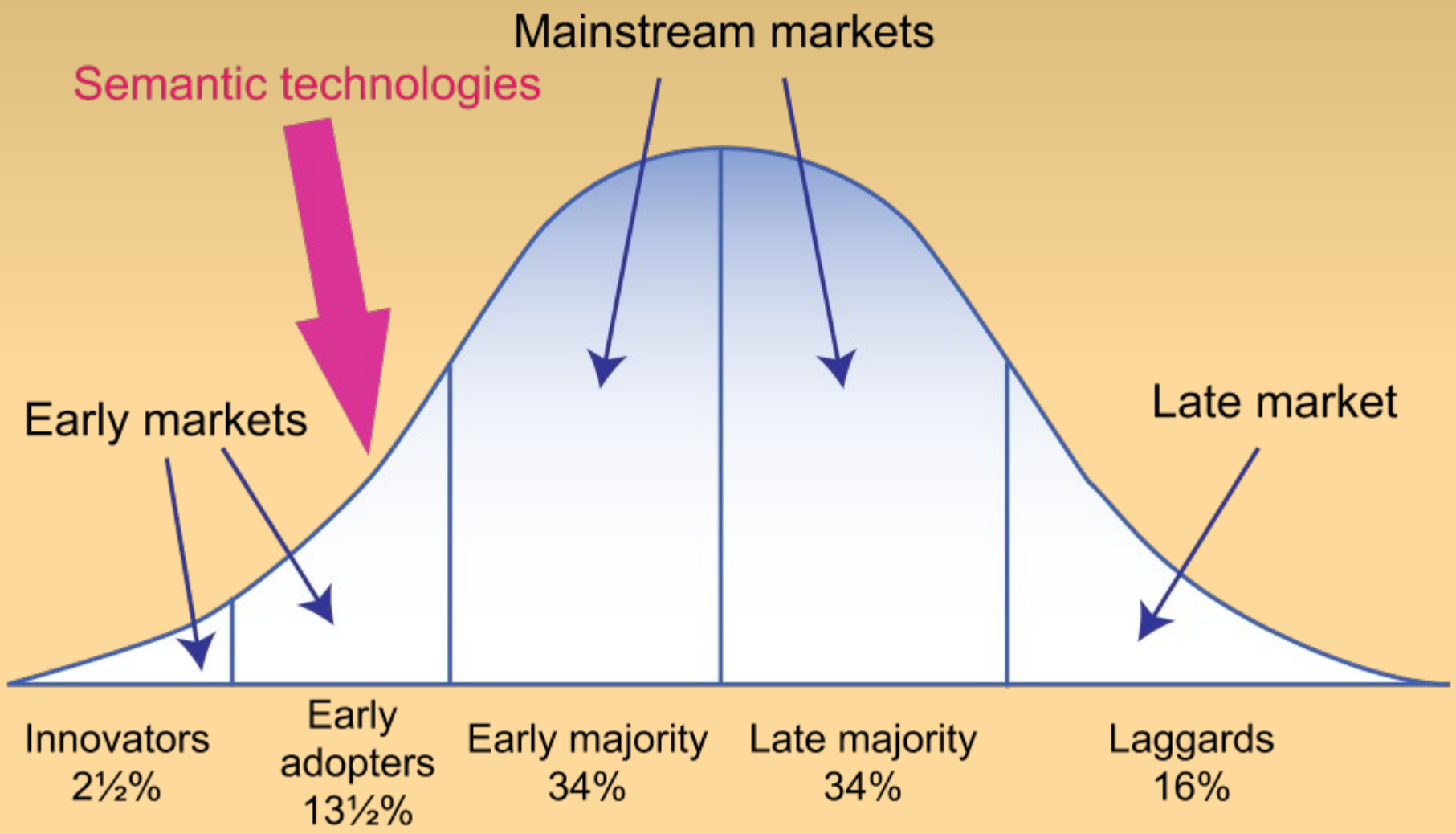
What is being done today?

Ivan Herman, W3C

Deutsche Telekom Workshop
Darmstadt, Germany
2009-12-14

Technology adoption life cycle



The 2007 Gartner predictions

During the next 10 years, Web-based technologies will improve the ability to embed semantic structures [… it] will occur in multiple evolutionary steps…

By 2017, we expect the vision of the Semantic Web […] to coalesce […] and the majority of Web pages are decorated with some form of semantic hypertext.

By 2012, 80% of public Web sites will use some level of semantic hypertext to create SW documents […] 15% of public Web sites will use more extensive Semantic Web-based ontologies to create semantic databases

(note: "semantic hypertext" refers to, eg, RDFa, microformats with possible GRDDL, etc.)

"Finding and Exploiting Value in Semantic Web Technologies on the Web", Gartner Report, May 2007

The “corporate” landscape is moving

- Major companies offer (or will offer) Semantic Web tools or systems using Semantic Web: Adobe, Oracle, IBM, HP, Software AG, GE, Northrop Gruman, Altova, Microsoft, Dow Jones, …
- Others are using it (or consider using it) as part of their own operations: Novartis, Pfizer, Telefónica, …
- Some of the names of active participants in W3C SW related groups: HP, Agfa, SRI International, Fair Isaac Corp., Oracle, Boeing, IBM, Chevron, Siemens, Nokia, Pfizer, Sun, Eli Lilly, Deutsche Telekom, …

Lots of Tools (*not* an exhaustive list!)

- Categories:
  - Triple Stores
  - Inference engines
  - Converters
  - Search engines
  - Middleware
  - CMS
  - Semantic Web browsers
  - Development environments
  - Semantic Wikis
  - …
- Some names:
  - Jena, AllegroGraph, Mulgara, Sesame, flickurl, …
  - TopBraid Suite, Virtuoso environment, Falcon, Drupal 7, Redland, Pellet, …
  - Disco, Oracle 11g, RacerPro, IODT, Ontobroker, OWLIM, Talis Platform, …
  - RDF Gateway, RDFLib, Open Anzo, DartGrid, Zitgist, Ontotext, Protégé, …
  - Thetus publisher, SemanticWorks, SWI-Prolog, RDFStore…
  - …

May start with specific communities

- The needs of a deployment application area:
  - have serious problem or opportunity
  - have the intellectual interest to pick up new things
  - have motivation to fix the problem
  - its data connects to other application areas
  - have an influence as a showcase for others
- The high energy physics community played this role for the Web in the 90's

Some deployment communities

- Major communities pick the technology up: digital libraries, defense, eGovernment, energy sector, financial services, health care, oil and gas industry, life sciences, publishing …
  - Health care and life science sector is also active at W3C
    - also at W3C, in the form of an Interest Group

Some deployment communities

- Semantic Web also appear in the “Web 2.0/Web 3.0” applications (whatever that means)
  - exchange of social data
  - personal “space” applications
  - dynamic Web site backends
  - multimedia asset management
  - etc

W3C's use case collection

- W3C is actively collecting SW use cases and case studies
  - use case: prototype applications within the enterprise
  - case study: deployed applications, either in an enterprise, community, governmental, etc sites

SWEO's use case collection

- At present there are
  - 24 case studies and 12 use cases (March 2009)
  - from countries around the globe
  - activity areas include: automotive, broadcasting, financial institution, health care, oil & gas industry, pharmaceutical, public and governmental institutions, publishing, telecommunications, …
  - usage areas include: data integration, portals with improved local search, business organization, B2B integration, …
- Remember this URI:
  `http://www.w3.org/2001/sw/UseCases/`

So how do applications look like?

Application patterns

- It is fairly difficult to “categorize” applications
- With this caveat, some of the application patterns:
  - data integration
  - intelligent (specialized) Web sites (portals) with improved local search
  - content and knowledge organization
  - knowledge representation, decision support
  - X2X integration (often combined with Web Services)
  - data registries, repositories
  - collaboration tools (eg, social network applications)

Applications are not always very complex…

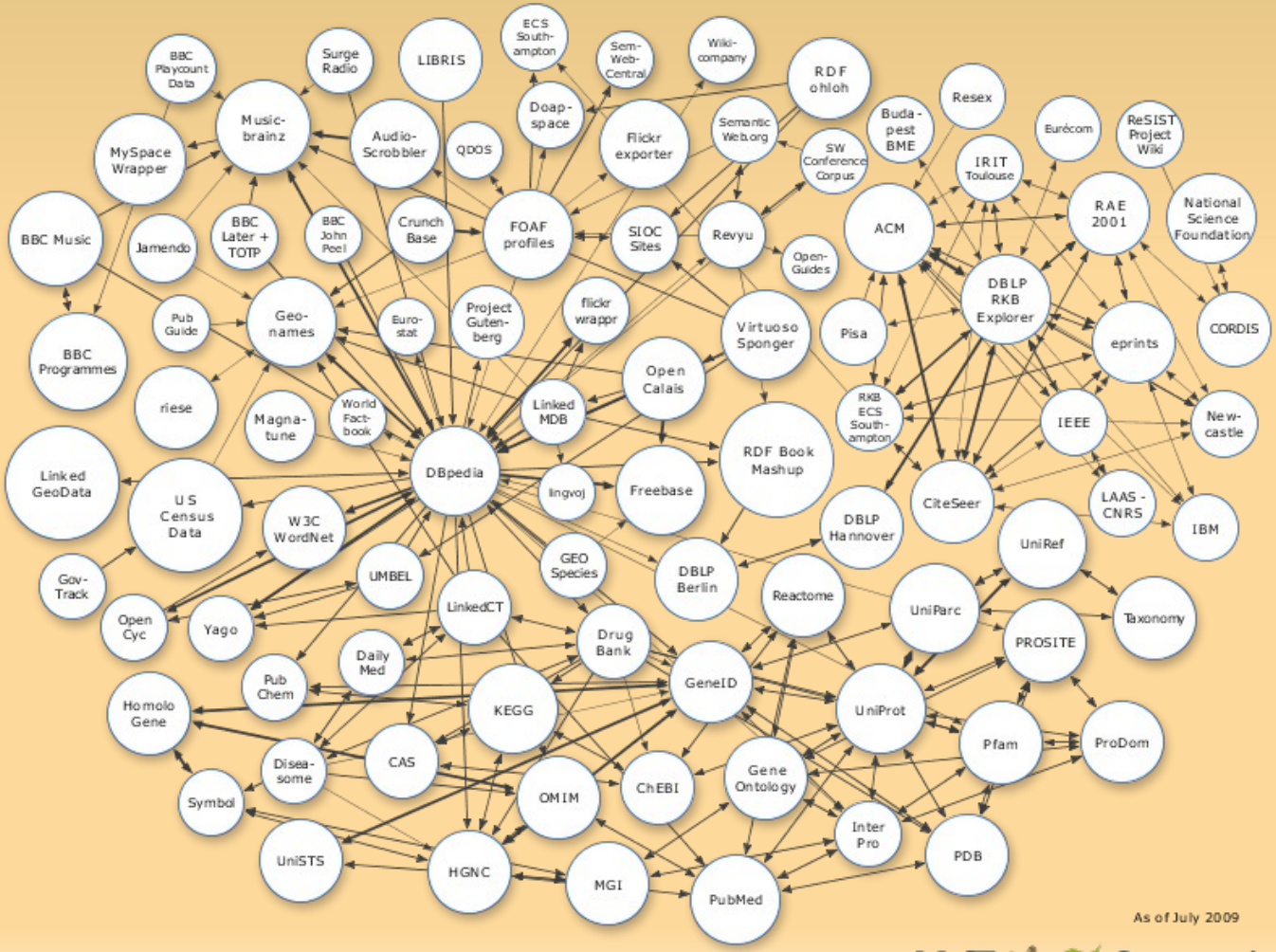
- Eg: simple semantic annotations of data provides easy integration (eg, with MusicBrainz, Wikipedia, geographic data sets, etc)
- What is needed: some simple vocabularies, simple annotation
  - annotation an be generated by a server automatically, or
  - added by the user via some user interface
- This extra data can be in some microformats, in RDFa, …

To “seed” a Web of Data...

- Data has to be published, ready for integration
- And this is now happening!
  - Linked Open Data project
  - eGovernmental initiatives in, eg, UK, USA, France,...
  - Various institutions publishing their data

Linking Open Data Project

- Goal: “expose” open datasets in RDF
- *Set RDF links among the data items* from different datasets
- Set up SPARQL endpoints
- Billions triples, millions of “links”



W3C Semantic Web

Example data source: DBpedia

- DBpedia is a community effort to
  - extract structured (“infobox”) information from Wikipedia
  - provide a SPARQL endpoint to the dataset
  - interlink the DBpedia dataset with other datasets on the Web

Extracting structured data from Wikipedia

Amsterdam

The Keizersgracht at dusk

Location of Amsterdam

Coordinates: 52°22'23"N 4°53'32"E

Country	Netherlands
Province	North Holland
Government	
- Type	Municipality
- Mayor	Job Cohen[1] (PvdA)
- Aldermen	Lodewijk Asscher Carolien Gehrels Tjeerd Herrema Maarten van Poelgeest Marijke Vos
- Secretary	Erik Gerritsen
Area [2][3]	
- City	219 km² (84.6 sq mi)
- Land	166 km² (64.1 sq mi)
- Water	53 km² (20.5 sq mi)
- Urban	1,003 km² (387.3 sq mi)
- Metro	1,815 km² (700.8 sq mi)
Elevation [4]	2 m (7 ft)
Population (1 October 2008)[5][6]	
- City	755,269
- Density	4,459/km² (11,548.8/sq mi)
- Urban	1,364,422
- Metro	2,158,372
- Demonym	Amsterdammer
Time zone	CET (UTC+1)
- Summer (DST)	CEST (UTC+2)
Postcodes	1011 – 1109
Area code(s)	020

Website: www.amsterdam.nl

```
@prefix dbpedia <http://dbpedia.org/resource/>.
@prefix dbterm  <http://dbpedia.org/property/>.

dbpedia:Amsterdam
   dbterm:officialName "Amsterdam" ;
   dbterm:longd "4" ;
   ...
   dbterm:leaderName dbpedia:Job_Cohen ;
   ...
   dbterm:areaTotalKm "219" ;
   ...
dbpedia:ABN_AMRO
   dbterm:location dbpedia:Amsterdam ;
   ...
```

W3C Semantic Web

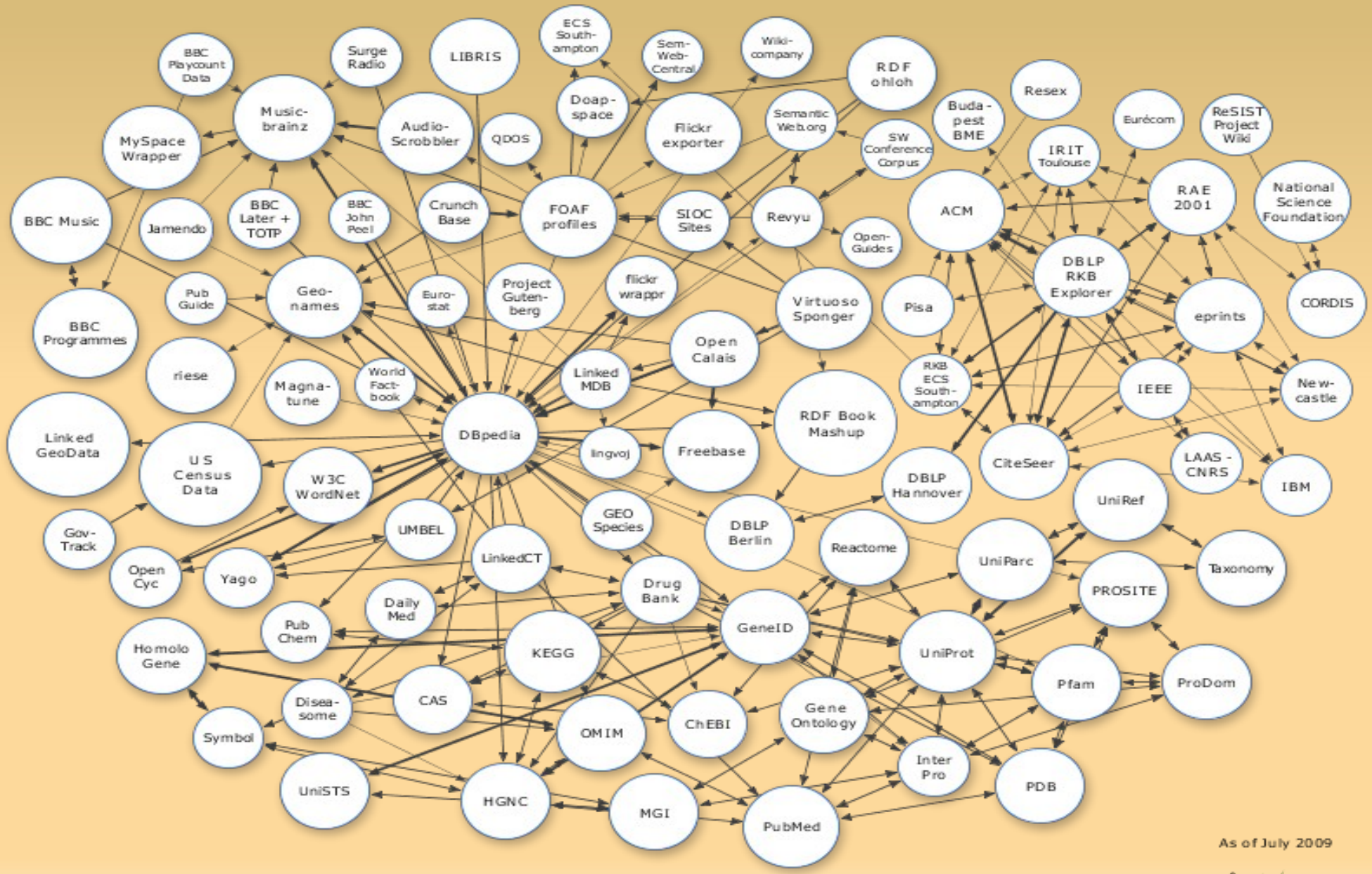
Automatic links among open datasets

```
<http://dbpedia.org/resource/Amsterdam>
  owl:sameAs <http://rdf.freebase.com/ns/...> ;
  owl:sameAs <http://sws.geonames.org/2759793> ;
  ...
```

```
<http://sws.geonames.org/2759793>
  owl:sameAs <http://dbpedia.org/resource/Amsterdam>
  wgs84_pos:lat "52.3666667" ;
  wgs84_pos:long "4.8833333" ;
  geo:inCountry <http://www.geonames.org/countries/#NL> ;
 ...
```

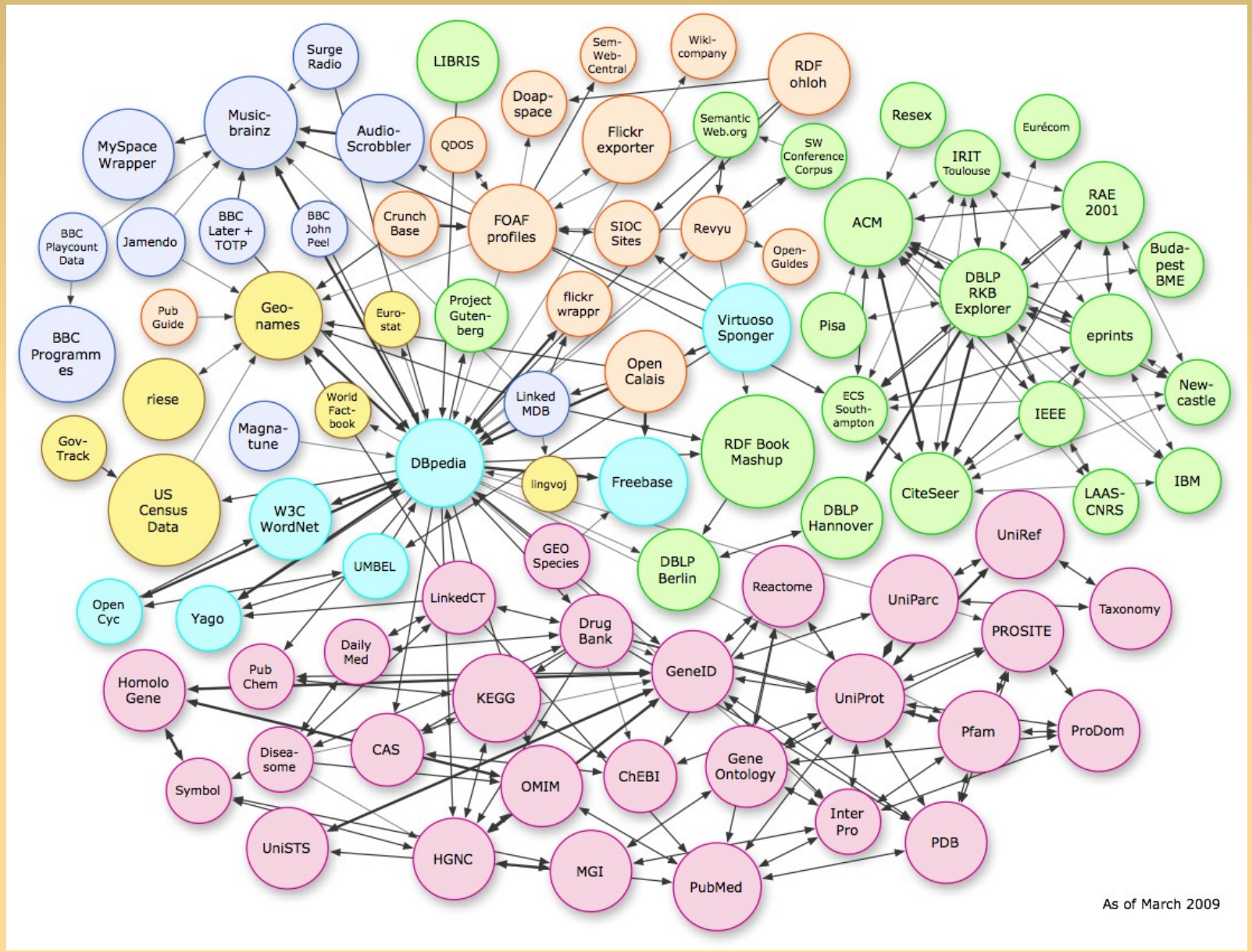
Processors can switch automatically from one to the other…

Linking Open Data Project (cont)

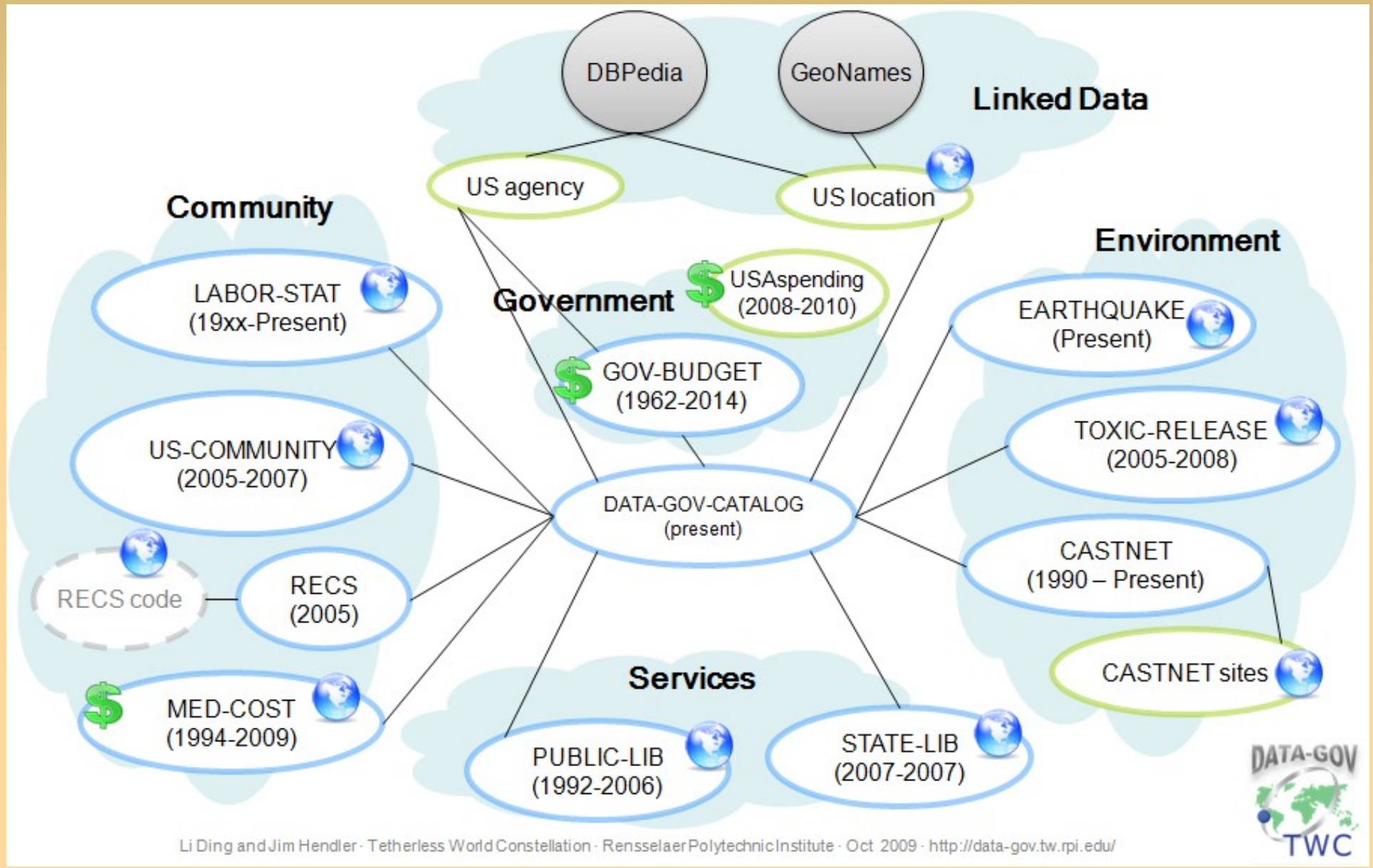


W3C Semantic Web

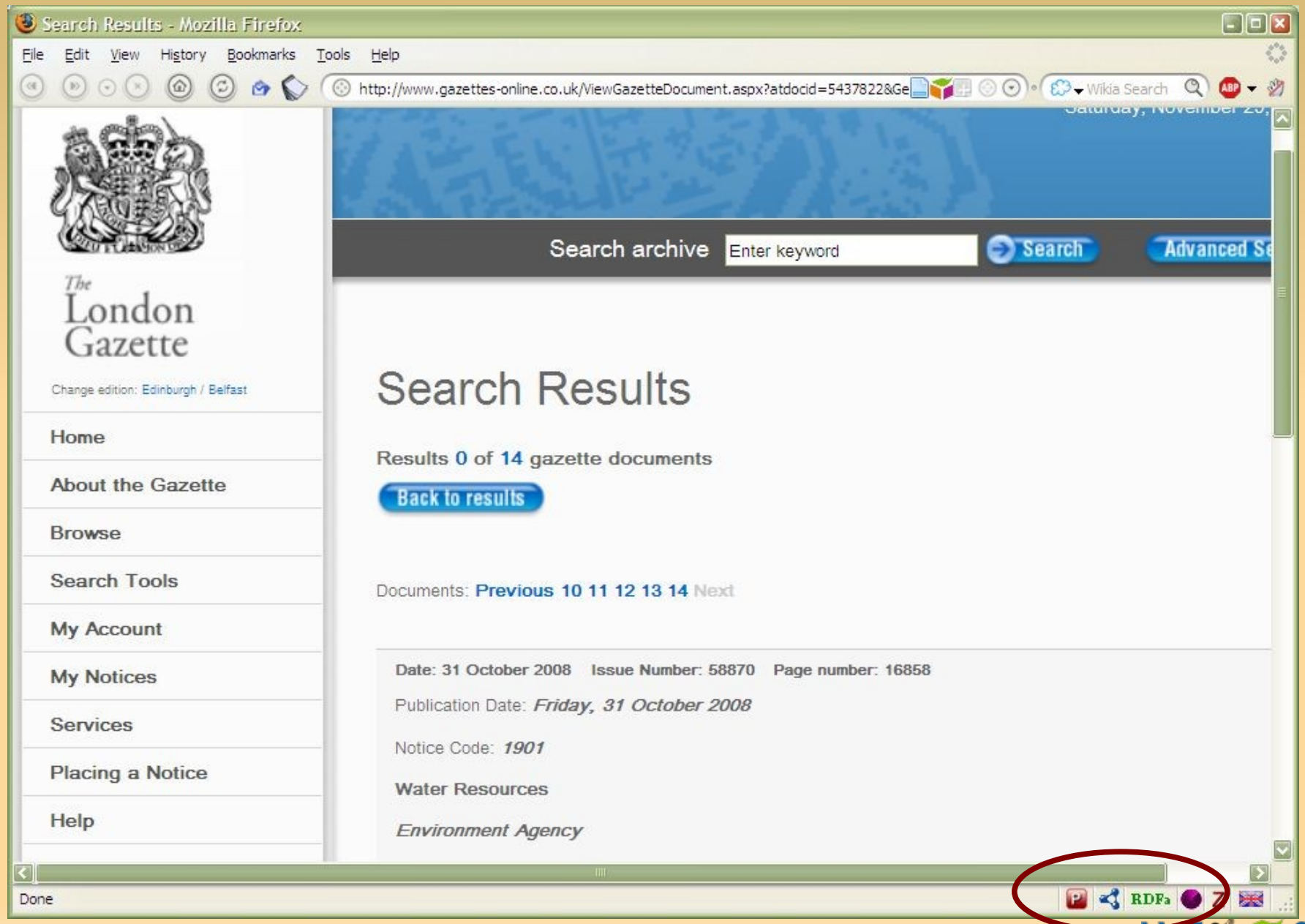
Linking Open Data Project (cont)



Linked Open eGov Data

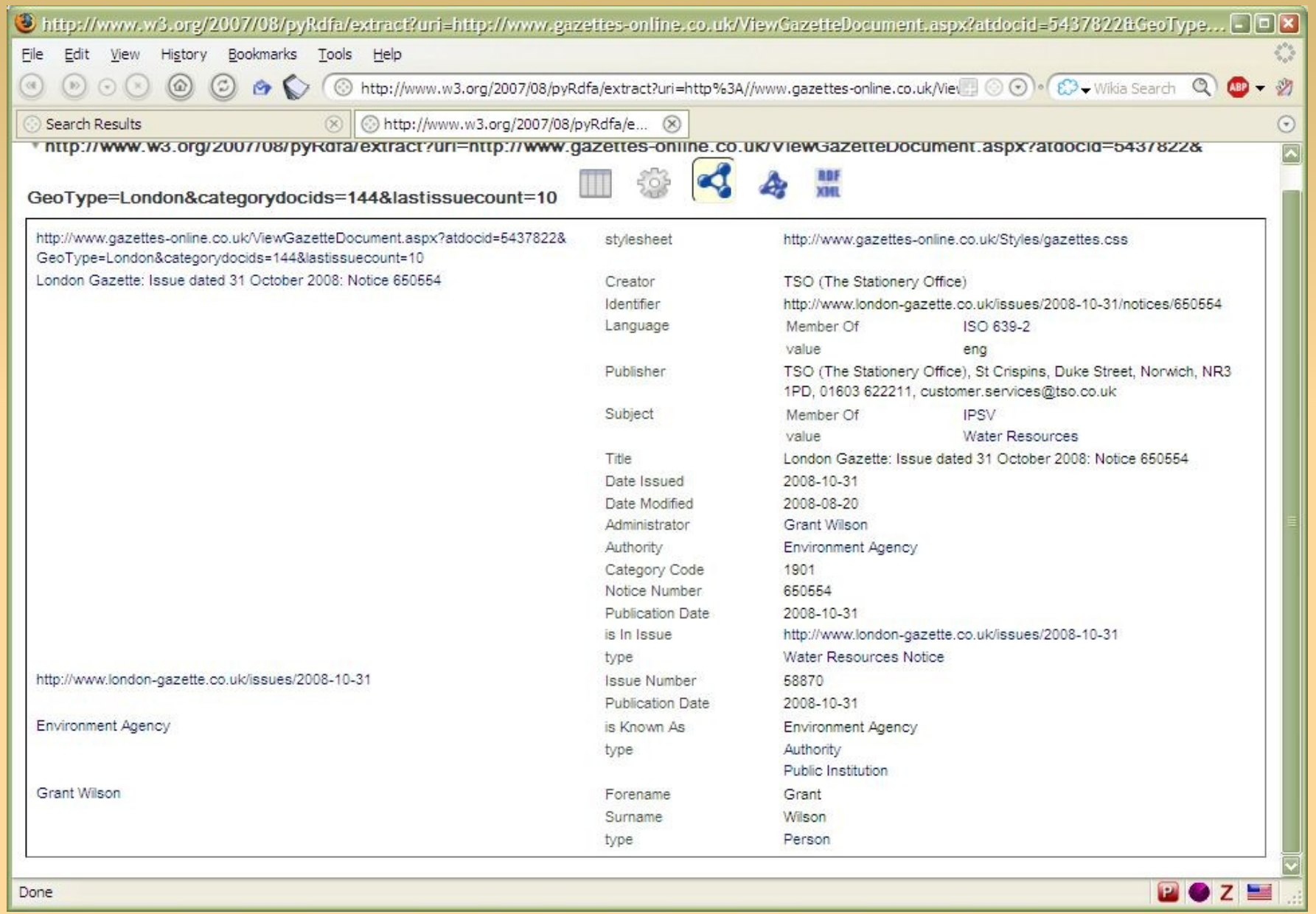


Publication of data (with RDFa): London Gazette

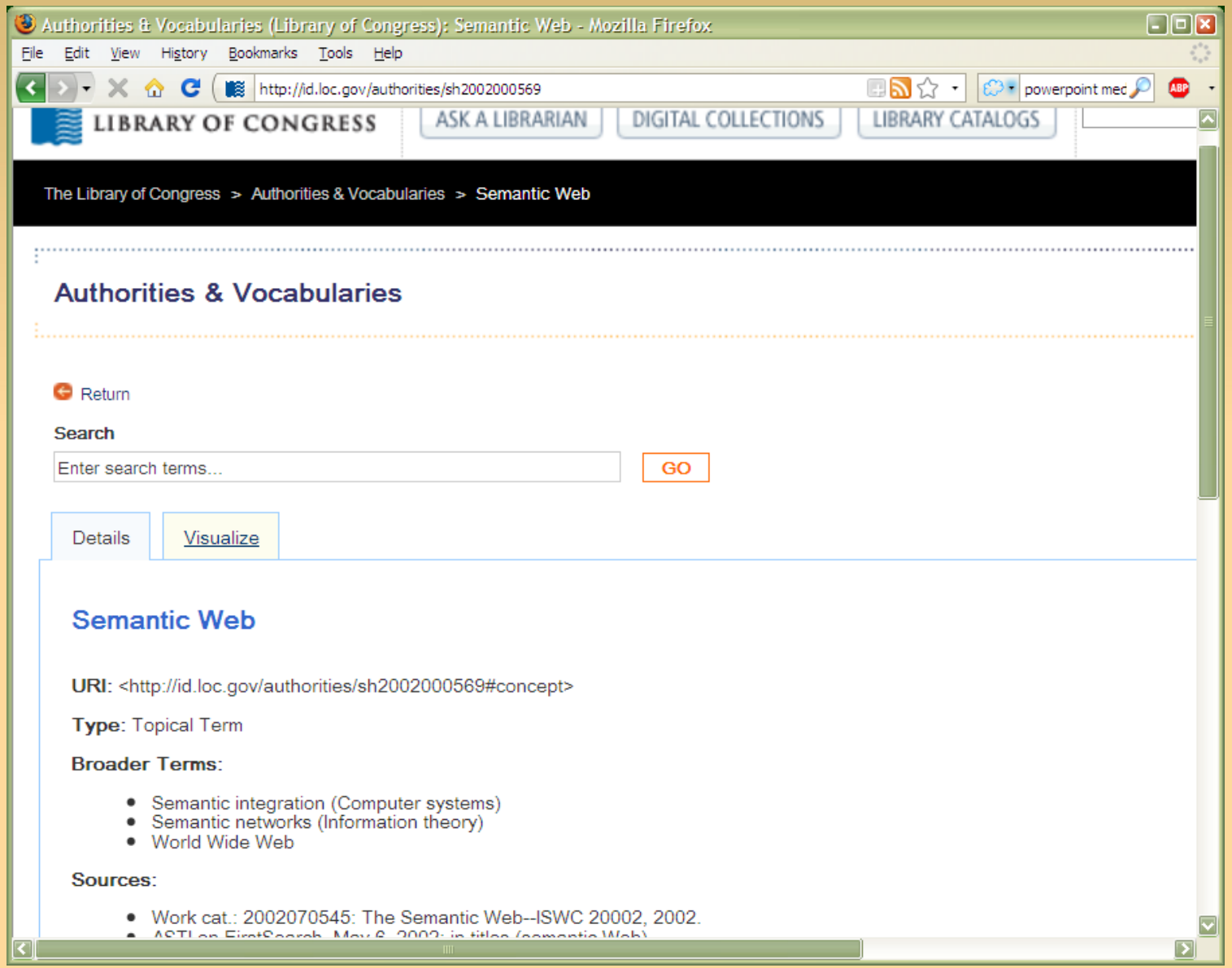


W3C Semantic Web

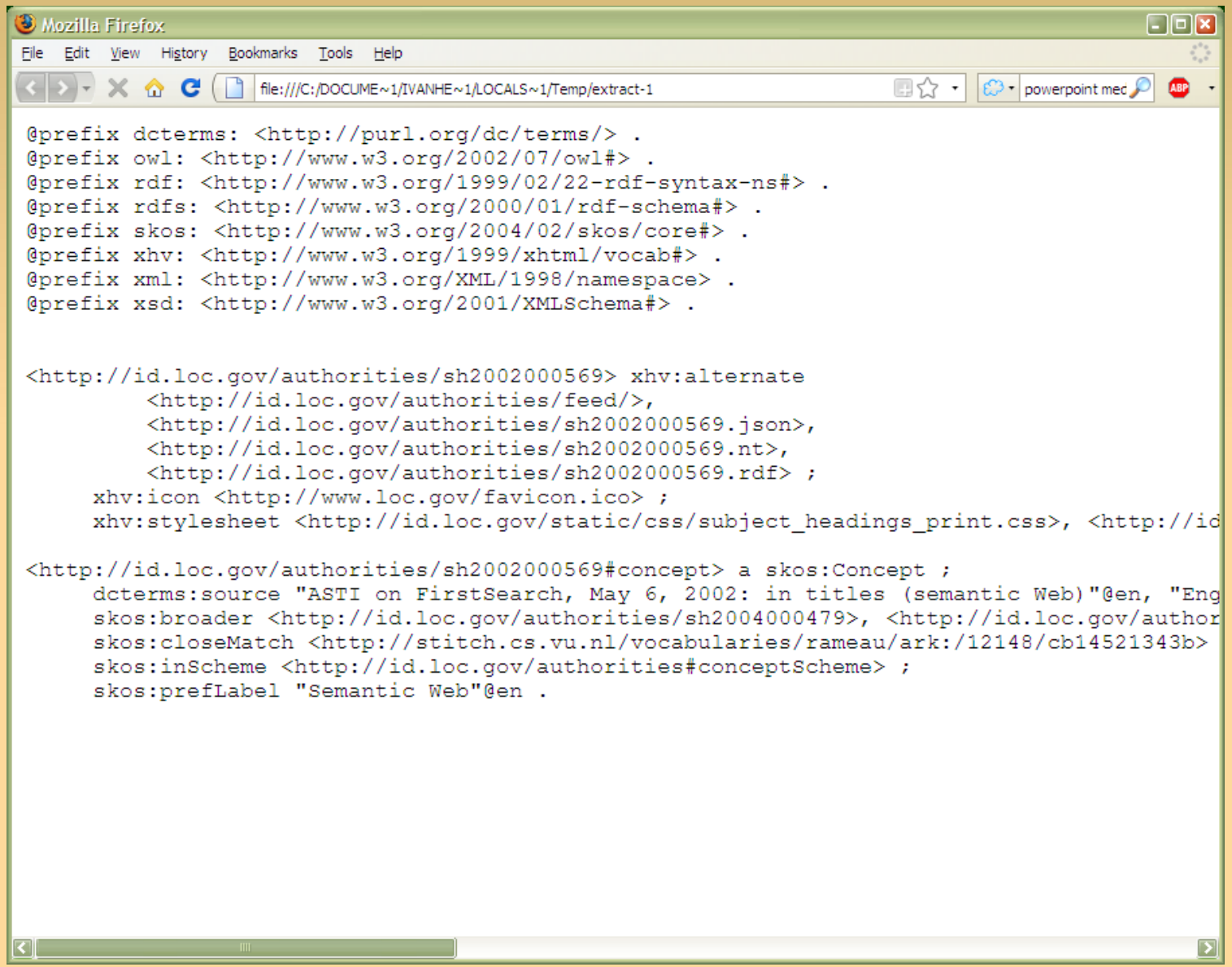
Publication of data (with RDFa): London Gazette



Publication of data (with RDFa & SKOS): Library of Congress Subject Headings



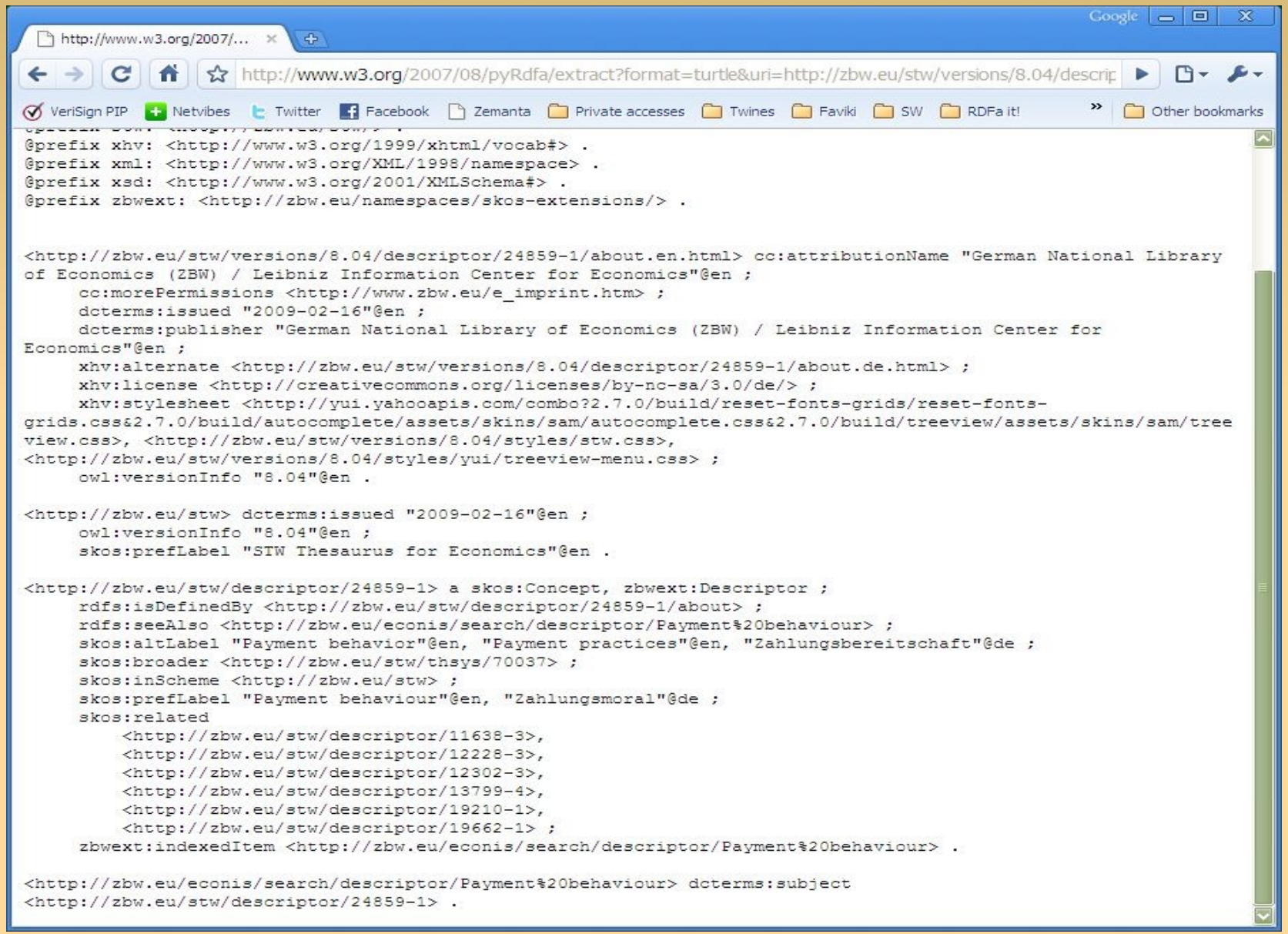
Publication of data (with RDFa & SKOS): Library of Congress Subject Headings



Publication of data (with RDFa & SKOS): Economics Thesaurus

Courtesy of Timo Borst and Joachim Neubert, German Nat. Libr. of Economics, (SWEO Case Study)

Publication of data (with RDFa & SKOS): Economics Thesaurus

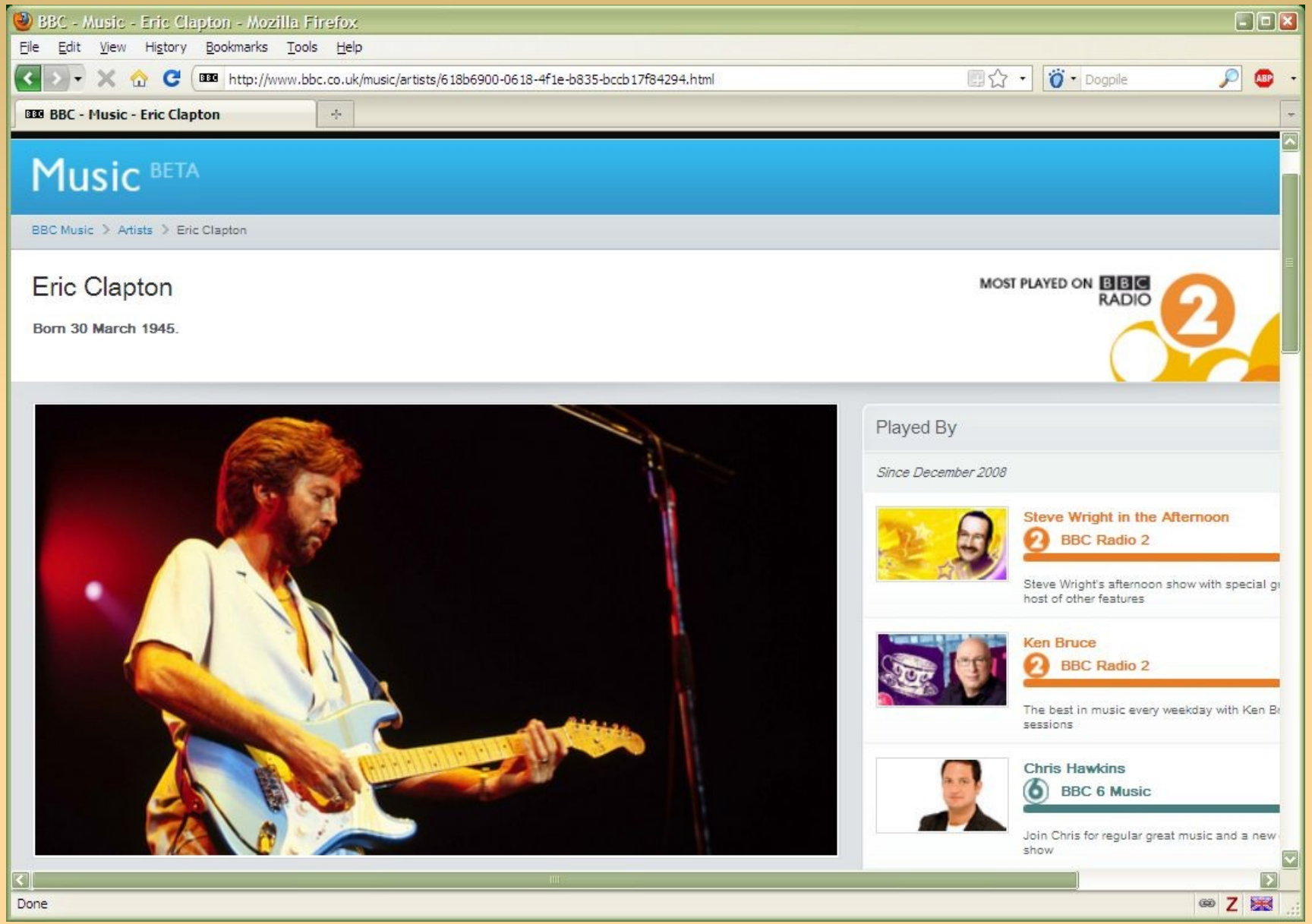



```
@prefix xhv: <http://www.w3.org/1999/xhtml/vocab#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix zbwext: <http://zbw.eu/namespaces/skos-extensions/> .


<http://zbw.eu/stw/versions/8.04/descriptor/24859-1/about.en.html> cc:attributionName "German National Library
of Economics (ZBW) / Leibniz Information Center for Economics"@en ;
     cc:morePermissions <http://www.zbw.eu/e_imprint.htm> ;
     dcterms:issued "2009-02-16"@en ;
     dcterms:publisher "German National Library of Economics (ZBW) / Leibniz Information Center for
Economics"@en ;
     xhv:alternate <http://zbw.eu/stw/versions/8.04/descriptor/24859-1/about.de.html> ;
     xhv:license <http://creativecommons.org/licenses/by-nc-sa/3.0/de/> ;
     xhv:stylesheet <http://yui.yahooapis.com/combo?2.7.0/build/reset-fonts-grids/reset-fonts-
grids.css&2.7.0/build/autocomplete/assets/skins/sam/autocomplete.css&2.7.0/build/treeview/assets/skins/sam/tree
view.css>, <http://zbw.eu/stw/versions/8.04/styles/stw.css>,
<http://zbw.eu/stw/versions/8.04/styles/yui/treeview-menu.css> ;
     owl:versionInfo "8.04"@en .

<http://zbw.eu/stw> dcterms:issued "2009-02-16"@en ;
     owl:versionInfo "8.04"@en ;
     skos:prefLabel "STW Thesaurus for Economics"@en .

<http://zbw.eu/stw/descriptor/24859-1> a skos:Concept, zbwext:Descriptor ;
     rdfs:isDefinedBy <http://zbw.eu/stw/descriptor/24859-1/about> ;
     rdfs:seeAlso <http://zbw.eu/econis/search/descriptor/Payment%20behaviour> ;
     skos:altLabel "Payment behavior"@en, "Payment practices"@en, "Zahlungsbereitschaft"@de ;
     skos:broader <http://zbw.eu/stw/thsys/70037> ;
     skos:inScheme <http://zbw.eu/stw> ;
     skos:prefLabel "Payment behaviour"@en, "Zahlungsmoral"@de ;
     skos:related
         <http://zbw.eu/stw/descriptor/11638-3>,
         <http://zbw.eu/stw/descriptor/12228-3>,
         <http://zbw.eu/stw/descriptor/12302-3>,
         <http://zbw.eu/stw/descriptor/13799-4>,
         <http://zbw.eu/stw/descriptor/19210-1>,
         <http://zbw.eu/stw/descriptor/19662-1> ;
     zbwext:indexedItem <http://zbw.eu/econis/search/descriptor/Payment%20behaviour> .

<http://zbw.eu/econis/search/descriptor/Payment%20behaviour> dcterms:subject
<http://zbw.eu/stw/descriptor/24859-1> .
```

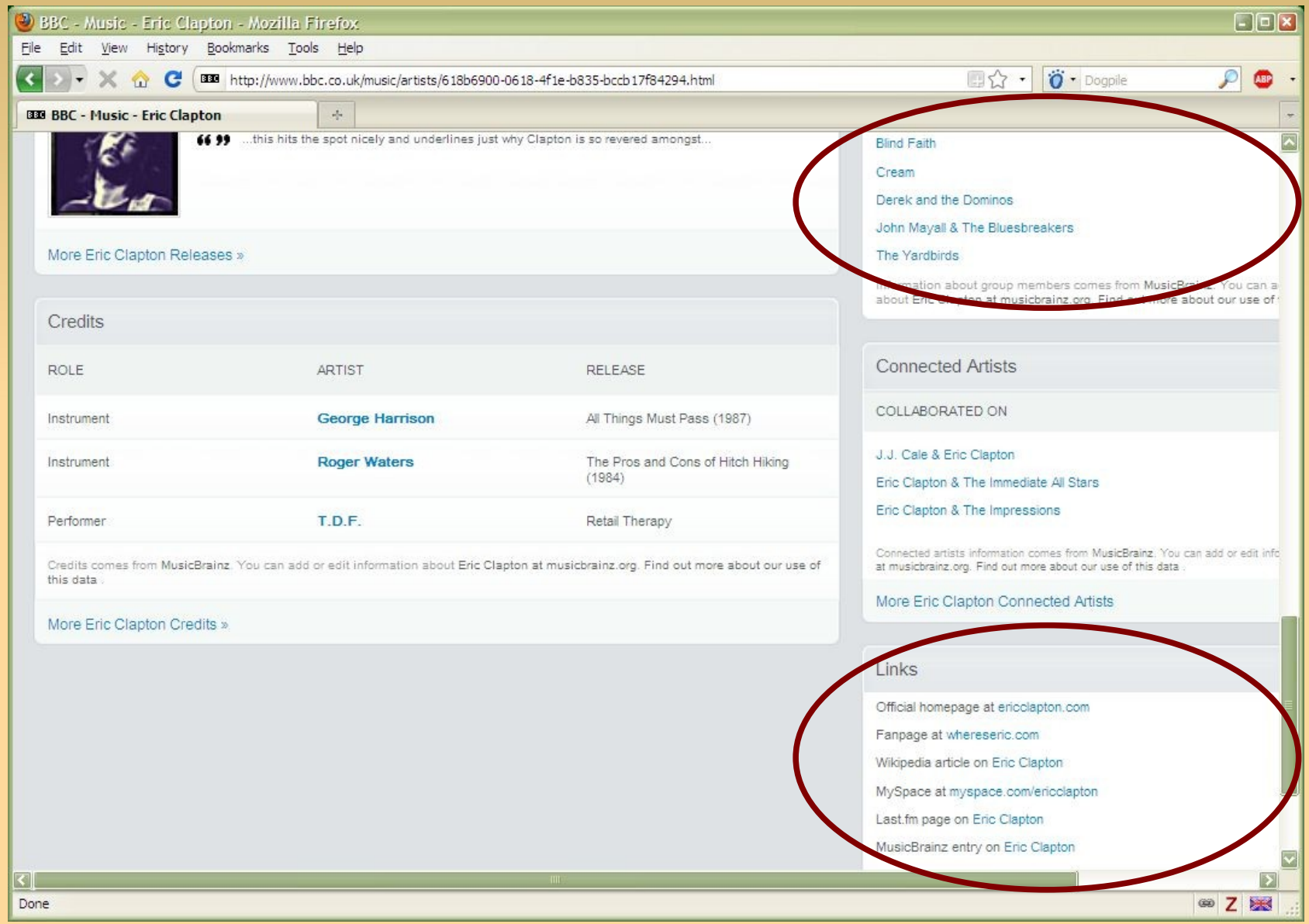
Courtesy of Timo Borst and Joachim Neubert, German Nat. Libr. of Economics, (SWEO Case Study)

Applications using this data come to the fore...

Using the LOD to build Web site: BBC

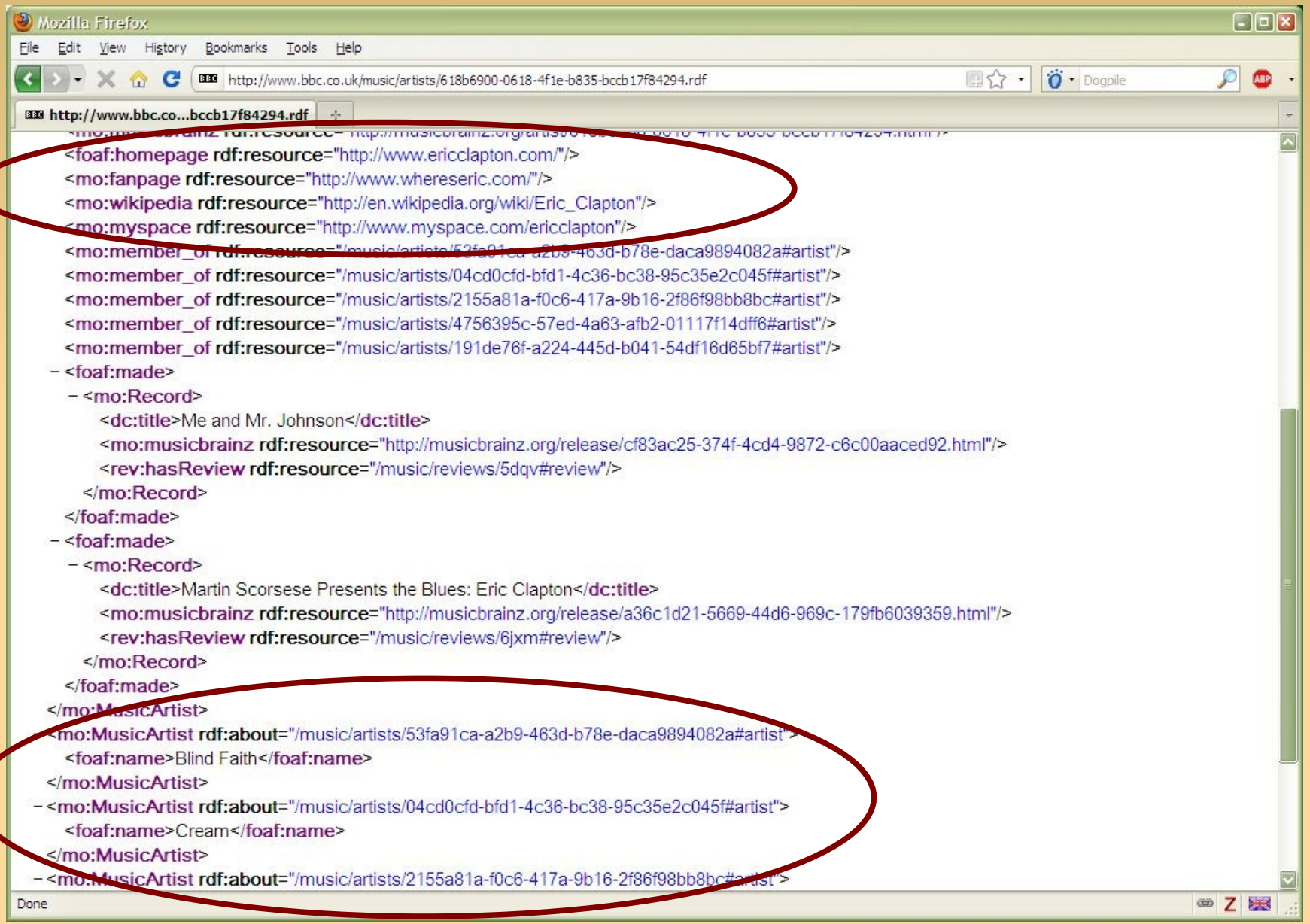


Using the LOD to build Web site: BBC



W3C Semantic Web

Using the LOD to build Web site: BBC



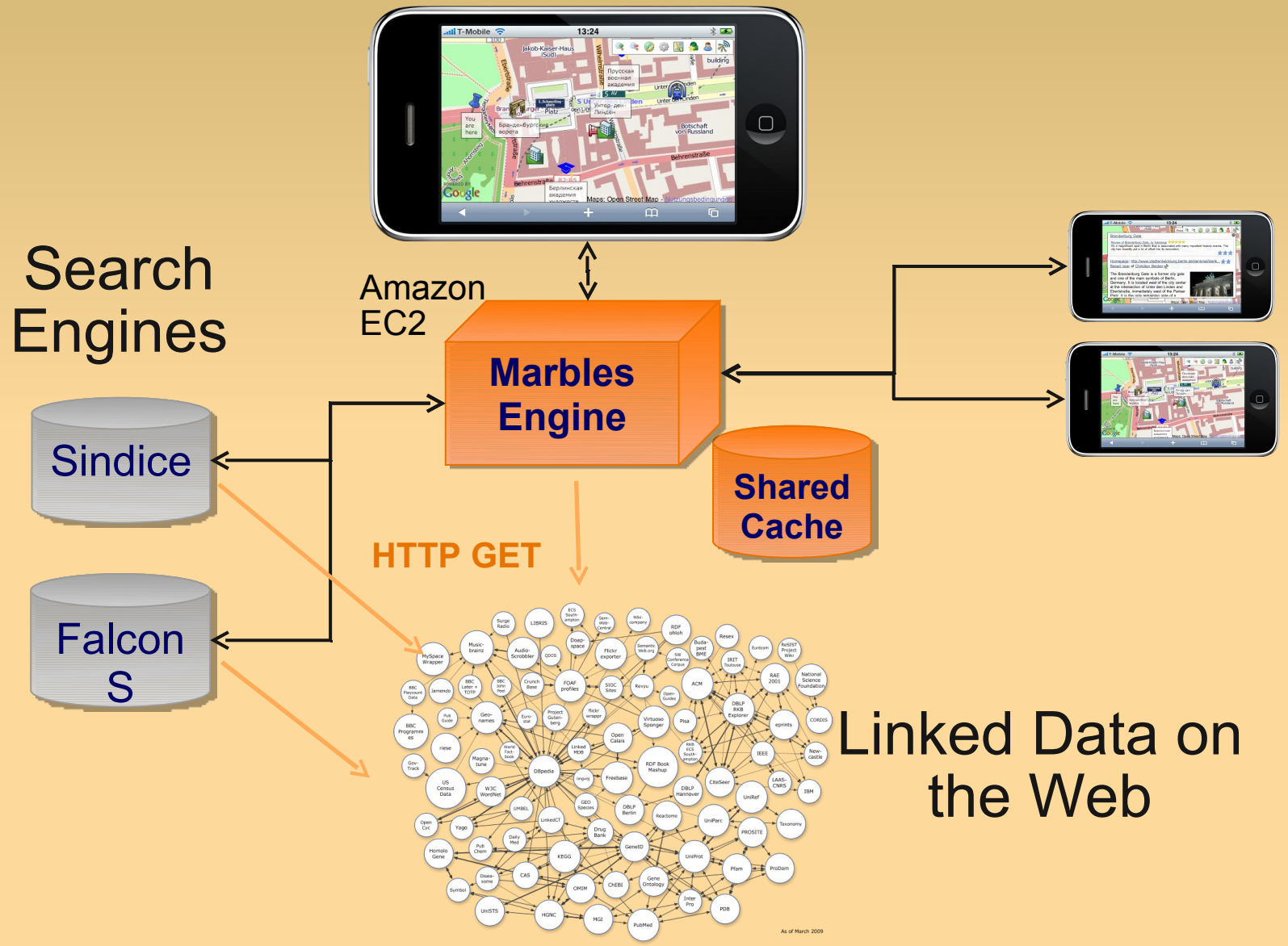
Using the LOD cloud on an iPhone

Courtesy of Chris Bizer and Christian Becker, Freie Universität, Berlin

Using the LOD cloud on an iPhone

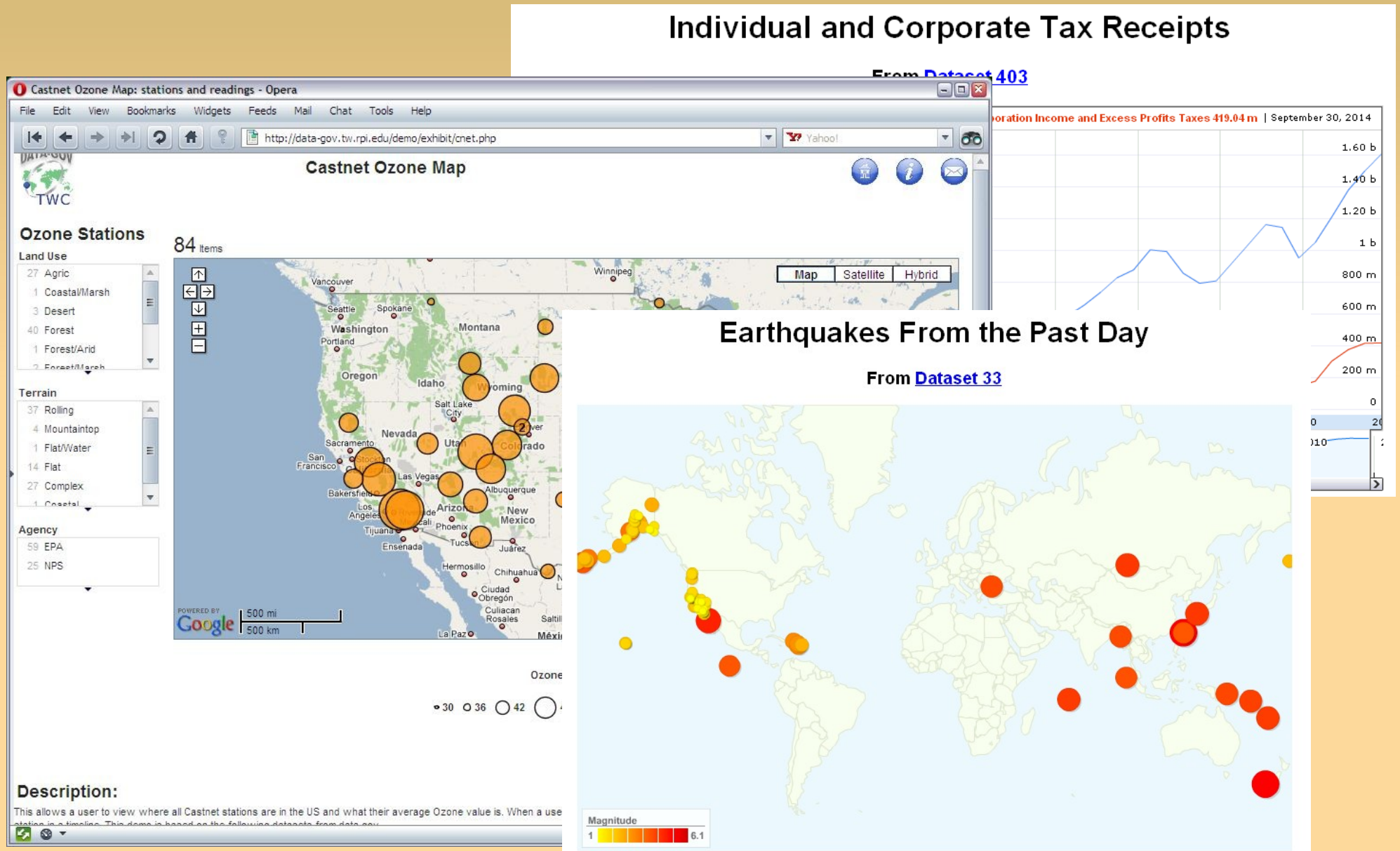
Courtesy of Chris Bizer and Christian Becker, Freie Universität, Berlin

Using the LOD cloud on an iPhone



Courtesy of Chris Bizer and Christian Becker, Freie Universität, Berlin

You publish the raw data, we use it…

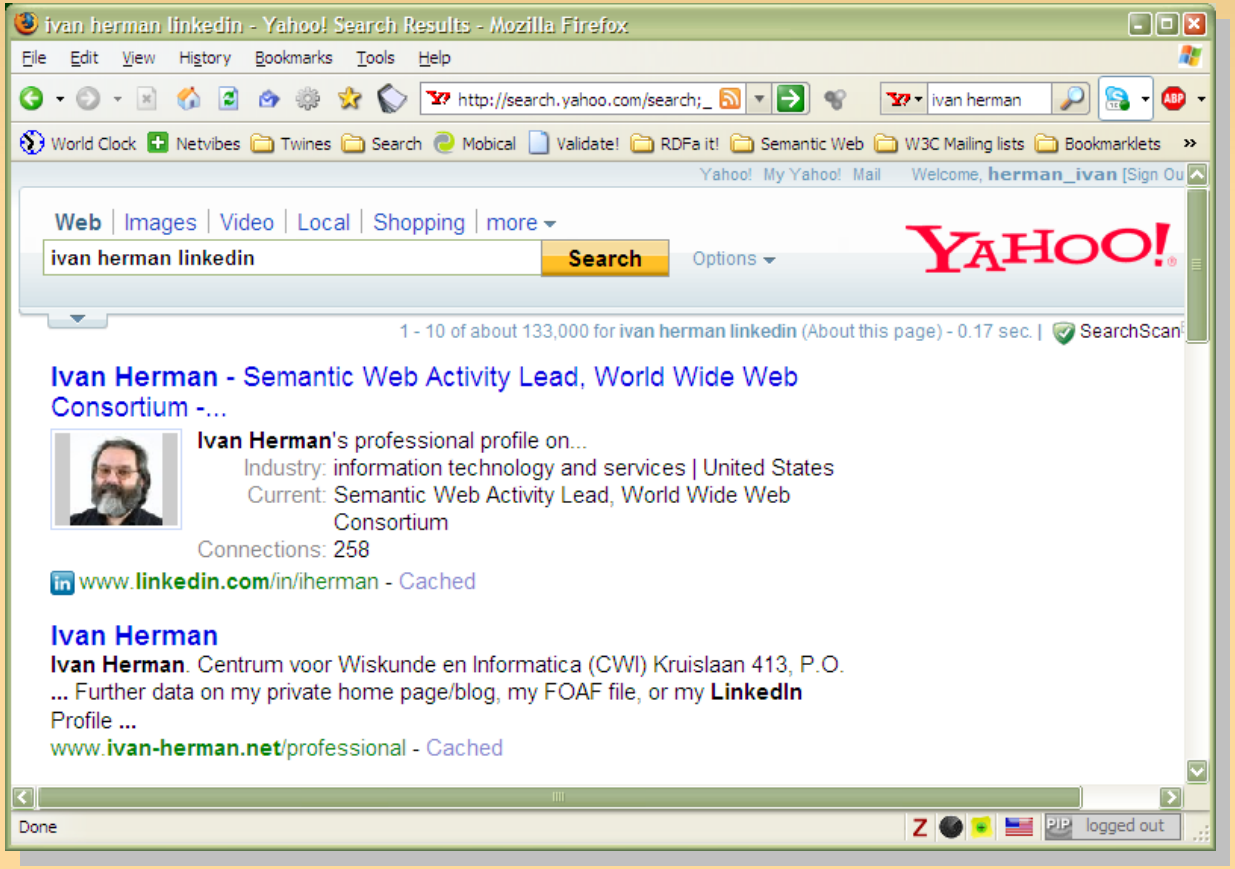


Examples from RPI's Data-gov Wiki, Jim Hendler & al.

W3C Semantic Web

Yahoo's SearchMonkey

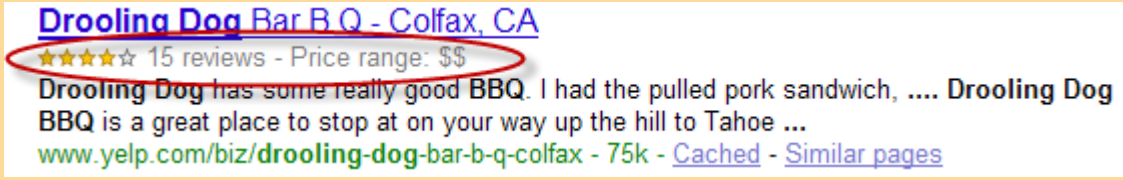
- Search based results may be customized via small applications
- Metadata embedded in pages (in RDFa, eRDF, etc) are reused
- Publishers can export extra (RDF) data via other formats
  - Yahoo Boss also indexes these



Courtesy of Peter Mika, Yahoo! Research, (SWEO Case Study)

Google's rich sniplet

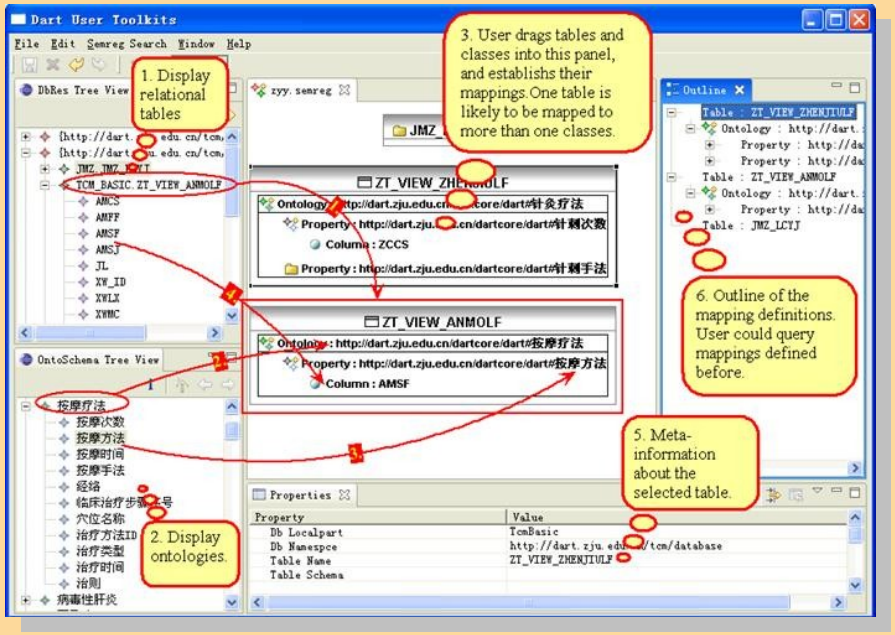
- Embedded metadata (in microformat or RDFa) is used to improve search result page
  - at the moment only a few vocabularies are recognized, but that will evolve over the years

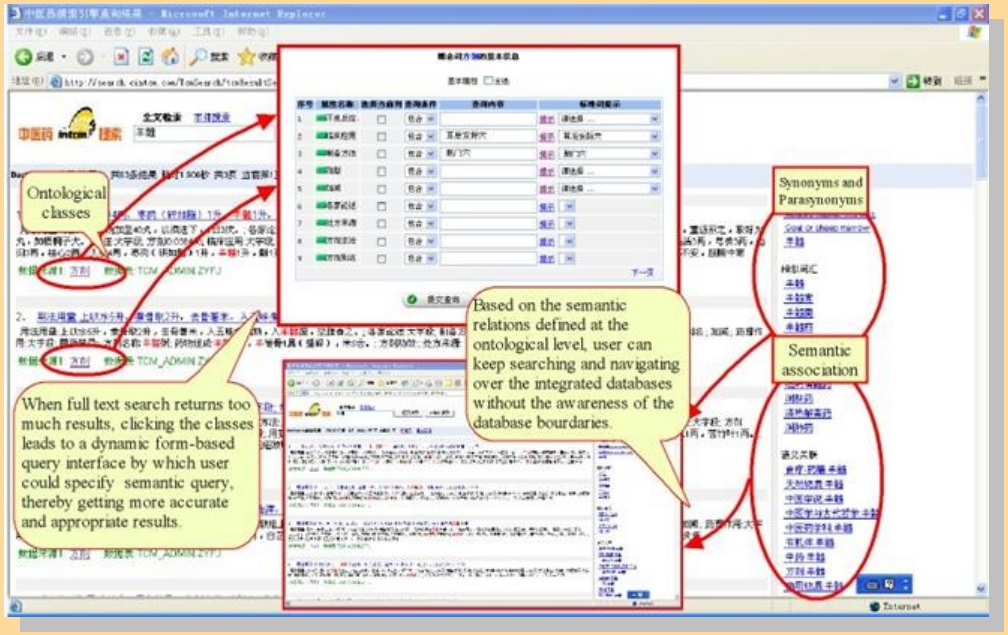


Other applications examples…

Integrate knowledge for Chinese Medicine

- Integration of a large number of TCM databases
  - around 80 databases, around 200,000 records each
- A visual tools for the end users
  - mapping, query building

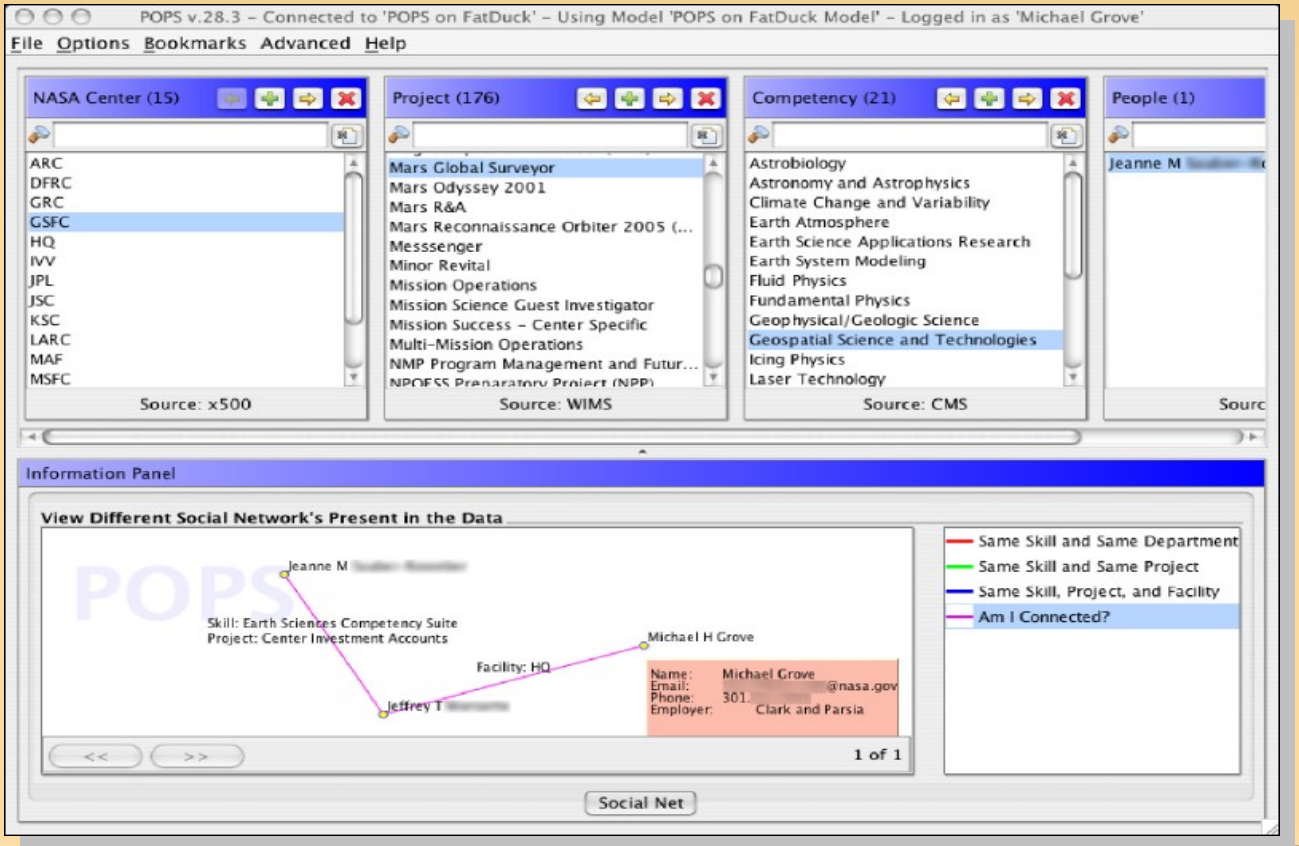




Courtesy of Huajun Chen, Zhejiang University, (SWEO Case Study)

Find the right experts at NASA

- Expertise locater for nearly 70,000 NASA civil servants
  - over 6 or 7 geographically distributed databases, data sources, and web services…

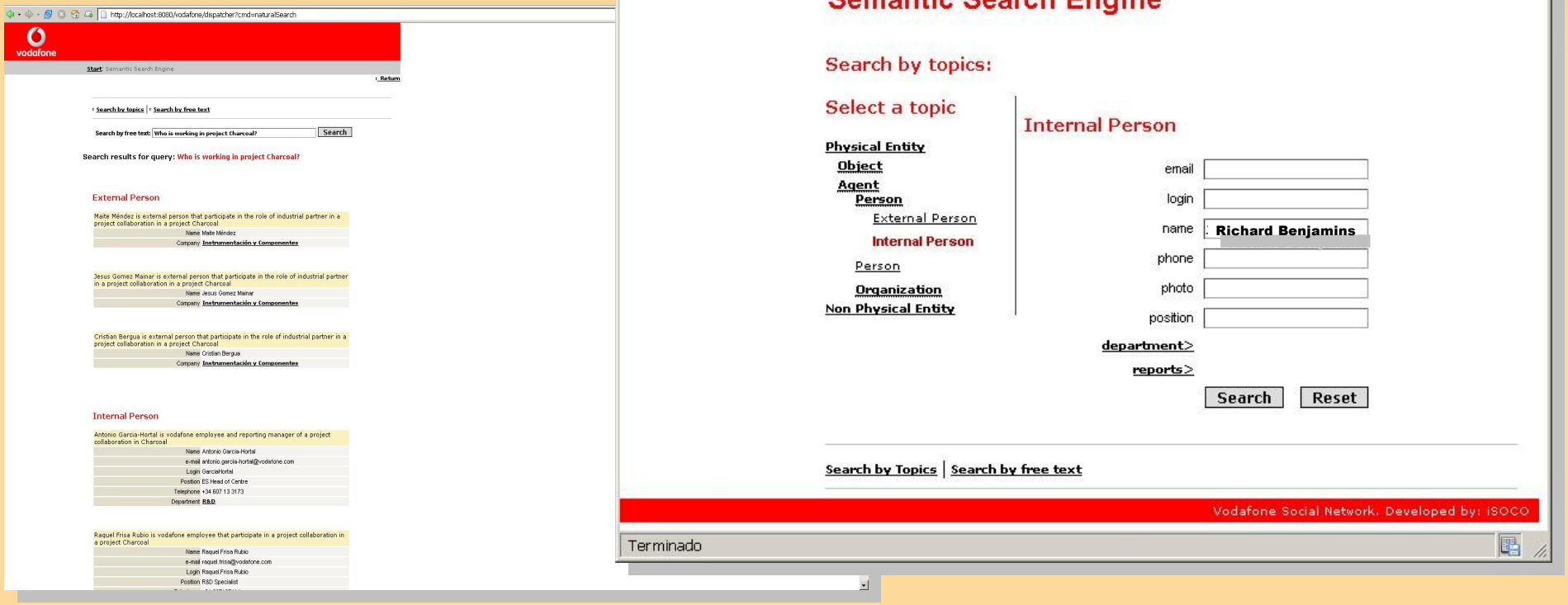


Michael Grove, Clark & Parsia, LLC, and Andrew Schain, NASA, (SWEO Case Study)

W3C Semantic Web

Find the right experts at Vodafone

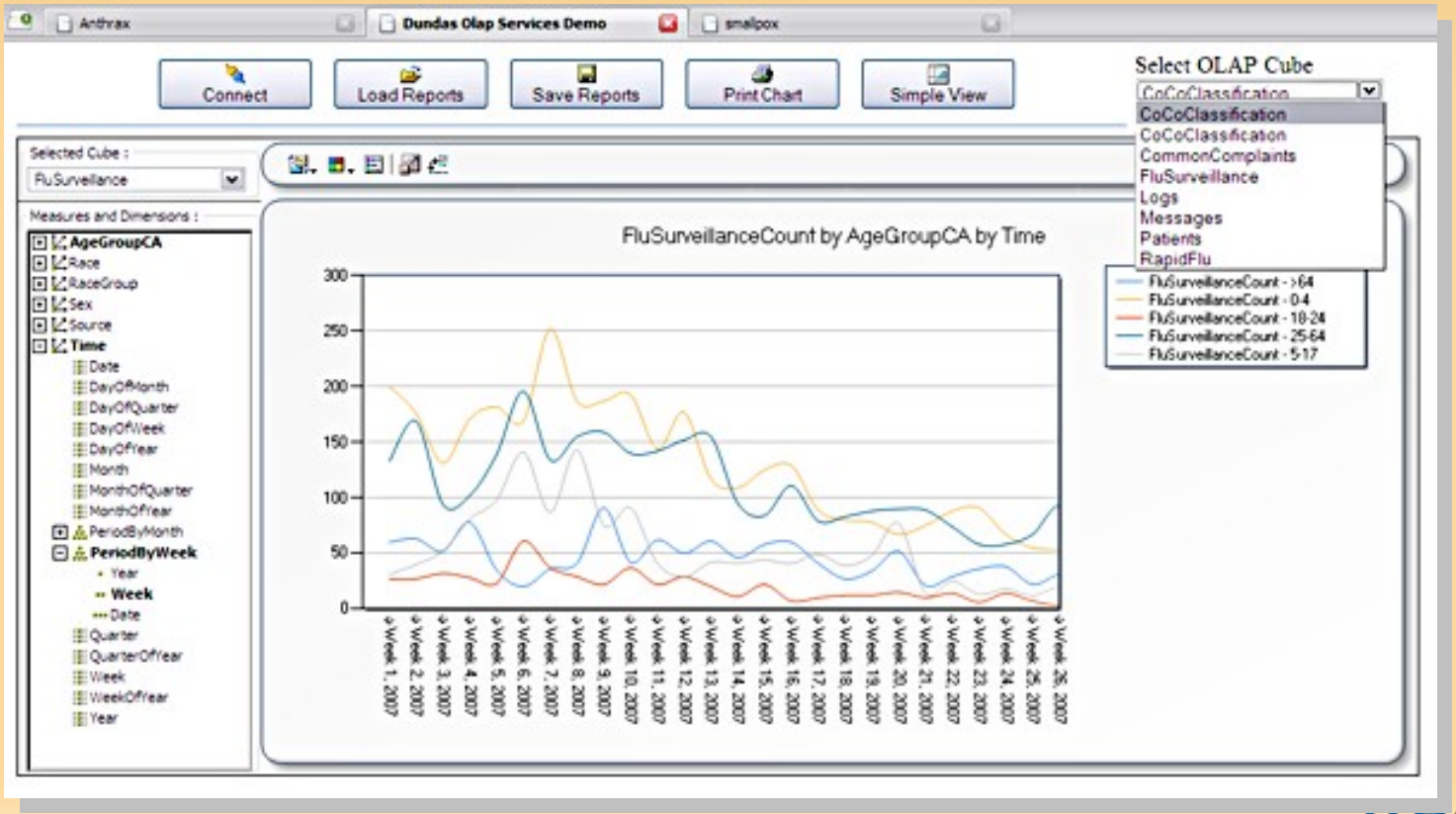
- Very similar to the NASA application, though with different technologies…



Courtesy of Juan José Fúster, Vodafone, and Richard Benjamins, iSOCO, (SWEO Use Case)

Public health surveillance (Sapphire)

- Integrated biosurveillance system (biohazards, bioterrorism, disease control, etc)
- Integrates multiple data sources
  - new data can be added easily



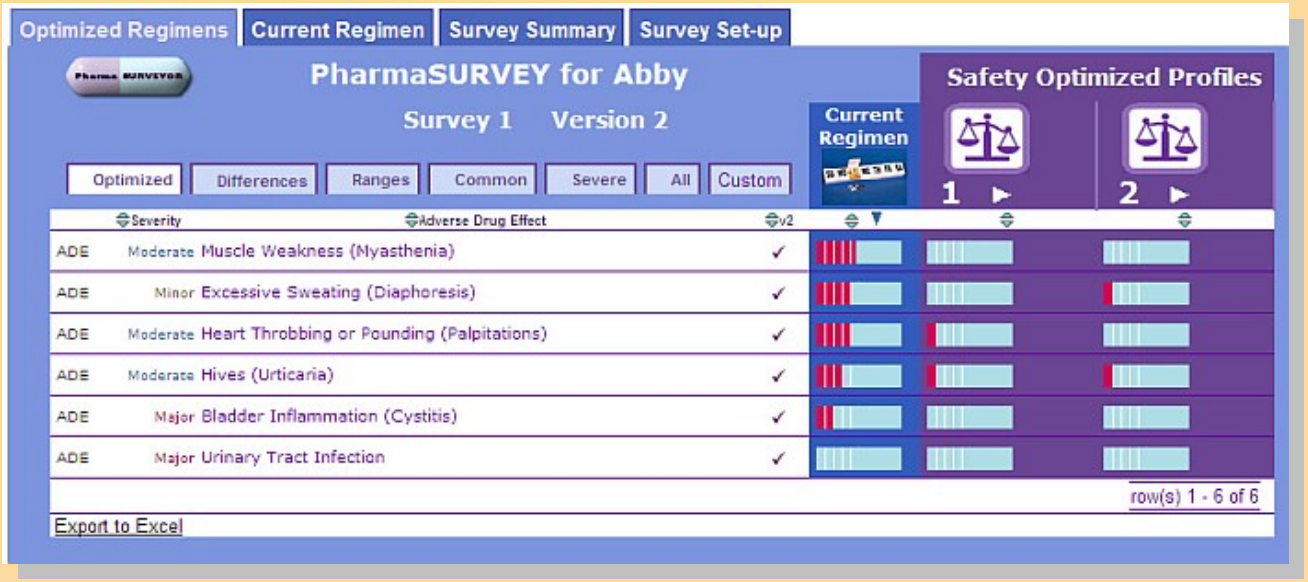
Courtesy of Parsa Mirhaji, School of Health Information Sciences, Un. of Texas (SWEO Case Study)

A frequent paradigm: intelligent portals

- “Portals” collecting data and presenting them to users
- They can be public or behind corporate firewalls Portal’s internal organization makes use of semantic data, ontologies
  - integration with external and internal data
  - better queries, often based on controlled vocabularies or ontologies…

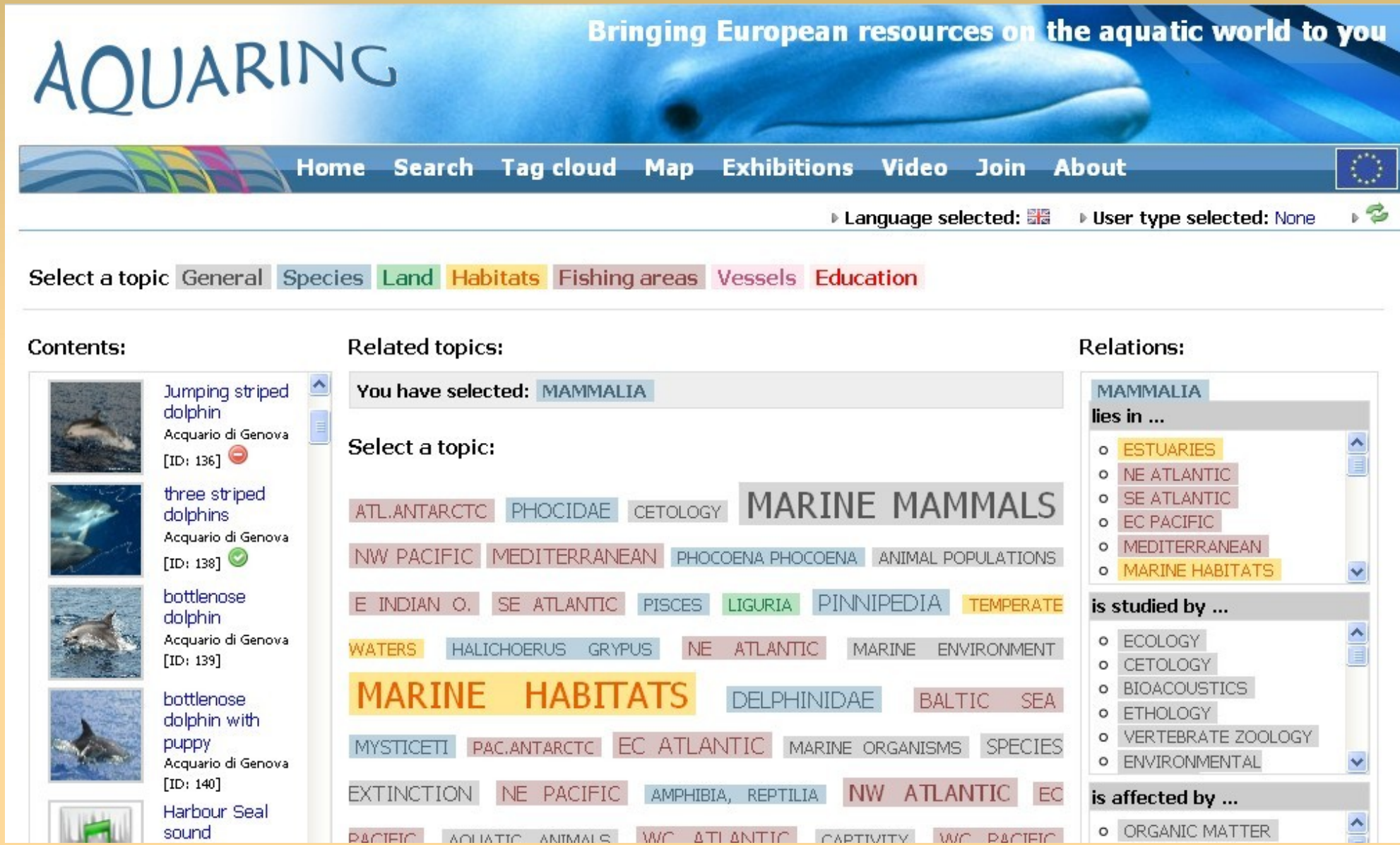
Help in choosing the right drug regimen

- Help in finding the best drug regimen for a specific case, per patient
- Integrate data from various sources (patients, physicians, Pharma, researchers, ontologies, etc)
- Data (eg, regulation, drugs) change often, but the tool is much more resistant against change



Courtesy of Erick Von Schweber, PharmaSURVEYOR Inc., (SWEO Use Case)

Portal to aquatic resources

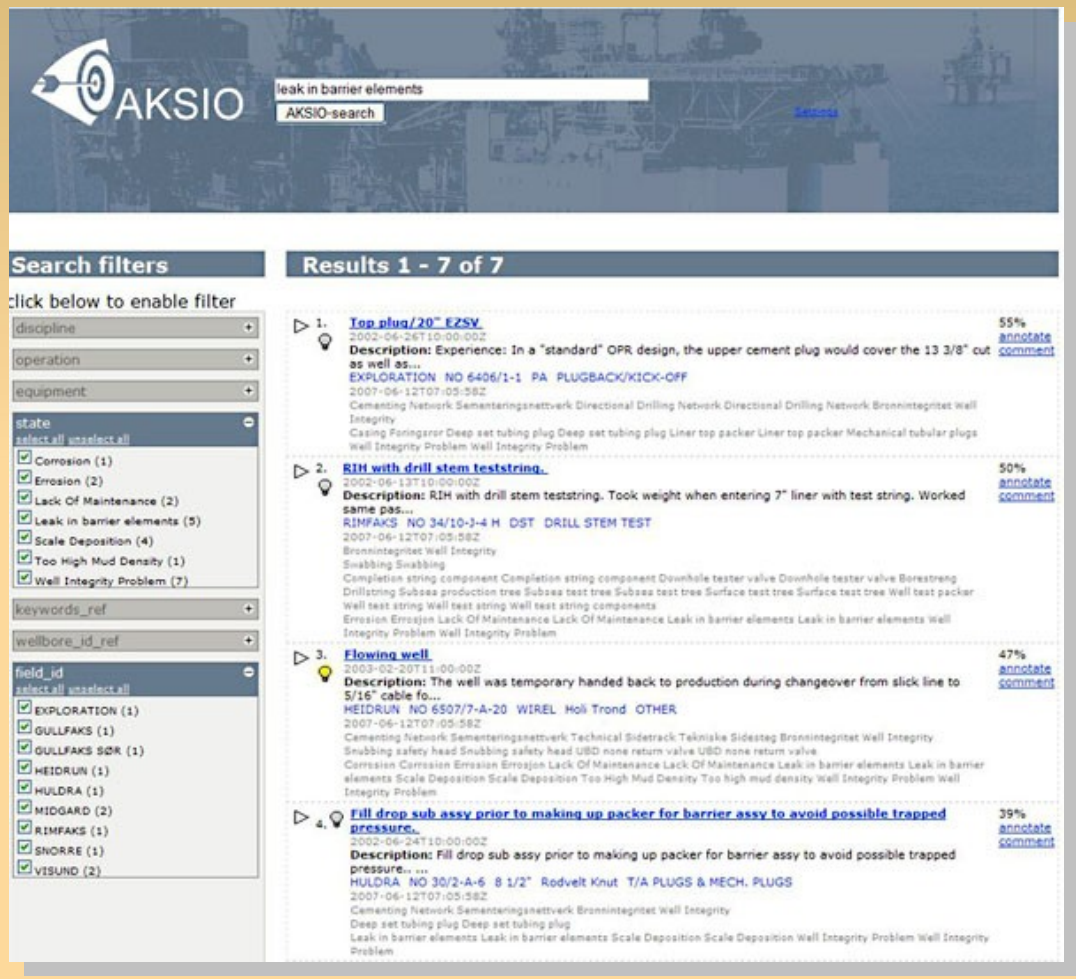


Courtesy of Marta González Rodríguez, Tecnalia, (SWEO Case Study)

W3C Semantic Web

Help for deep sea drilling operations

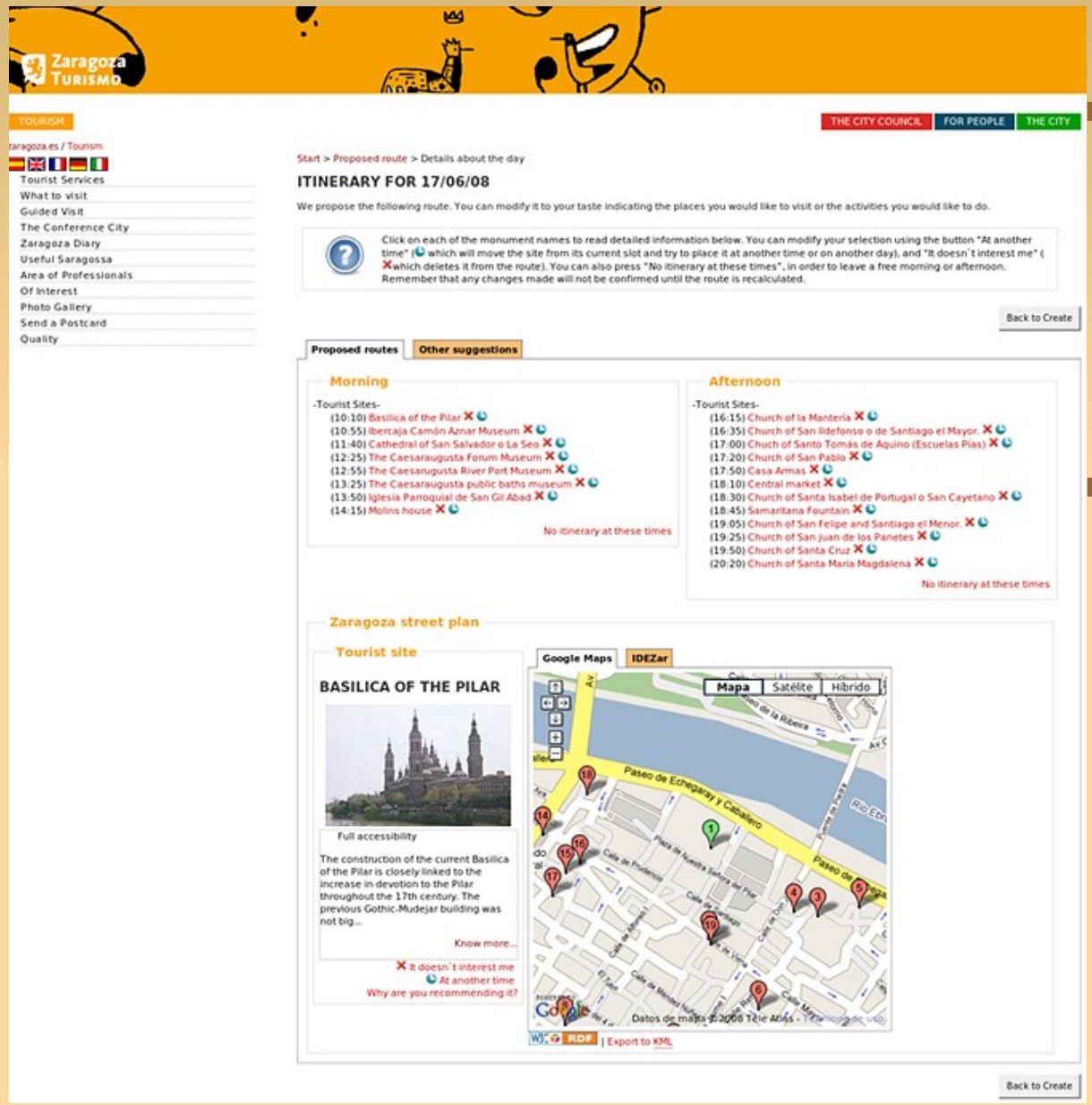
- Integration of experience and data in the planning and operation of deep sea drilling processes
- Discover relevant experiences that could affect current or planned drilling operations
  - uses an ontology backed search engine



Courtesy of David Norheim and Roar Fjellheim, Computas AS (SWEO Use Case)

W3C Semantic Web

eTourism: provide personalized itinerary

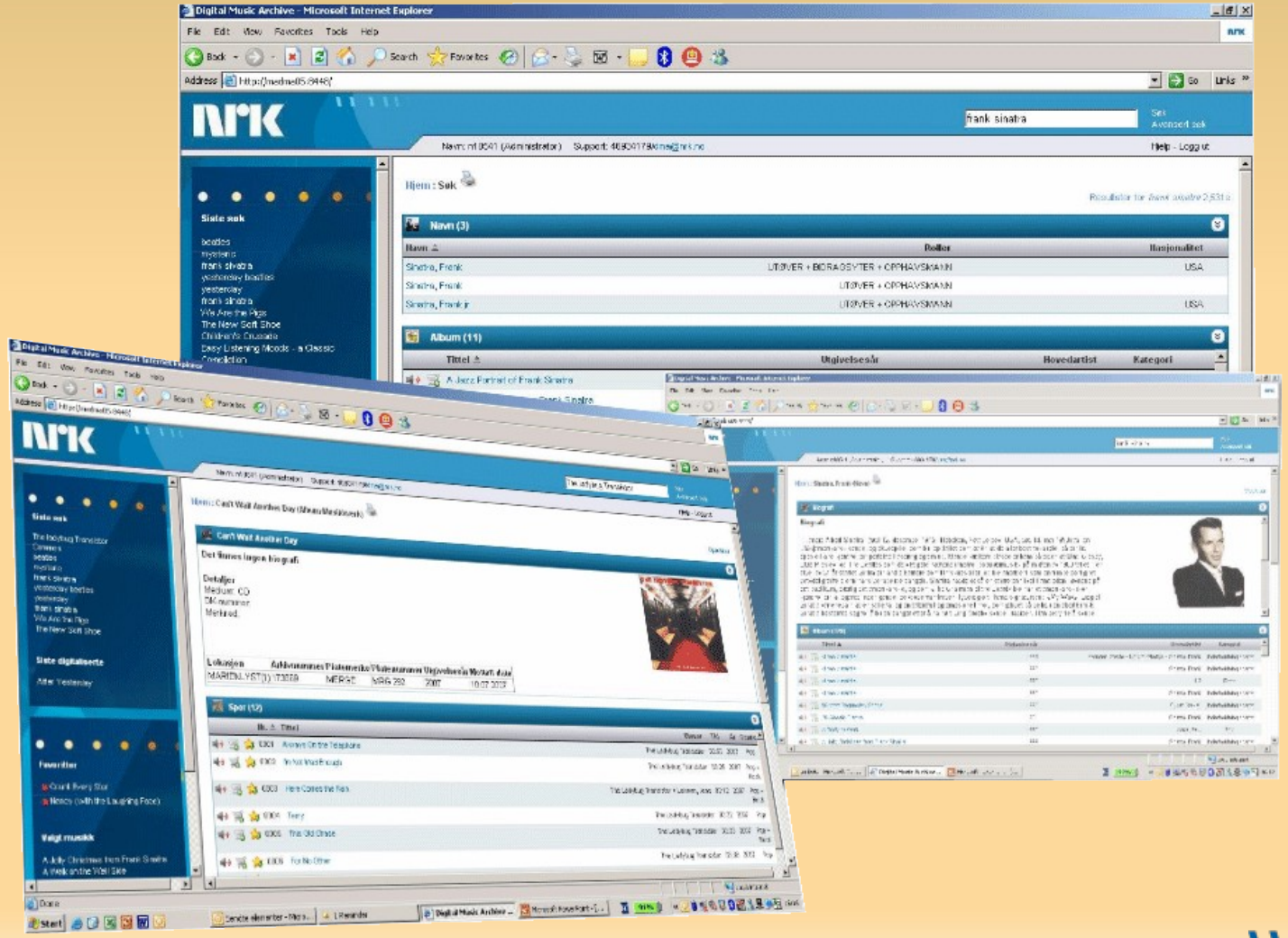


- Integration of relevant data in Zaragoza (using RDF and ontologies)
- Use rules on the RDF data to provide a proper itinerary

Courtesy of Jesús Fernández, Mun. of Zaragoza, and Antonio Campos, CTIC (SWEO Use Case)

Digital music asset portal at NRK

- Used by program production to find the right music in the archive for a specific show



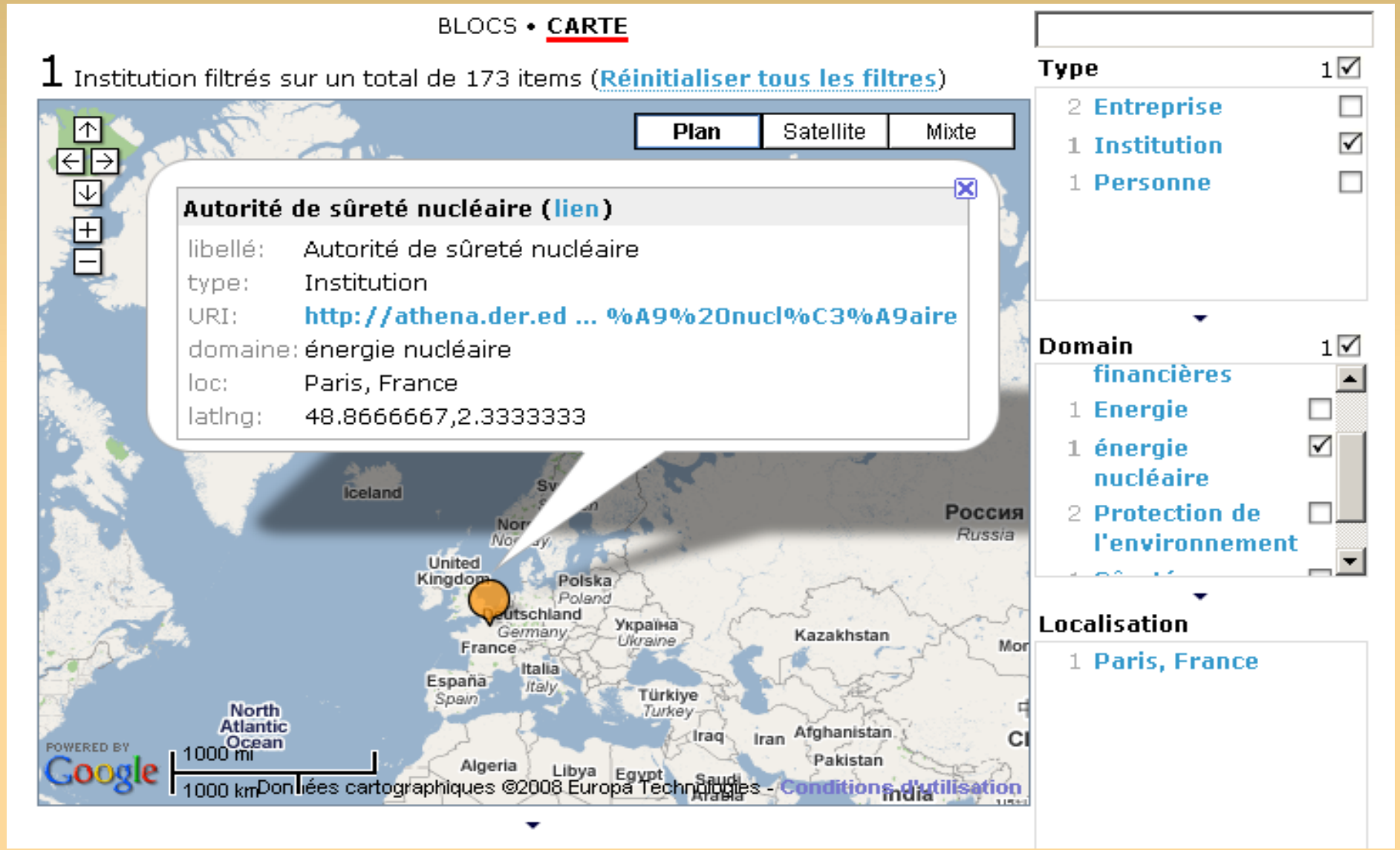
Courtesy of Robert Engels, ESIS, and Jon Roar Tønnesen, NRK (SWEO Case Study)

Integration of "social" software data

- Internal usage of wikis, blogs, RSS, etc, at EDF
  - goal is to manage the flow of information better
- Items are integrated via
  - RDF as a unifying format
  - simple vocabularies like SIOC, FOAF, MOAT (all public)
  - internal data is *combined* with linked open data like Geonames
  - SPARQL is used for internal queries
- Details are hidden from end users (via plugins, extra layers, etc)

Courtesy of A. Passant, EDF R&D and LaLIC, Université Paris-Sorbonne, (SWEO Case Study)

Integration of “social” software data

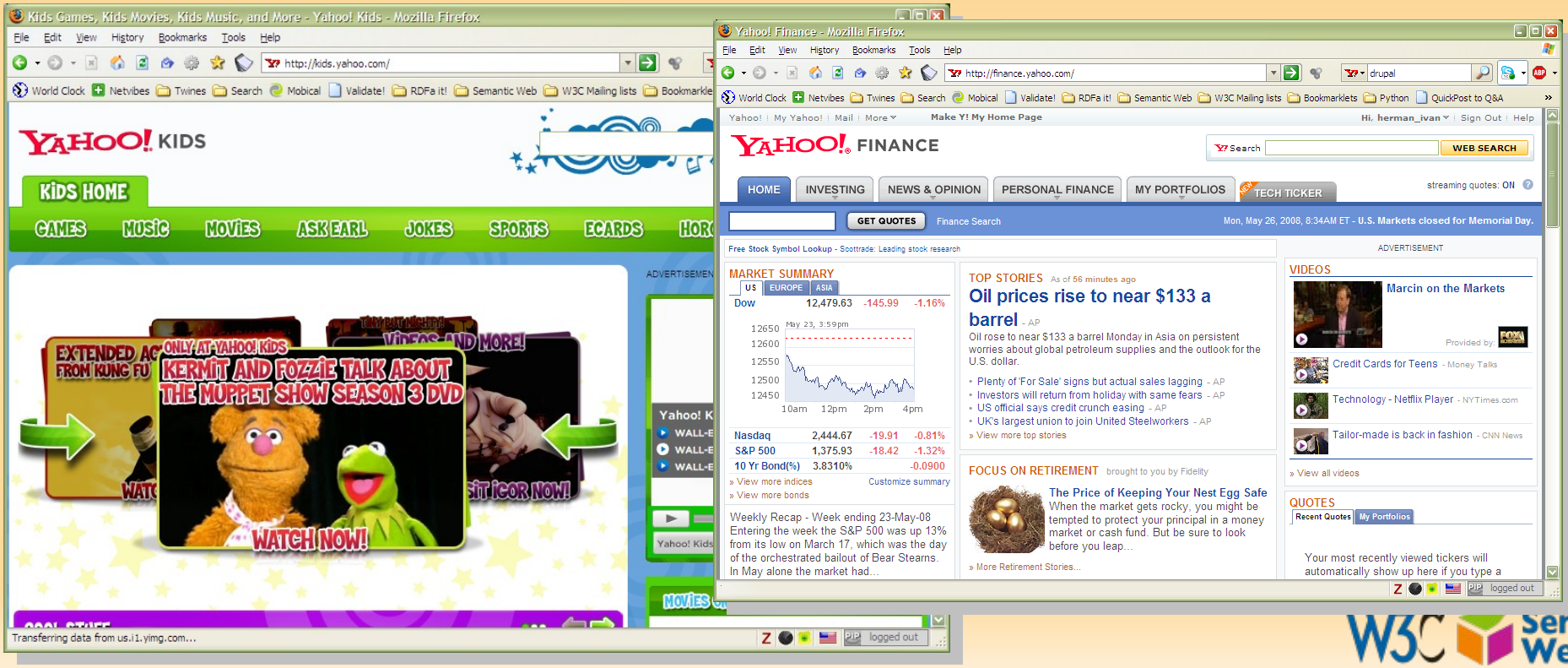


Courtesy of A. Passant, EDF R&D and LaLIC, Université Paris-Sorbonne, (SWEO Case Study)

W3C Semantic Web

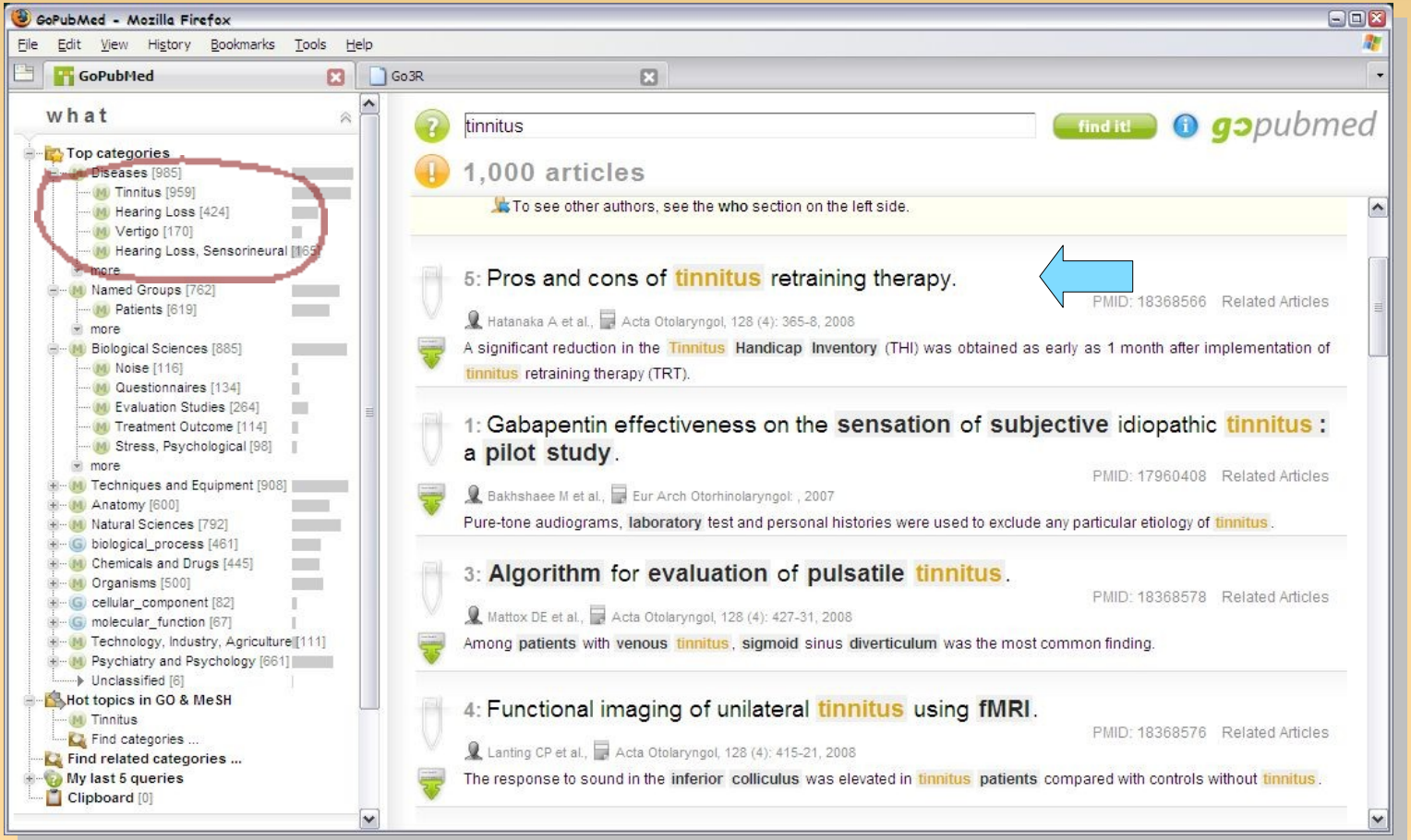
Yahoo! portals

- "Back-end" is built using SW tools
  - common (RDF) data model for data, metadata, relationships,…
  - constraints expressed in OWL, Rules
  - uses public (DC, PRISM) and private vocabularies



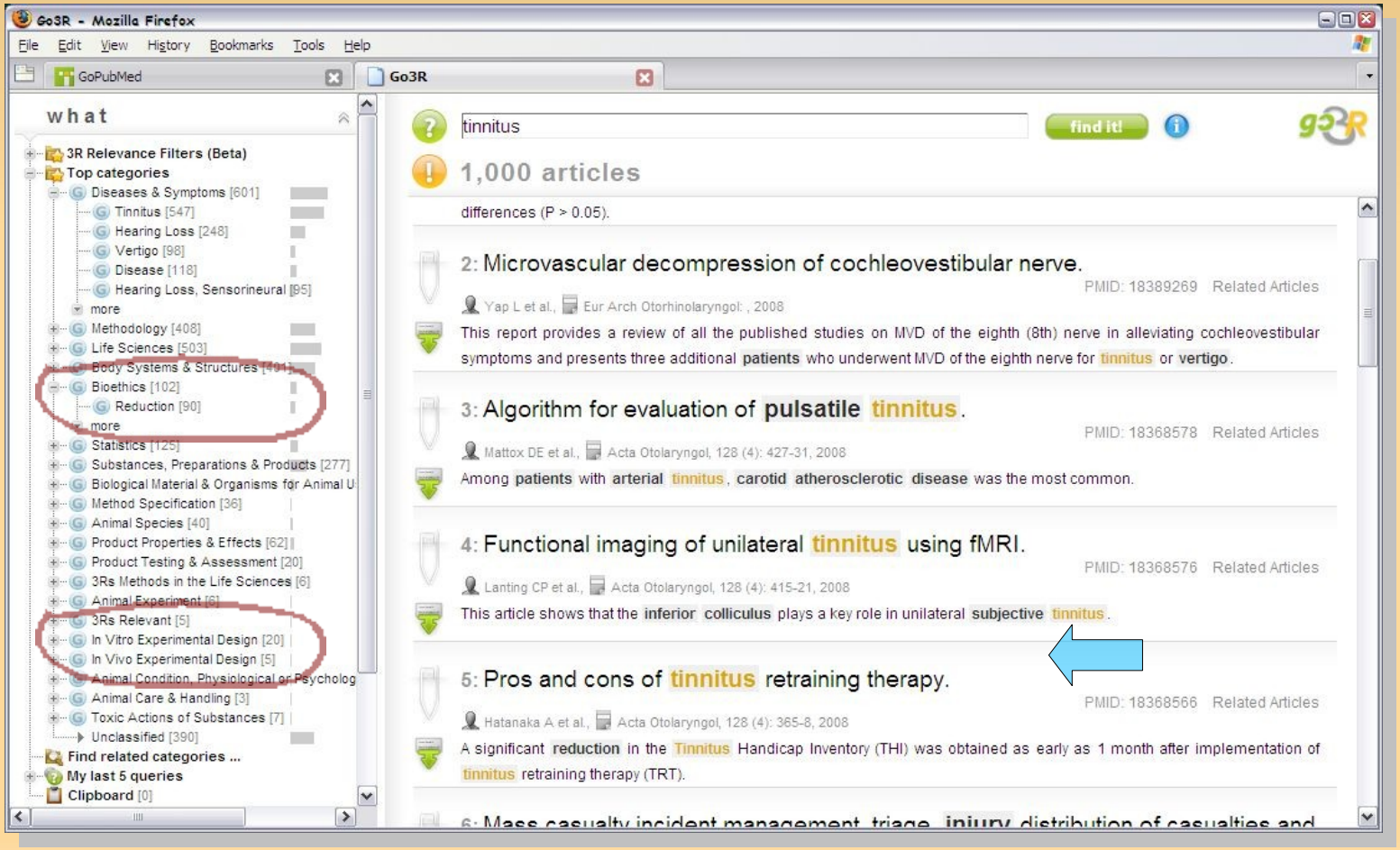
Improved Search via Ontology (GoPubMed)

- Search results are re-ranked using ontologies
- Related terms are highlighted, usable for further search

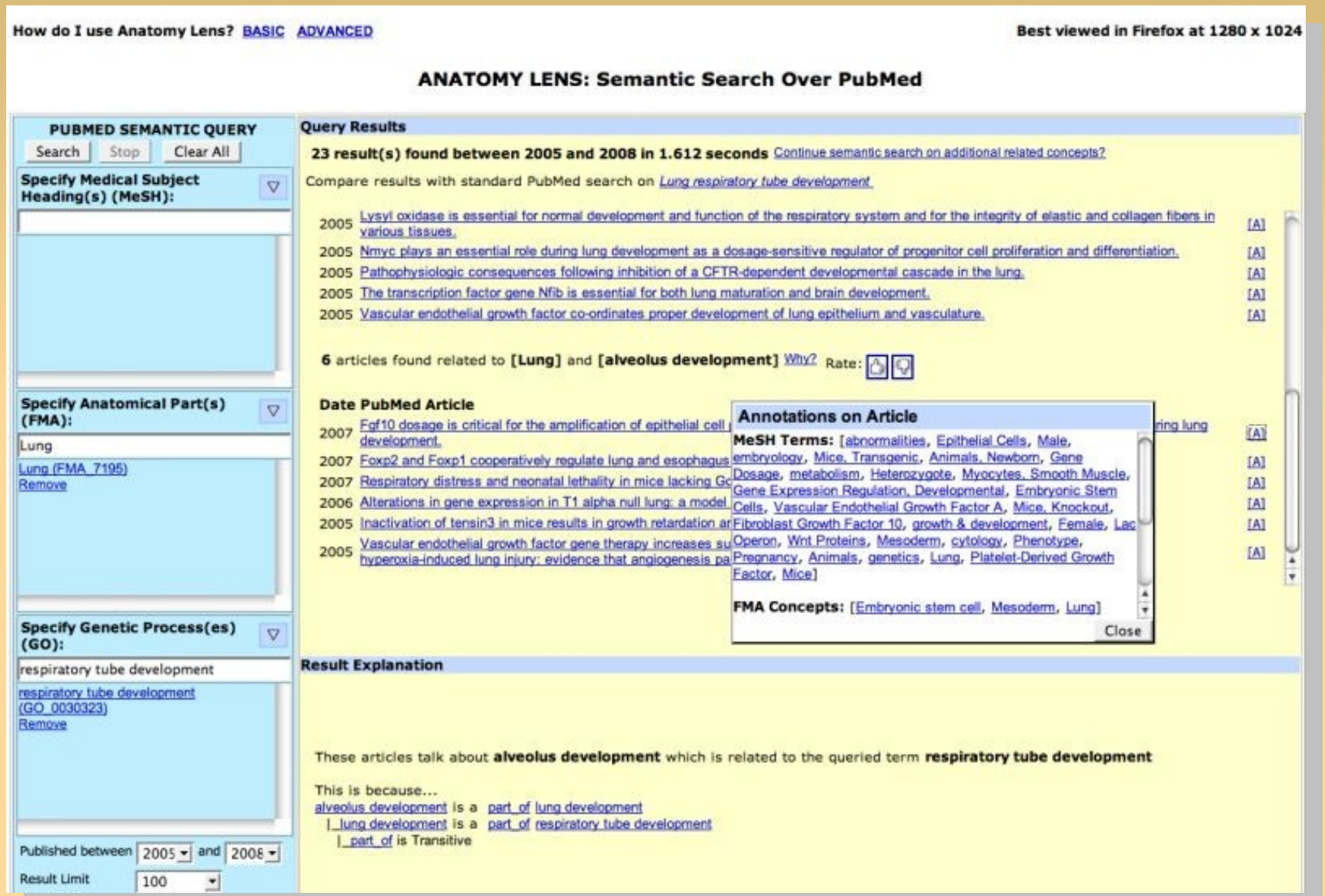


Improved Search via Ontology (Go3R)

- Same dataset, different ontology
  - (ontology is on non-animal experimentation)



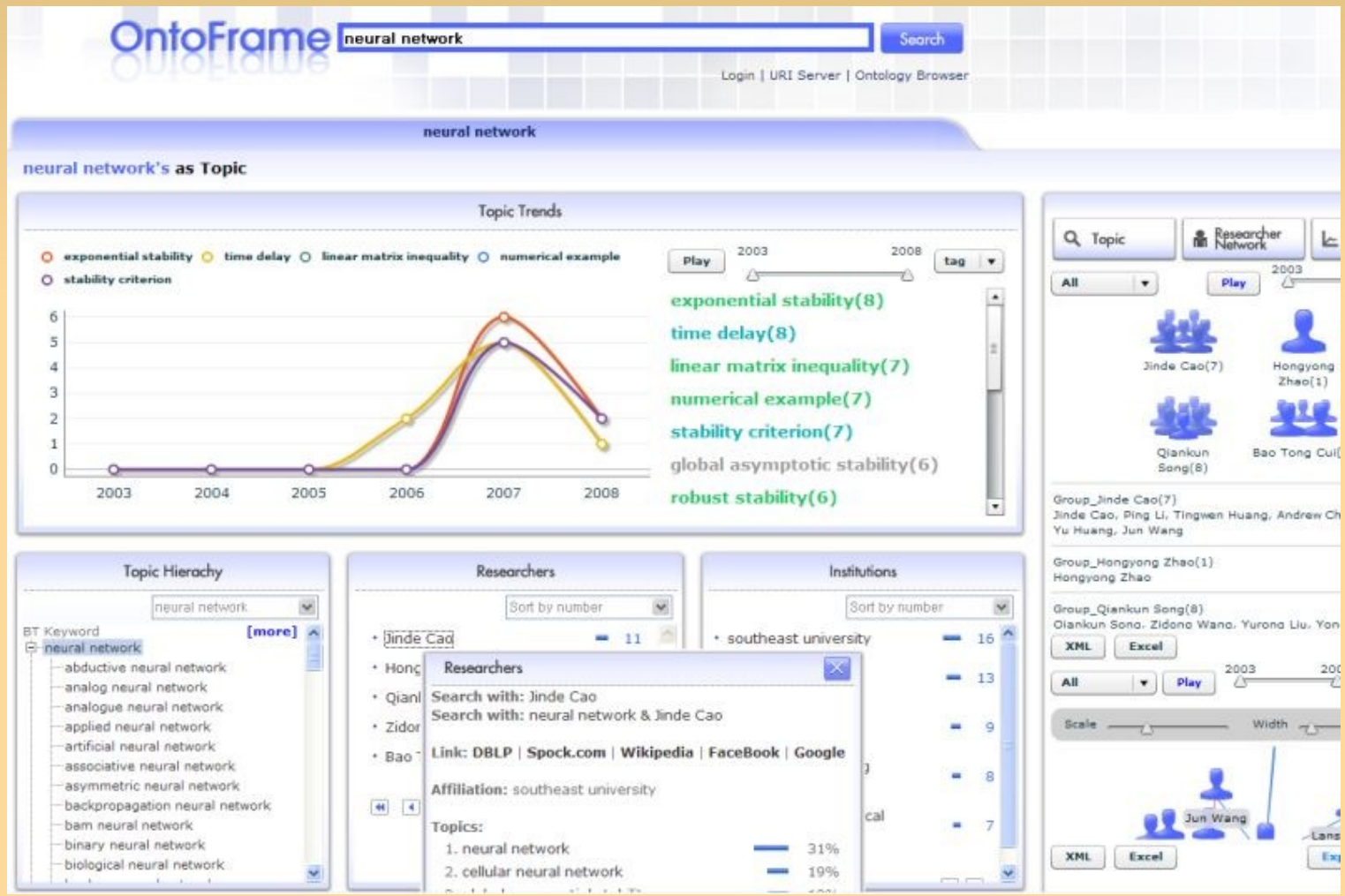
Same problem, different solution…



Courtesy of Kavitha Srinivas, IBM J Watson Research Center

Portal for researchers

- "Ontology-based" search is combined with keyword search to find information on scientific publications

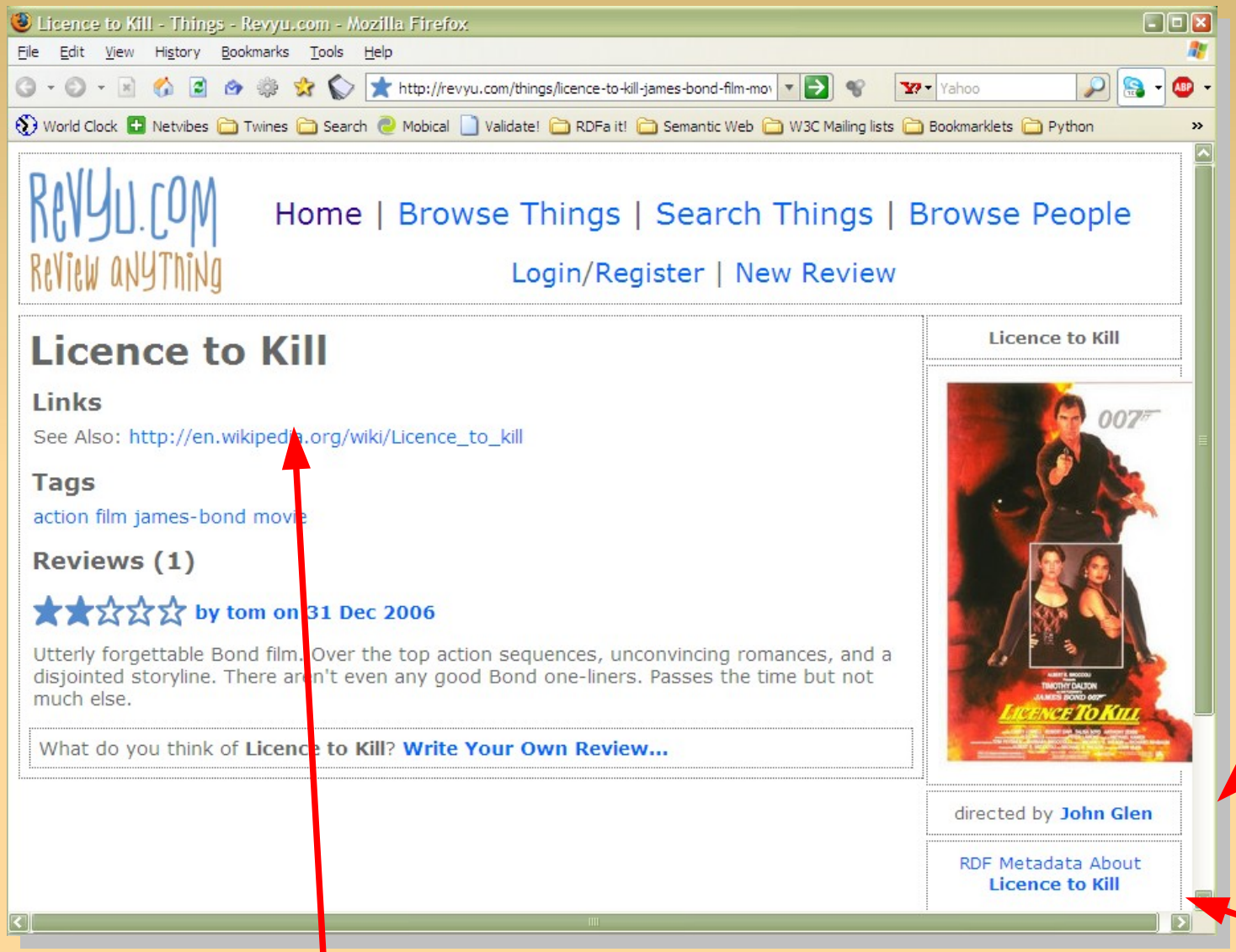


Courtesy of Hanming Young et al, KISTI (SWEO Use Case)

New type of Web 2.0 applications

- New Web 2.0 applications come every day
- Some begin to look at Semantic Web as possible technology to improve their operation
  - more structured tagging, making use of external services
  - providing extra information to users
  - etc.
- Some examples: Twine, Revyu, Faviki, …

“Review Anything”



enhance output with linked data

data in RDF

links to, eg, (DB/Wiki)Pedia

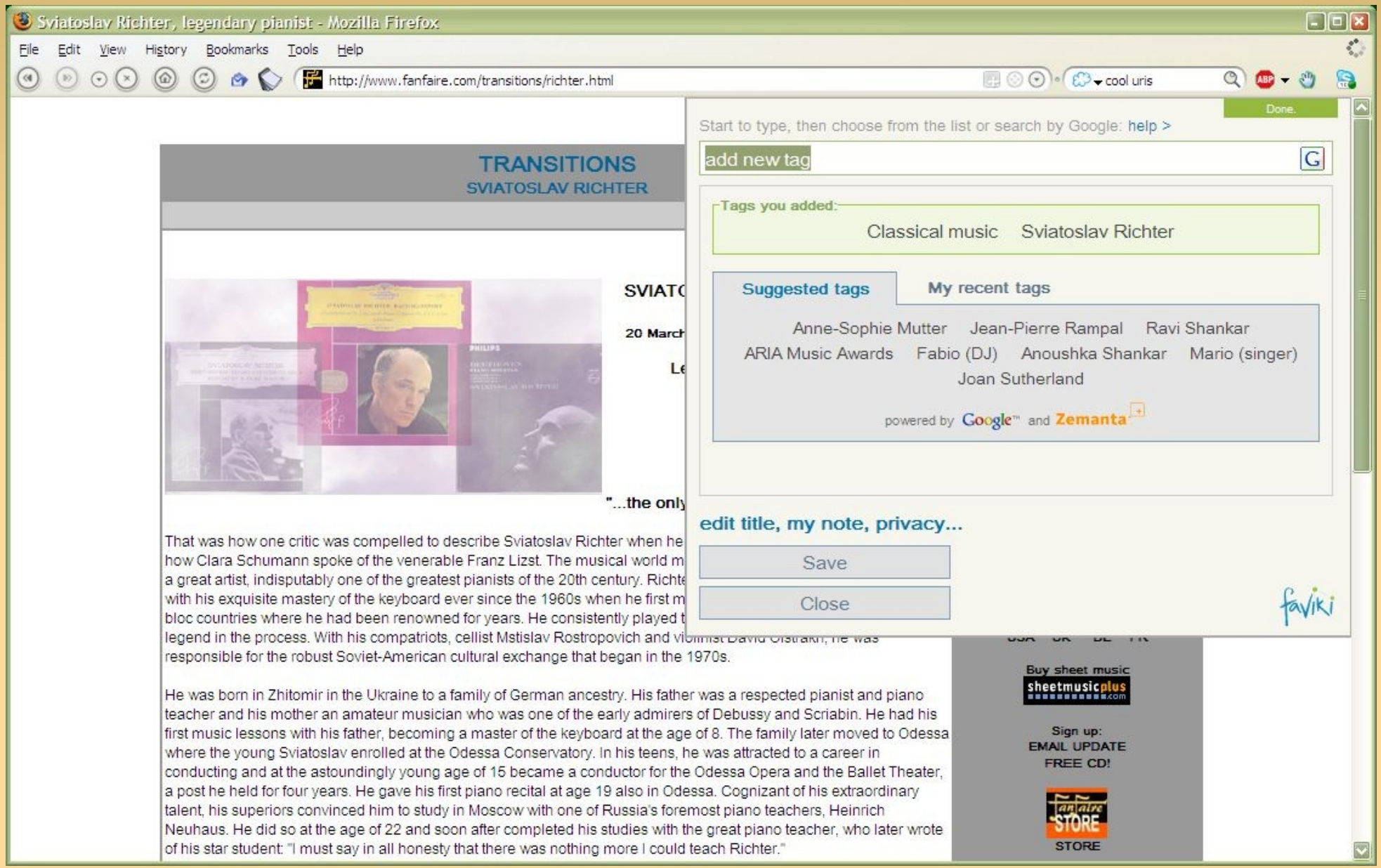
W3C Semantic Web

Faviki: social bookmarking, semantic tagging

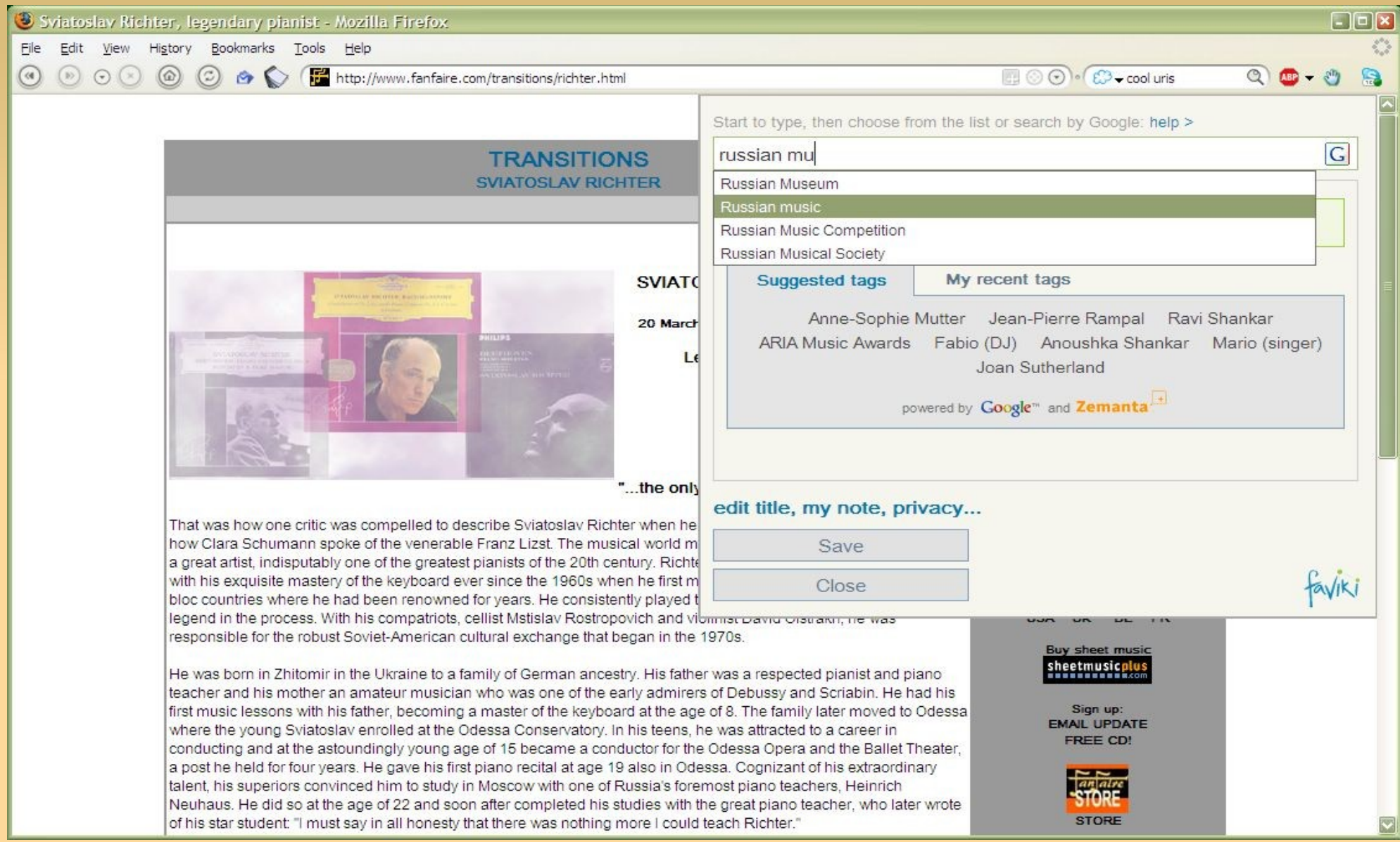
- Social bookmarking system (a bit like del.icio.us) but with a controlled set of tags
  - tags are terms extracted from Wikipedia/DBpedia
  - tags are categorized using the relationships stored in DBpedia
  - tags can be multilingual, DBpedia providing the linguistic bridge
- The tagging process itself is done via a user interface hiding the complexities

Courtesy of Vuk Milicic, Faviki, (SWEO Case Study)

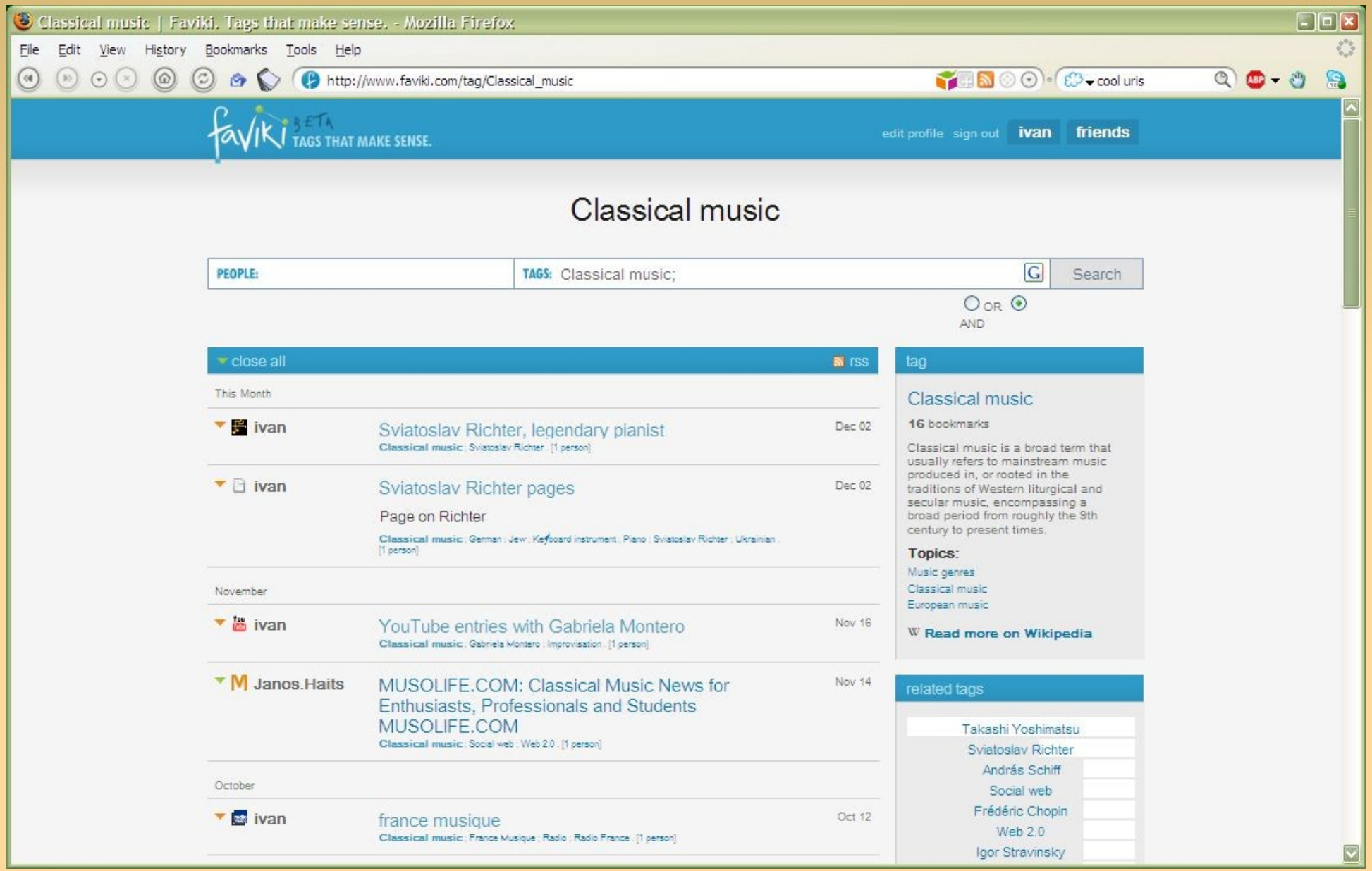
Faviki Example



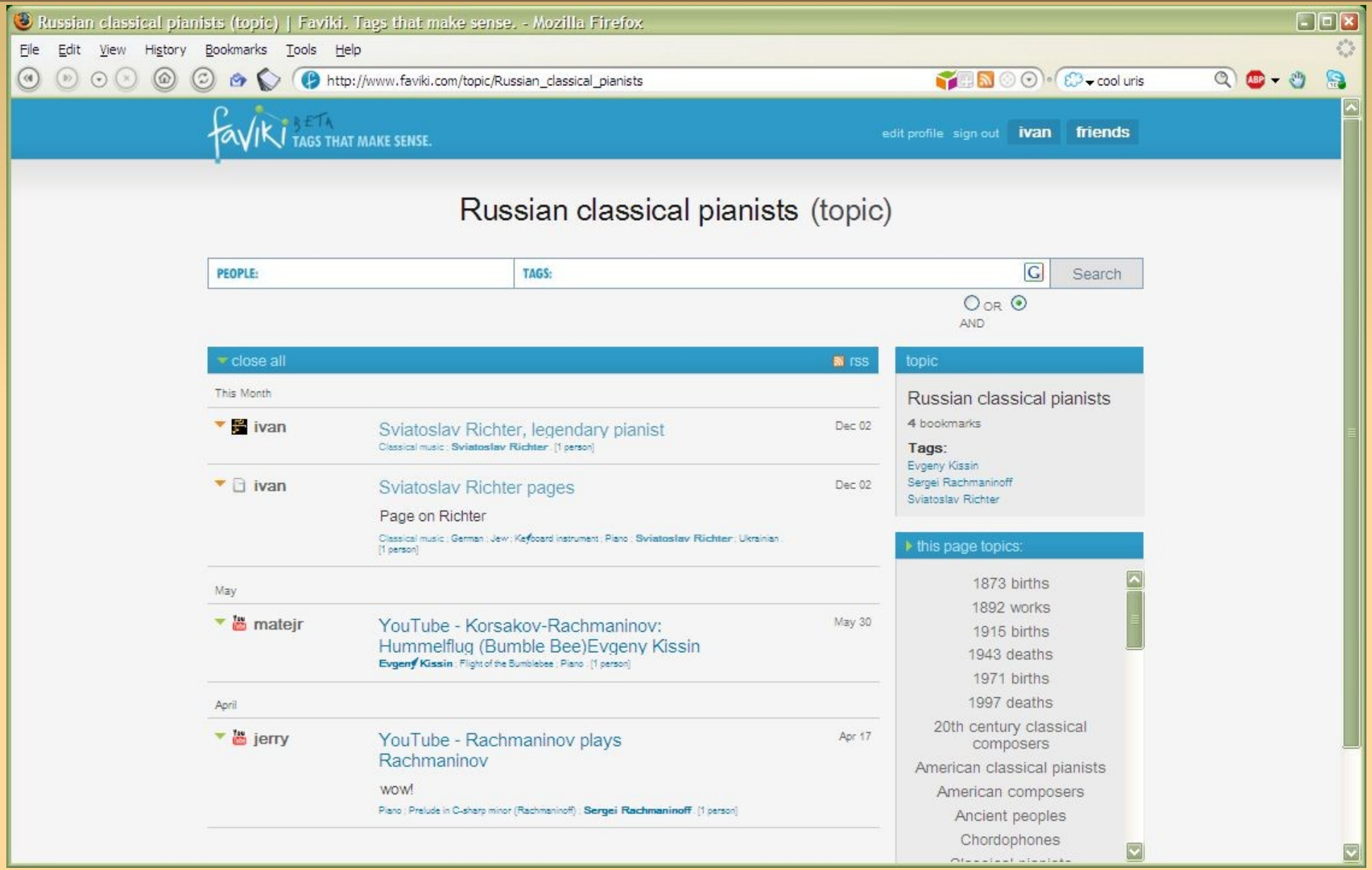
Faviki Example



Faviki Example



Faviki Example



Other application areas come to the fore

- Content management
- Business intelligence
- Collaborative user interfaces
- Sensor-based services
- Linking virtual communities
- Grid infrastructure
- Multimedia data management
- Etc

CEO guide for SW: the “DO-s”

- **Start small**: Test the Semantic Web waters with a pilot project […] before investing large sums of time and money.
- **Check credentials**: A lot of systems integrators don't really have the skills to deal with Semantic Web technologies. Get someone who's savy in semantics.
- **Expect training challenges**: It often takes people a while to understand the technology. […]
- **Find an ally**: It can be hard to articulate the potential benefits, so find someone with a problem that can be solved with the Semantic Web and make that person a partner.

Source: BusinessWeek Online, April 2007

CEO guide for SW: the “DON’T-s”

- **Go it alone**: The Semantic Web is complex, and it's best to get help. […]
- **Forget privacy**: Just because you can gather and correlate data about employees doesn’t mean you should. Set usage guidelines to safeguard employee privacy.
- **Expect perfection**: While these technologies will help you find and correlate information more quickly, they’re far from perfect. Nothing can help if data are unreliable in the first place.
- **Be impatient**: One early adopter at NASA says that the potential benefits can justify the investments in time, money, and resources, but there must be a multi-year commitment to have any hope of success

Source: BusinessWeek Online, April 2007

Thank you for your attention!

- These slides are publicly available via:

`http://www.w3.org/2009/Talks/1214-Darmstadt/`