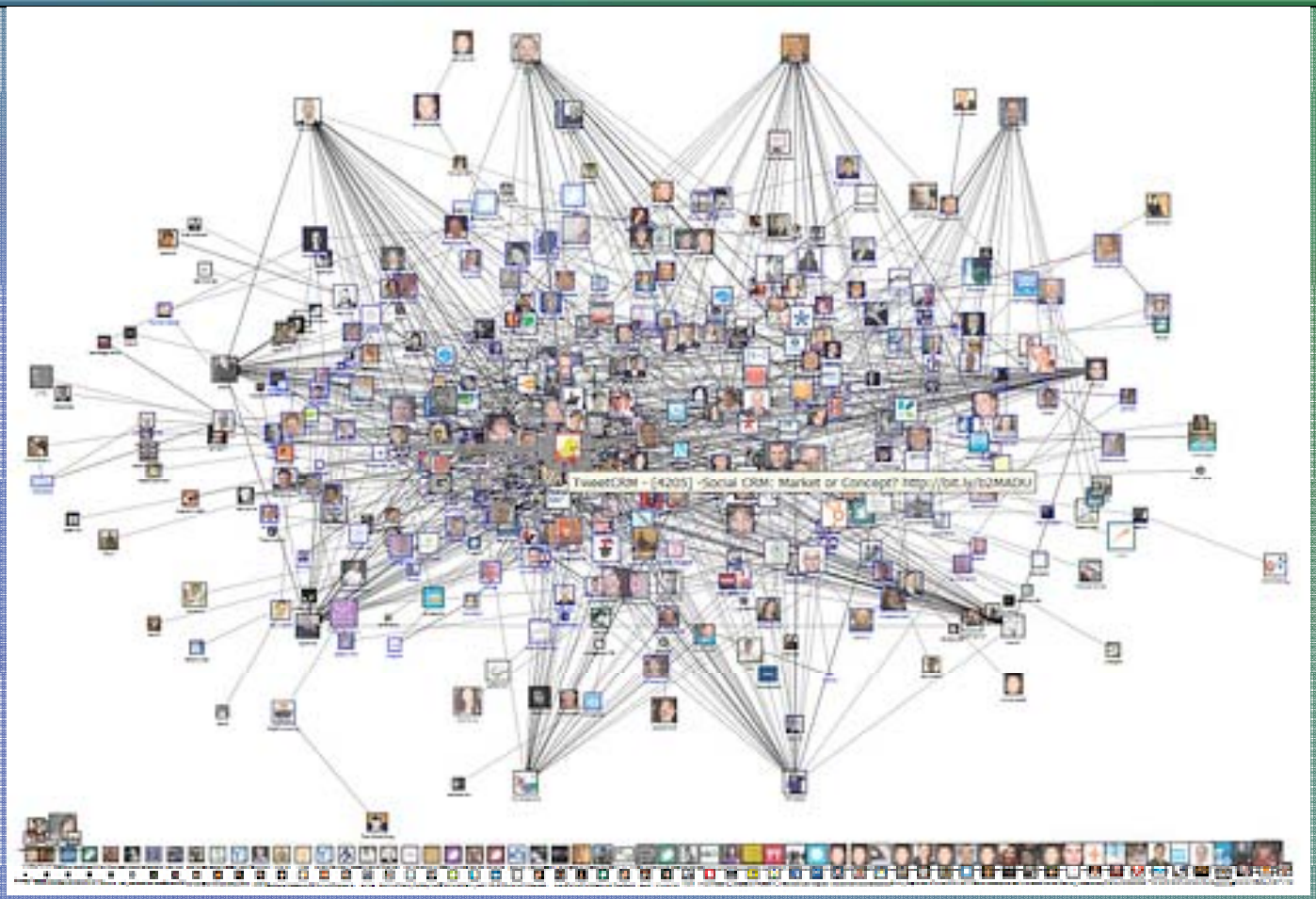
Facebook incentivizing use of RDFa like buttons

Tetherless World Constellation



OGP creates a fast-growing, multiply-labeled, network

Tetherless World Constellation



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

Important Real World Use Case: 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 …

June30,2009

Putting Govt Data online- Data.gov.uk beta

January 19, 2010

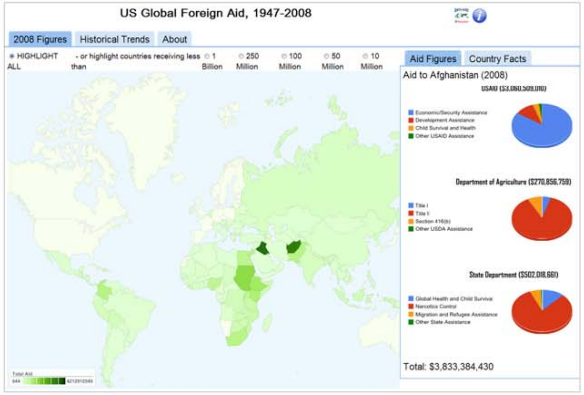
data.gov.uk online

~6000 Data Set

Government Mashups and Applications

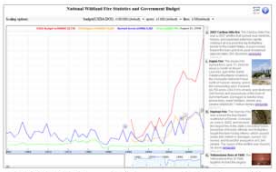
Tetherless World Constellation

(a) White House visitor search

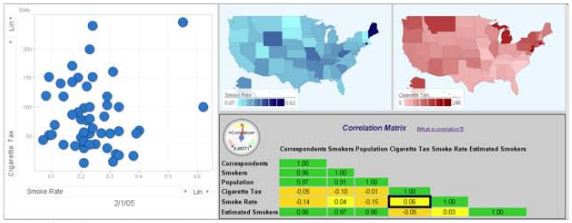


(b) US-UK Foreign Aid Comparison

(c) Agency Budget and NYTimes

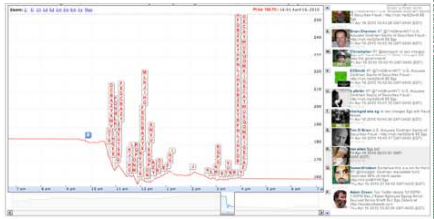


(d) Wildland fire and DBpedia



(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

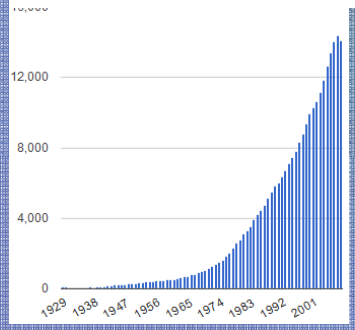
(k) [RSS] data.gov updates

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

Linking GDP of the US and China

Tetherless World Constellation

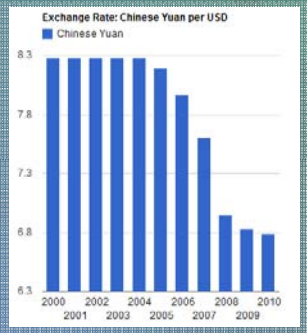
GDP of the US (Billion Dollar)



Federal Reserve Statistical Release

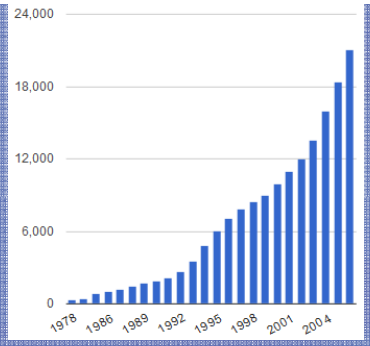
H.10

Foreign Exchange Rates (Weekly)



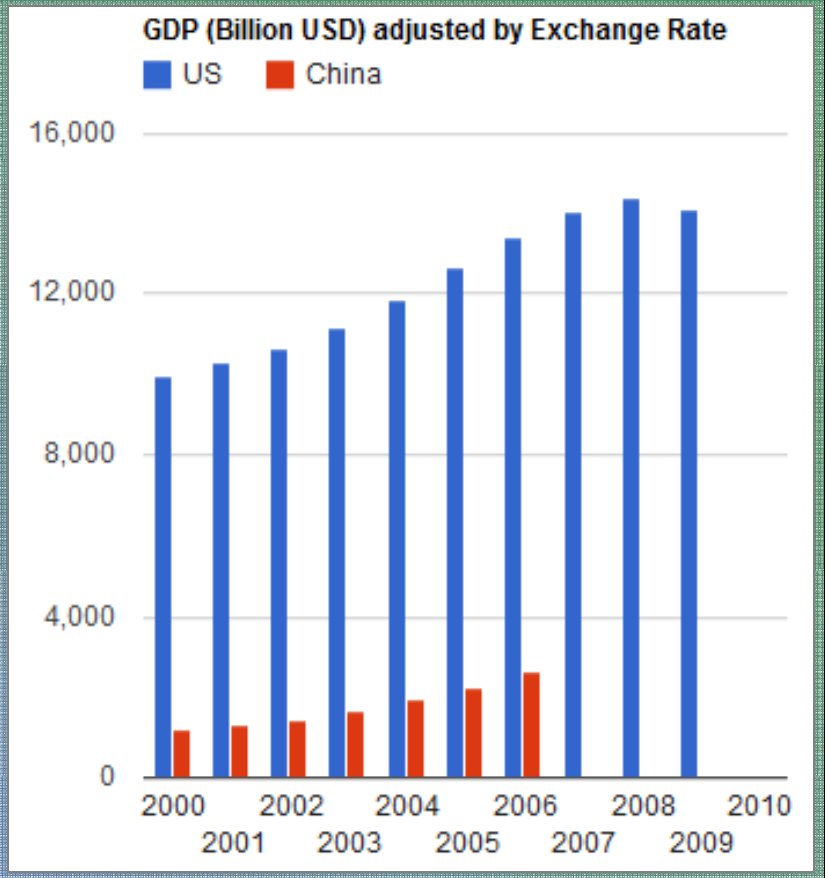
中华人民共和国国家统计局
National Bureau of Statistics of China

GDP of China (Billion Chinese Yuan)



TWC LOGD

Rensselaer



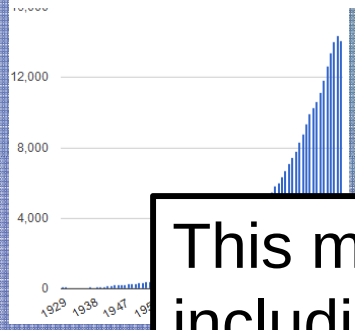
[Temporal Mashup] bea.gov + federalreserve.gov +stats.gov.cn

Linking GDP of the US and China

Tetherless World Constellation

BEA
BUREAU OF ECONOMIC ANALYSIS
U.S. DEPARTMENT OF COMMERCE

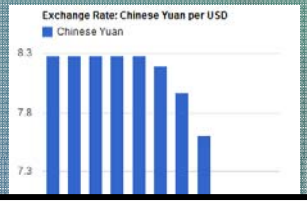
GDP of the US (Billion Dollar)



Federal Reserve Statistical Release

H.10

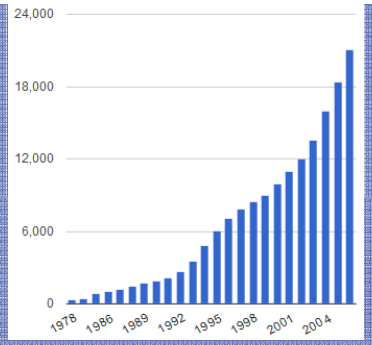
Foreign Exchange Rates (Weekly)



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

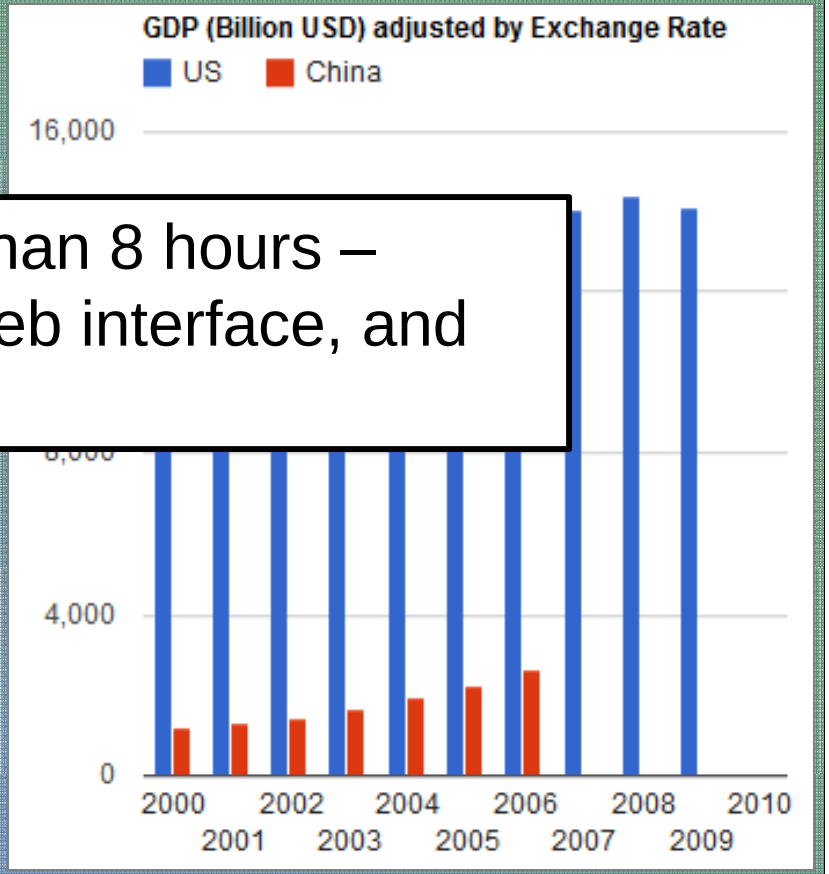
中华人民共和国国家统计局
National Bureau of Statistics of China

GDP of China (Billion Chinese Yuan)



TWC LOGD

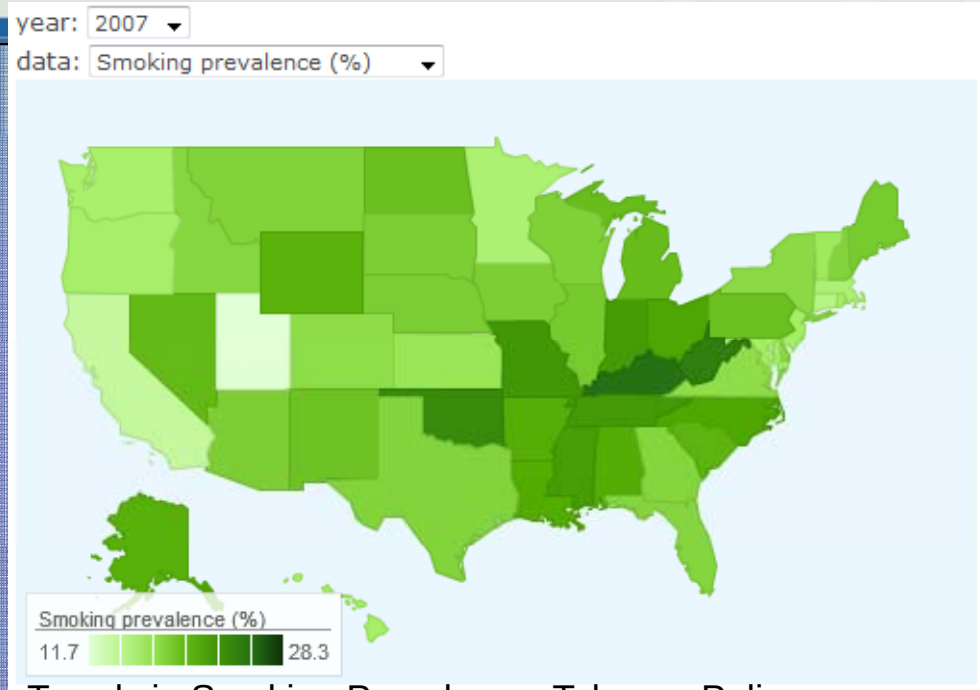
Rensselaer



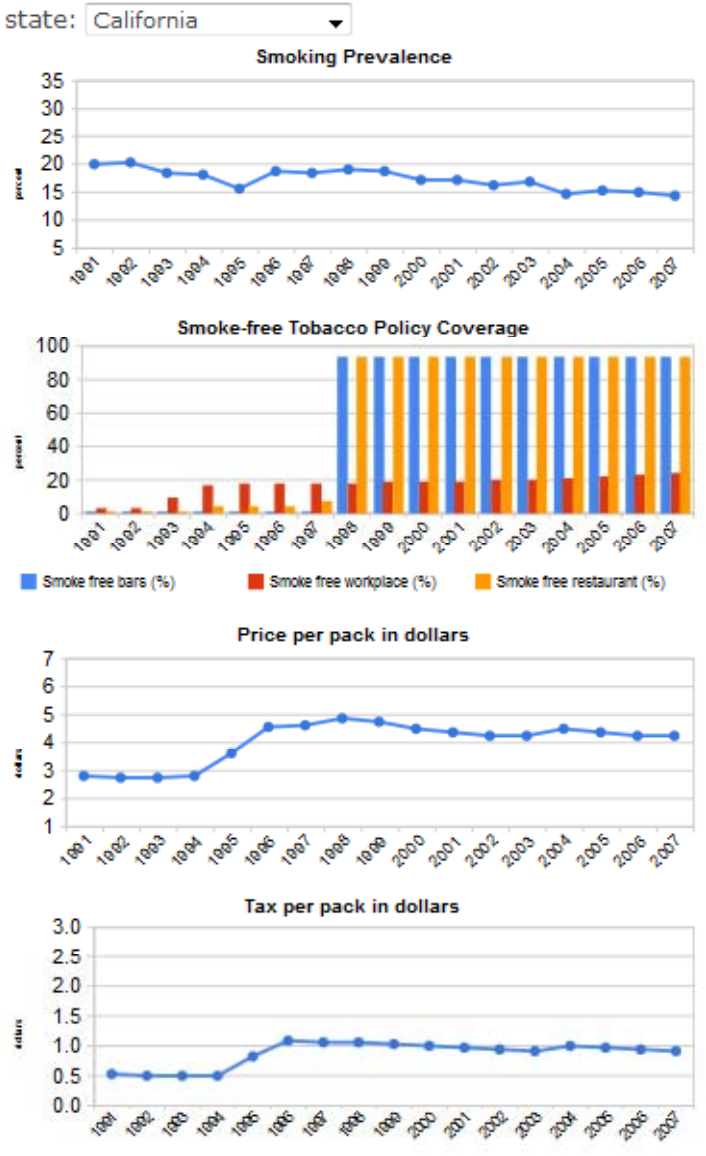
[Temporal Mashup] bea.gov + federalreserve.gov +stats.gov.cn

Mashups allow comparisons that single data sets cannot

Tetherless World Constellation



Trends in Smoking Prevalence, Tobacco Policy Coverage and Tobacco Prices (1991-2007)



Extensible Mashups via Linked Data

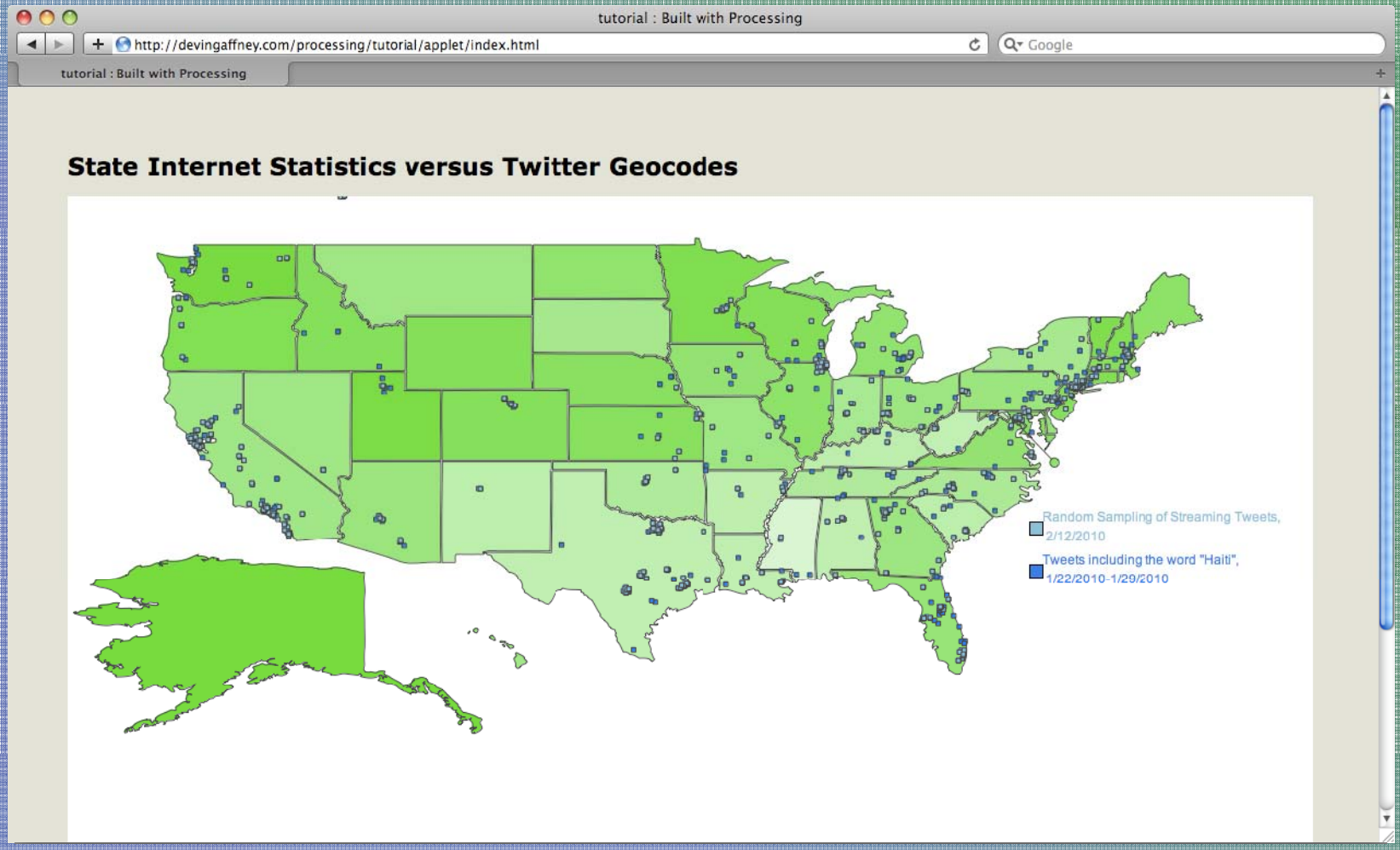
- ❑ Diverse datasets from NIH
- ❑ Potentially linking to “unemployment rate”

Accountable Mashups via Provenance

- ❑ Annotate datasets used in demos
- ❑ Feedback users’ comment to gov contact (e.g. %)

Govt data linked to Social Media Metadata

Tetherless World Constellation

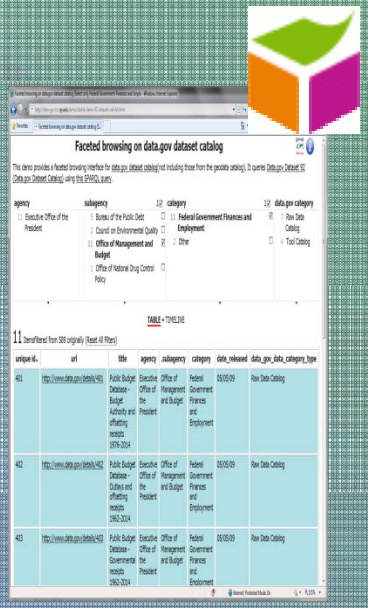


There is a lot of workflow information in the mix

Tetherless World Constellation

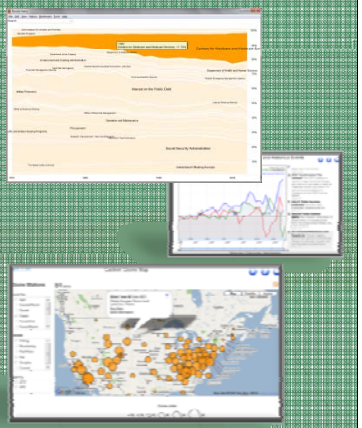
Convert

derive



Access

derive



Enhance

create

revision

Version

SemDiff

derive

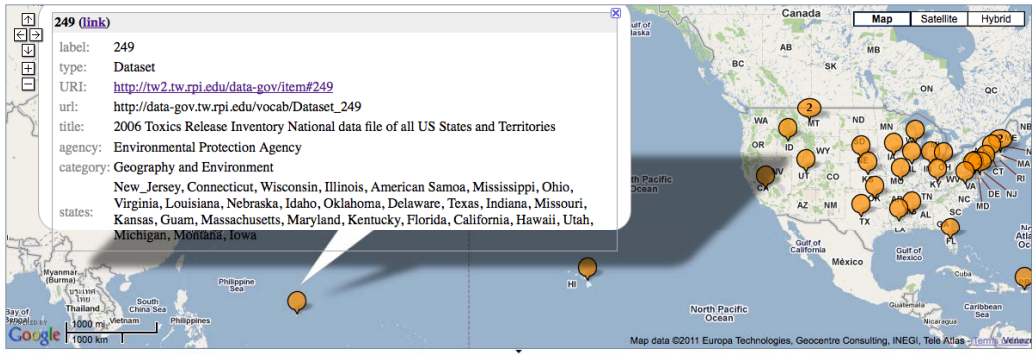
A Web Science Challenge

Tetherless World Constellation

State-related Dataset Faceted Search

Datasets	States	1 ☑	Agency	Category
1 249	17 Alabama ☐		2 Environmental Protection Agency	2 Geography and Environment
1 307	12 Alaska ☐			
	2 **American Samoa** ☑			
	6 Arizona ☐			
	10 Arkansas ☐			
	9 California ☐			

2 Dataset filtered from 21 originally (Reset All Filters)



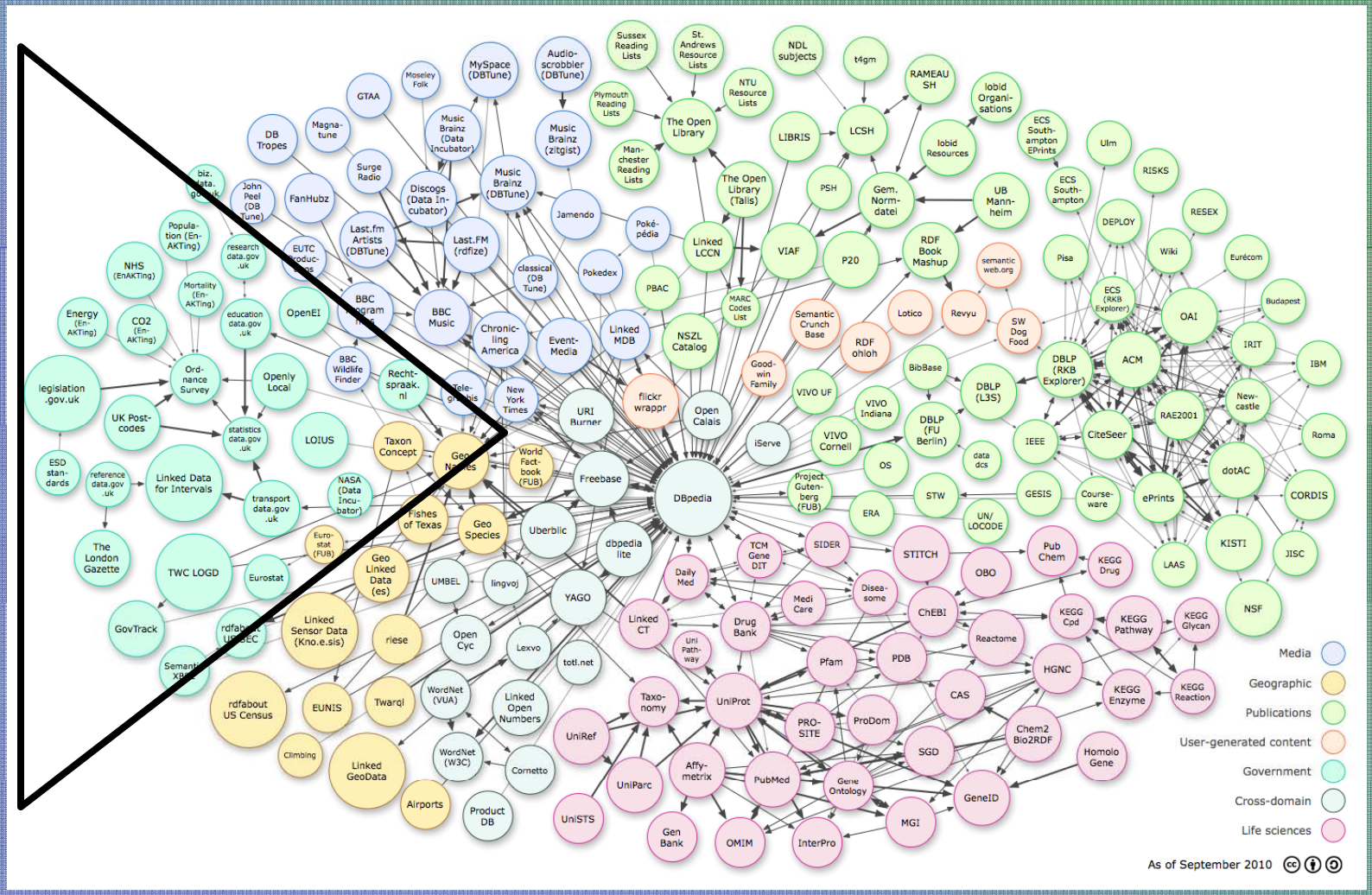
2 Dataset filtered from 21 originally (Reset All Filters)

Dataset ID	URL	Title	Agency	Category	State-related Predicates
249	http://data-gov.tw.rpi.edu/vocab/Dataset_249	2006 Toxics Release Inventory National data file of all US States and Territories	Environmental Protection Agency	Geography and Environment	• /191/stack_air_emissions_release_pounds • /191/mailing_state • /191/facility_state • /191/total_transferred_off_site_for_further_waste_management • /191/other_on_site_waste_management • /191/total_stack_air_emissions
307	http://data-gov.tw.rpi.edu/vocab/Dataset_307	2007 Toxics Release Inventory National data file of all US States and Territories	Environmental Protection Agency	Geography and Environment	• /191/mailing_state • /191/stack_air_emissions_release_pounds • /191/facility_state • /191/total_stack_air_emissions

How can we search for data?

Effective open govt requires exploiting the linked open govt network

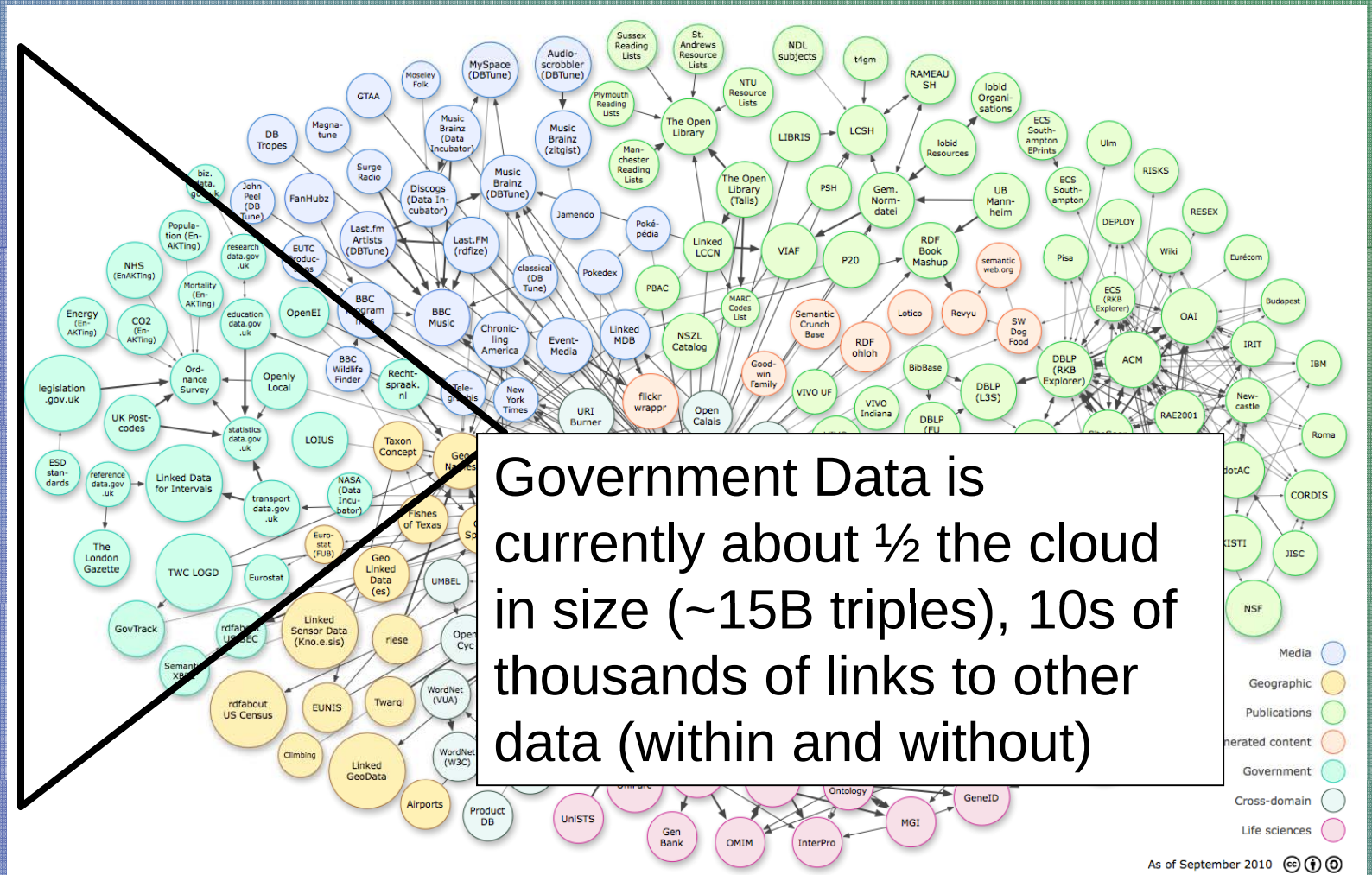
Tetherless World Constellation



http://linkeddata.org/

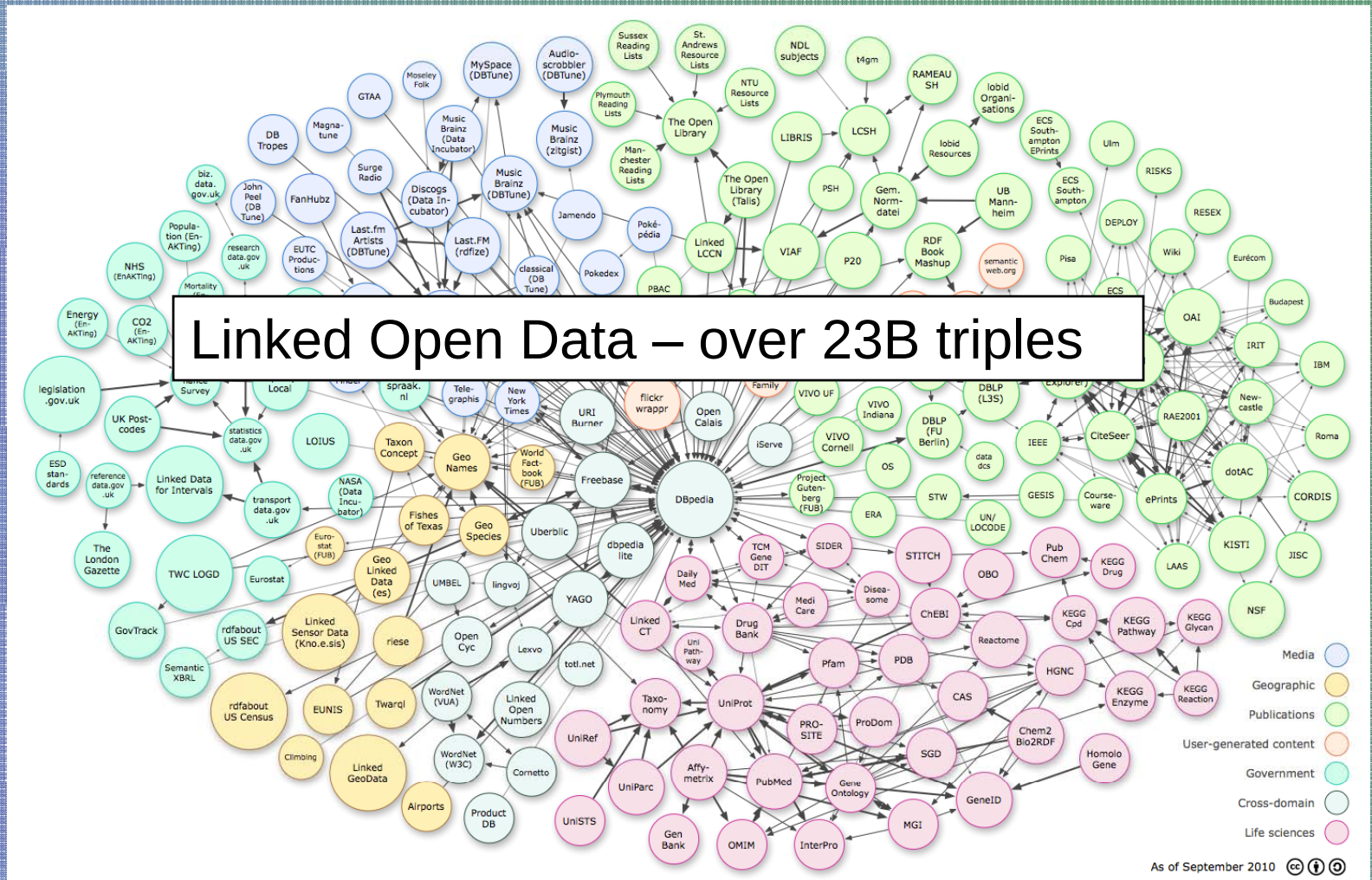
Effective open govt requires exploiting the linked open govt network

Tetherless World Constellation



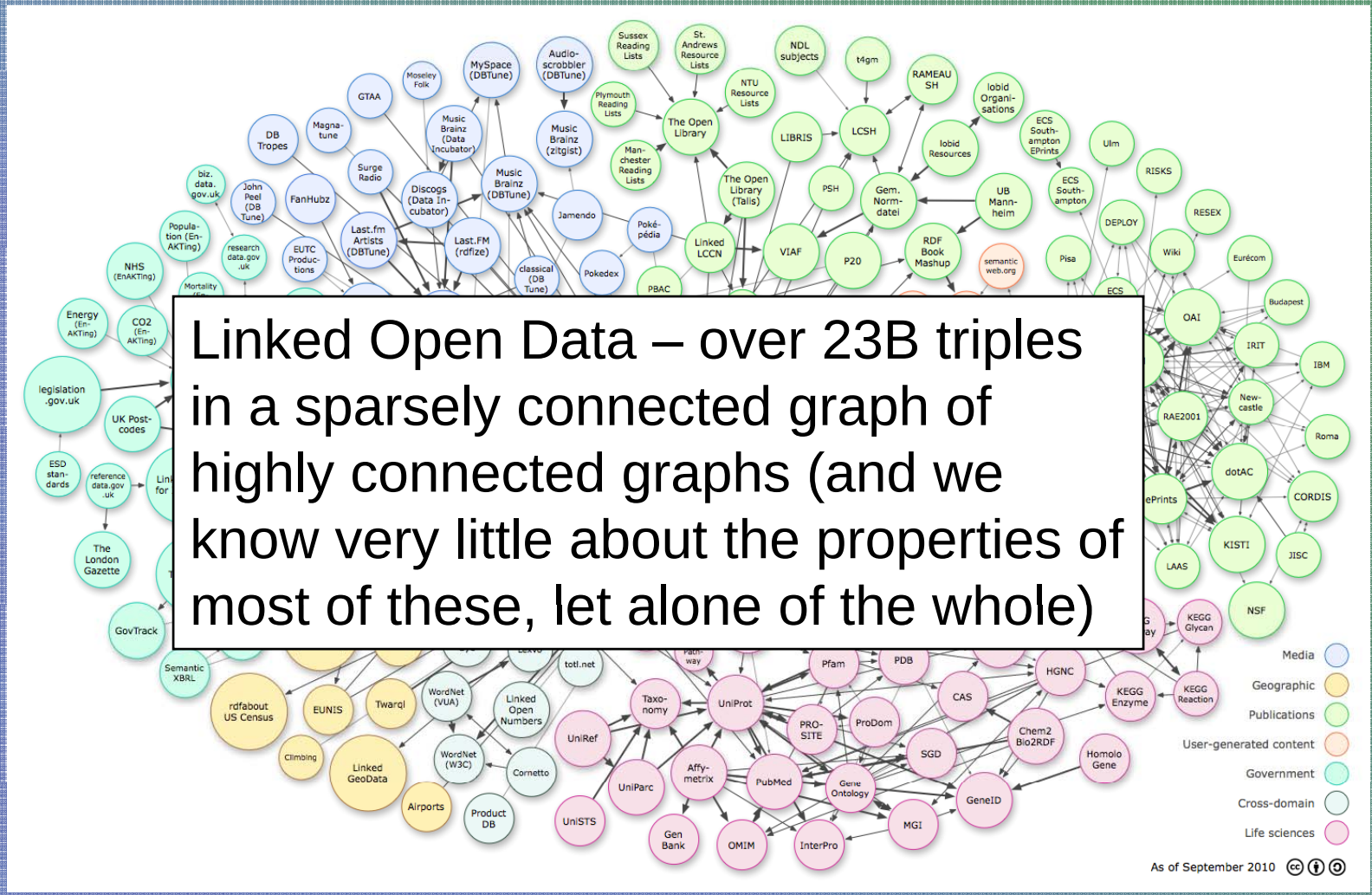
Linked Open Data on the Web

Tetherless World Constellation



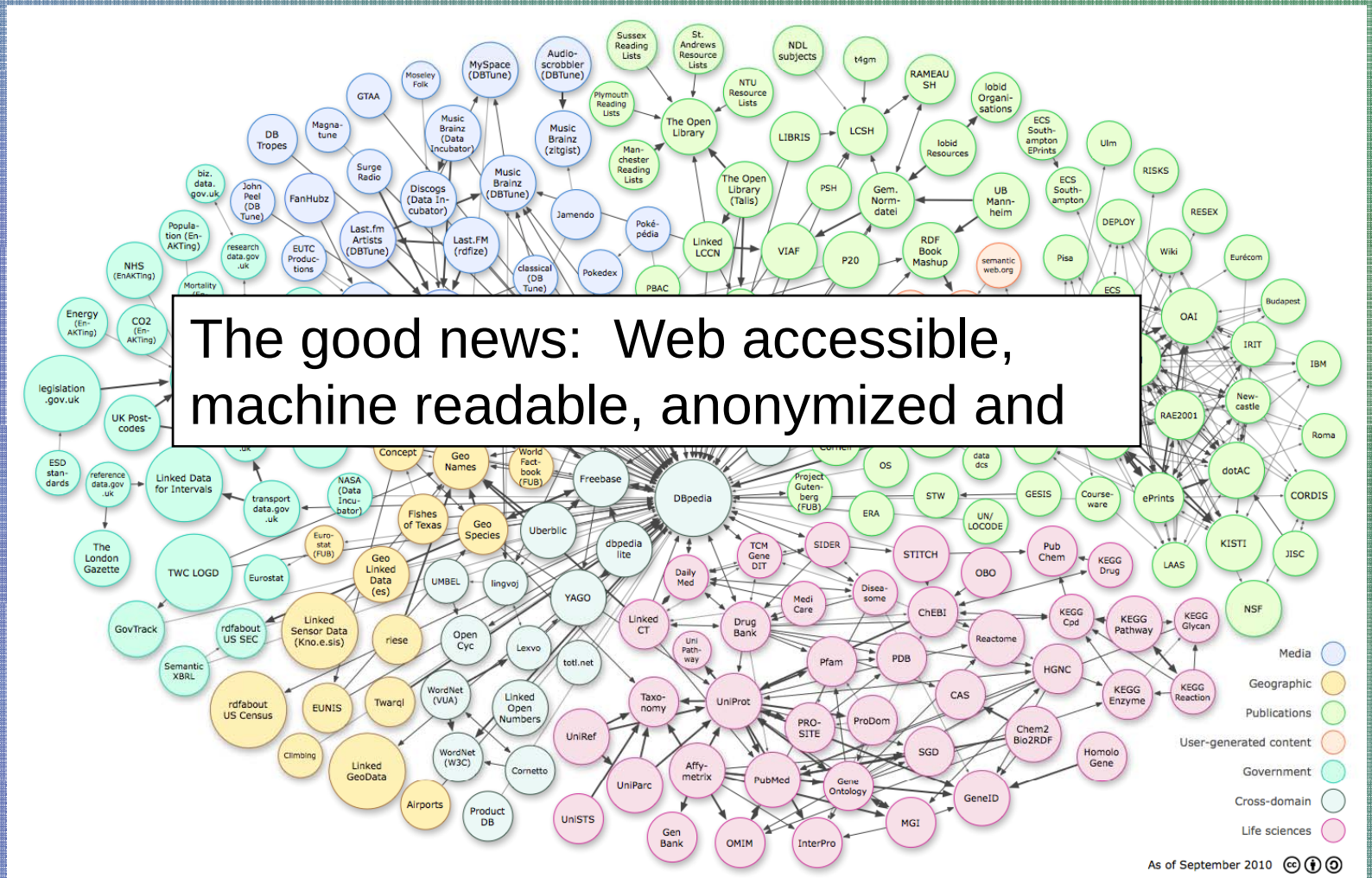
Linked open data network

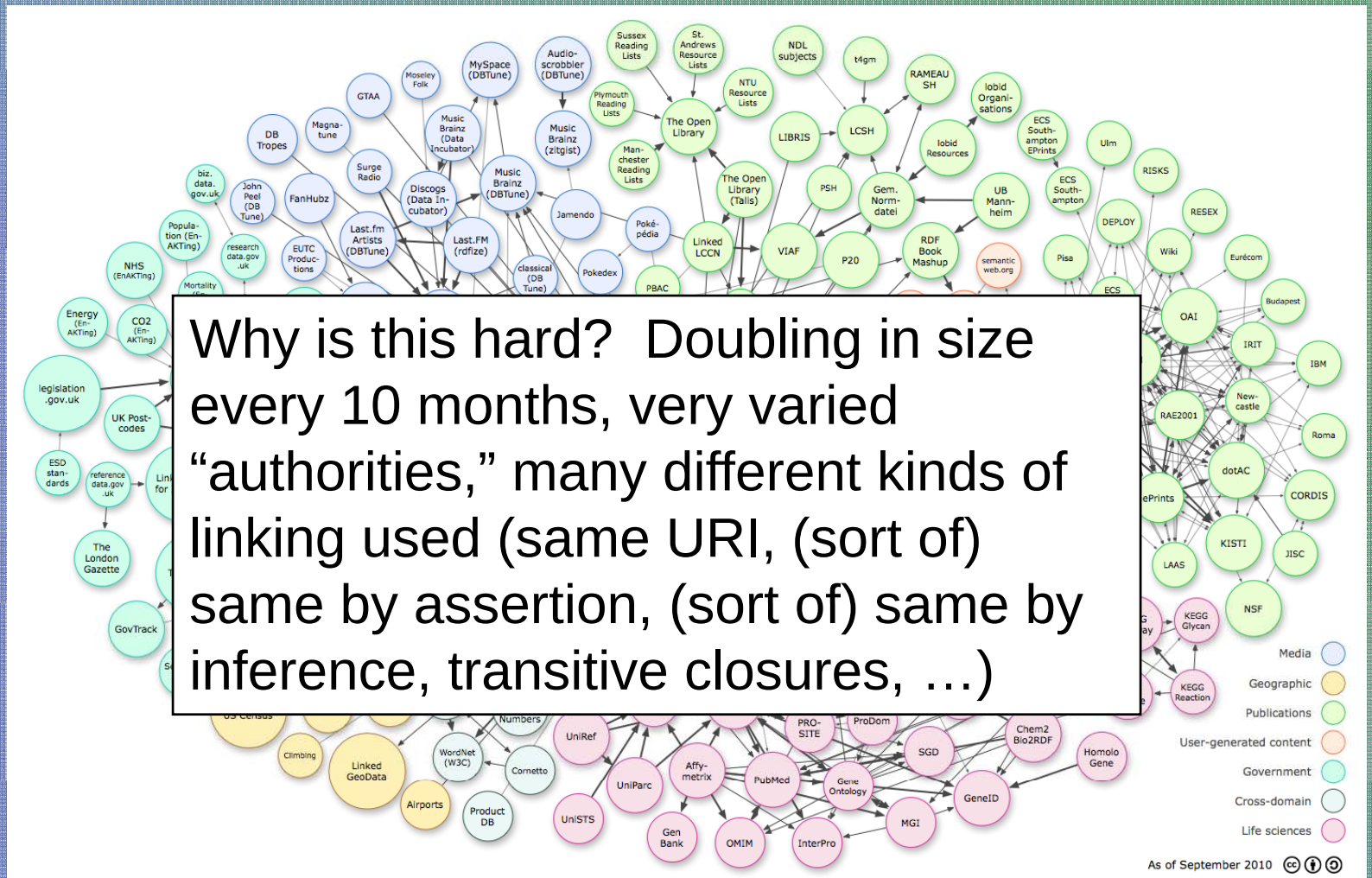
Tetherless World Constellation



Linked open data network

Tetherless World Constellation




Why is this hard? Doubling in size every 10 months, very varied “authorities,” many different kinds of linking used (same URI, (sort of) same by assertion, (sort of) same by inference, transitive closures, …)
Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences
As of September 2010

A new buzzword...

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.

And a new commercial motivator

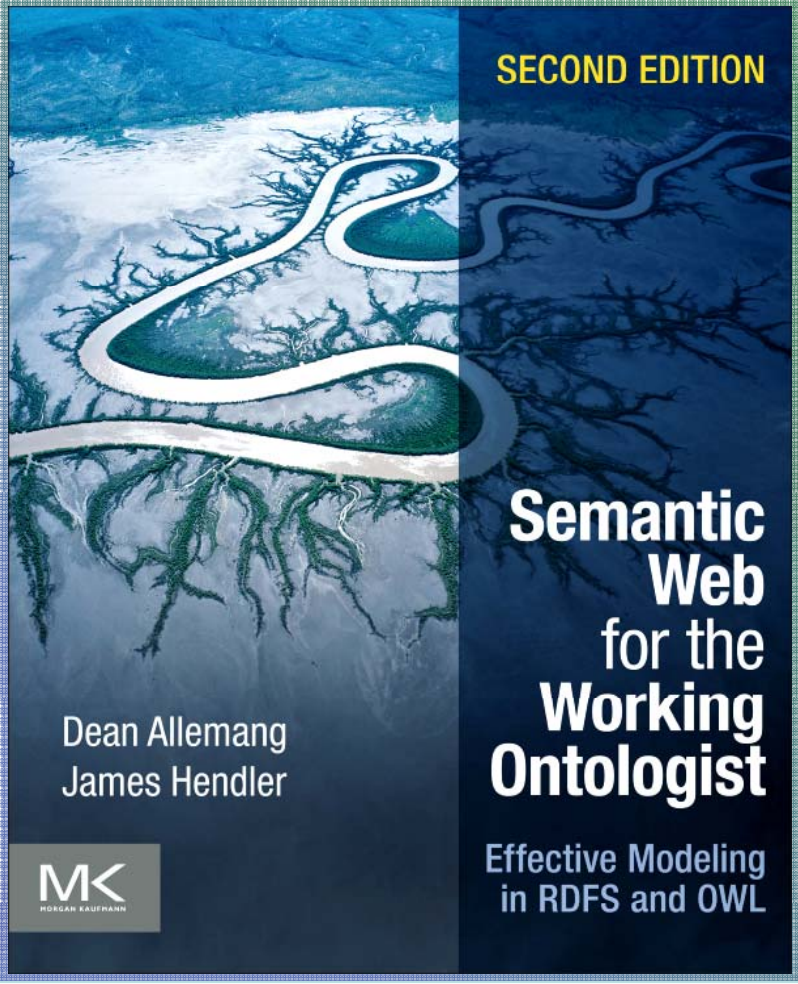
Tetherless World Constellation

- Web: Google
  - Underlying theory: exploit the Web graph
- Web 2.0: Facebook, YouTube, Twitter…
    - Underlying theory: exploit the Social network
  - Web 3.0: (Your company here)
    - Underlying theory: exploit the “semantics” in all these graphs

- Semantic Web is real
- Growing at a fast pace
- Producing lots of interesting networks
- That no one is really analyzing from a network science perspective
- Which could hugely help those of us trying to use this for some really hard real world problems
    - For example, open govt data

Questions?

Tetherless World Constellation



New edition includes OGP, Data.gov, …