9th ReasoningWeb Summer School 2013

July 30th, 2013, Mannheim, Germany

Linked Data on the Web

Prof. Dr. Christian Bizer

Dr. Heiko Paulheim

University of Mannheim

Hallo

- **Prof. Dr. Christian Bizer**
- **Professor for Information Systems**
- **Research Interests:**
  - Global Data Spaces
  - Linked Data Technologies
  - Data- and Web Mining
- **eMail: chris@informatik.uni-mannheim.de**

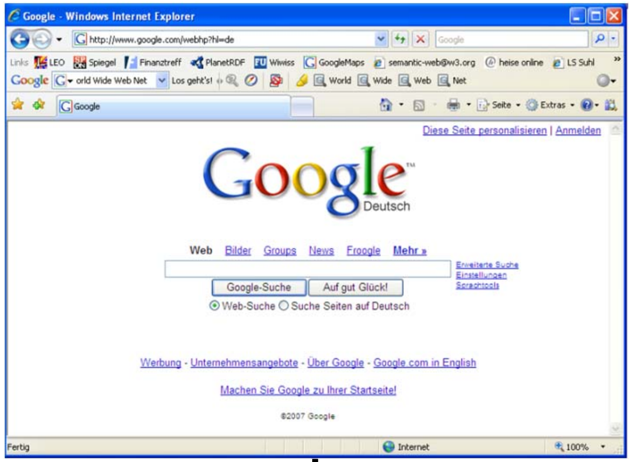
Hello

- **Dr. Heiko Paulheim**
- **Postdoctoral Researcher**
- **Research Interests:**
  - Data Mining and Machine Learning on/with Linked Data
  - Ontology and Schema Matching
  - Data Quality
- **eMail: heiko@informatik.uni-mannheim.de**

Outline

1. **Foundations of Linked Data**
    - What is the vision and goal?
2. **The Web of Linked Data**
    - What data is out there?
3. **How to publish and consume Linked Data?**
    - Tasks and Tools
    - Sharing the Integration Effort
4. **Alternative Web Data Publication Formats**
    - RDFa, Microdata, Microformats
5. **Challenges involved in using Web Data**
6. **Building Knowledge-intensive Applications**

What does the classic Web offer us?


Google - Windows Internet Explorer
http://www.google.com/webhp?hl=de
Links LEO Spiegel Finanztreff PlanetRDF Wiwiss GoogleMaps semantic-web@w3.org heise online LS Suhl
orld Wide Web Net Los geht's! World Wide Web Net
Google
Seite Extras
Fertig
Internet
100%

Diese Seite personalisieren | Anmelden

Google Deutsch

Web Bilder Groups News Froogle Mehr »

Google-Suche Auf gut Glück!

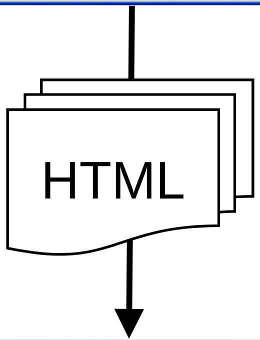
Erweiterte Suche
Einstellungen
Sprachtools

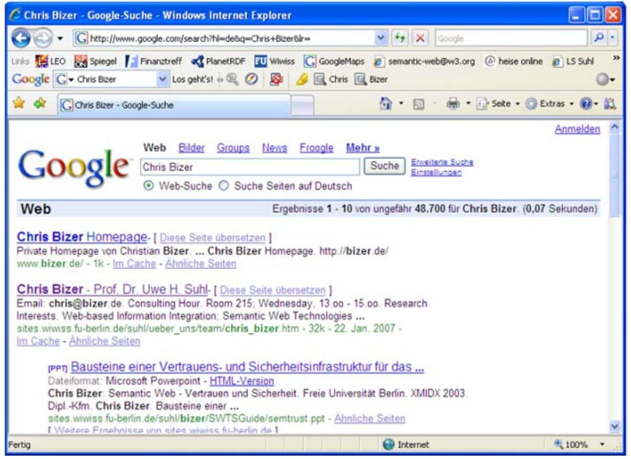
● Web-Suche ○ Suche Seiten auf Deutsch

Werbung - Unternehmensangebote - Über Google - Google.com in English

Machen Sie Google zu Ihrer Startseite!

©2007 Google




Chris Bizer - Google-Suche - Windows Internet Explorer
http://www.google.com/search?hl=de&q=Chris+Bizer&lr=
Links LEO Spiegel Finanztreff PlanetRDF Wiwiss GoogleMaps semantic-web@w3.org heise online LS Suhl
Chris Bizer Los geht's! Chris Bizer
Chris Bizer - Google-Suche
Seite Extras
Fertig
Internet
100%

Anmelden

Google

Web Bilder Groups News Froogle Mehr »

Chris Bizer Suche

Erweiterte Suche
Einstellungen

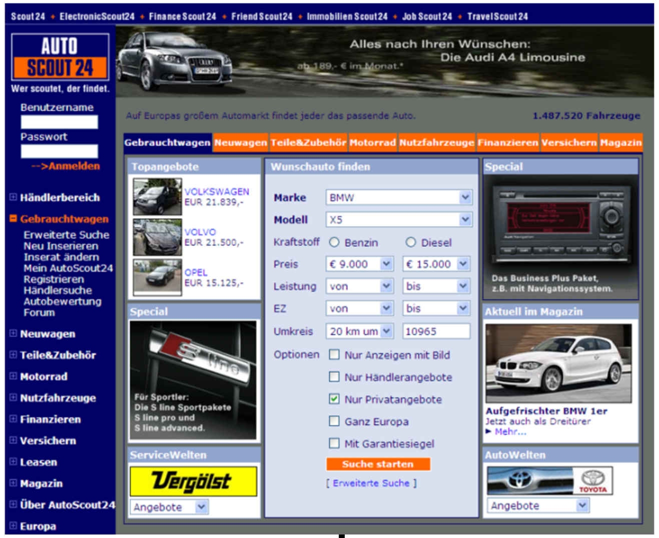
● Web-Suche ○ Suche Seiten auf Deutsch

Web Ergebnisse **1** - **10** von ungefähr **48.700** für **Chris Bizer**. (**0,07** Sekunden)

Chris Bizer Homepage- [Diese Seite übersetzen]
Private Homepage von Christian **Bizer**. ... **Chris Bizer** Homepage. http://**bizer**.de/
www.**bizer**.de/ - 1k - Im Cache - Ähnliche Seiten

Chris Bizer - Prof. Dr. Uwe H. Suhl- [Diese Seite übersetzen]
Email: **chris@bizer**.de. Consulting Hour. Room 215; Wednesday, 13.oo - 15.oo. Research Interests. Web-based Information Integration; Semantic Web Technologies ...
sites.wiwiss.fu-berlin.de/suhl/ueber_uns/team/**chris_bizer**.htm - 32k - 22. Jan. 2007 -
Im Cache - Ähnliche Seiten

[PPT] Bausteine einer Vertrauens- und Sicherheitsinfrastruktur für das ...
Dateiformat: Microsoft Powerpoint - HTML-Version
Chris Bizer: Semantic Web - Vertrauen und Sicherheit. Freie Universität Berlin. XMIDX 2003. Dipl.-Kfm. **Chris Bizer**. Bausteine einer ...
sites.wiwiss.fu-berlin.de/suhl/**bizer**/SWTSGuide/semtrust.ppt - Ähnliche Seiten
[Weitere Ergebnisse von sites.wiwiss.fu-berlin.de]


Scout24 ◆ ElectronicScout24 ◆ FinanceScout24 ◆ FriendScout24 ◆ ImmobilienScout24 ◆ JobScout24 ◆ TravelScout24
AUTO SCOUT 24
Wer scoutet, der findet.
Alles nach Ihren Wünschen:
Die Audi A4 Limousine
ab 189,- € im Monat.*
Vergölst
TOYOTA

Benutzername

Passwort

-->Anmelden

⊞ Händlerbereich
⊟ Gebrauchtwagen
Erweiterte Suche
Neu Inserieren
Inserat ändern
Mein AutoScout24
Registrieren
Händlersuche
Autobewertung
Forum
⊞ Neuwagen
⊞ Teile&Zubehör
⊞ Motorrad
⊞ Nutzfahrzeuge
⊞ Finanzieren
⊞ Versichern
⊞ Leasen
⊞ Magazin
⊞ Über AutoScout24
⊞ Europa

Auf Europas großem Automarkt findet jeder das passende Auto. 1.487.520 Fahrzeuge

Gebrauchtwagen | Neuwagen | Teile&Zubehör | Motorrad | Nutzfahrzeuge | Finanzieren | Versichern | Magazin

Topangebote
VOLKSWAGEN EUR 21.839,-
VOLVO EUR 21.500,-
OPEL EUR 15.125,-

Special
Für Sportler:
Die S line Sportpakete
S line pro und
S line advanced.

ServiceWelten
Angebote

Wunschauto finden

Marke	BMW	
Modell	X5	
Kraftstoff	○ Benzin	○ Diesel
Preis	€ 9.000	€ 15.000
Leistung	von	bis
EZ	von	bis
Umkreis	20 km um	10965

Optionen
□ Nur Anzeigen mit Bild
□ Nur Händlerangebote
☑ Nur Privatangebote
□ Ganz Europa
□ Mit Garantiesiegel

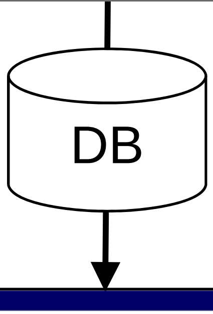
Suche starten

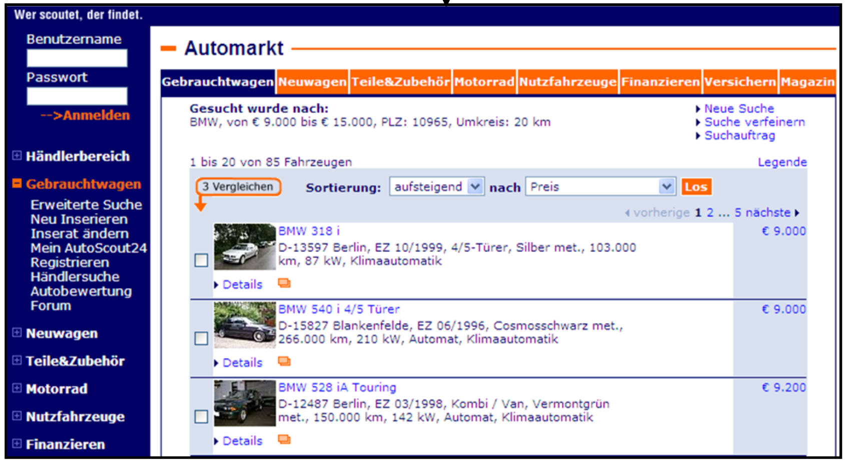
[Erweiterte Suche]

Special
Das Business Plus Paket,
z.B. mit Navigationssystem.

Aktuell im Magazin
Aufgefrischter BMW 1er
Jetzt auch als Dreitürer
▸ Mehr...

AutoWelten
Angebote





Benutzername

Passwort

-->Anmelden

⊞ Händlerbereich
⊟ Gebrauchtwagen
Erweiterte Suche
Neu Inserieren
Inserat ändern
Mein AutoScout24
Registrieren
Händlersuche
Autobewertung
Forum
⊞ Neuwagen
⊞ Teile&Zubehör
⊞ Motorrad
⊞ Nutzfahrzeuge
⊞ Finanzieren

Automarkt

Gebrauchtwagen | Neuwagen | Teile&Zubehör | Motorrad | Nutzfahrzeuge | Finanzieren | Versichern | Magazin

Gesucht wurde nach:
BMW, von € 9.000 bis € 15.000, PLZ: 10965, Umkreis: 20 km

▸ Neue Suche
▸ Suche verfeinern
▸ Suchauftrag

1 bis 20 von 85 Fahrzeugen Legende

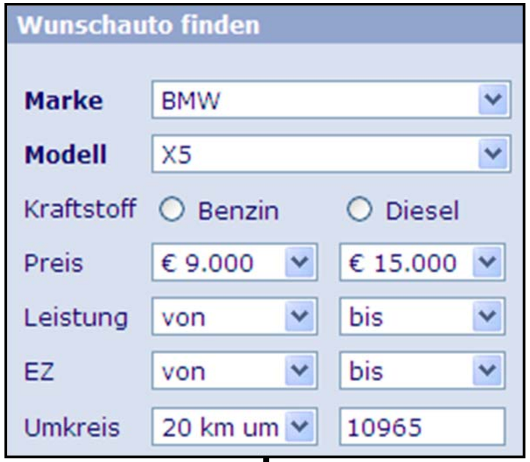
3 Vergleichen **Sortierung:** aufsteigend **nach** Preis Los

◂ vorherige **1** 2 ... 5 nächste ▸

□	BMW 318 i D-13597 Berlin, EZ 10/1999, 4/5-Türer, Silber met., 103.000 km, 87 kW, Klimaautomatik ▸ Details	€ 9.000
□	BMW 540 i 4/5 Türer D-15827 Blankenfelde, EZ 06/1996, Cosmosschwarz met., 266.000 km, 210 kW, Automat, Klimaautomatik ▸ Details	€ 9.000
□	BMW 528 iA Touring D-12487 Berlin, EZ 03/1998, Kombi / Van, Vermontgrün met., 150.000 km, 142 kW, Automat, Klimaautomatik ▸ Details	€ 9.200

What do we actually want?

Use the Web like a single, global database



Wunschauto finden

Marke BMW

Modell X5

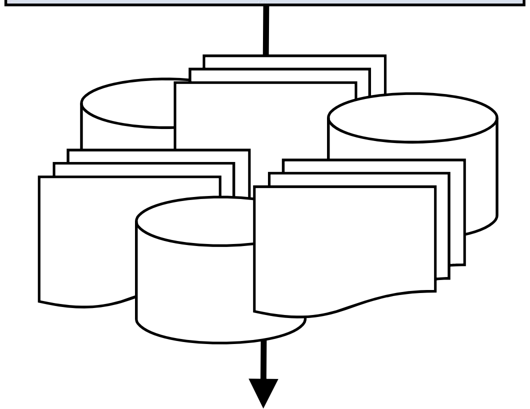
Kraftstoff ○ Benzin ○ Diesel

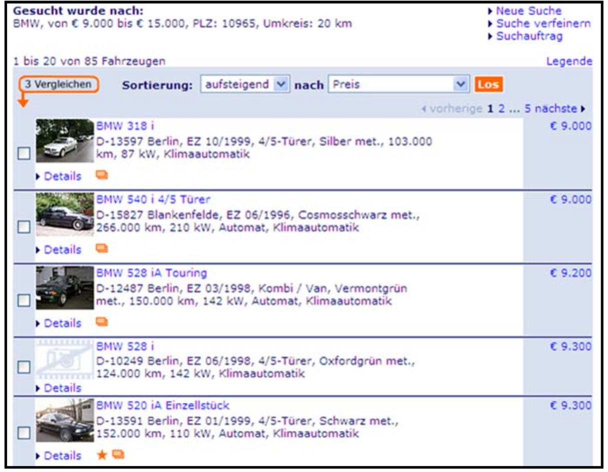
Preis € 9.000 € 15.000

Leistung von bis

EZ von bis

Umkreis 20 km um 10965





Gesucht wurde nach:
BMW, von € 9.000 bis € 15.000, PLZ: 10965, Umkreis: 20 km

- Neue Suche
- Suche verfeinern
- Suchauftrag

1 bis 20 von 85 Fahrzeugen

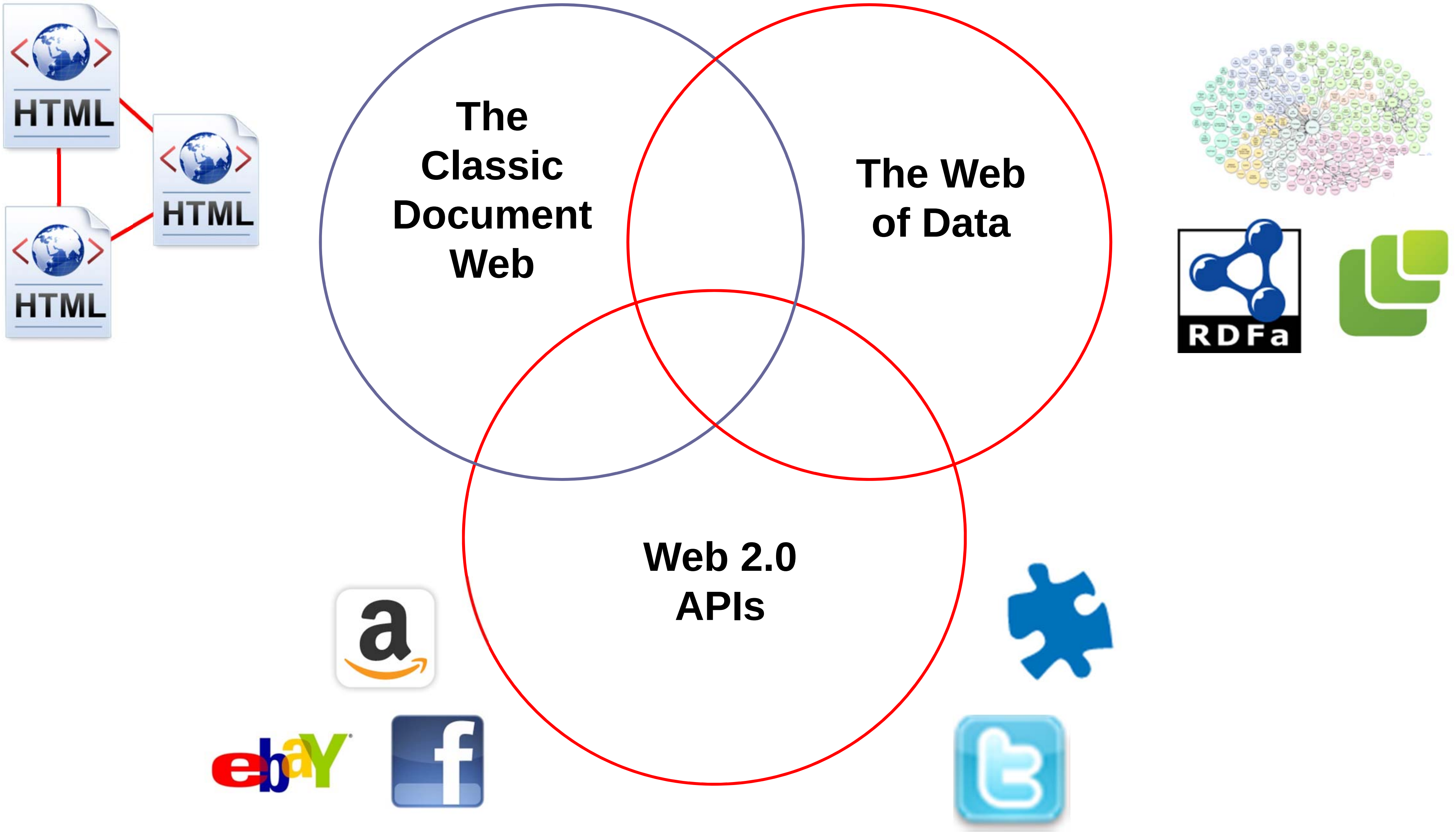
Legende

3 Vergleichen **Sortierung:** aufsteigend **nach** Preis Los

vorherige **1** 2 ... 5 nächste

□	BMW 318 i D-13597 Berlin, EZ 10/1999, 4/5-Türer, Silber met., 103.000 km, 87 kW, Klimaautomatik Details	€ 9.000
□	BMW 540 i 4/5 Türer D-15827 Blankenfelde, EZ 06/1996, Cosmosschwarz met., 266.000 km, 210 kW, Automat, Klimaautomatik Details	€ 9.000
□	BMW 528 iA Touring D-12487 Berlin, EZ 03/1998, Kombi / Van, Vermontgrün met., 150.000 km, 142 kW, Automat, Klimaautomatik Details	€ 9.200
□	BMW 528 i D-10249 Berlin, EZ 06/1998, 4/5-Türer, Oxfordgrün met., 124.000 km, 142 kW, Klimaautomatik Details	€ 9.300
□	BMW 520 iA Einzellstück D-13591 Berlin, EZ 01/1999, 4/5-Türer, Schwarz met., 152.000 km, 110 kW, Automat, Klimaautomatik Details ★	€ 9.300

Access to structured Data on the Web

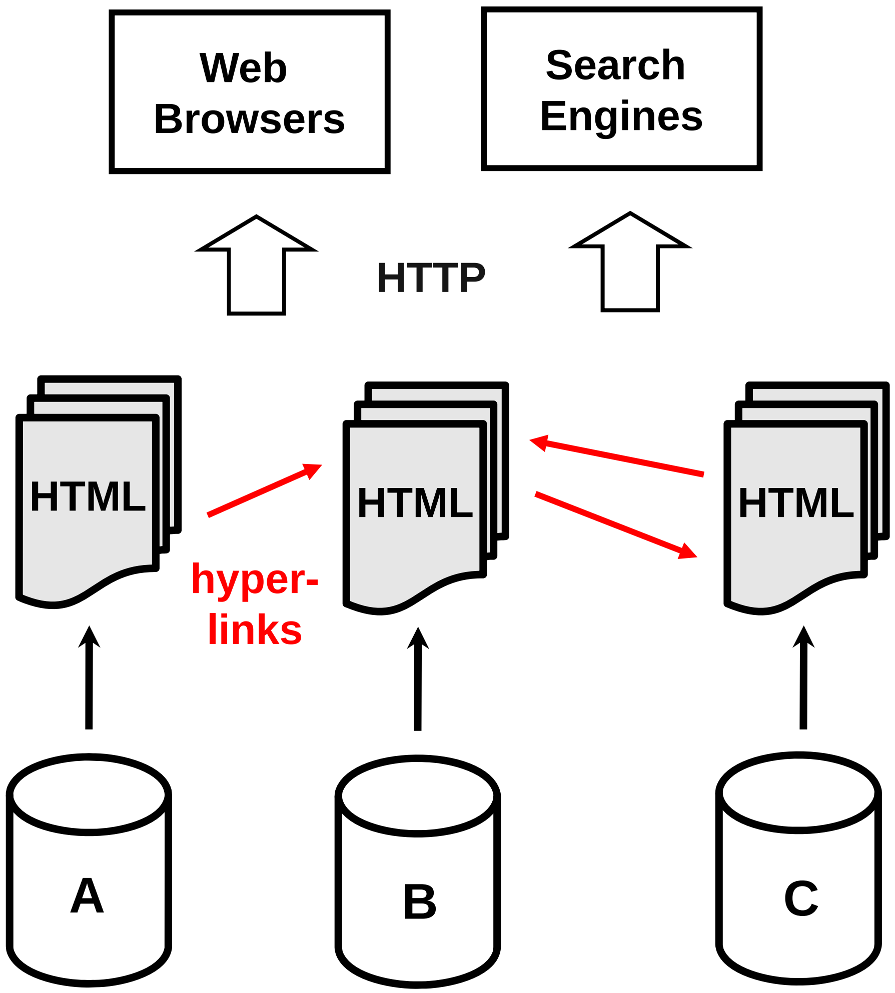


The Classic Document Web

The Web of Data

Web 2.0 APIs

Architecture of the classic Document Web

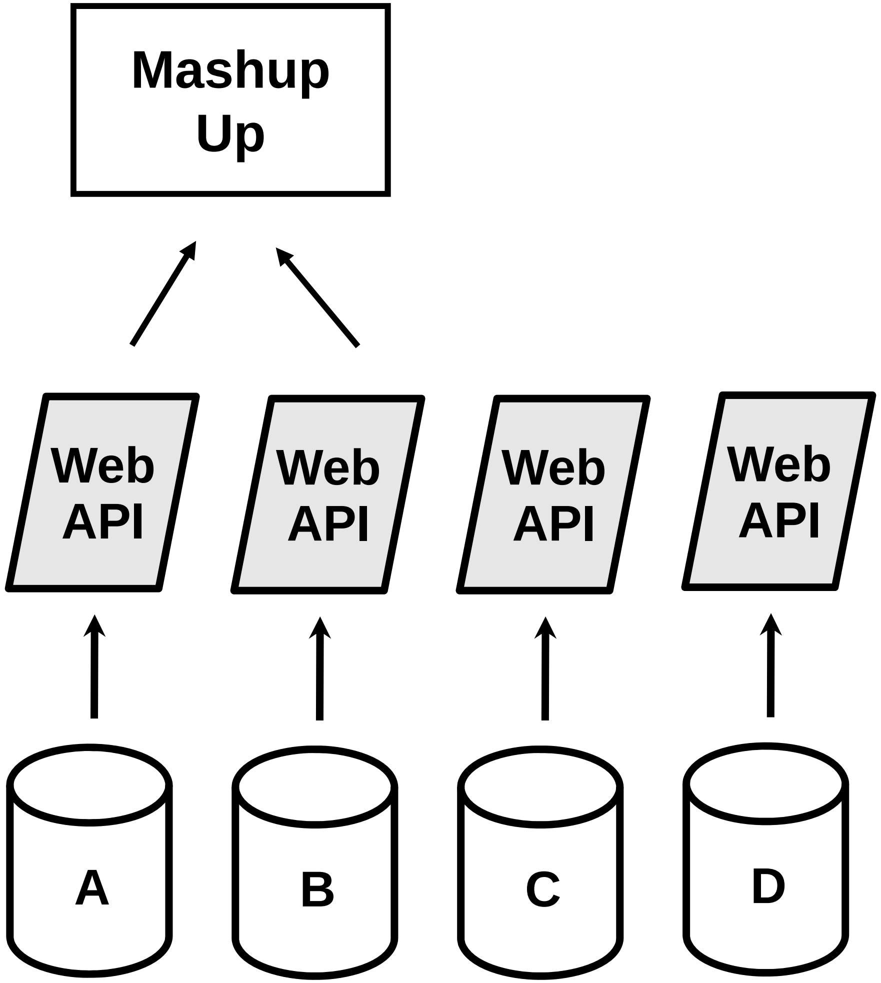


Single global information space

Small set of simple standards

1. **HTML as document format**
2. **HTTP URLs as**
   - globally unique IDs
   - retrieval mechanism
3. **Hyperlinks to connect everything**

Web 2.0 APIs and Mashups

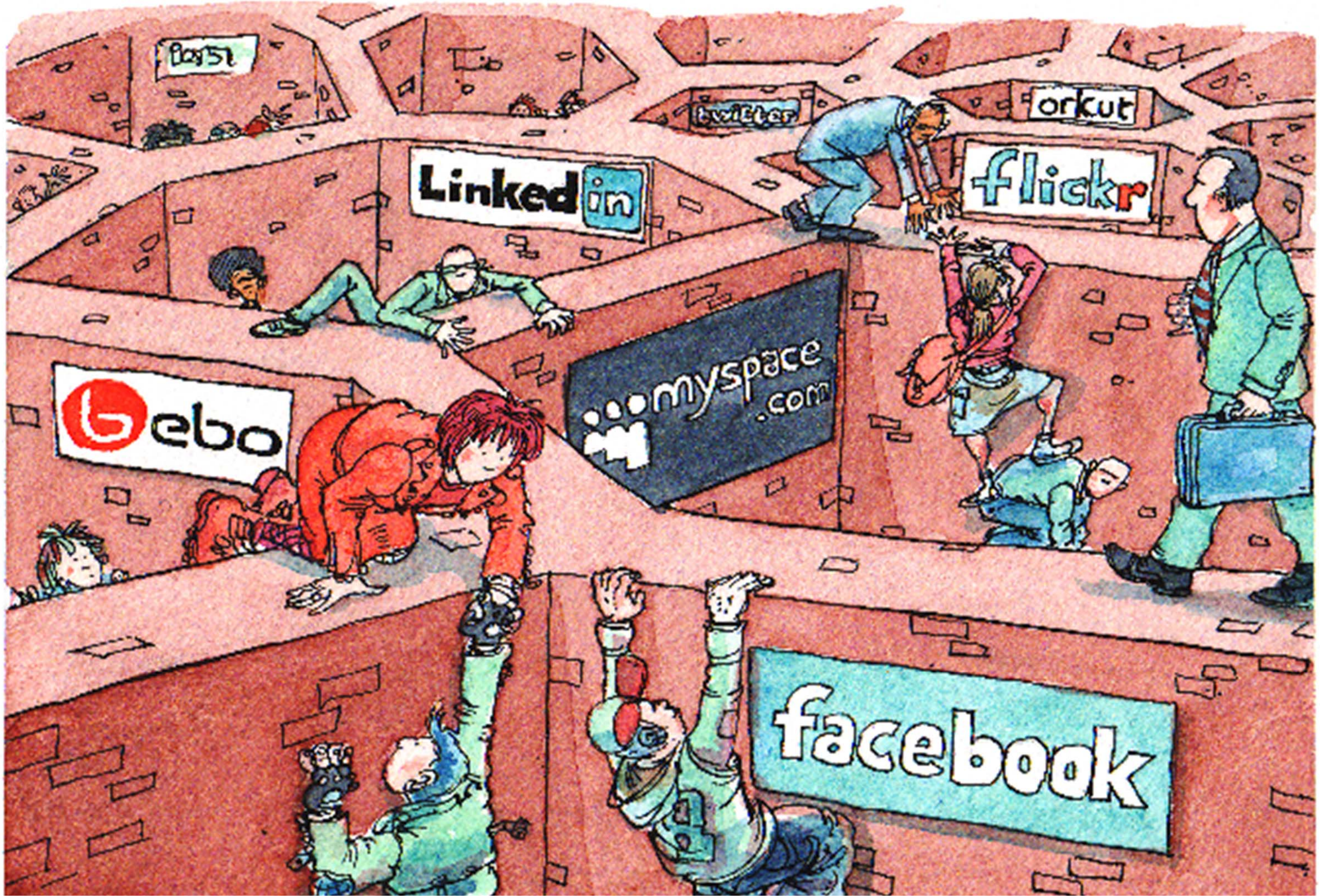


No single global dataspace

Shortcomings

1. APIs have proprietary interfaces
2. No hyperlinks between data items within different APIs
3. Mashups are based on a fixed set of data sources

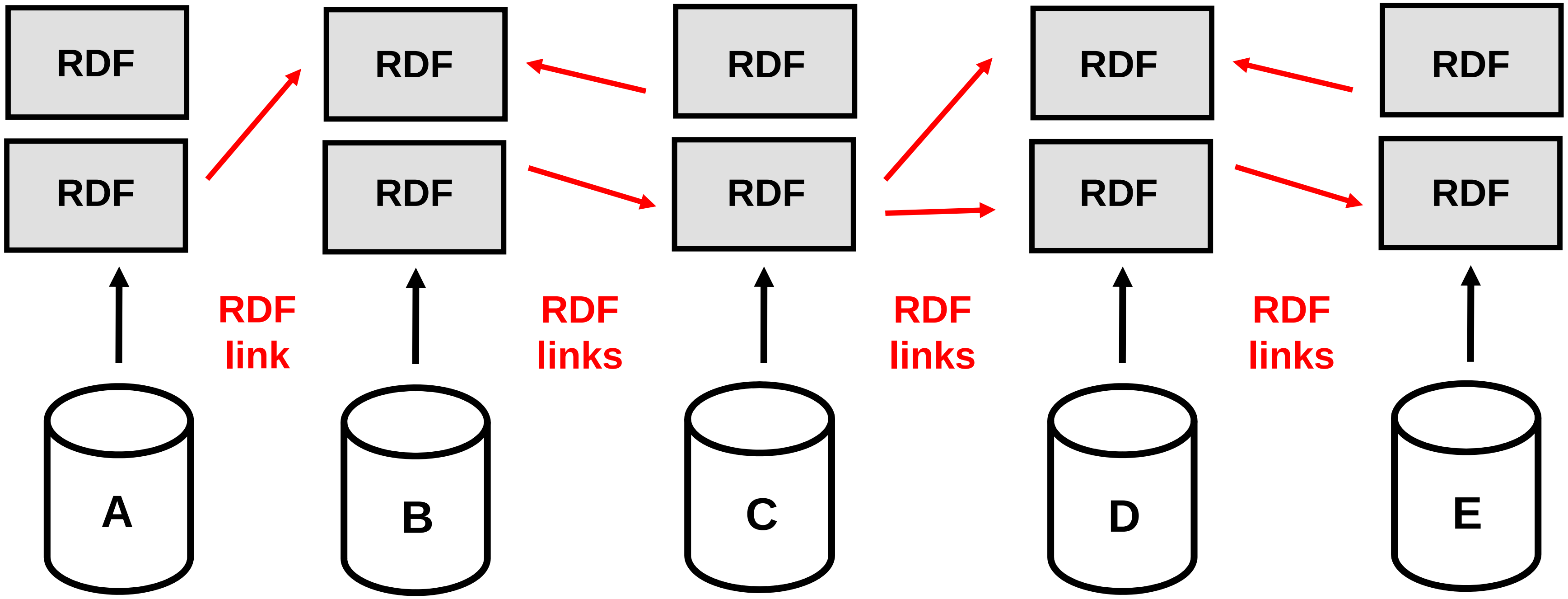
Web APIs slice the Web into Walled Gardens



Linked Data

Extend the Web with a single global dataspace

1. by using RDF to publish structured data on the Web
2. by setting links between data items within different data sources.



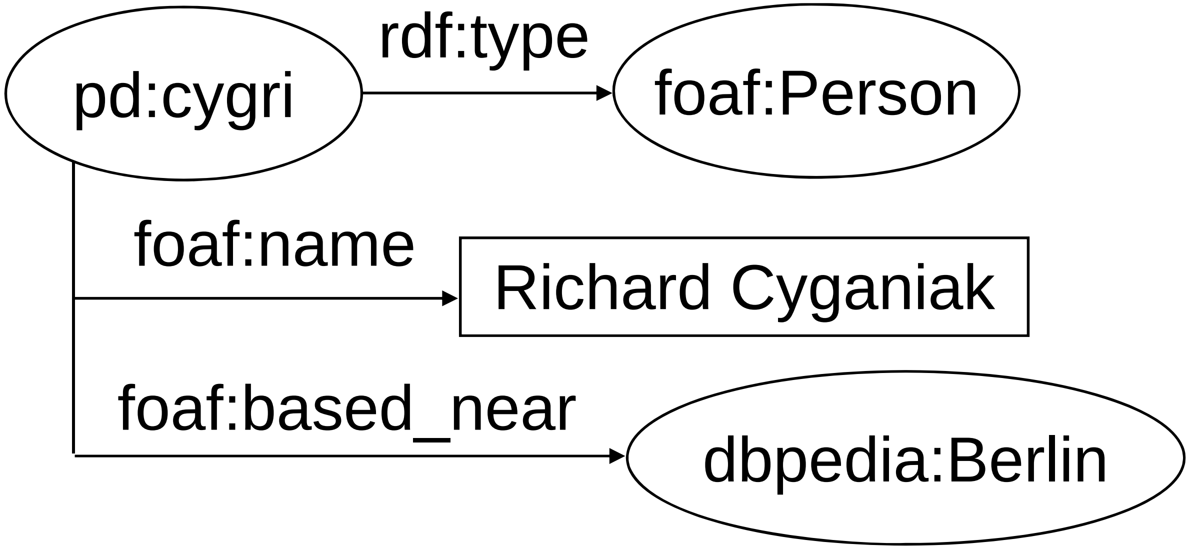
Linked Data

Set of best practices for publishing structured data on the Web in accordance with the general architecture of the Web.

1. **Use URIs as names for things.**
2. **Use HTTP URIs so that people can look up those names.**
3. **When someone looks up a URI, provide useful RDF information.**
4. **Include RDF statements that link to other URIs so that they can discover related things.**

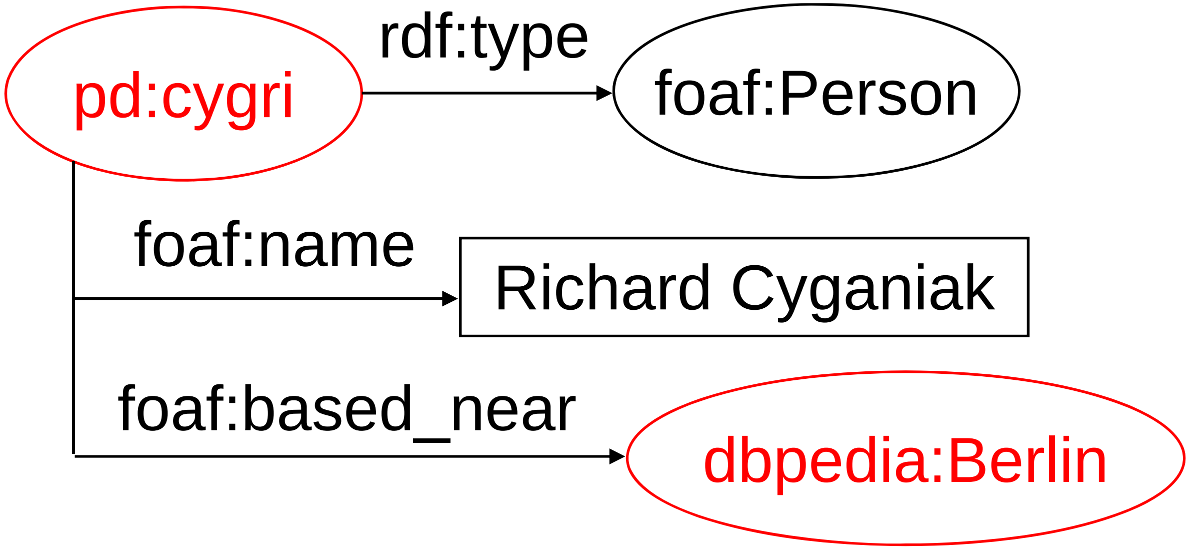
Tim Berners-Lee, http://www.w3.org/DesignIssues/LinkedData.html, 2006

The Basis: RDF Data Model



Flexible graph-based data model.

Data items are identified with HTTP URIs

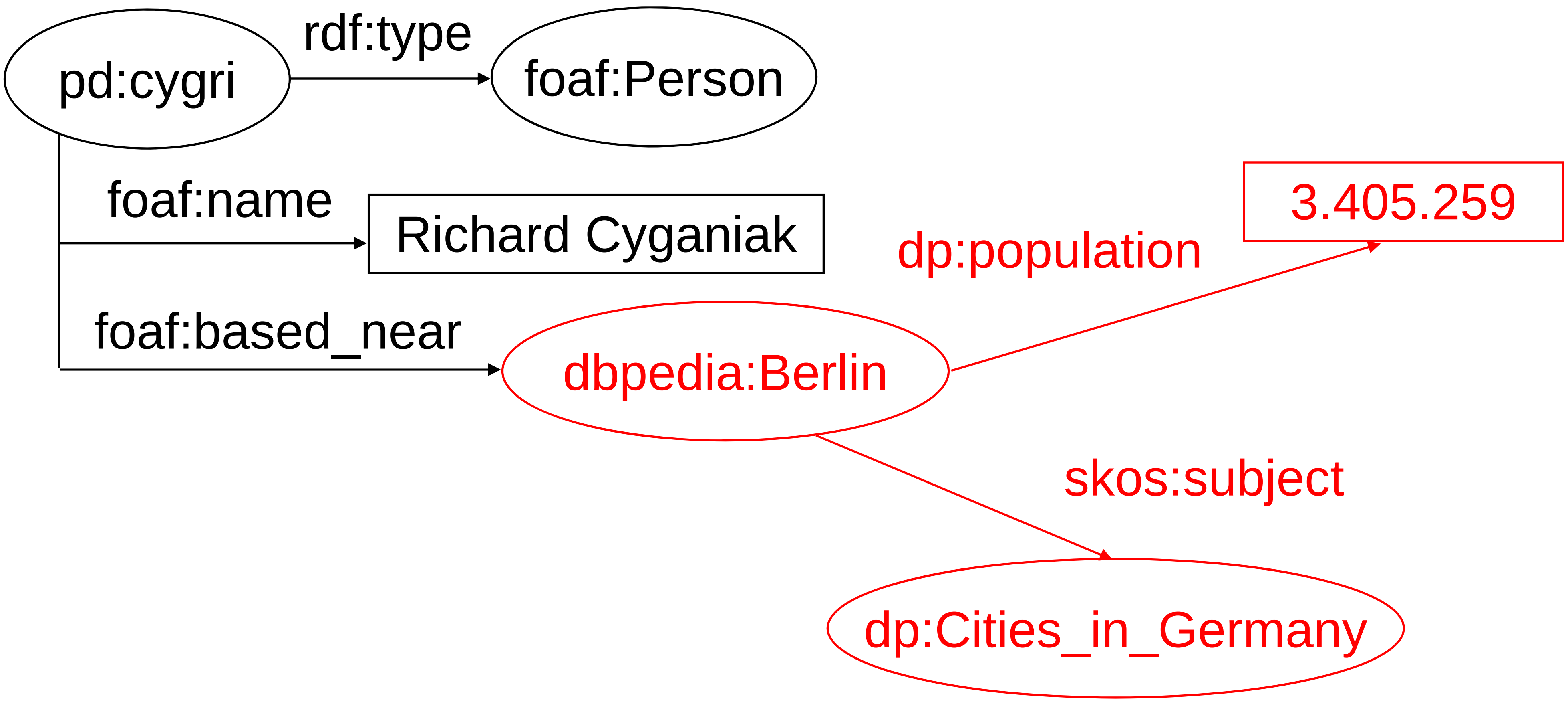


HTTP URIs take the role of global primary keys.

pd:cygri = http://richard.cyganiak.de/foaf.rdf#cygri

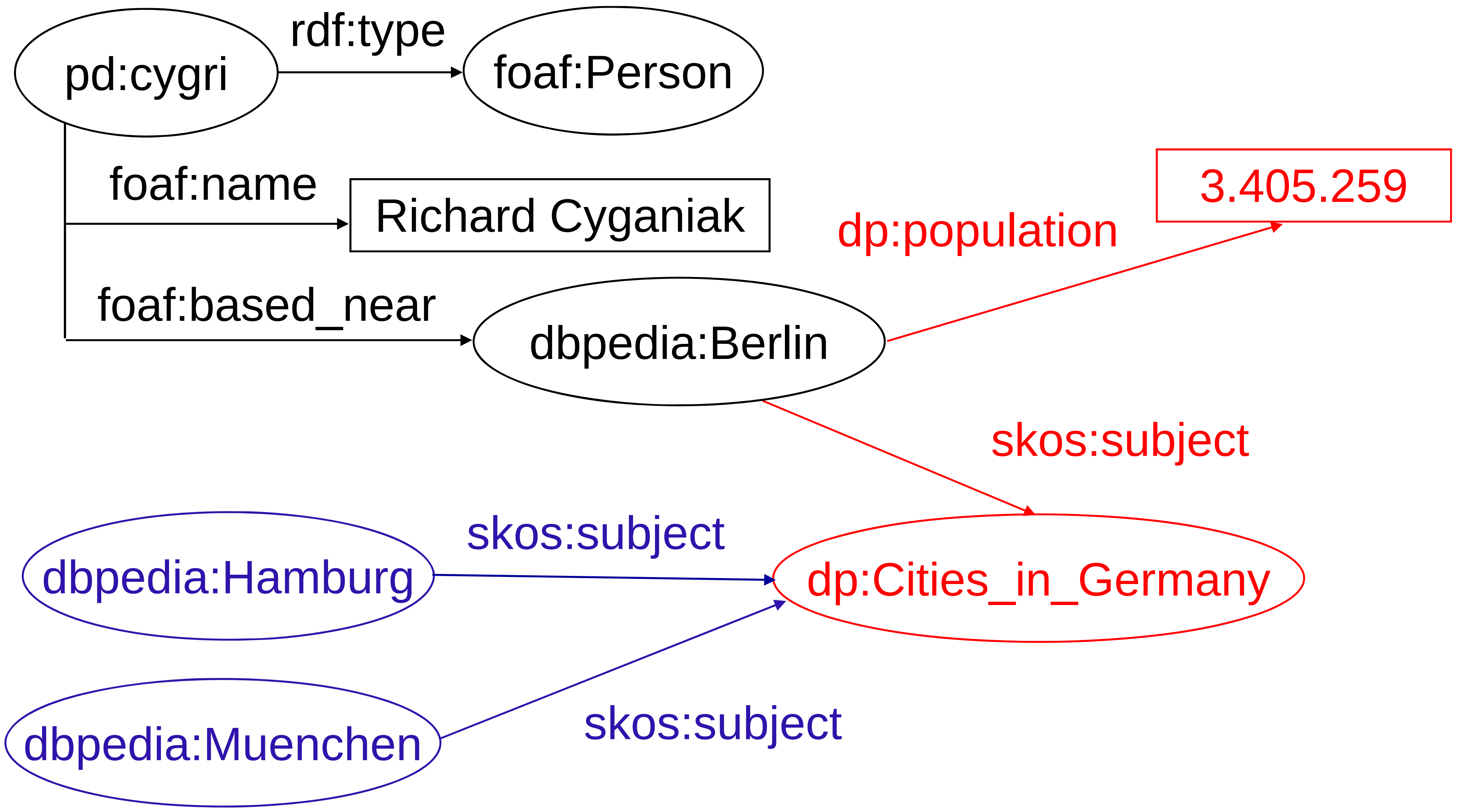
dbpedia:Berlin = http://dbpedia.org/resource/Berlin

Resolving URIs over the Web



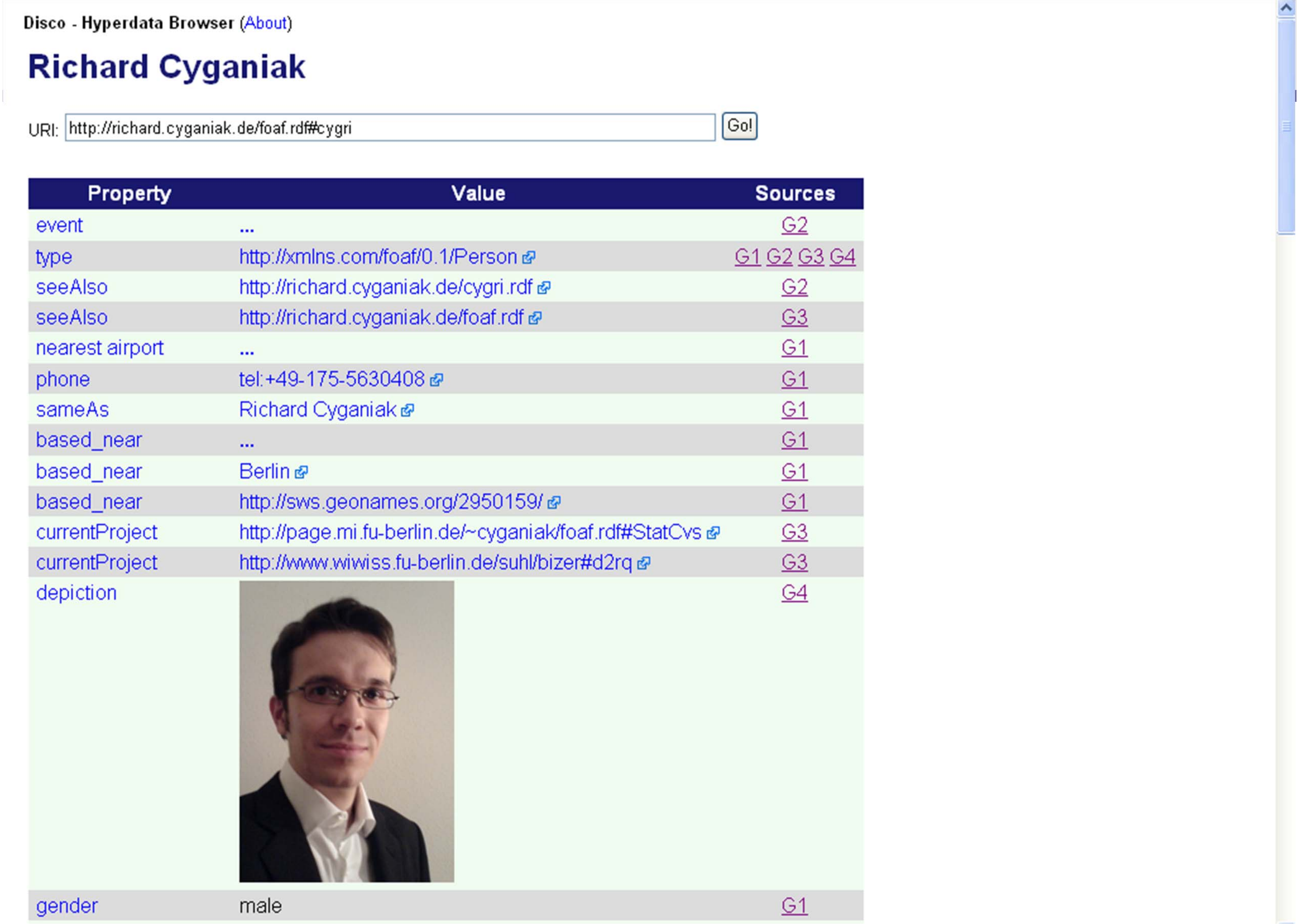
The HTTP protocol brings together identification and retrieval again.

Following Links deeper into the Web



Properties of the Web of Linked Data

- **Global, distributed dataspace build on a simple set of standards**
  - RDF, URIs, HTTP
- **Entities are connected by links**
  - creating a global data graph that spans data sources and
  - enables the discovery of new data sources
- **Provides for data-coexistence**
  - Everyone can publish data to the Web of Linked Data
  - Everyone can express their personal view on things
- **The Web of Linked Data can be used by generic applications**
  - Linked Data Browsers
  - Linked Data Search Engines

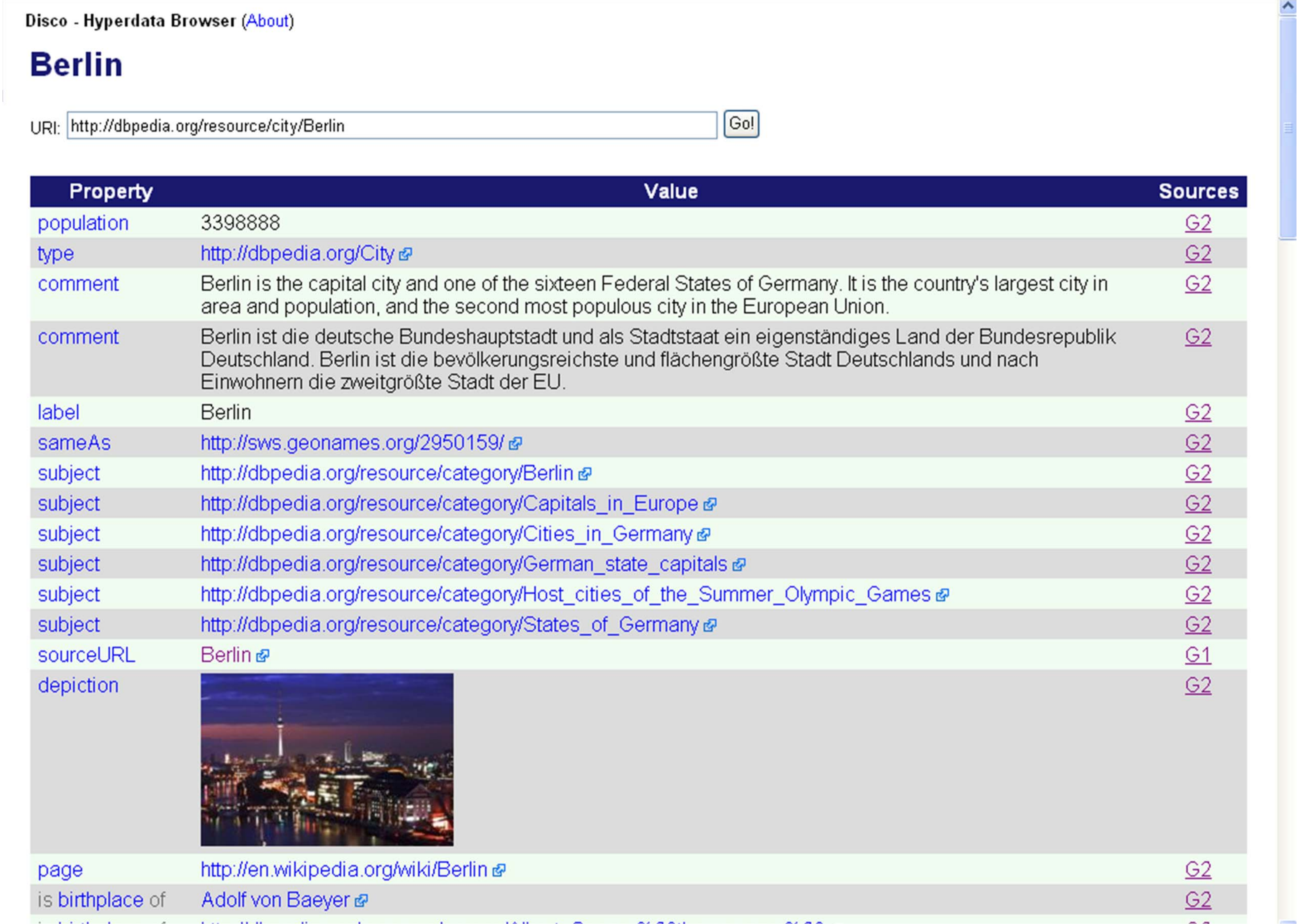


Disco - Hyperdata Browser (About)

Richard Cyganiak

URI: http://richard.cyganiak.de/foaf.rdf#cygri Go!

Property	Value	Sources
event	...	G2
type	http://xmlns.com/foaf/0.1/Person	G1 G2 G3 G4
seeAlso	http://richard.cyganiak.de/cygri.rdf	G2
seeAlso	http://richard.cyganiak.de/foaf.rdf	G3
nearest airport	...	G1
phone	tel:+49-175-5630408	G1
sameAs	Richard Cyganiak	G1
based_near	...	G1
based_near	Berlin	G1
based_near	http://sws.geonames.org/2950159/	G1
currentProject	http://page.mi.fu-berlin.de/~cyganiak/foaf.rdf#StatCvs	G3
currentProject	http://www.wiwiss.fu-berlin.de/suhl/bizer#d2rq	G3
depiction		G4
gender	male	G1

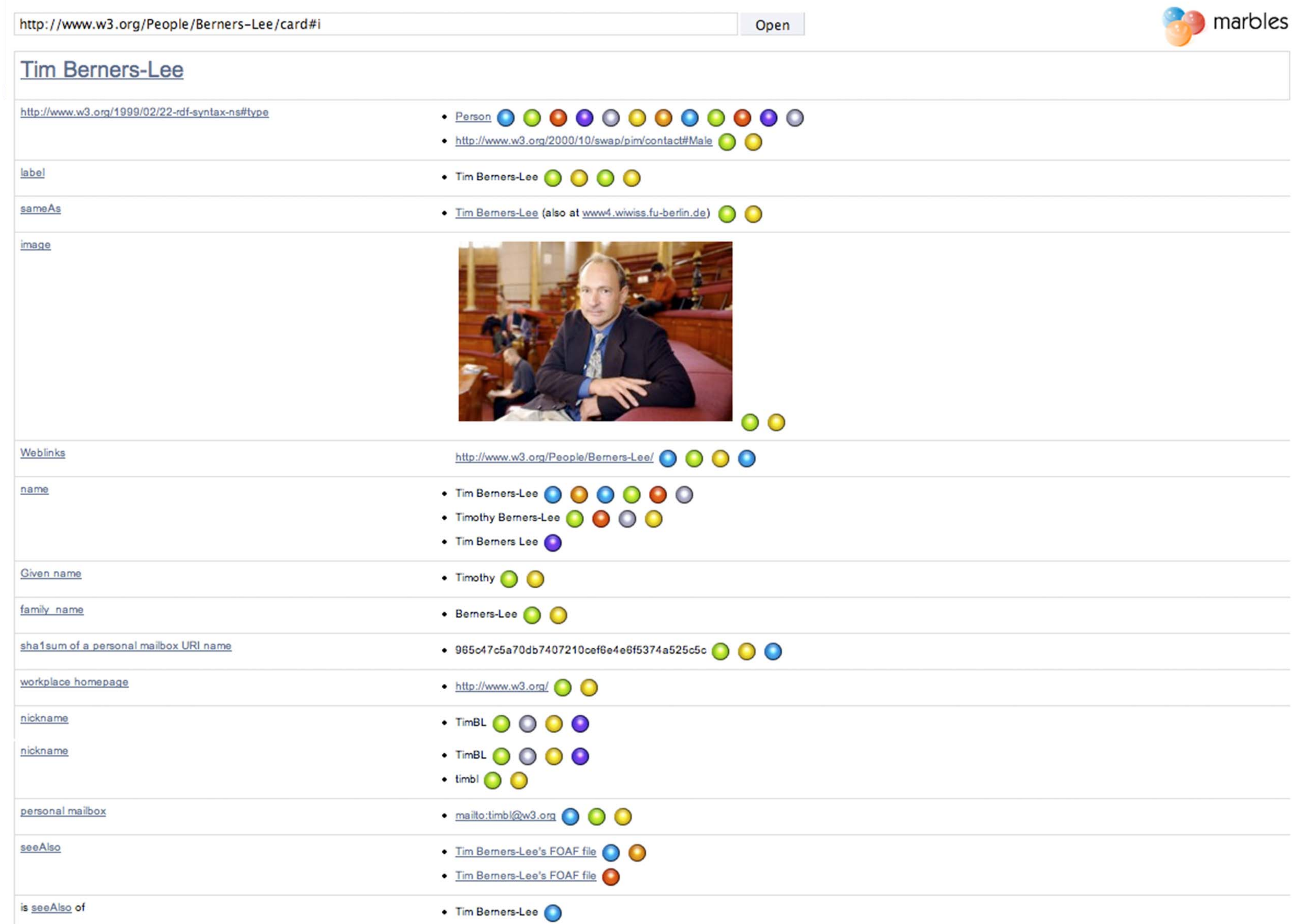


Disco - Hyperdata Browser (About)

Berlin

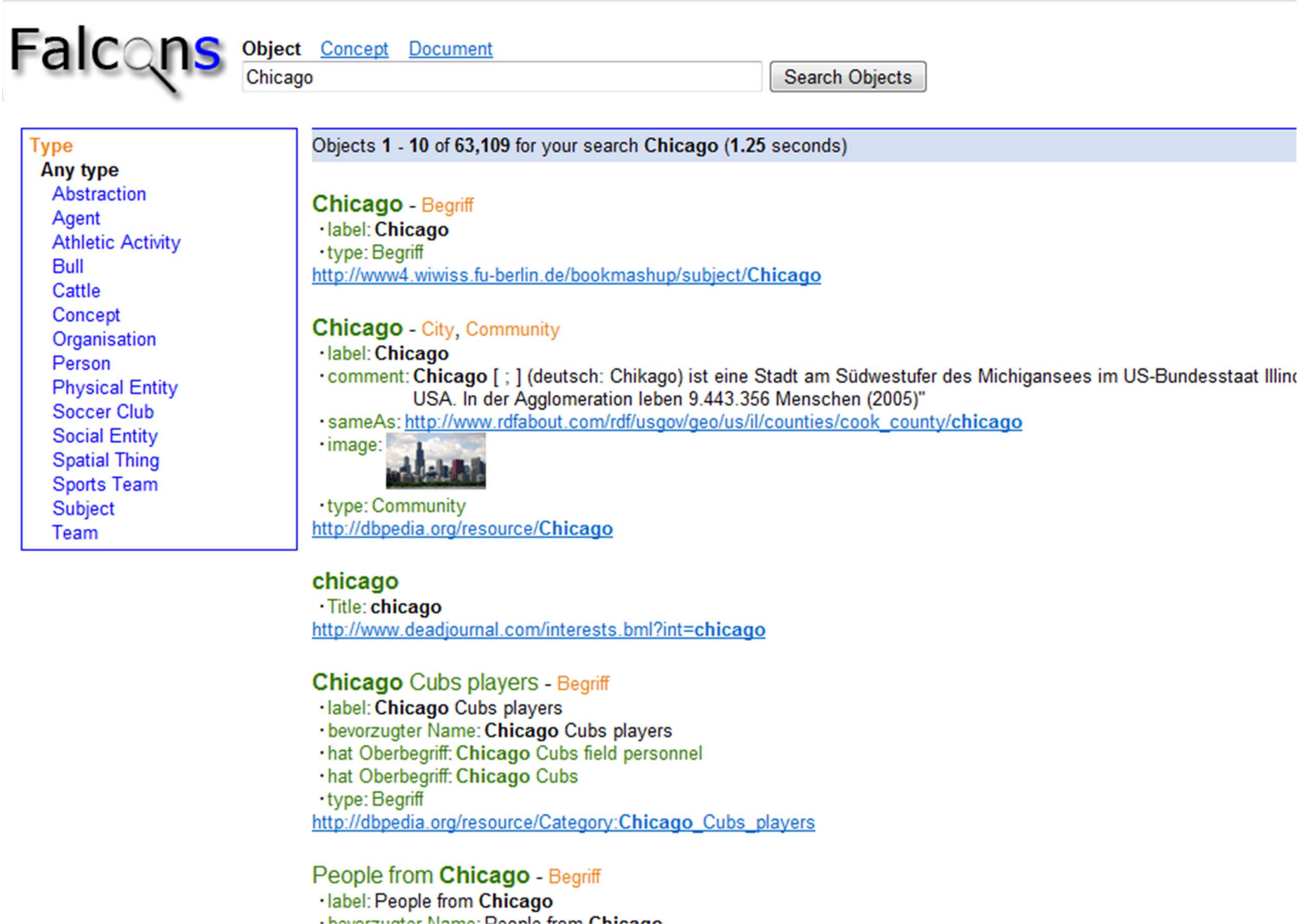
URI: http://dbpedia.org/resource/city/Berlin Go!

Property	Value	Sources
population	3398888	G2
type	http://dbpedia.org/City	G2
comment	Berlin is the capital city and one of the sixteen Federal States of Germany. It is the country's largest city in area and population, and the second most populous city in the European Union.	G2
comment	Berlin ist die deutsche Bundeshauptstadt und als Stadtstaat ein eigenständiges Land der Bundesrepublik Deutschland. Berlin ist die bevölkerungsreichste und flächengrößte Stadt Deutschlands und nach Einwohnern die zweitgrößte Stadt der EU.	G2
label	Berlin	G2
sameAs	http://sws.geonames.org/2950159/	G2
subject	http://dbpedia.org/resource/category/Berlin	G2
subject	http://dbpedia.org/resource/category/Capitals_in_Europe	G2
subject	http://dbpedia.org/resource/category/Cities_in_Germany	G2
subject	http://dbpedia.org/resource/category/German_state_capitals	G2
subject	http://dbpedia.org/resource/category/Host_cities_of_the_Summer_Olympic_Games	G2
subject	http://dbpedia.org/resource/category/States_of_Germany	G2
sourceURL	Berlin	G1
depiction		G2
page	http://en.wikipedia.org/wiki/Berlin	G2
is birthplace of	Adolf von Baeyer	G2



Tim Berners-Lee

http://www.w3.org/1999/02/22-rdf-syntax-ns#type	• Person ● ● ● ● ● ● ● ● ● ● ● ● • http://www.w3.org/2000/10/swap/pim/contact#Male ● ●
label	• Tim Berners-Lee ● ● ● ●
sameAs	• Tim Berners-Lee (also at www4.wiwiss.fu-berlin.de) ● ●
image	● ●
Weblinks	http://www.w3.org/People/Berners-Lee/ ● ● ● ●
name	• Tim Berners-Lee ● ● ● ● ● ● • Timothy Berners-Lee ● ● ● ● • Tim Berners Lee ●
Given name	• Timothy ● ●
family name	• Berners-Lee ● ●
sha1sum of a personal mailbox URI name	• 965c47c5a70db7407210cef6e4e6f5374a525c5c ● ● ●
workplace homepage	• http://www.w3.org/ ● ●
nickname	• TimBL ● ● ● ●
nickname	• TimBL ● ● ● ● • timbl ● ●
personal mailbox	• mailto:timbl@w3.org ● ● ●
seeAlso	• Tim Berners-Lee's FOAF file ● ● • Tim Berners-Lee's FOAF file ●
is seeAlso of	• Tim Berners-Lee ●



Chicago

Objects **1** - **10** of **63,109** for your search **Chicago** (**1.25** seconds)

Chicago - Begriff
- label: **Chicago**
- type: Begriff

http://www4.wiwiss.fu-berlin.de/bookmashup/subject/**Chicago**

Chicago - City, Community
- label: **Chicago**
- comment: **Chicago** [;] (deutsch: Chikago) ist eine Stadt am Südwestufer des Michigansees im US-Bundesstaat Illinc
  USA. In der Agglomeration leben 9.443.356 Menschen (2005)"
- sameAs: http://www.rdfabout.com/rdf/usgov/geo/us/il/counties/cook_county/**chicago**
- image:
- type: Community

http://dbpedia.org/resource/**Chicago**

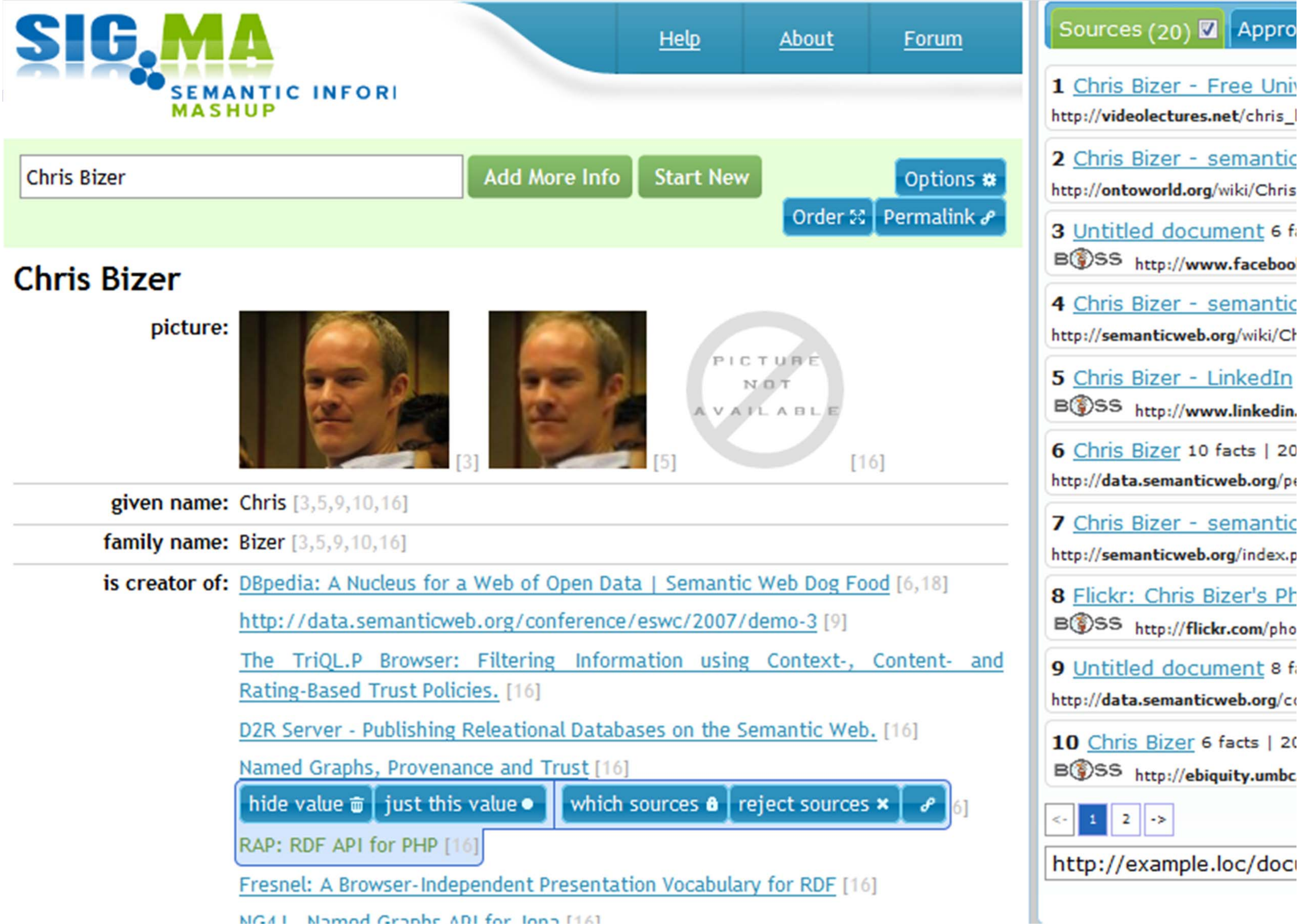
chicago
- Title: **chicago**

http://www.deadjournal.com/interests.bml?int=**chicago**

Chicago Cubs players - Begriff
- label: **Chicago** Cubs players
- bevorzugter Name: **Chicago** Cubs players
- hat Oberbegriff: **Chicago** Cubs field personnel
- hat Oberbegriff: **Chicago** Cubs
- type: Begriff

http://dbpedia.org/resource/Category:**Chicago**_Cubs_players

People from **Chicago** - Begriff
- label: People from **Chicago**
- bevorzugter Name: People from **Chicago**



Chris Bizer

Chris Bizer

picture: [3] [5] [16]

given name: Chris [3,5,9,10,16]

family name: Bizer [3,5,9,10,16]

is creator of: DBpedia: A Nucleus for a Web of Open Data | Semantic Web Dog Food [6,18]

http://data.semanticweb.org/conference/eswc/2007/demo-3 [9]

The TriQL.P Browser: Filtering Information using Context-, Content- and Rating-Based Trust Policies. [16]

D2R Server - Publishing Releational Databases on the Semantic Web. [16]

Named Graphs, Provenance and Trust [16]

6]

RAP: RDF API for PHP [16]

Fresnel: A Browser-Independent Presentation Vocabulary for RDF [16]

NG4J - Named Graphs API for Jena [16]

1 Chris Bizer - Free Uni
http://videolectures.net/chris_

2 Chris Bizer - semantic
http://ontoworld.org/wiki/Chris

3 Untitled document 6 f
BOSS http://www.faceboo

4 Chris Bizer - semantic
http://semanticweb.org/wiki/Ch

5 Chris Bizer - LinkedIn
BOSS http://www.linkedin.

6 Chris Bizer 10 facts | 20
http://data.semanticweb.org/pe

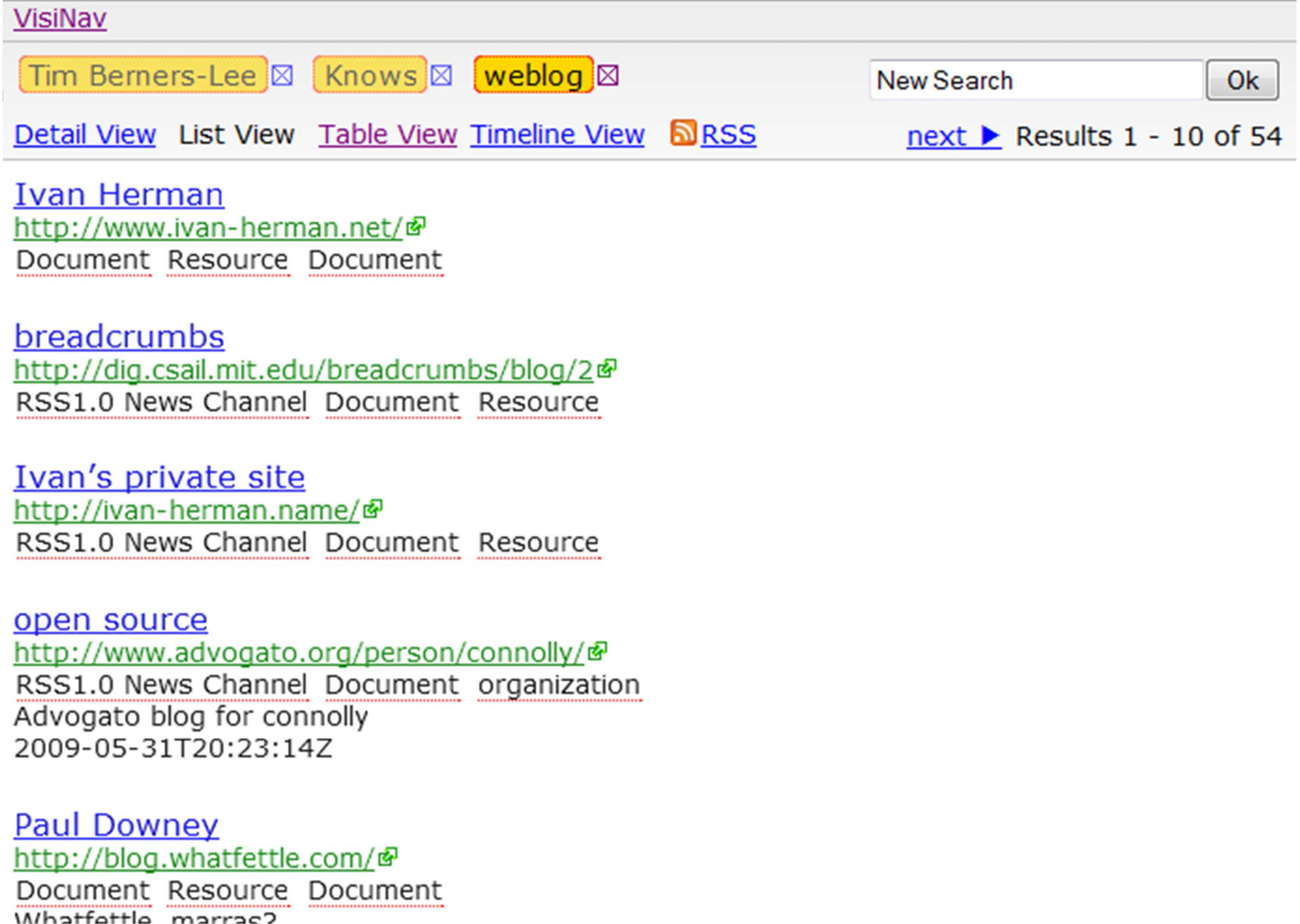
7 Chris Bizer - semantic
http://semanticweb.org/index.p

8 Flickr: Chris Bizer's Ph
BOSS http://flickr.com/pho

9 Untitled document 8 f
http://data.semanticweb.org/c

10 Chris Bizer 6 facts | 20
BOSS http://ebiquity.umbc

http://example.loc/doc



Results 1 - 10 of 54

Ivan Herman
http://www.ivan-herman.net/
Document Resource Document

breadcrumbs
http://dig.csail.mit.edu/breadcrumbs/blog/2
RSS1.0 News Channel Document Resource

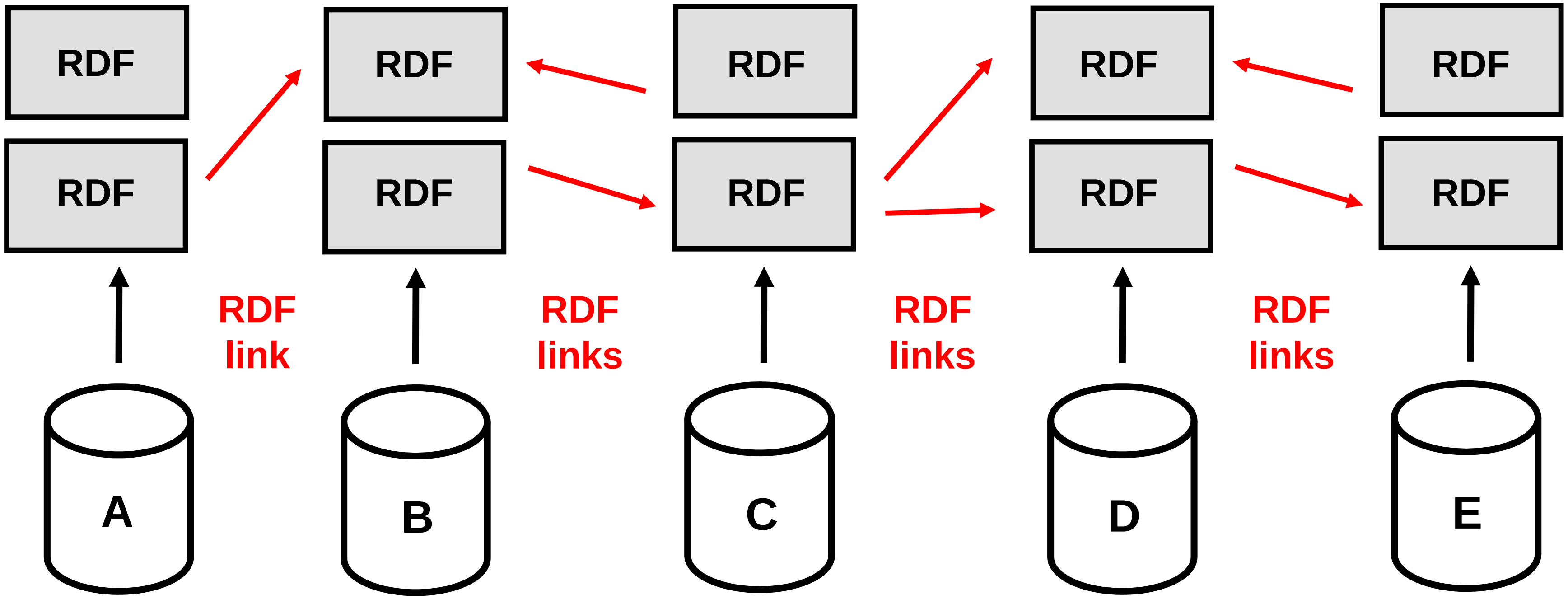
Ivan's private site
http://ivan-herman.name/
RSS1.0 News Channel Document Resource

open source
http://www.advogato.org/person/connolly/
RSS1.0 News Channel Document organization
Advogato blog for connolly
2009-05-31T20:23:14Z

Paul Downey
http://blog.whatfettle.com/
Document Resource Document
Whatfettle marras?

2. Linked Data Deployment on the Web

- Is this real?

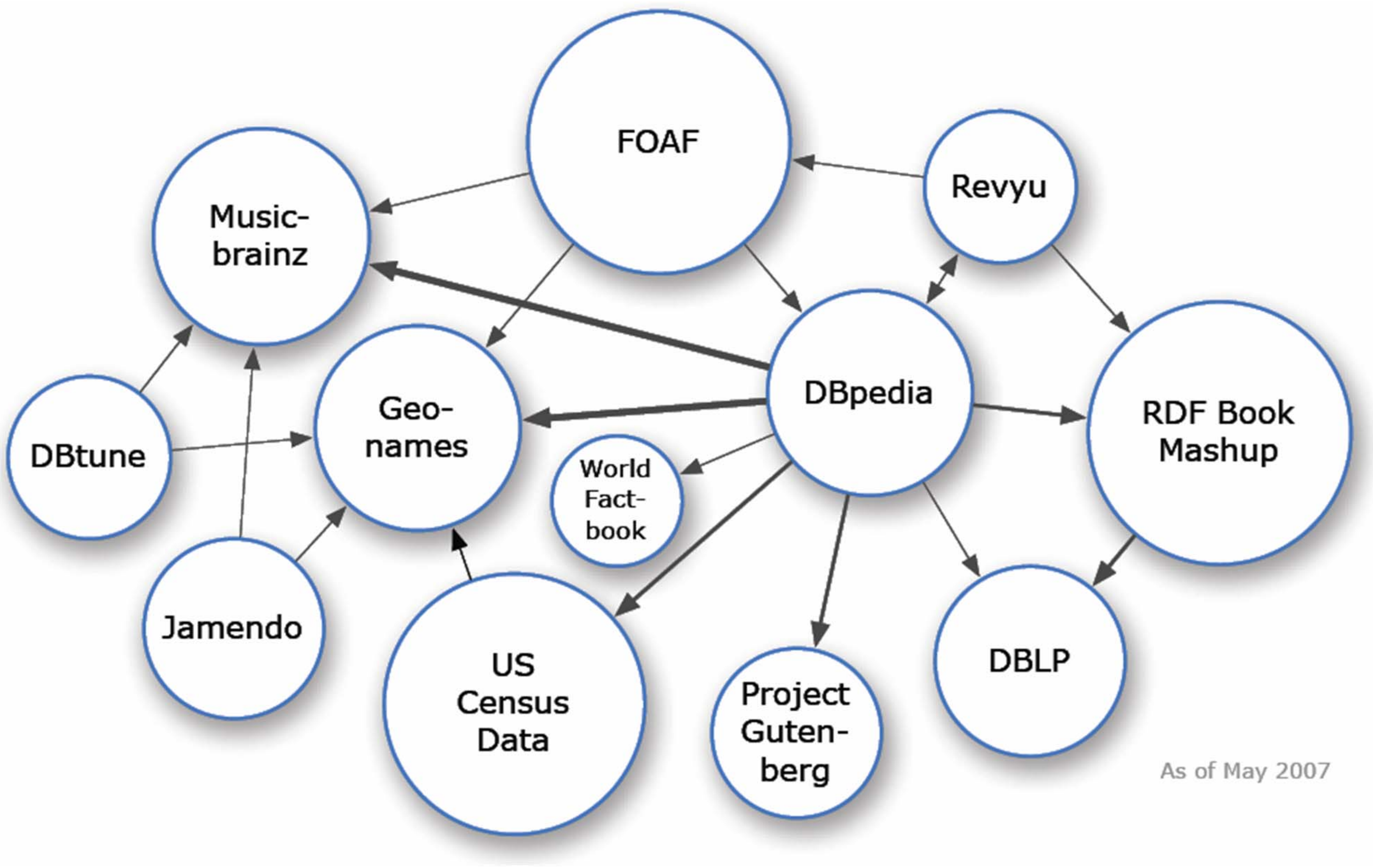


W3C Linking Open Data Project

LINKINGOPENDATA
W3C SWEO Community Project

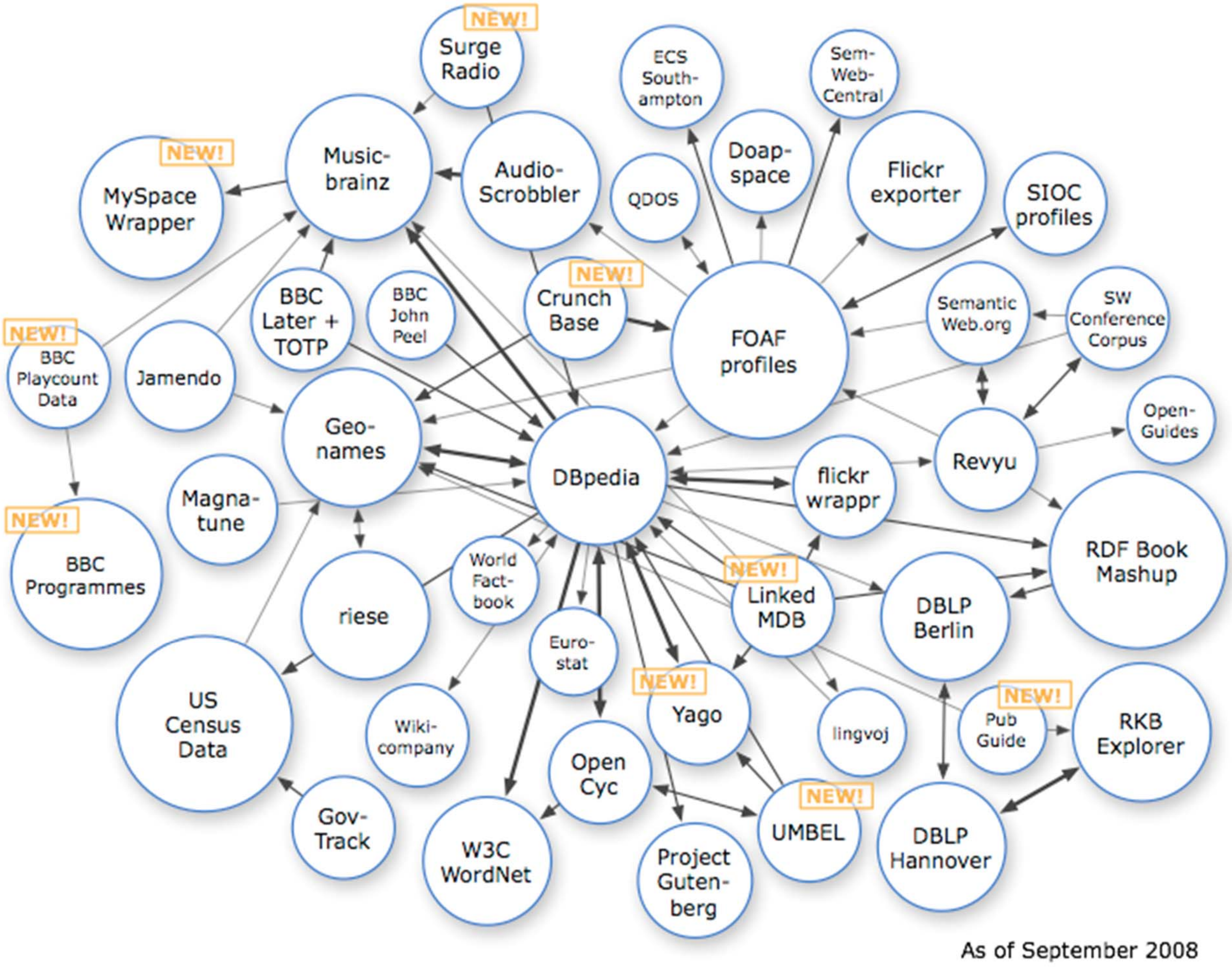
- **Grassroots community effort to**
  - publish existing open license datasets as Linked Data on the Web
  - interlink things between different data sources

LOD Datasets on the Web: May 2007

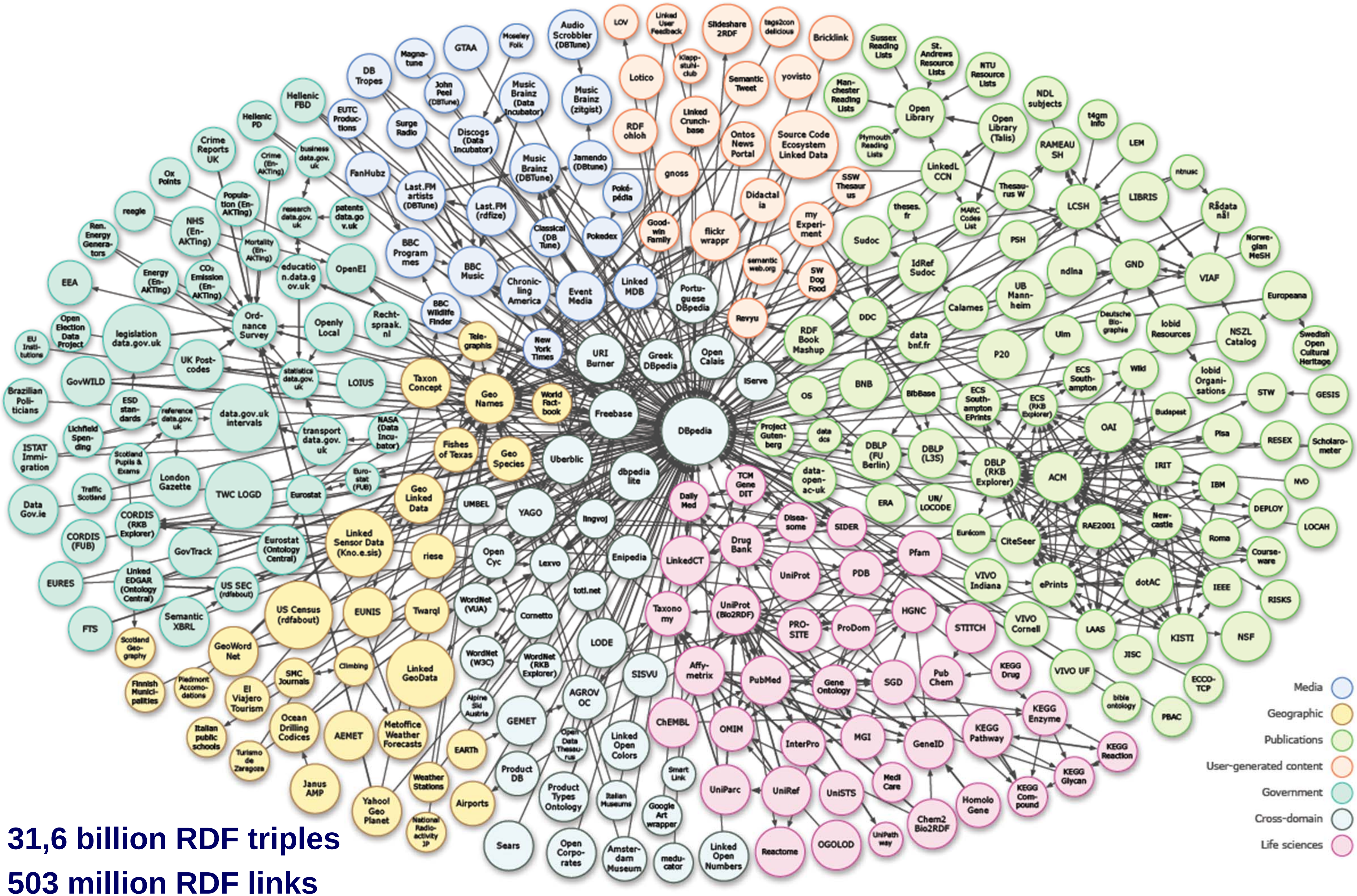


- Over 500 million RDF triples
- Around 120,000 RDF links between data sources

LOD Datasets on the Web: September 2008



LOD Datasets on the Web: November 2011



- 31,6 billion RDF triples
- 503 million RDF links

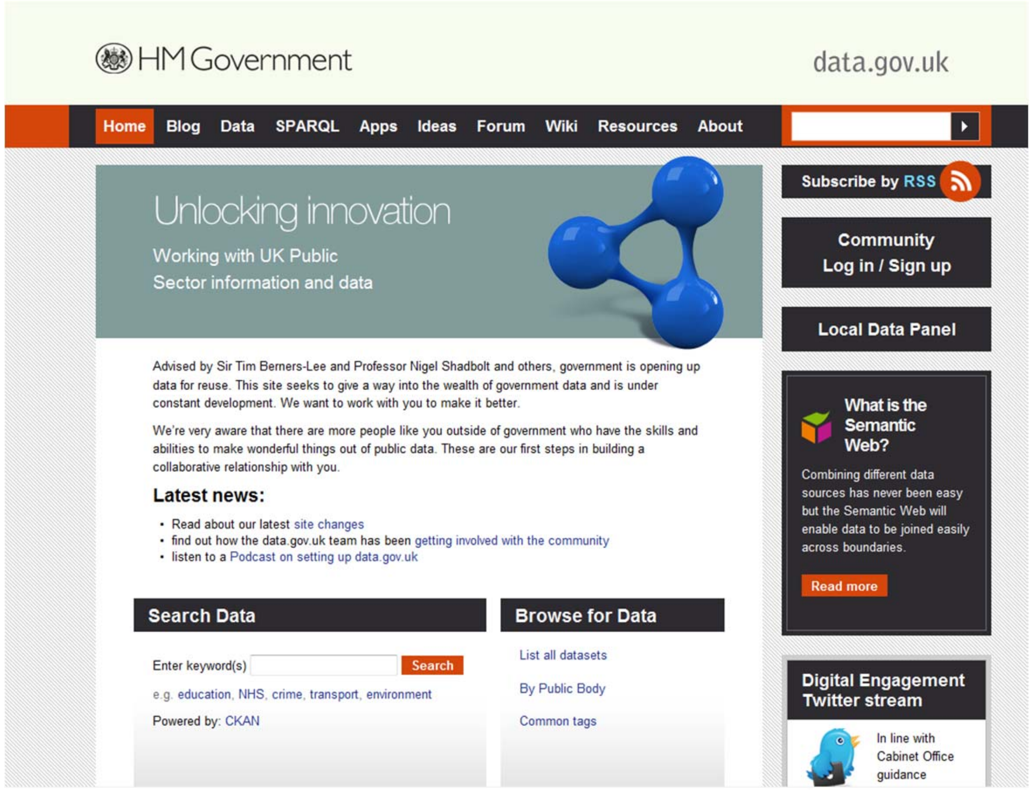
Distribution by Topical Domain (Nov 2011)

Domain	Data Sets	Triples	Percent	RDF Links	Percent
Media	25	1,841,852,061	5.82 %	50,440,705	10.01 %
Geographic	31	6,145,532,484	19.43 %	35,812,328	7.11 %
Government	49	13,315,009,400	**42.09 %**	19,343,519	3.84 %
Library	**87**	2,950,720,693	9.33 %	139,925,218	**27.76 %**
Cross-domain	41	4,184,635,715	13.23 %	63,183,065	12.54 %
Life sciences	41	3,036,336,004	9.60 %	191,844,090	**38.06 %**
User content	20	134,127,413	0.42 %	3,449,143	0.68 %
SUM	**295**	**31,634,213,770**		**503,998,829**	

More statistics

http://lod-cloud.net/state/

Uptake in the Government Domain



Unlocking innovation

Working with UK Public Sector information and data

Advised by Sir Tim Berners-Lee and Professor Nigel Shadbolt and others, government is opening up data for reuse. This site seeks to give a way into the wealth of government data and is under constant development. 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.

Latest news:

- Read about our latest site changes
- find out how the data.gov.uk team has been getting involved with the community
- listen to a Podcast on setting up data.gov.uk

Search Data

Enter keyword(s)

e.g. education, NHS, crime, transport, environment

Powered by: CKAN

Browse for Data

List all datasets

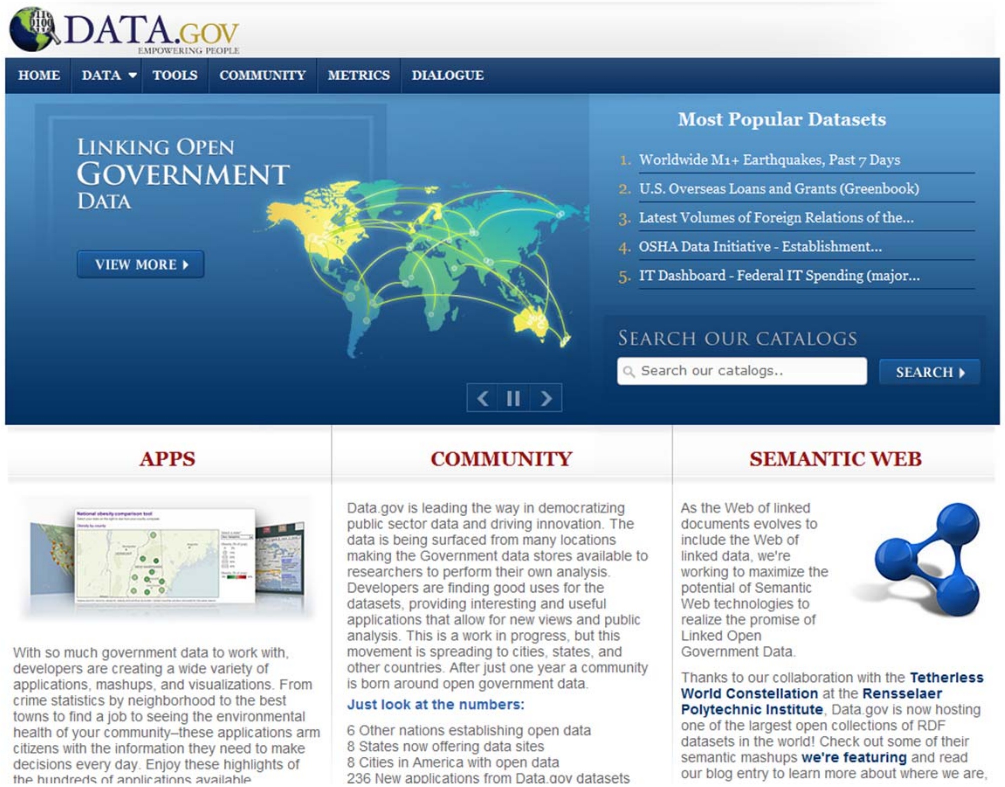
By Public Body

Common tags

What is the Semantic Web?

Combining different data sources has never been easy but the Semantic Web will enable data to be joined easily across boundaries.

In line with Cabinet Office guidance



LINKING OPEN GOVERNMENT DATA

Most Popular Datasets

1. Worldwide M1+ Earthquakes, Past 7 Days
2. U.S. Overseas Loans and Grants (Greenbook)
3. Latest Volumes of Foreign Relations of the...
4. OSHA Data Initiative - Establishment...
5. IT Dashboard - Federal IT Spending (major...

SEARCH OUR CATALOGS

Search our catalogs..

APPS

With so much government data to work with, developers are creating a wide variety of applications, mashups, and visualizations. From crime statistics by neighborhood to the best towns to find a job to seeing the environmental health of your community–these applications arm citizens with the information they need to make decisions every day. Enjoy these highlights of the hundreds of applications available

COMMUNITY

Data.gov is leading the way in democratizing public sector data and driving innovation. The data is being surfaced from many locations making the Government data stores available to researchers to perform their own analysis. Developers are finding good uses for the datasets, providing interesting and useful applications that allow for new views and public analysis. This is a work in progress, but this movement is spreading to cities, states, and other countries. After just one year a community is born around open government data.

Just look at the numbers:

6 Other nations establishing open data
8 States now offering data sites
8 Cities in America with open data
236 New applications from Data.gov datasets

SEMANTIC WEB

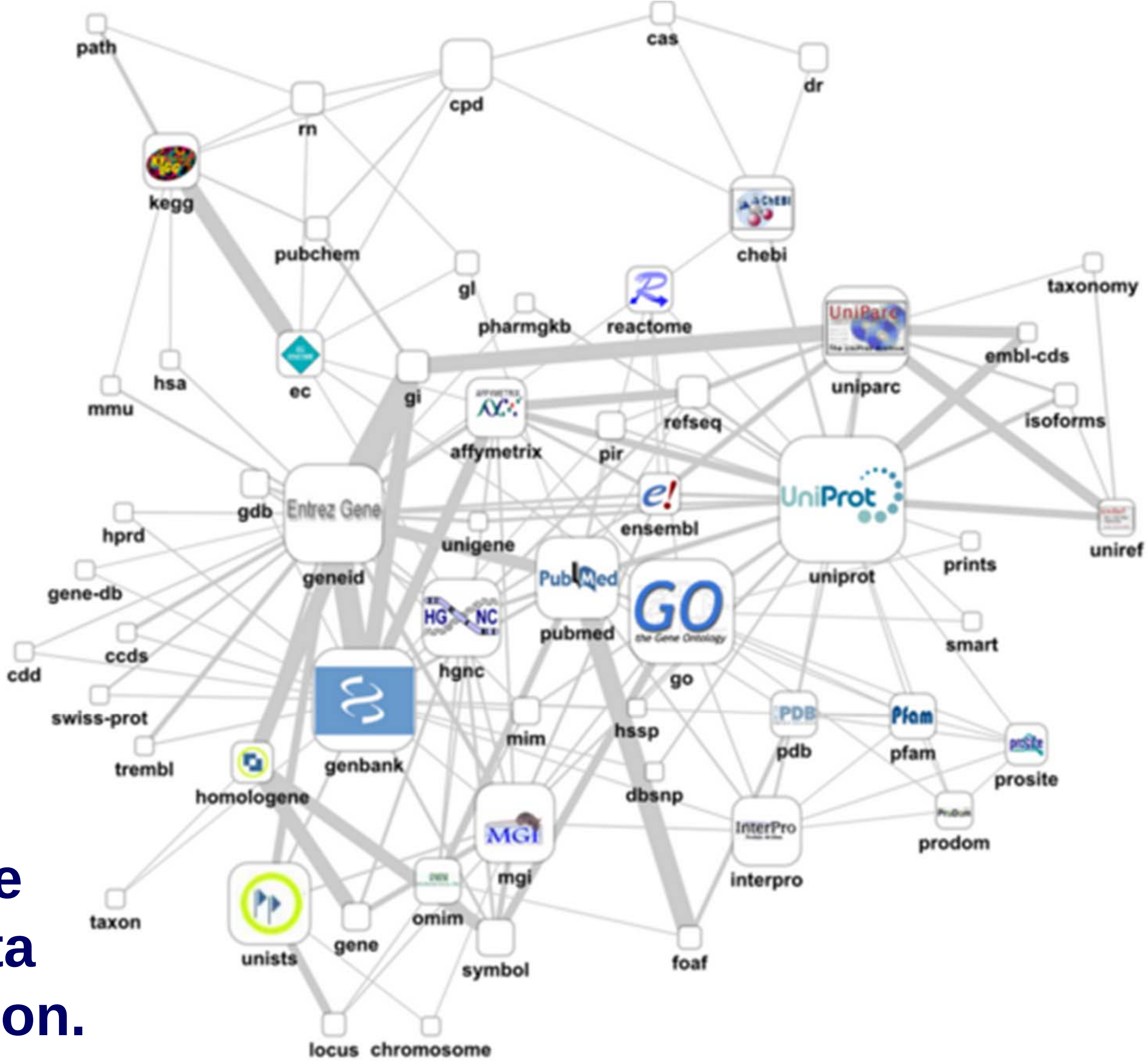
As the Web of linked documents evolves to include the Web of linked data, we're working to maximize the potential of Semantic Web technologies to realize the promise of Linked Open Government Data.

Thanks to our collaboration with the **Tetherless World Constellation** at the **Rensselaer Polytechnic Institute**, Data.gov is now hosting one of the largest open collections of RDF datasets in the world! Check out some of their semantic mashups **we're featuring** and read our blog entry to learn more about where we are,

- The EU is also starting to publish Linked Data
- Various other national efforts

Uptake in Life Sciences

- W3C Linking Open Drug Data Effort
- Bio2RDF Project



- **Goal:** Smoothly integrate internal and external data in a pay-as-you-go-fashion.

Uptake in the Libraries Community

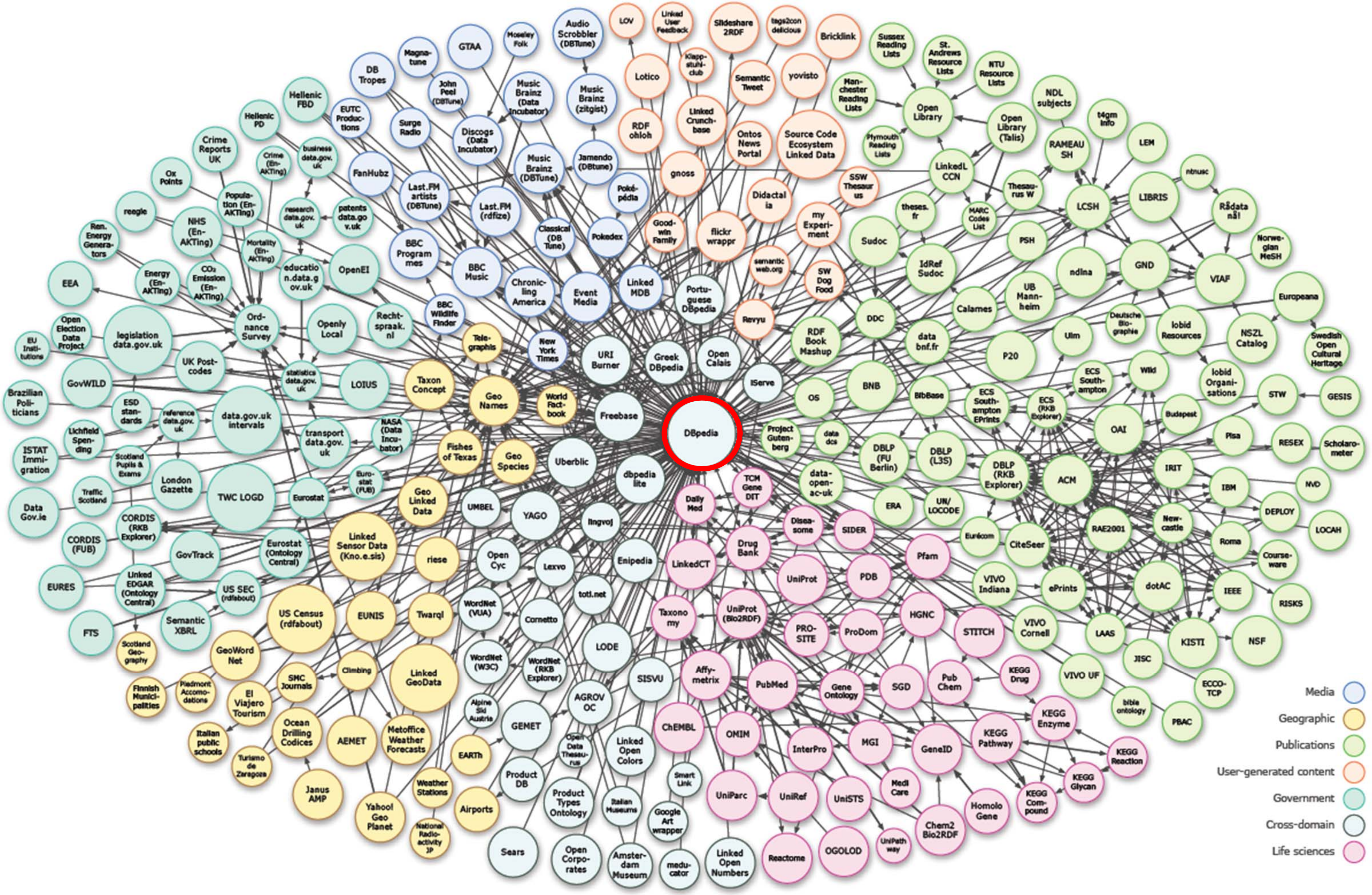
Institutions publishing Linked Data

- Library of Congress (subject headings)
- German National Library (PND dataset and subject headings)
- Swedish National Library (Libris - catalog)
- Hungarian National Library (OPAC and Digital Library)
- Europeana Digital Library just released data about 4 million artifacts

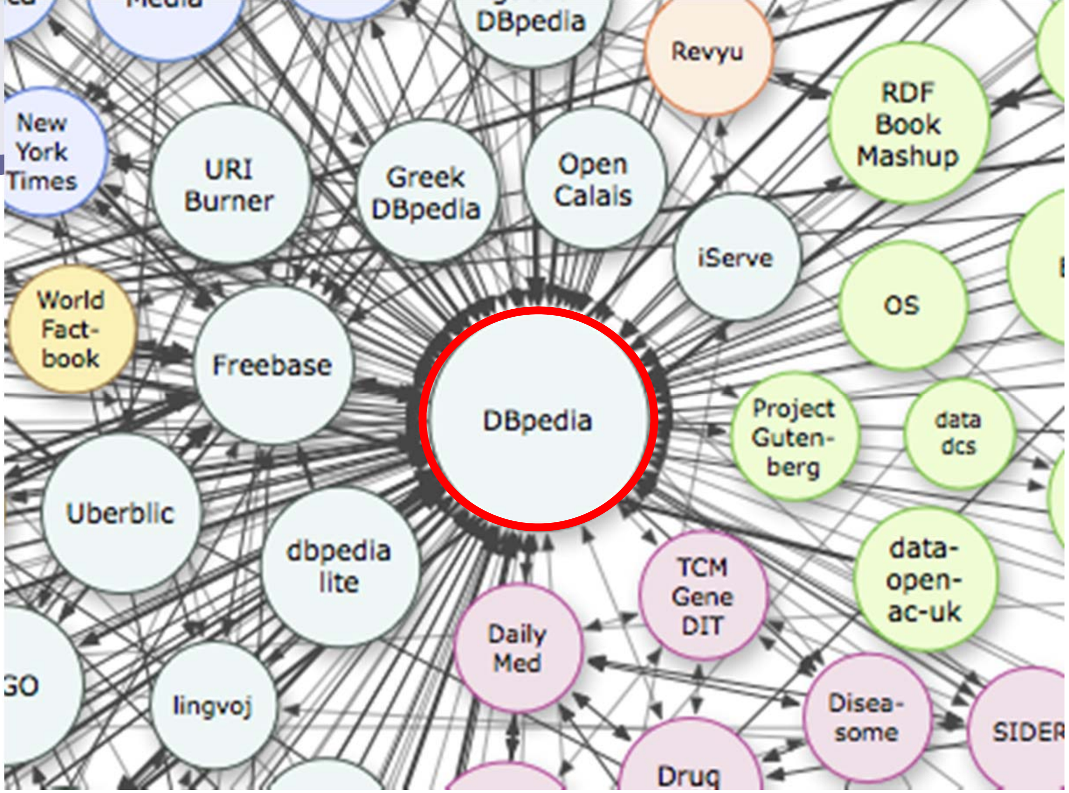
Goals:

1. Integrate Library Catalogs on global scale.
2. Interconnect resources between repositories
   (by topic, by location, by historical period, by ...).

Excursus: DBpedia



Excursus: DBpedia



- **DBpedia is a community effort**
  - **to extract structured information from Wikipedia**
  - **make this data available on the Web under an open license**
- **Contributors**
  - **University of Mannheim (Germany)**
  - **Universität Leipzig (Germany)**
  - **OpenLink Software (UK)**

UNIVERSITÄT MANNHEIM

UNIVERSITÄT LEIPZIG

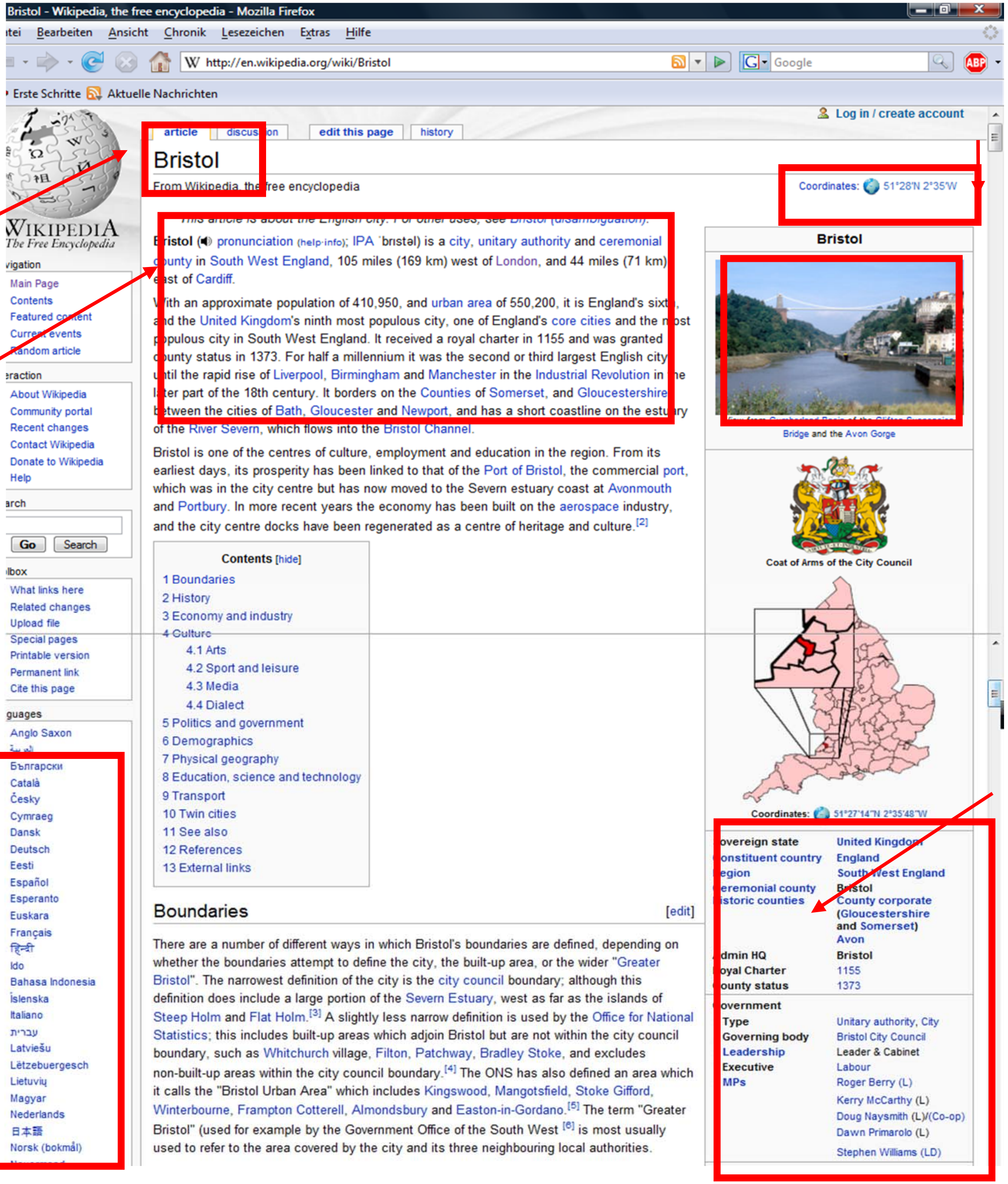
OPENLINK SOFTWARE

Structured Data within Wikipedia

Title

Description

Language Links



Bristol

From Wikipedia, the free encyclopedia

Coordinates: 51°28′N 2°35′W

This article is about the English city. For other uses, see Bristol (disambiguation).

ristol (pronunciation (help·info); IPA ˈbrɪstəl) is a city, unitary authority and ceremonial unty in South West England, 105 miles (169 km) west of London, and 44 miles (71 km) st of Cardiff.

ith an approximate population of 410,950, and urban area of 550,200, it is England's sixt , d the United Kingdom's ninth most populous city, one of England's core cities and the ost pulous city in South West England. It received a royal charter in 1155 and was granted unty status in 1373. For half a millennium it was the second or third largest English city til the rapid rise of Liverpool, Birmingham and Manchester in the Industrial Revolution in he ter part of the 18th century. It borders on the Counties of Somerset, and Gloucestershire tween the cities of Bath, Gloucester and Newport, and has a short coastline on the estu ry of the River Severn, which flows into the Bristol Channel.

Bristol is one of the centres of culture, employment and education in the region. From its earliest days, its prosperity has been linked to that of the Port of Bristol, the commercial port, which was in the city centre but has now moved to the Severn estuary coast at Avonmouth and Portbury. In more recent years the economy has been built on the aerospace industry, and the city centre docks have been regenerated as a centre of heritage and culture.[2]

Contents [hide]

1 Boundaries
2 History
3 Economy and industry
4 Culture
4.1 Arts
4.2 Sport and leisure
4.3 Media
4.4 Dialect
5 Politics and government
6 Demographics
7 Physical geography
8 Education, science and technology
9 Transport
10 Twin cities
11 See also
12 References
13 External links

Boundaries [edit]

There are a number of different ways in which Bristol's boundaries are defined, depending on whether the boundaries attempt to define the city, the built-up area, or the wider "Greater Bristol". The narrowest definition of the city is the city council boundary; although this definition does include a large portion of the Severn Estuary, west as far as the islands of Steep Holm and Flat Holm.[3] A slightly less narrow definition is used by the Office for National Statistics; this includes built-up areas which adjoin Bristol but are not within the city council boundary, such as Whitchurch village, Filton, Patchway, Bradley Stoke, and excludes non-built-up areas within the city council boundary.[4] The ONS has also defined an area which it calls the "Bristol Urban Area" which includes Kingswood, Mangotsfield, Stoke Gifford, Winterbourne, Frampton Cotterell, Almondsbury and Easton-in-Gordano.[5] The term "Greater Bristol" (used for example by the Government Office of the South West [6] is most usually used to refer to the area covered by the city and its three neighbouring local authorities.

overeign state	United Kingdom
onstituent country	England
egion	South West England
eremonial county	Bristol
istoric counties	County corporate (Gloucestershire and Somerset) Avon
dmin HQ	Bristol
oyal Charter	1155
ounty status	1373
overnment	
Type	Unitary authority, City
Governing body	Bristol City Council
Leadership	Leader & Cabinet
Executive	Labour
MPs	Roger Berry (L) Kerry McCarthy (L) Doug Naysmith (L)/(Co-op) Dawn Primarolo (L) Stephen Williams (LD)

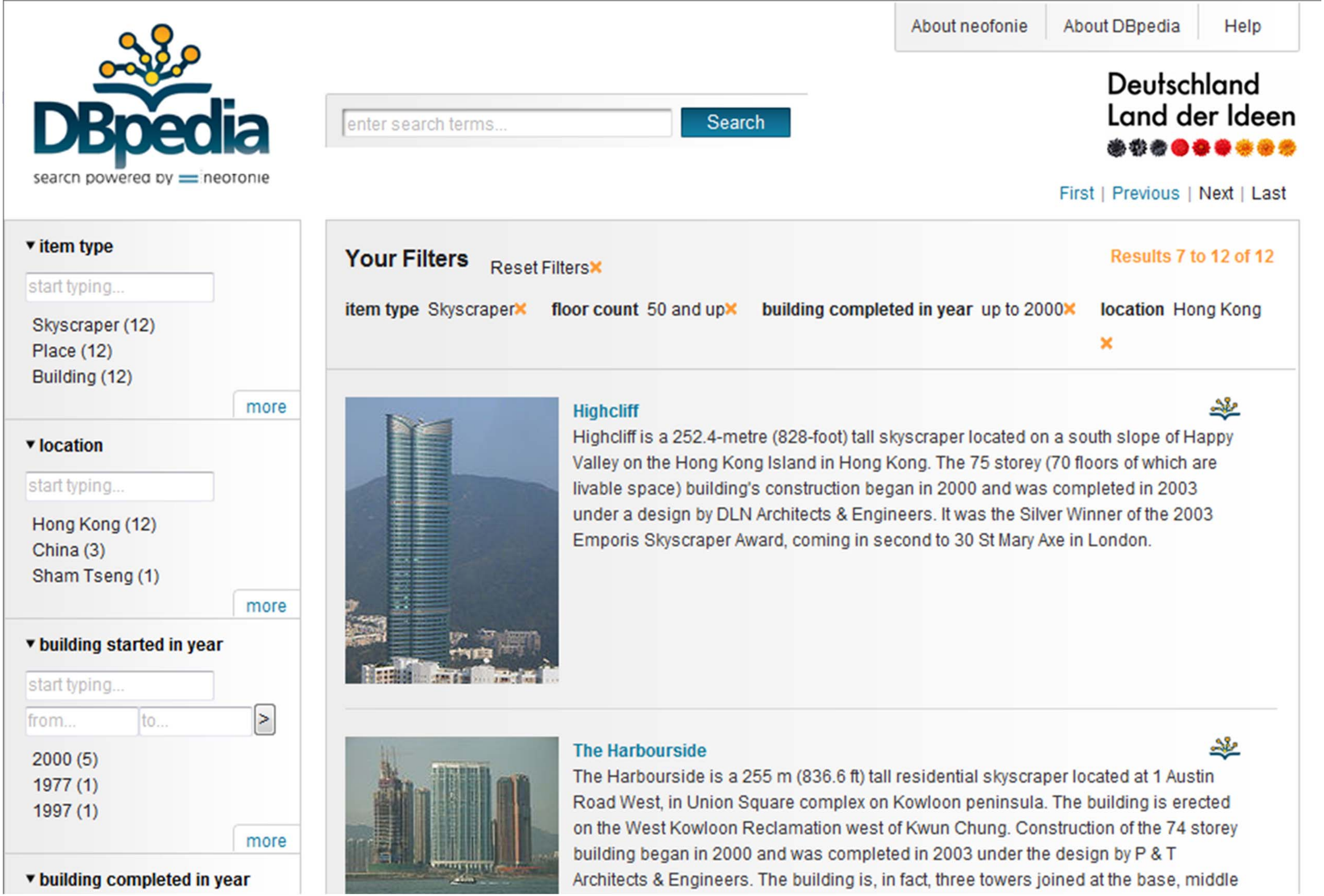
Geo-Coordinates

Images

Infoboxes

The DBpedia 3.8 Knowledge Base

- describes 3.77 million things, out of which 2.35 million are classified in a consistent ontology
  - 764,000 persons
  - 573,000 places
  - 192,000 organizations
  - 112,000 music albums
- Altogether 1.89 billion pieces of information (RDF triples)
  - 8,000,000 links to images
  - 24,000,000 links to external web pages
  - 27,200,000 external links into other RDF datasets
- DBpedia Internationalization Effort
  - provides data from 111 Wikipedia language editions for download



About neofonie
About DBpedia
Help
DBpedia
search powered by neofonie
enter search terms...
Search
Deutschland Land der Ideen
First | Previous | Next | Last

▾ item type

start typing...

Skyscraper (12)
Place (12)
Building (12)

more

▾ location

start typing...

Hong Kong (12)
China (3)
Sham Tseng (1)

more

▾ building started in year

start typing...

from... to... >

2000 (5)
1977 (1)
1997 (1)

more

▾ building completed in year

Your Filters Reset Filters✖

Results 7 to 12 of 12

item type Skyscraper✖ **floor count** 50 and up✖ **building completed in year** up to 2000✖ **location** Hong Kong ✖

Highcliff

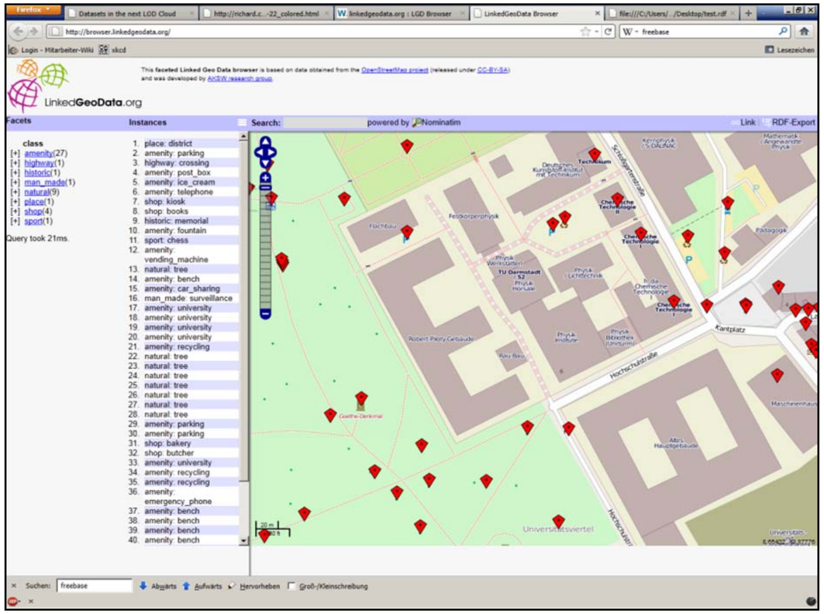
Highcliff is a 252.4-metre (828-foot) tall skyscraper located on a south slope of Happy Valley on the Hong Kong Island in Hong Kong. The 75 storey (70 floors of which are livable space) building's construction began in 2000 and was completed in 2003 under a design by DLN Architects & Engineers. It was the Silver Winner of the 2003 Emporis Skyscraper Award, coming in second to 30 St Mary Axe in London.

The Harbourside

The Harbourside is a 255 m (836.6 ft) tall residential skyscraper located at 1 Austin Road West, in Union Square complex on Kowloon peninsula. The building is erected on the West Kowloon Reclamation west of Kwun Chung. Construction of the 74 storey building began in 2000 and was completed in 2003 under the design by P & T Architects & Engineers. The building is, in fact, three towers joined at the base, middle

Other Examples of Linked Data Sets

- Linked Geo Data
  - Linked Data version of Open Street Maps
  - millions of places
- Linked Movie Database
  - data about movies, actors and directors
  - 40,000 films
- Music Brainz
  - musicians, albums
  - 22,000 albums, 40,000 musicians
- DBLP
  - computer science papers
  - 1.6 million articles



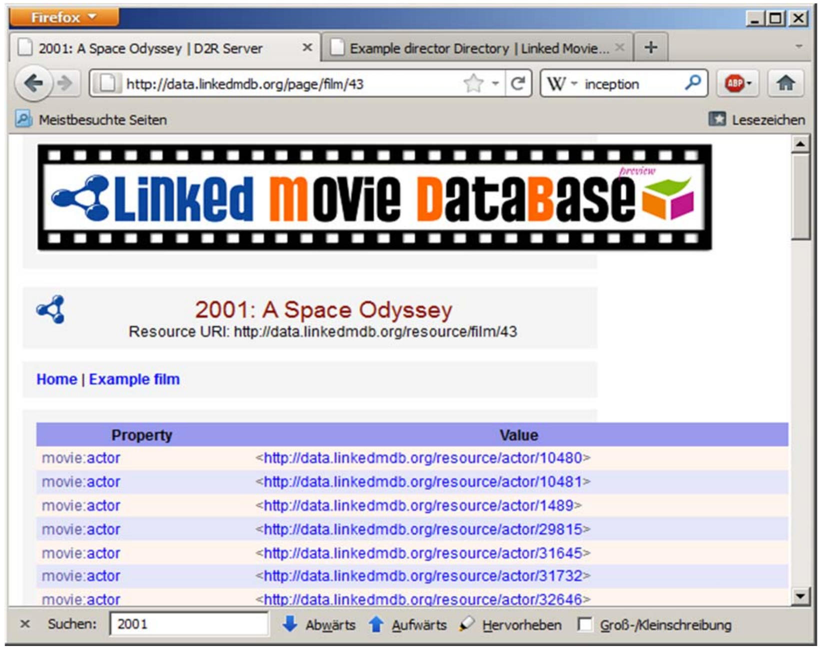
Firefox
Datasets in the next LOD Cloud
http://richard.c...-22_colored.html
linkedgeodata.org : LGD Browser
LinkedGeoData Browser
http://browser.linkedgeodata.org/
Login - Mitarbeiter-Wiki
xkcd
LinkedGeoData.org
Facets
Instances
Search:
powered by Nominatim
Link
RDF-Export

This faceted Linked Geo Data browser is based on data obtained from the OpenStreetMap project (released under CC-BY-SA) and was developed by AKSW research group.

class
[+] amenity(27)
[+] highway(1)
[+] historic(1)
[+] man_made(1)
[+] natural(9)
[+] place(1)
[+] shop(4)
[+] sport(1)

Query took 21ms.

1. place: district
2. amenity: parking
3. highway: crossing
4. amenity: post_box
5. amenity: ice_cream
6. amenity: telephone
7. shop: kiosk
8. shop: books
9. historic: memorial
10. amenity: fountain
11. sport: chess
12. amenity: vending_machine
13. natural: tree
14. amenity: bench
15. amenity: car_sharing
16. man_made: surveillance
17. amenity: university
18. amenity: university
19. amenity: university
20. amenity: university
21. amenity: recycling
22. natural: tree
23. natural: tree
24. natural: tree
25. natural: tree
26. natural: tree
27. natural: tree
28. natural: tree
29. amenity: parking
30. amenity: parking
31. shop: bakery
32. shop: butcher
33. amenity: university
34. amenity: recycling
35. amenity: recycling
36. amenity: emergency_phone
37. amenity: bench
38. amenity: bench
39. amenity: bench
40. amenity: bench



Firefox
2001: A Space Odyssey | D2R Server
Example director Directory | Linked Movie...
http://data.linkedmdb.org/page/film/43
inception
Meistbesuchte Seiten
Lesezeichen
Linked Movie DataBase preview

2001: A Space Odyssey
Resource URI: http://data.linkedmdb.org/resource/film/43

Home | Example film

Property	Value
movie:actor	<http://data.linkedmdb.org/resource/actor/10480>
movie:actor	<http://data.linkedmdb.org/resource/actor/10481>
movie:actor	<http://data.linkedmdb.org/resource/actor/1489>
movie:actor	<http://data.linkedmdb.org/resource/actor/29815>
movie:actor	<http://data.linkedmdb.org/resource/actor/31645>
movie:actor	<http://data.linkedmdb.org/resource/actor/31732>
movie:actor	<http://data.linkedmdb.org/resource/actor/32646>

Questions so far?

3. How to Publish and Consume Linked Data?

Tasks involved in Publishing Linked Data:

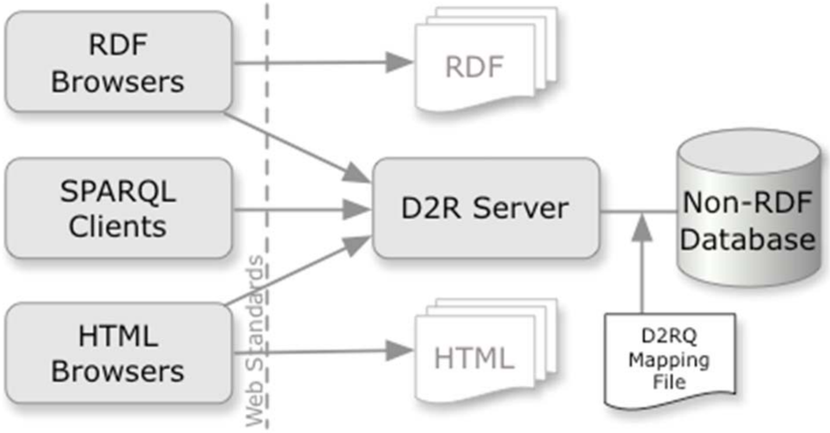
1. **Make data available as RDF via HTTP**
2. **Set RDF links pointing at other data sources**
3. **Make your data self-descriptive**

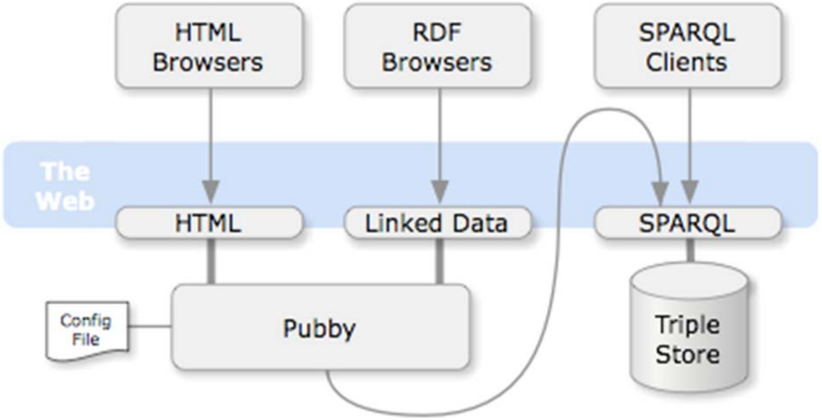
- **Tom Heath and Christian Bizer:**
  Linked Data: Evolving the Web into a Global Data Space
  http://linkeddatabook.com/

3.1 Make Data available as RDF via HTTP

Ready to use tools (examples)

1. **D2R Server**
   - provides for mapping relational databases into RDF and for serving them as Linked Data
2. **Pubby**
   - Linked Data Frontend for SPARQL Endpoints
3. **More tools**
   - http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/PublishingTools





3.2 Set RDF links pointing at other data sources

- Examples of RDF links

```
<http://dbpedia.org/resource/Berlin> owl:sameAs
<http://sws.geonames.org/2950159> .
```

```
<http://example-bookshop.com/book006251587X> owl:sameAs
<http://www4.wiwiss.fu-berlin.de/bookmashup/books/006251587X> .
```

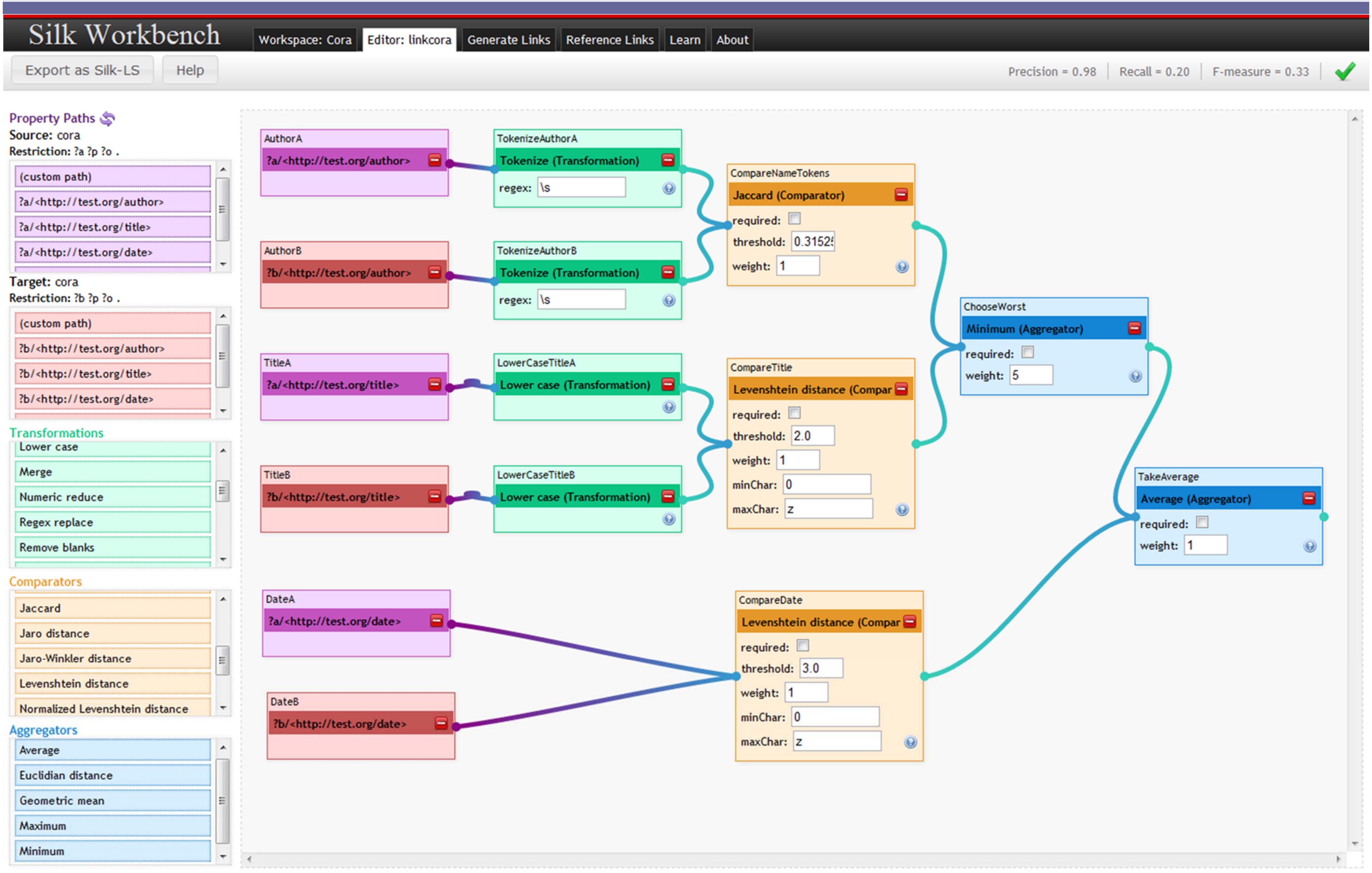
How to generate RDF links?

1. **Pattern-based Approaches**
   - Exploit naming conventions within URIs (for instance ISBNs, Gen IDs, …)
2. **Similarity-based Approaches**
   - Compare items within different data sources using various similarity metrics

Link Generation Tools

- **Silk – Link Discovery Framework**
  - provides a user interface for specifying link conditions which may combine different similarity metrics
- **More tools**
  - http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/EquivalenceMining

A Silk Linkage Rule



Precision = 0.98 | Recall = 0.20 | F-measure = 0.33 ✓

Property Paths

Source: cora

Restriction: ?a ?p ?o .

(custom path)
?a/<http://test.org/author>
?a/<http://test.org/title>
?a/<http://test.org/date>

Target: cora

Restriction: ?b ?p ?o .

AuthorA: ?a/<http://test.org/author>

AuthorB: ?b/<http://test.org/author>

TitleA: ?a/<http://test.org/title>

TitleB: ?b/<http://test.org/title>

DateA: ?a/<http://test.org/date>

DateB: ?b/<http://test.org/date>

TokenizeAuthorA: Tokenize (Transformation) — regex: \s

TokenizeAuthorB: Tokenize (Transformation) — regex: \s

LowerCaseTitleA: Lower case (Transformation)

LowerCaseTitleB: Lower case (Transformation)

CompareNameTokens: Jaccard (Comparator) — required: □ threshold: 0.3152! weight: 1

CompareTitle: Levenshtein distance (Compar — required: □ threshold: 2.0 weight: 1 minChar: 0 maxChar: z

CompareDate: Levenshtein distance (Compar — required: □ threshold: 3.0 weight: 1 minChar: 0 maxChar: z

ChooseWorst: Minimum (Aggregator) — required: □ weight: 5

TakeAverage: Average (Aggregator) — required: □ weight: 1

3.3 Make your Data Self-Descriptive

- **Increase the usefulness of your data and ease data integration**
- **Aspects of self-descriptiveness**
  1. Enable clients to retrieve the schema
  2. Reuse terms from common vocabularies / ontologies
  3. Publish schema mappings for proprietary terms
  4. Provide provenance metadata
  5. Provide licensing metadata

- **Statistics about the compliance with these best practices**
  - http://lod-cloud.net/state/

Enable Clients to retrieve the Schema

Clients can resolve the URIs that identify vocabulary terms in order to get their RDFS or OWL definitions.

Some data on the Web

```
<http://richard.cyganiak.de/foaf.rdf#cygri>
      foaf:name "Richard Cyganiak" ;
      rdf:type <http://xmlns.com/foaf/0.1/Person> .
```

Resolve unknown term

`http://xmlns.com/foaf/0.1/Person`

RDFS or OWL definition

```
<http://xmlns.com/foaf/0.1/Person>
           rdf:type owl:Class ;
           rdfs:label "Person";
           rdfs:subClassOf <http://xmlns.com/foaf/0.1/Agent> ;
           rdfs:subClassOf <http://xmlns.com/wordnet/1.6/Agent> .
```

Reuse Terms from Common Vocabularies

Common Vocabularies

- **Friend-of-a-Friend** for describing people and their social network
- **SIOC** for describing forums and blogs
- **SKOS** for representing topic taxonomies
- **Organization Ontology** for describing the structure of organizations
- **GoodRelations** provides terms for describing products and business entities
- **Music Ontology** for describing artists, albums, and performances
- **Review Vocabulary** provides terms for representing reviews

Common sources of identifiers (URIs) for real world objects

- **LinkedGeoData** and **Geonames** locations
- **GeneID** and **UniProt** life science identifiers
- **DBpedia** wide range of things

Usage of Common Vocabularies in the LOD Cloud

- Some terms from non-proprietary vocabularies:
  191 (64.75 %) of the 295 sources
- Only proprietary vocabularies:
  104 (35.25 %) of the 295 sources
- Common Vocabularies

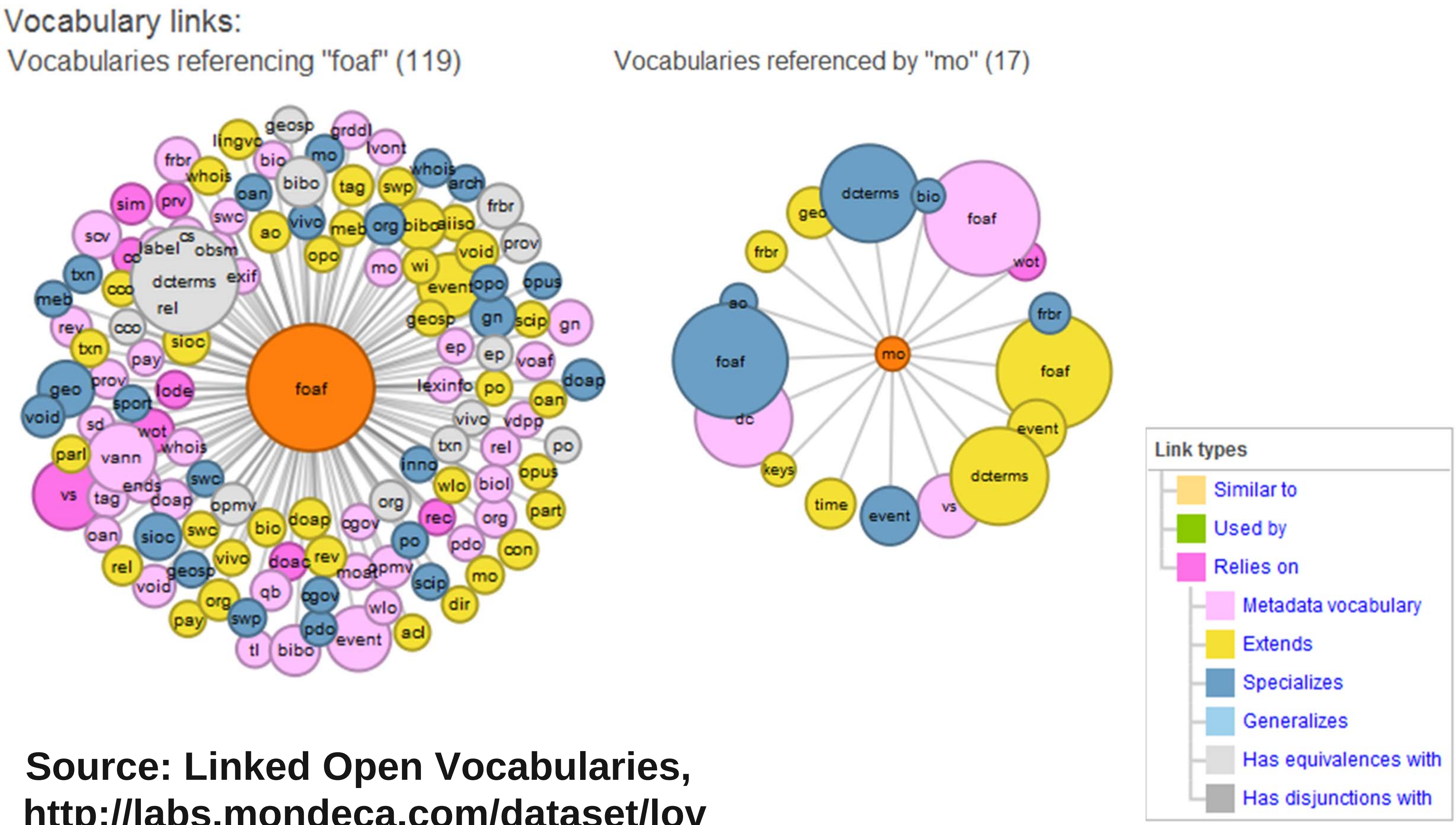
dc	92 (31.19 %)
foaf	81 (27.46 %)
skos	58 (19.66 %)
geo	25 (8.47 %)
akt	17 (5.76 %)
bibo	14 (4.75 %)
mo	13 (4.41 %)
vcard	10 (3.39 %)
sioc	10 (3.39 %)
cc	8 (2.71 %)

Publish Schema Mappings on the Web

```
<http://dbpedia.org/ontology/Person>
owl:equivalentClass
<http://xmlns.com/foaf/0.1/Person> .
```

- **Terms for representing correspondences**
  - owl:equivalentClass, owl:equivalentProperty,
  - rdfs:subClassOf, rdfs:subPropertyOf
  - skos:broadMatch, skos:narrowMatch

Deployment of Vocabulary Links



Vocabulary links:

Vocabularies referencing "foaf" (119)

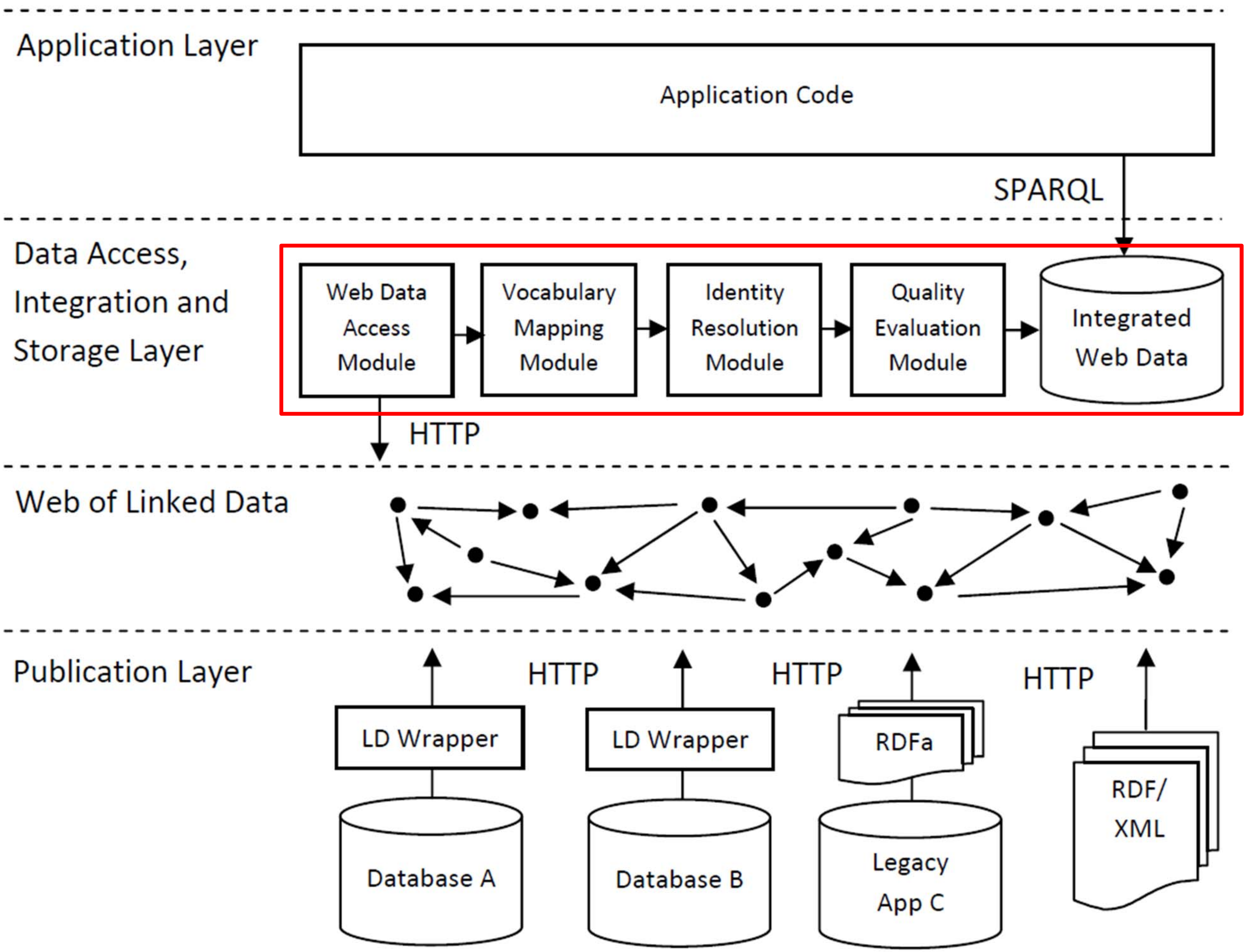
Vocabularies referenced by "mo" (17)

Link types

- 🟧 Similar to
- 🟩 Used by
- 🟪 Relies on
  - 🟪 Metadata vocabulary
  - 🟨 Extends
  - 🟦 Specializes
  - 🟦 Generalizes
  - ⬜ Has equivalences with
  - ⬛ Has disjunctions with

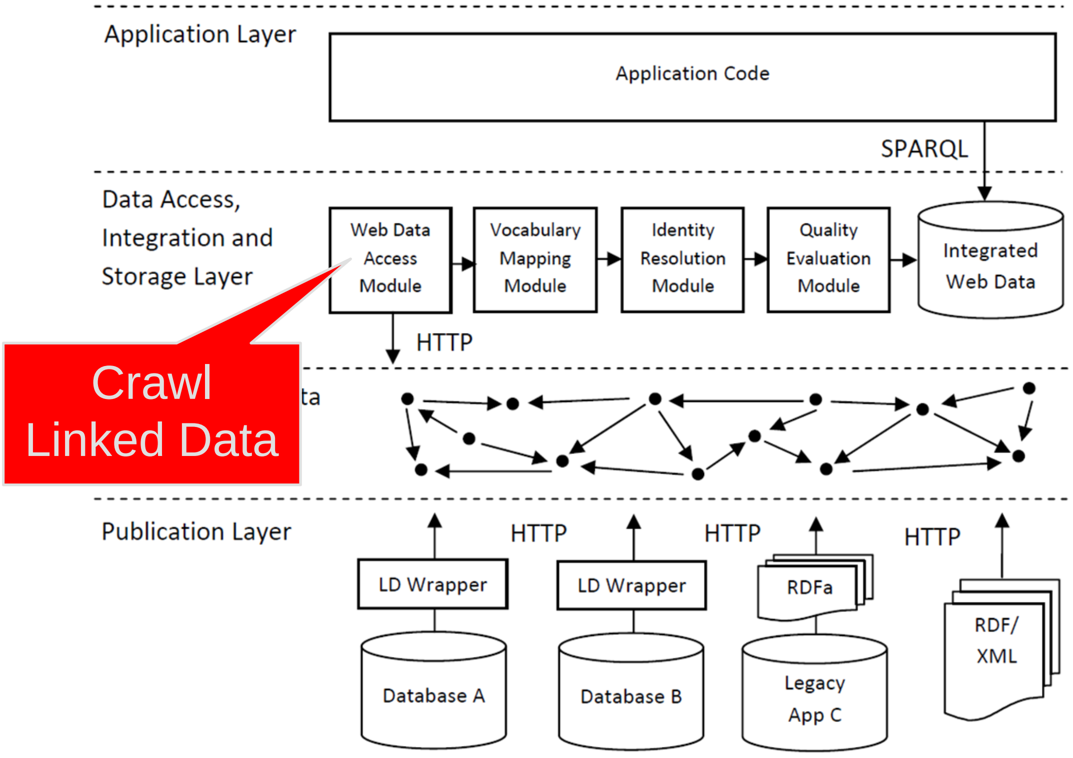
Source: Linked Open Vocabularies, http://labs.mondeca.com/dataset/lov

3.3 Tasks involved in Consuming Linked Data



LDspider

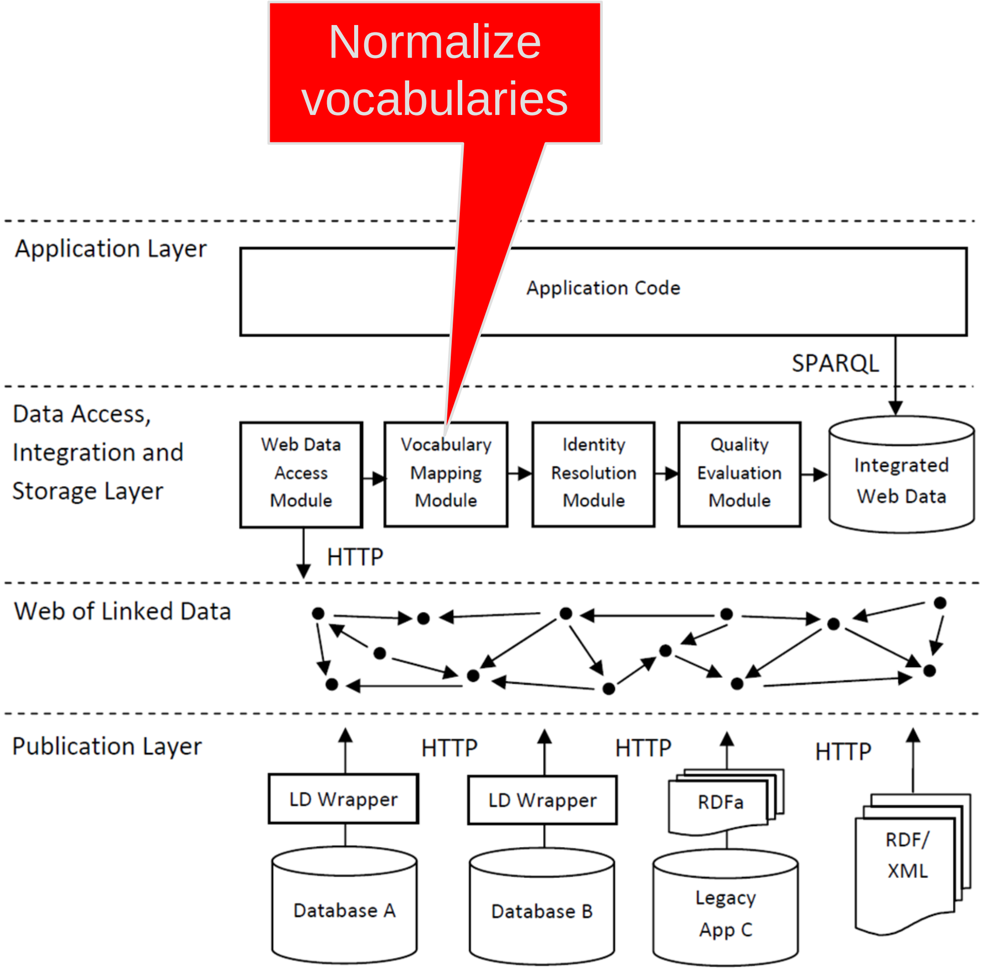
- Flexible open-source Linked Data crawler
- Crawls RDF/XML and RDFa
- https://code.google.com/p/ldspider/



Crawl Linked Data

R2R Framework

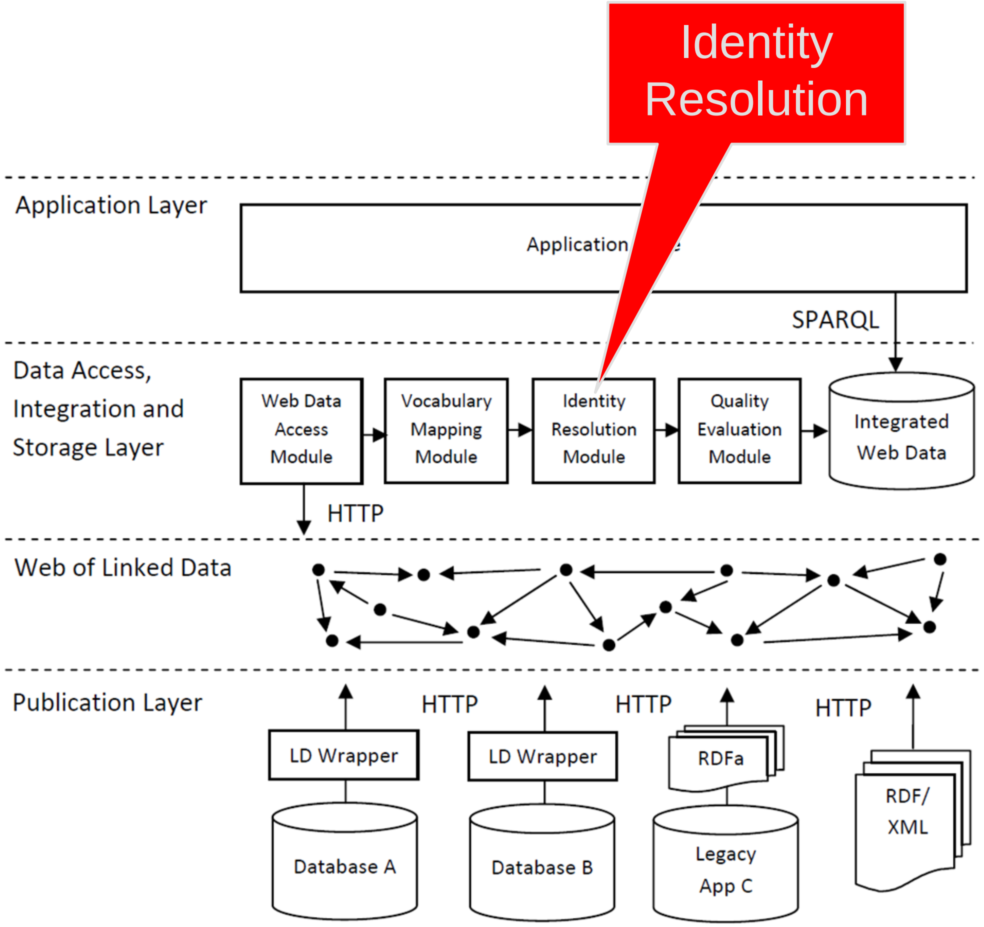
- Tool for translating RDF data between different vocabularies
- http://wifo5-03.informatik.uni-mannheim.de/bizer/r2r/
- Alternative:
  Use SPARQL Construct
  queries to translate data



Normalize vocabularies

Silk Server

- Add missing links while consuming Linked Data
- Designed to work together with LDspider



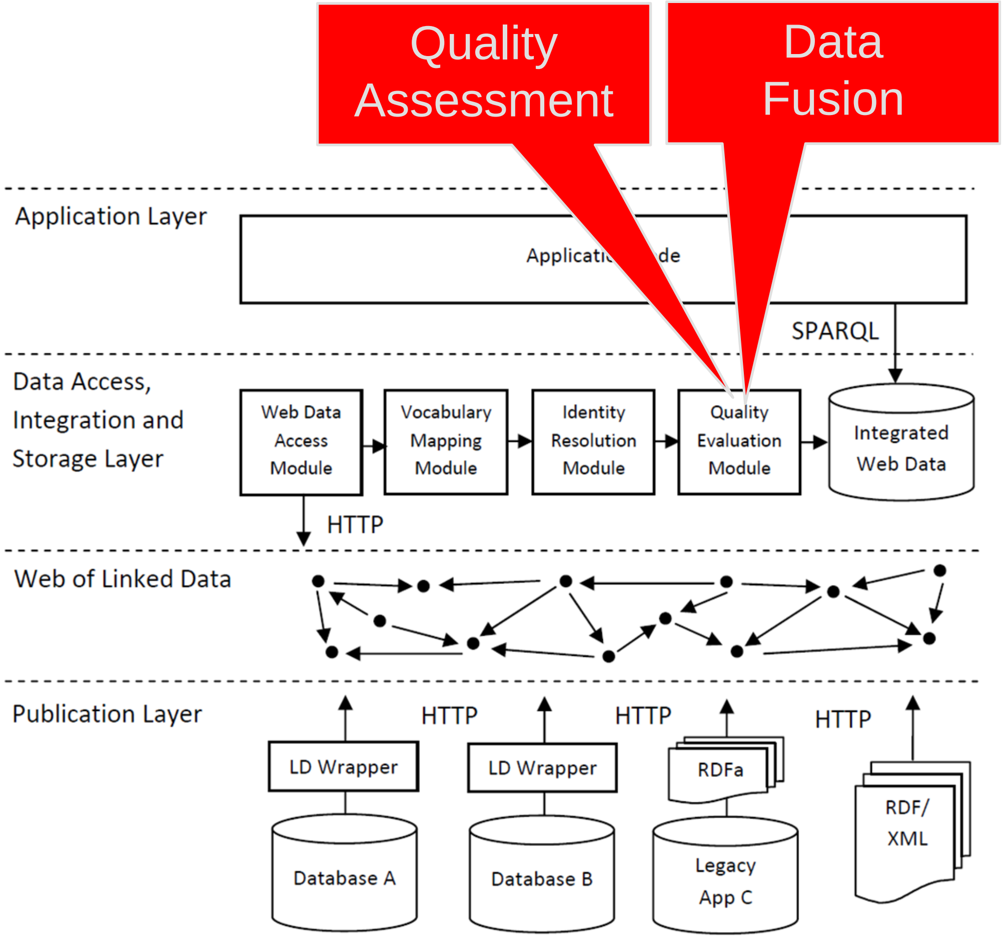
Sieve Framework and WIQA Browser

Sieve Framework

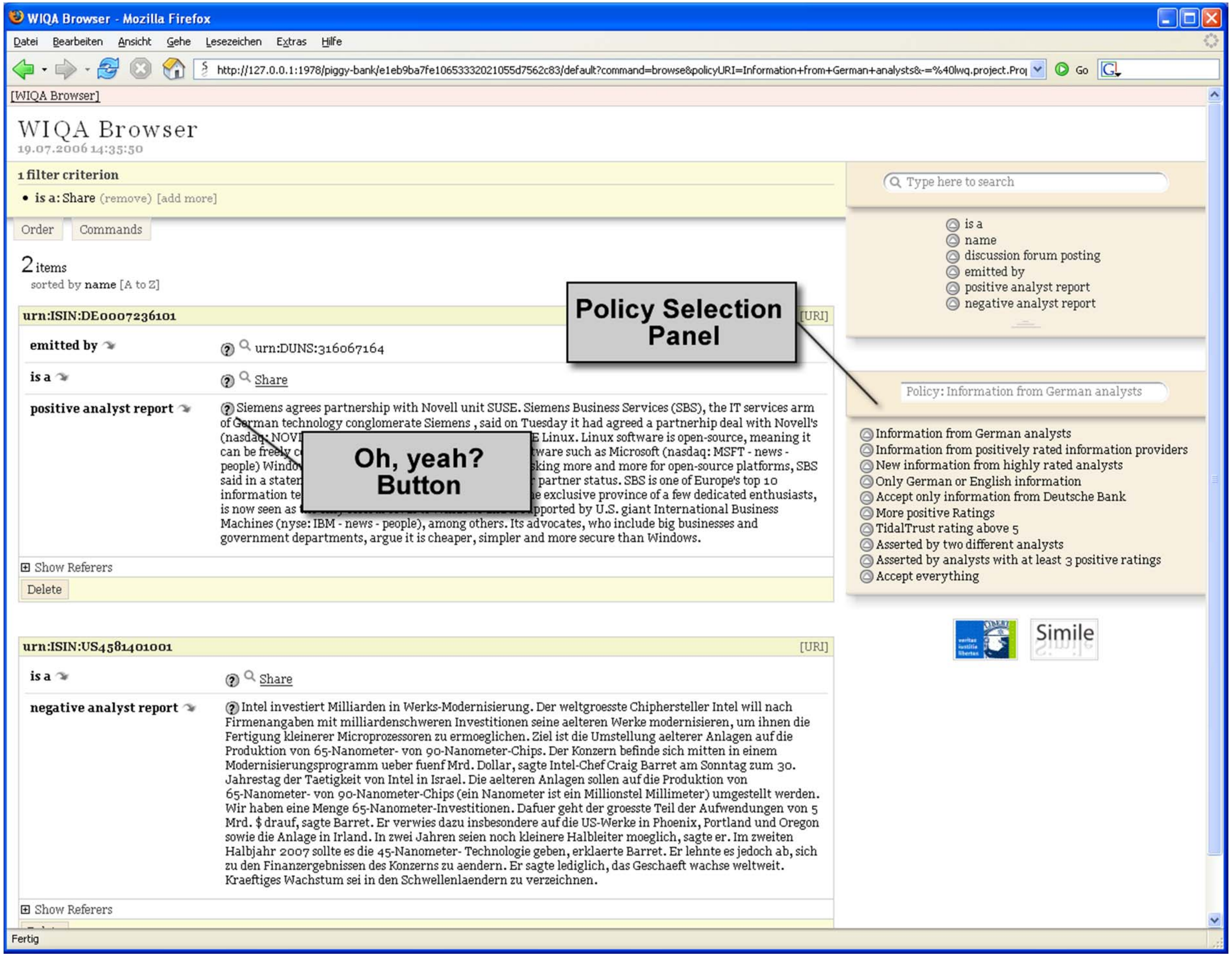
- Allows you to filter Web data using different data quality assessment policies
- Allows you to fuse data from different sources
- http://sieve.wbsg.de/

WIQA Browser

- Enables you to interactively employ different quality assessment policies
- Produces explanations about filtering decisions
- http://wifo5-03.informatik.uni-mannheim.de/bizer/wiqa/browser/



The WIQA Browser



WIQA Browser - Mozilla Firefox
Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe
http://127.0.0.1:1978/piggy-bank/e1eb9ba7fe10653332021055d7562c83/default?command=browse&policyURI=Information+from+German+analysts&-=%40lwq.project.Pro
Go
[WIQA Browser]

WIQA Browser
19.07.2006 14:35:50

1 filter criterion

- is a: Share (remove) [add more]

2 items
sorted by name [A to Z]

urn:ISIN:DE0007236101 [URI]

emitted by	urn:DUNS:316067164
is a	Share
positive analyst report	Siemens agrees partnership with Novell unit SUSE. Siemens Business Services (SBS), the IT services arm of German technology conglomerate Siemens , said on Tuesday it had agreed a partnerhip deal with Novell's (nasdaq: NOVI … E Linux. Linux software is open-source, meaning it can be freely c … tware such as Microsoft (nasdaq: MSFT - news - people) Windo … sking more and more for open-source platforms, SBS said in a staten … r partner status. SBS is one of Europe's top 10 information te … e exclusive province of a few dedicated enthusiasts, is now seen as t … pported by U.S. giant International Business Machines (nyse: IBM - news - people), among others. Its advocates, who include big businesses and government departments, argue it is cheaper, simpler and more secure than Windows.

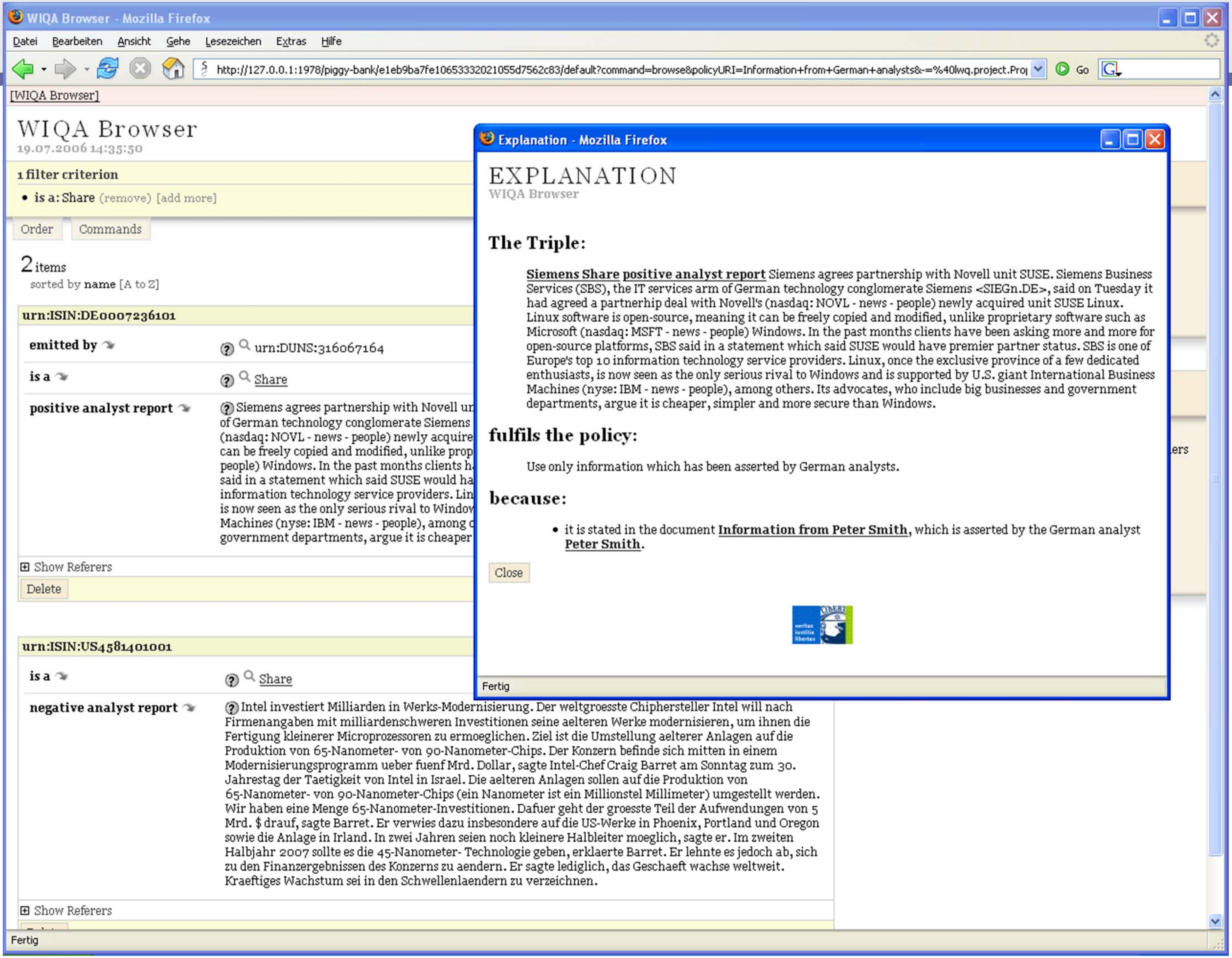
urn:ISIN:US4581401001 [URI]

is a	Share
negative analyst report	Intel investiert Milliarden in Werks-Modernisierung. Der weltgroesste Chiphersteller Intel will nach Firmenangaben mit milliardenschweren Investitionen seine aelteren Werke modernisieren, um ihnen die Fertigung kleinerer Microprozessoren zu ermoeglichen. Ziel ist die Umstellung aelterer Anlagen auf die Produktion von 65-Nanometer- von 90-Nanometer-Chips. Der Konzern befinde sich mitten in einem Modernisierungsprogramm ueber fuenf Mrd. Dollar, sagte Intel-Chef Craig Barret am Sonntag zum 30. Jahrestag der Taetigkeit von Intel in Israel. Die aelteren Anlagen sollen auf die Produktion von 65-Nanometer- von 90-Nanometer-Chips (ein Nanometer ist ein Millionstel Millimeter) umgestellt werden. Wir haben eine Menge 65-Nanometer-Investitionen. Dafuer geht der groesste Teil der Aufwendungen von 5 Mrd. $ drauf, sagte Barret. Er verwies dazu insbesondere auf die US-Werke in Phoenix, Portland und Oregon sowie die Anlage in Irland. In zwei Jahren seien noch kleinere Halbleiter moeglich, sagte er. Im zweiten Halbjahr 2007 sollte es die 45-Nanometer- Technologie geben, erklaerte Barret. Er lehnte es jedoch ab, sich zu den Finanzergebnissen des Konzerns zu aendern. Er sagte lediglich, das Geschaeft wachse weltweit. Kraeftiges Wachstum sei in den Schwellenlaendern zu verzeichnen.

- is a
- name
- discussion forum posting
- emitted by
- positive analyst report
- negative analyst report

Policy: Information from German analysts

- Information from German analysts
- Information from positively rated information providers
- New information from highly rated analysts
- Only German or English information
- Accept only information from Deutsche Bank
- More positive Ratings
- TidalTrust rating above 5
- Asserted by two different analysts
- Asserted by analysts with at least 3 positive ratings
- Accept everything


WIQA Browser - Mozilla Firefox
Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe
http://127.0.0.1:1978/piggy-bank/e1eb9ba7fe10653332021055d7562c83/default?command=browse&policyURI=Information+from+German+analysts&-=%40lwq.project.Pro|
Go
[WIQA Browser]

WIQA Browser
19.07.2006 14:35:50

1 filter criterion

- is a: Share (remove) [add more]

2 items
sorted by name [A to Z]

urn:ISIN:DE0007236101

emitted by urn:DUNS:316067164

is a Share

positive analyst report Siemens agrees partnership with Novell ur of German technology conglomerate Siemens (nasdaq: NOVL - news - people) newly acquire can be freely copied and modified, unlike prop people) Windows. In the past months clients h said in a statement which said SUSE would ha information technology service providers. Lin is now seen as the only serious rival to Windov Machines (nyse: IBM - news - people), among c government departments, argue it is cheaper

urn:ISIN:US4581401001

is a Share

negative analyst report Intel investiert Milliarden in Werks-Modernisierung. Der weltgroesste Chiphersteller Intel will nach Firmenangaben mit milliardenschweren Investitionen seine aelteren Werke modernisieren, um ihnen die Fertigung kleinerer Microprozessoren zu ermoeglichen. Ziel ist die Umstellung aelterer Anlagen auf die Produktion von 65-Nanometer- von 90-Nanometer-Chips. Der Konzern befinde sich mitten in einem Modernisierungsprogramm ueber fuenf Mrd. Dollar, sagte Intel-Chef Craig Barret am Sonntag zum 30. Jahrestag der Taetigkeit von Intel in Israel. Die aelteren Anlagen sollen auf die Produktion von 65-Nanometer- von 90-Nanometer-Chips (ein Nanometer ist ein Millionstel Millimeter) umgestellt werden. Wir haben eine Menge 65-Nanometer-Investitionen. Dafuer geht der groesste Teil der Aufwendungen von 5 Mrd. $ drauf, sagte Barret. Er verweis dazu insbesondere auf die US-Werke in Phoenix, Portland und Oregon sowie die Anlage in Irland. In zwei Jahren seien noch kleinere Halbleiter moeglich, sagte er. Im zweiten Halbjahr 2007 sollte es die 45-Nanometer- Technologie geben, erklaerte Barret. Er lehnte es jedoch ab, sich zu den Finanzergebnissen des Konzerns zu aendern. Er sagte lediglich, das Geschaeft wachse weltweit. Kraeftiges Wachstum sei in den Schwellenlaendern zu verzeichnen.

EXPLANATION
WIQA Browser

The Triple:

Siemens Share **positive analyst report** Siemens agrees partnership with Novell unit SUSE. Siemens Business Services (SBS), the IT services arm of German technology conglomerate Siemens <SIEGn.DE>, said on Tuesday it had agreed a partnerhip deal with Novell's (nasdaq: NOVL - news - people) newly acquired unit SUSE Linux. Linux software is open-source, meaning it can be freely copied and modified, unlike proprietary software such as Microsoft (nasdaq: MSFT - news - people) Windows. In the past months clients have been asking more and more for open-source platforms, SBS said in a statement which said SUSE would have premier partner status. SBS is one of Europe's top 10 information technology service providers. Linux, once the exclusive province of a few dedicated enthusiasts, is now seen as the only serious rival to Windows and is supported by U.S. giant International Business Machines (nyse: IBM - news - people), among others. Its advocates, who include big businesses and government departments, argue it is cheaper, simpler and more secure than Windows.

fulfils the policy:

Use only information which has been asserted by German analysts.

because:

- it is stated in the document **Information from Peter Smith**, which is asserted by the German analyst **Peter Smith**.

Naive Reasoning on Web Data does not work!

- **Experiment: Naive RDF Schema reasoning on DBpedia data**
  - What are the rdf:types of dbpedia:Germany?
  - Results: Place, Award, Populated Place, City, SportsTeam, Mountain, Agent, Organisation, Country, Stadium, RecordLabel, MilitaryUnit, Company, EducationalInstitution, PersonFunction, EthnicGroup, Architect, WineRegion, Language, MilitaryConflict, Settlement, RouteOfTransportation
- **What is going on here?**
  - DBpedia data is noisy as it was produced by many different people
  - With naïve reasoning one wrong statement is enough for a wrong conclusion
  - Germany example: 38,000 statements, 20 wrong types from 20 wrong statements (error rate of 0.05%)
- **Conclusion**
  - Always assess the quality of Web data before applying any reasoning
  - Alternatively use robust reasoning methods

    (for instance: Paulheim/Bizer: Type inference on noisy RDF data. ISWC 2013)

The Dataspace Vision

Alternative to classic data integration systems in order to cope with growing number of data sources.

- **Properties of dataspaces**
  - provide for data-coexistence
  - require no upfront investment into a global schema
  - give best effort answers to queries
  - rely on pay-as-you-go data integration

Franklin, M., Halevy, A., and Maier, D.: From Databases to Dataspaces A new Abstraction for Information Management, SIGMOD Rec. 2005.

Madhavan, J., et al.: Web-scale Data Integration: You Can Only Afford to Pay As You Go, CIDR 2007

Linked Data relies on the Pay-as-You-Go Idea

- for Identity Management
- for Schema/Vocabulary Management

Providing Integration Hints

- by publishing Identity Links on the Web

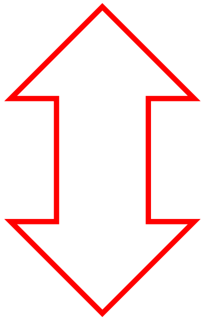
Identity Link

```
<http://www4.wiwiss.fu-berlin.de/is-group/resource/persons/Person4>
owl:sameAs
<http://dblp.l3s.de/d2r/resource/authors/Christian_Bizer> .
```

- You publish links pointing at other data sources.
- Somebody else publishes links pointing at your data source.

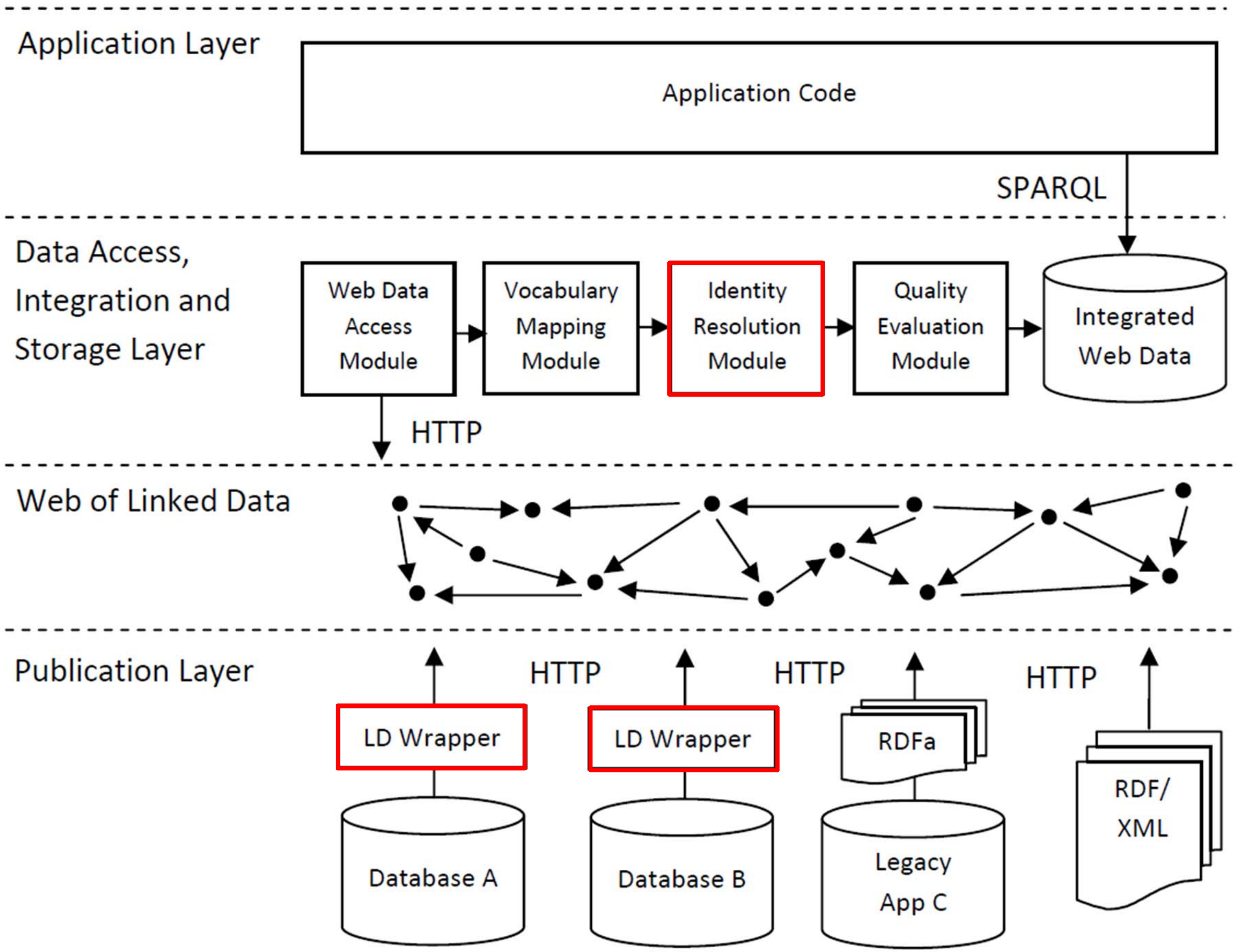
Effort Distribution between Publisher and Consumer

Consumer data mines identity links



Effort Distribution

Publishers or third parties provides identity links



Providing Integration Hints

- by publishing Vocabulary Links on the Web

Vocabulary Link

```
<http://xmlns.com/foaf/0.1/Person>
owl:equivalentClass
<http://dbpedia.org/ontology/Person> .
```

- Terms for expressing Correspondences
  - owl:equivalentClass, owl:equivalentProperty
  - rdfs:subClassOf, rdfs:subPropertyOf

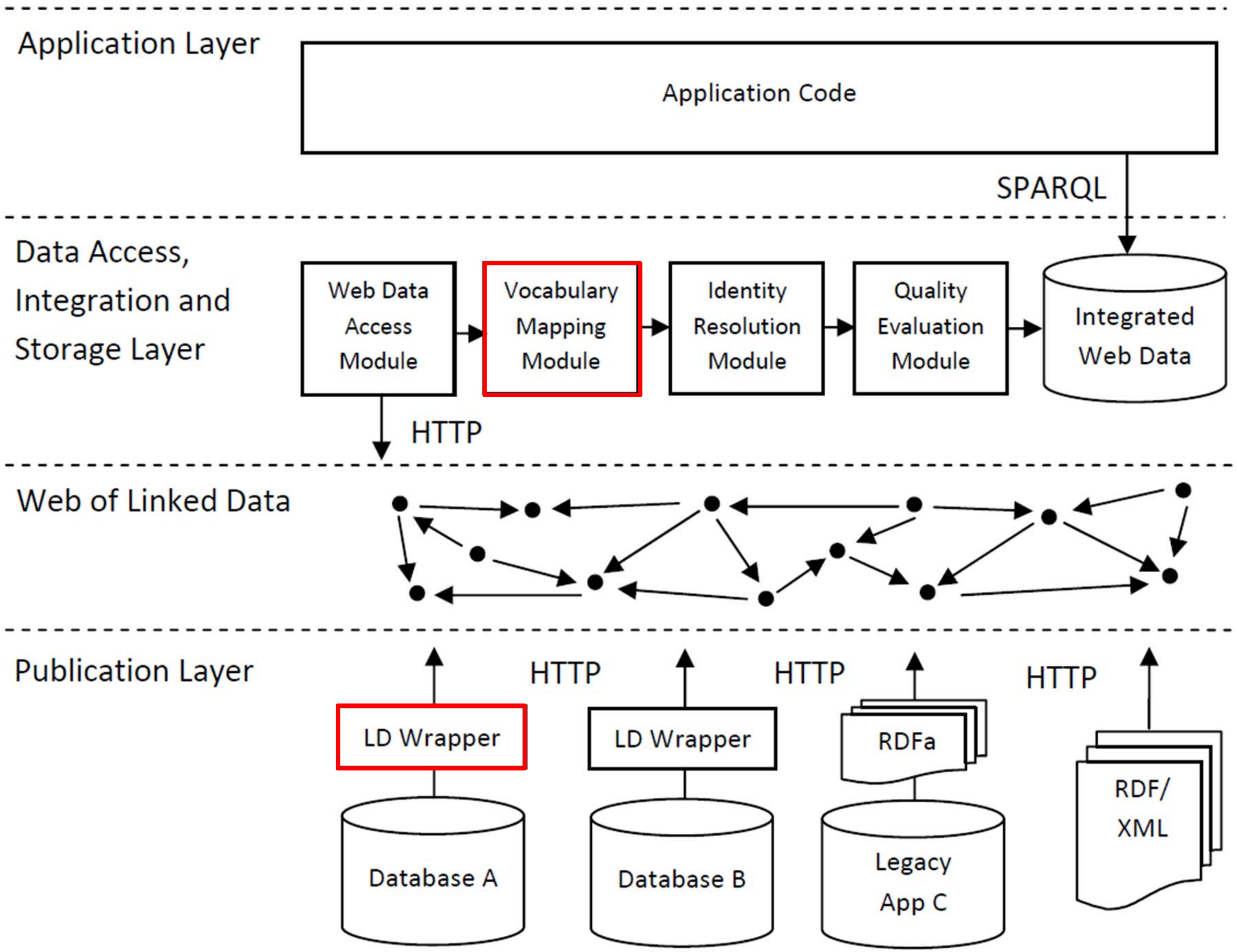
Effort Distribution between Publisher and Consumer

Consumer defines or data mines mappings

Effort Distribution

Publisher reuses vocabularies

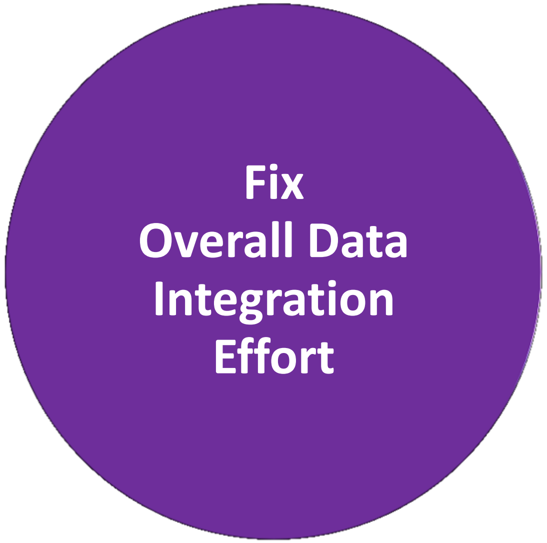
Publisher or third party publishes mappings

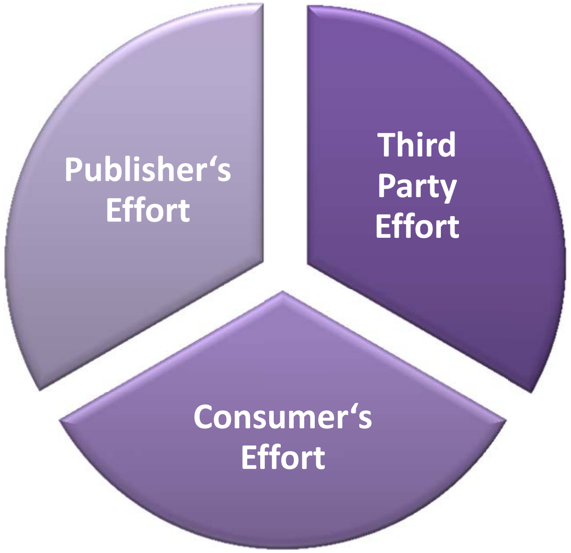


Somebody-Pays-As-You-Go

The overall data integration effort is split between the data publisher, the data consumer and third parties.

- **Data Publisher**
  - publishes data as RDF
  - sets identity links
  - reuses terms or publishes mappings
- **Third Parties**
  - set identity links pointing at your data
  - publish mappings to the Web
- **Data Consumer**
  - has to do the rest
  - using record linkage and schema matching techniques





4. Alternative Web Data Publishing Formats

More and more Websites semantically markup the content of their HTML pages.

Microformats

RDFa

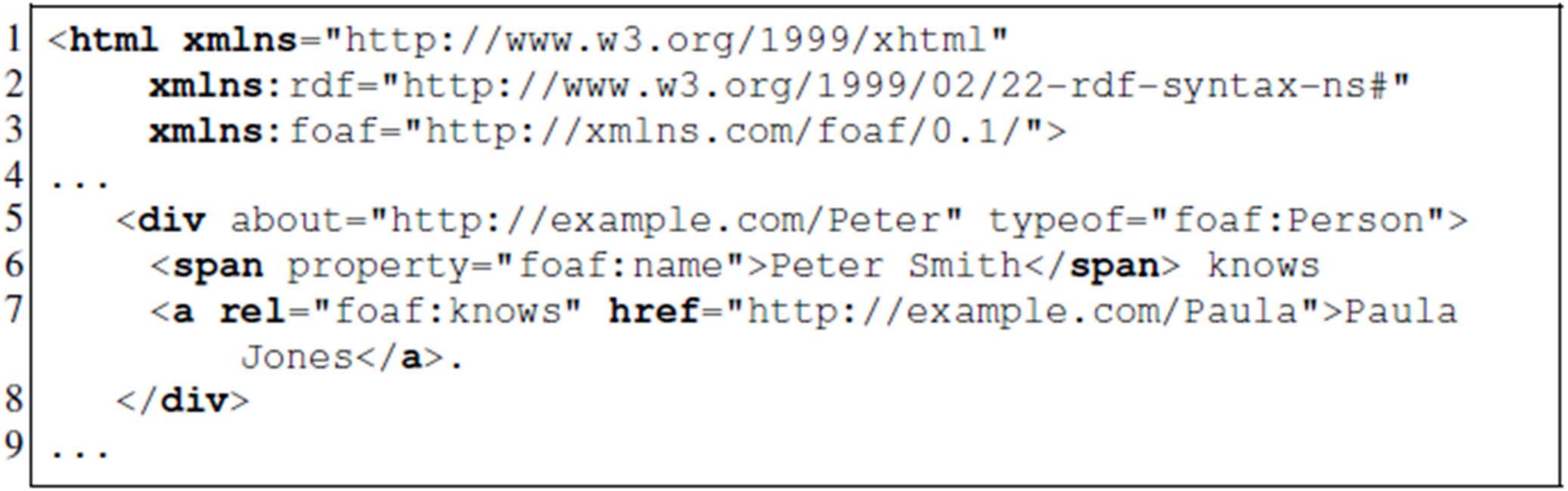
Microdata

Microformats

- **Microformat effort dates back to 2003**
- **Small set of fixed formats**
  - hcard : people, companies, organizations, and places
  - XFN : relationships between people
  - hCalendar : calendaring and events
  - hListing : small-ads; classifieds
  - hReview : reviews of products, businesses, events
- **Shortcoming of Microformats**
  - can not represent any kind of data.
- **indexed by Google and Yahoo since 2009**

RDFa

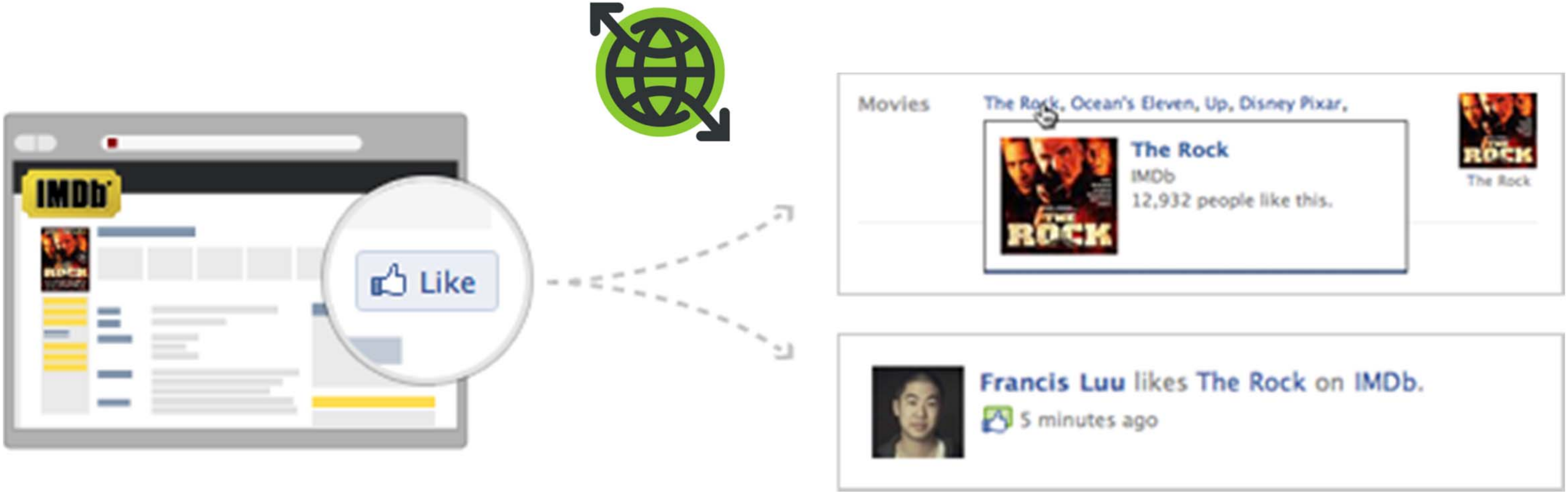
- serialization format for embedding RDF data into HTML pages
- proposed in 2004, W3C Recommendation in 2008
- can be used together with any vocabulary
- can assign URIs as global primary keys to entities



```
<html xmlns="http://www.w3.org/1999/xhtml"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:foaf="http://xmlns.com/foaf/0.1/">
...
   <div about="http://example.com/Peter" typeof="foaf:Person">
     <span property="foaf:name">Peter Smith</span> knows
     <a rel="foaf:knows" href="http://example.com/Paula">Paula
         Jones</a>.
   </div>
...
```

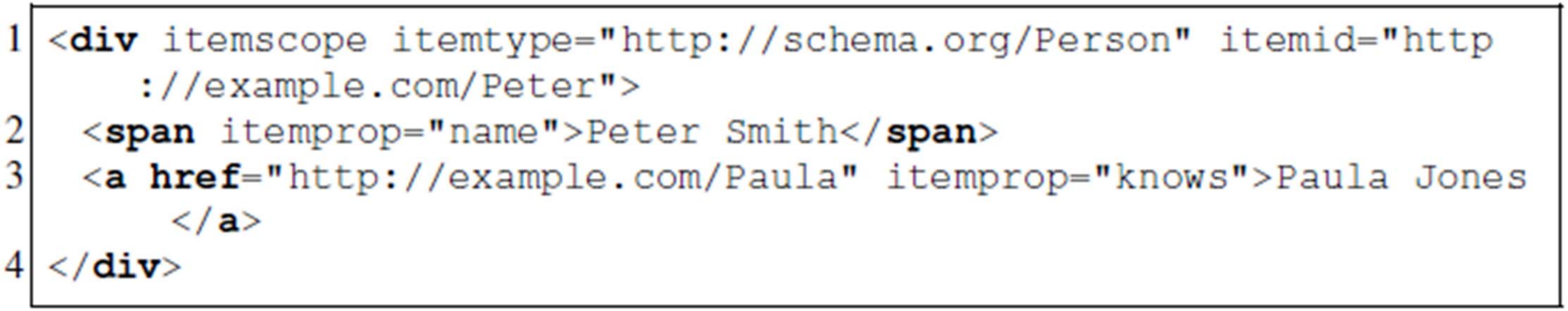
Open Graph Protocol

- allows site owners to determine how entities are described in Facebook
- relies on RDFa for encoding data in HTML pages
- available since April 2010


IMDb
Like
Movies
The Rock, Ocean's Eleven, Up, Disney Pixar,
The Rock
IMDb
12,932 people like this.
The Rock
Francis Luu likes The Rock on IMDb.
5 minutes ago

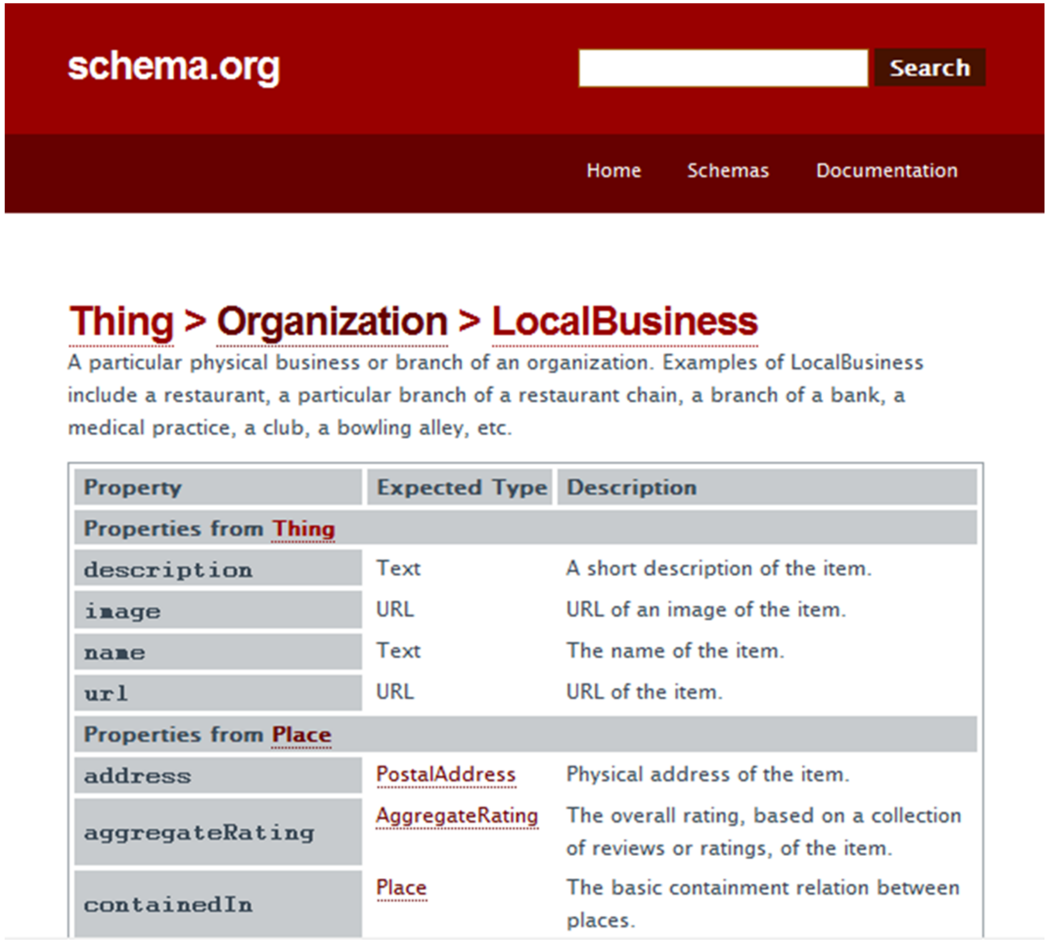
Microdata

- alternative technique for embedding structured data
- proposed in 2009 by WHATWG as part of HTML5 work
- tries to be simpler than RDFa (5 new attributes instead of 8)
- W3C currently tries to reconcile the two alternative proposals



```
<div itemscope itemtype="http://schema.org/Person" itemid="http
    ://example.com/Peter">
  <span itemprop="name">Peter Smith</span>
  <a href="http://example.com/Paula" itemprop="knows">Paula Jones
      </a>
</div>
```

Schema.org



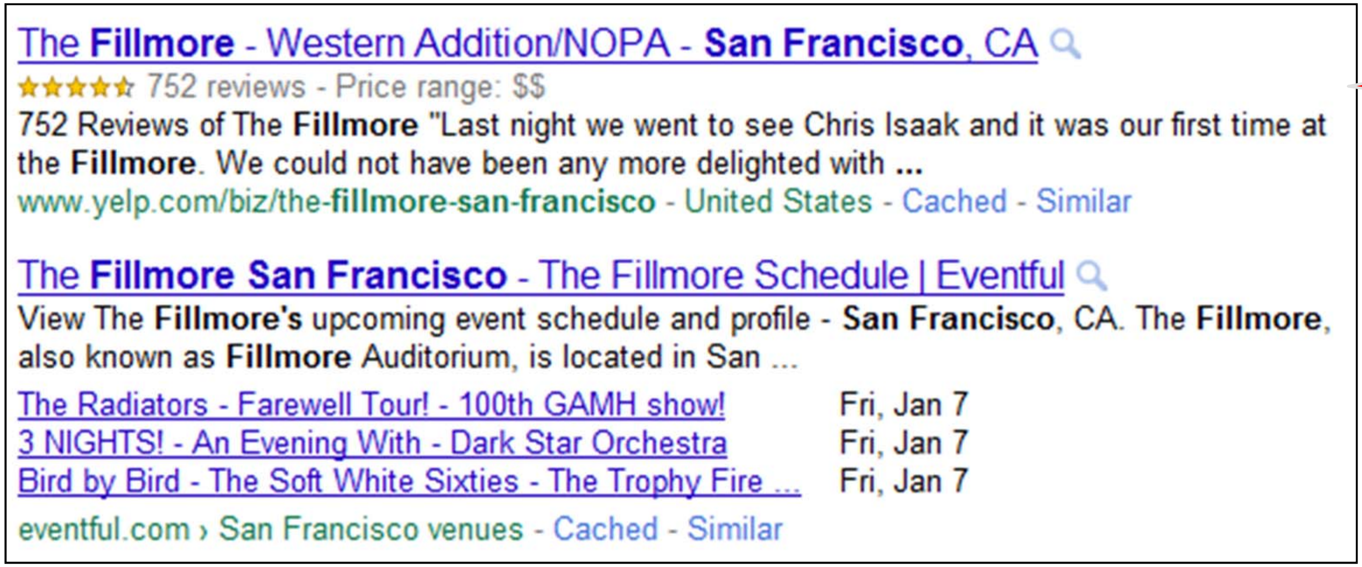
Thing > Organization > LocalBusiness

A particular physical business or branch of an organization. Examples of LocalBusiness include a restaurant, a particular branch of a restaurant chain, a branch of a bank, a medical practice, a club, a bowling alley, etc.

Property	Expected Type	Description
Properties from Thing		
description	Text	A short description of the item.
image	URL	URL of an image of the item.
name	Text	The name of the item.
url	URL	URL of the item.
Properties from Place		
address	PostalAddress	Physical address of the item.
aggregateRating	AggregateRating	The overall rating, based on a collection of reviews or ratings, of the item.
containedIn	Place	The basic containment relation between places.

- ask site owners to embed data to enrich search results.
- 200+ Types: Event, Organization, Person, Place, Product, Review
- Encoding: Microdata or alternatively RDFa

Usage of Schema.org Data @ Google



The **Fillmore** - Western Addition/NOPA - **San Francisco**, CA

★★★★☆ 752 reviews - Price range: $$

752 Reviews of The **Fillmore** "Last night we went to see Chris Isaak and it was our first time at the **Fillmore**. We could not have been any more delighted with **...**

www.yelp.com/biz/the-**fillmore-san-francisco** - United States - Cached - Similar

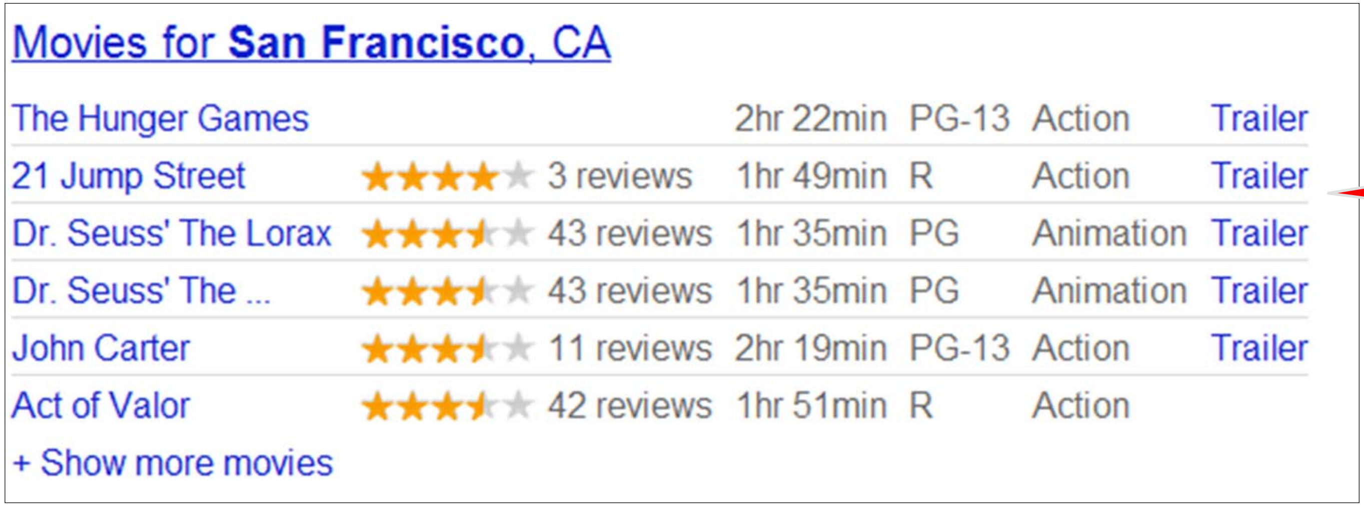
The **Fillmore San Francisco** - The Fillmore Schedule | Eventful

View The **Fillmore's** upcoming event schedule and profile - **San Francisco**, CA. The **Fillmore**, also known as **Fillmore** Auditorium, is located in San ...

The Radiators - Farewell Tour! - 100th GAMH show!	Fri, Jan 7
3 NIGHTS! - An Evening With - Dark Star Orchestra	Fri, Jan 7
Bird by Bird - The Soft White Sixties - The Trophy Fire ...	Fri, Jan 7

eventful.com › San Francisco venues - Cached - Similar

Data snippets within search results

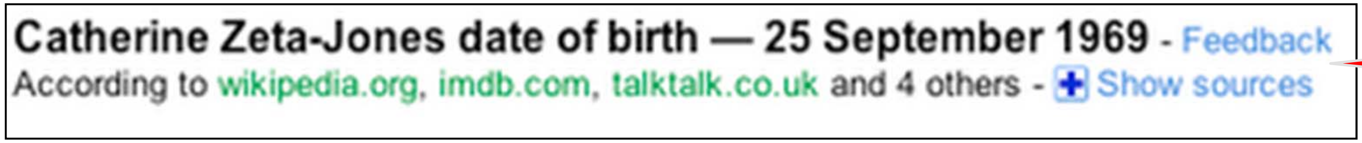


Movies for **San Francisco**, CA

The Hunger Games			2hr 22min	PG-13	Action	Trailer
21 Jump Street	★★★★☆	3 reviews	1hr 49min	R	Action	Trailer
Dr. Seuss' The Lorax	★★★½☆	43 reviews	1hr 35min	PG	Animation	Trailer
Dr. Seuss' The ...	★★★½☆	43 reviews	1hr 35min	PG	Animation	Trailer
John Carter	★★★½☆	11 reviews	2hr 19min	PG-13	Action	Trailer
Act of Valor	★★★½☆	42 reviews	1hr 51min	R	Action	

+ Show more movies

Data tables within search results

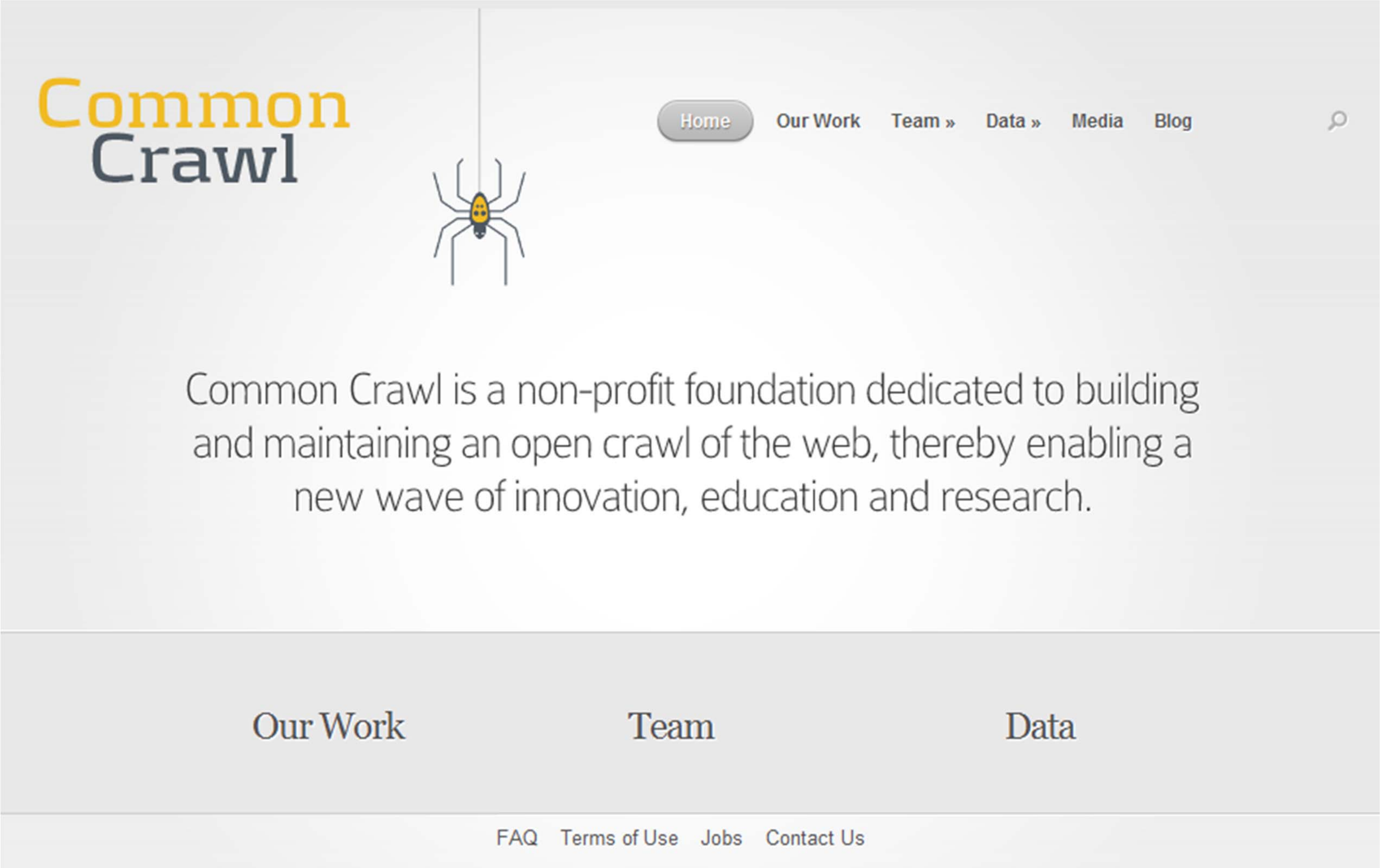


Catherine Zeta-Jones date of birth — 25 September 1969 - Feedback

According to wikipedia.org, imdb.com, talktalk.co.uk and 4 others - ➕ Show sources

Answers to fact queries

The Common Crawl



Common Crawl is a non-profit foundation dedicated to building and maintaining an open crawl of the web, thereby enabling a new wave of innovation, education and research.

Web Data Commons

WebDataCommons.org Project

- extracts all Microformat, Microdata, RDFa data from the Common Crawl
- provides the extracted data for free download

Two extractions runs

- 2009/2010 CC Corpus: 2.5 billion HTML pages → 5.1 billion RDF triples
- 2012 CC Corpus: 3.0 billion HTML pages → 7.3 billion RDF triples

Jointed project of

UNIVERSITÄT
MANNHEIM

KIT
Karlsruhe Institute of Technology

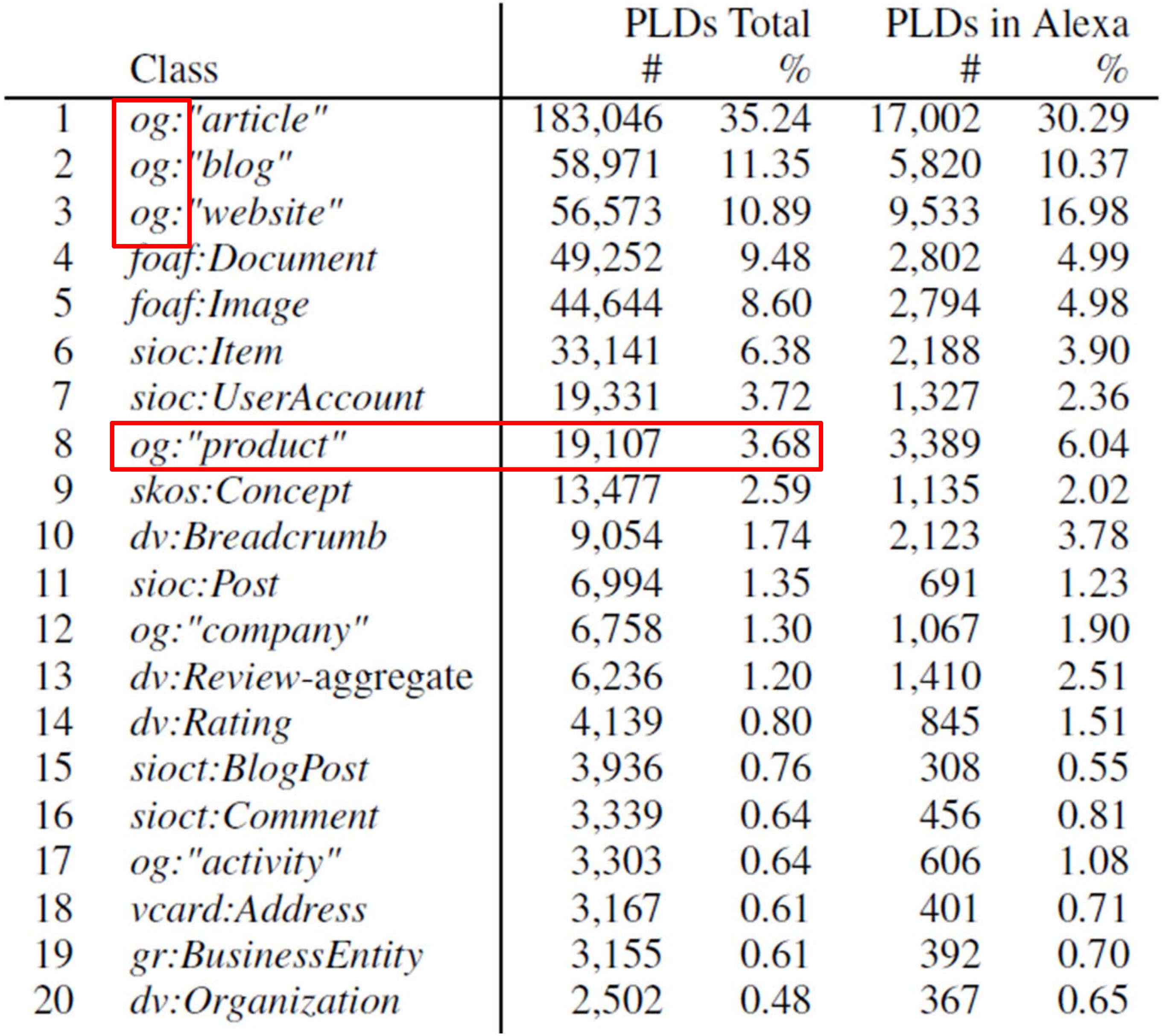
Websites containing Structured Data (CC 2012)

369 million of the 3 billion pages contain Microformat, Microdata or RDFa data (12.3%).

2.29 million websites (PLDs) out of 40.6 million provide Microformat, Microdata or RDFa data (5.65%)

RDFa Topics (CC 2012)

- **Top Classes:**
- **Topics**
  - CMS and Blog metadata
  - Product data
  - Ratings
  - Company listings

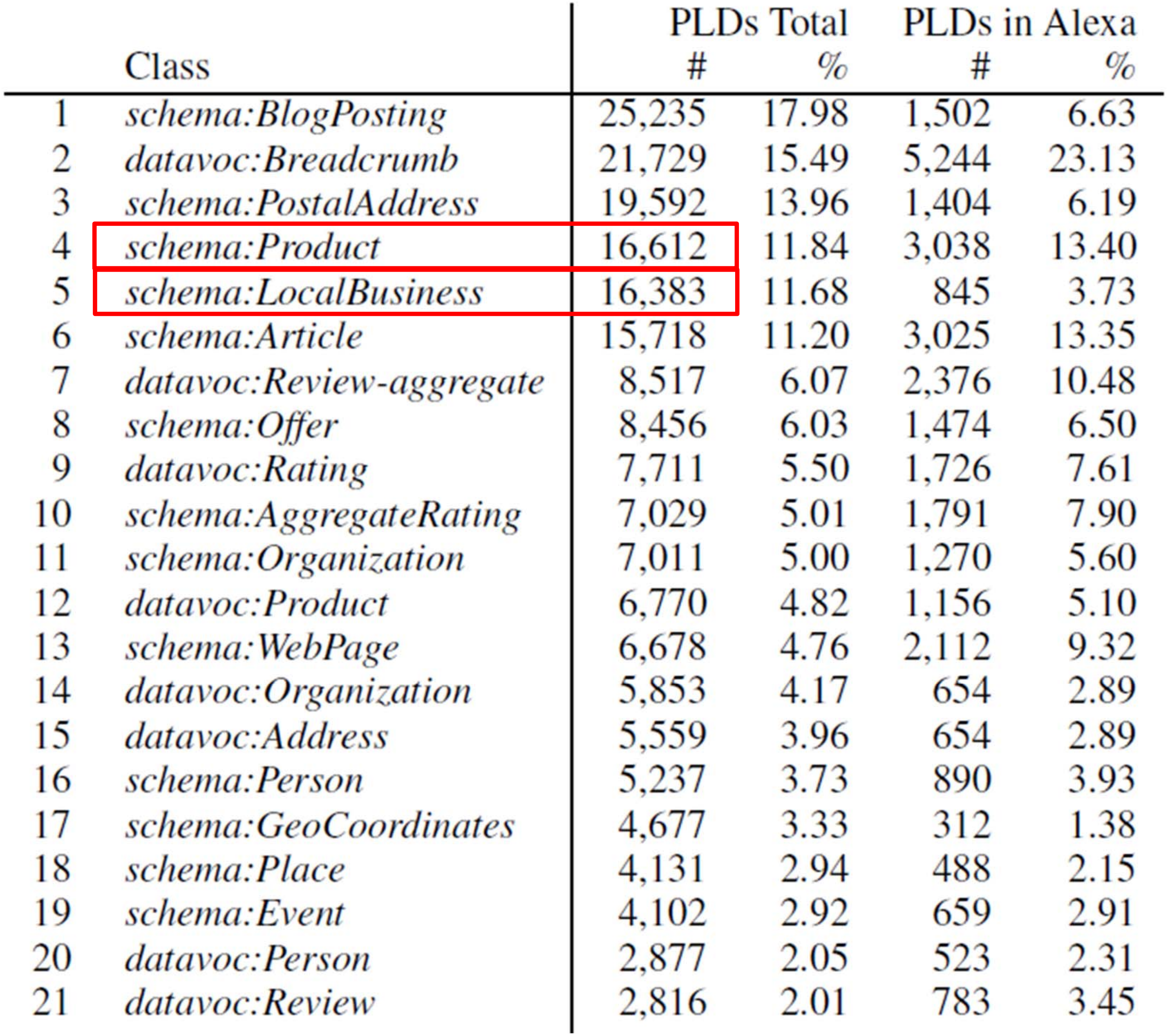


		PLDs Total		PLDs in Alexa	
	Class	#	%	#	%
1	*og:"article"*	183,046	35.24	17,002	30.29
2	*og:"blog"*	58,971	11.35	5,820	10.37
3	*og:"website"*	56,573	10.89	9,533	16.98
4	*foaf:Document*	49,252	9.48	2,802	4.99
5	*foaf:Image*	44,644	8.60	2,794	4.98
6	*sioc:Item*	33,141	6.38	2,188	3.90
7	*sioc:UserAccount*	19,331	3.72	1,327	2.36
8	*og:"product"*	19,107	3.68	3,389	6.04
9	*skos:Concept*	13,477	2.59	1,135	2.02
10	*dv:Breadcrumb*	9,054	1.74	2,123	3.78
11	*sioc:Post*	6,994	1.35	691	1.23
12	*og:"company"*	6,758	1.30	1,067	1.90
13	*dv:Review-aggregate*	6,236	1.20	1,410	2.51
14	*dv:Rating*	4,139	0.80	845	1.51
15	*sioct:BlogPost*	3,936	0.76	308	0.55
16	*sioct:Comment*	3,339	0.64	456	0.81
17	*og:"activity"*	3,303	0.64	606	1.08
18	*vcard:Address*	3,167	0.61	401	0.71
19	*gr:BusinessEntity*	3,155	0.61	392	0.70
20	*dv:Organization*	2,502	0.48	367	0.65

og = Facebook‘s Open Graph Protocol

Microdata Topics (CC 2012)

- **Top Classes:**
- **Topics**
  - CMS and Blog metadata
  - Navigational metadata
  - Products and offers
  - Business listings
  - Ratings

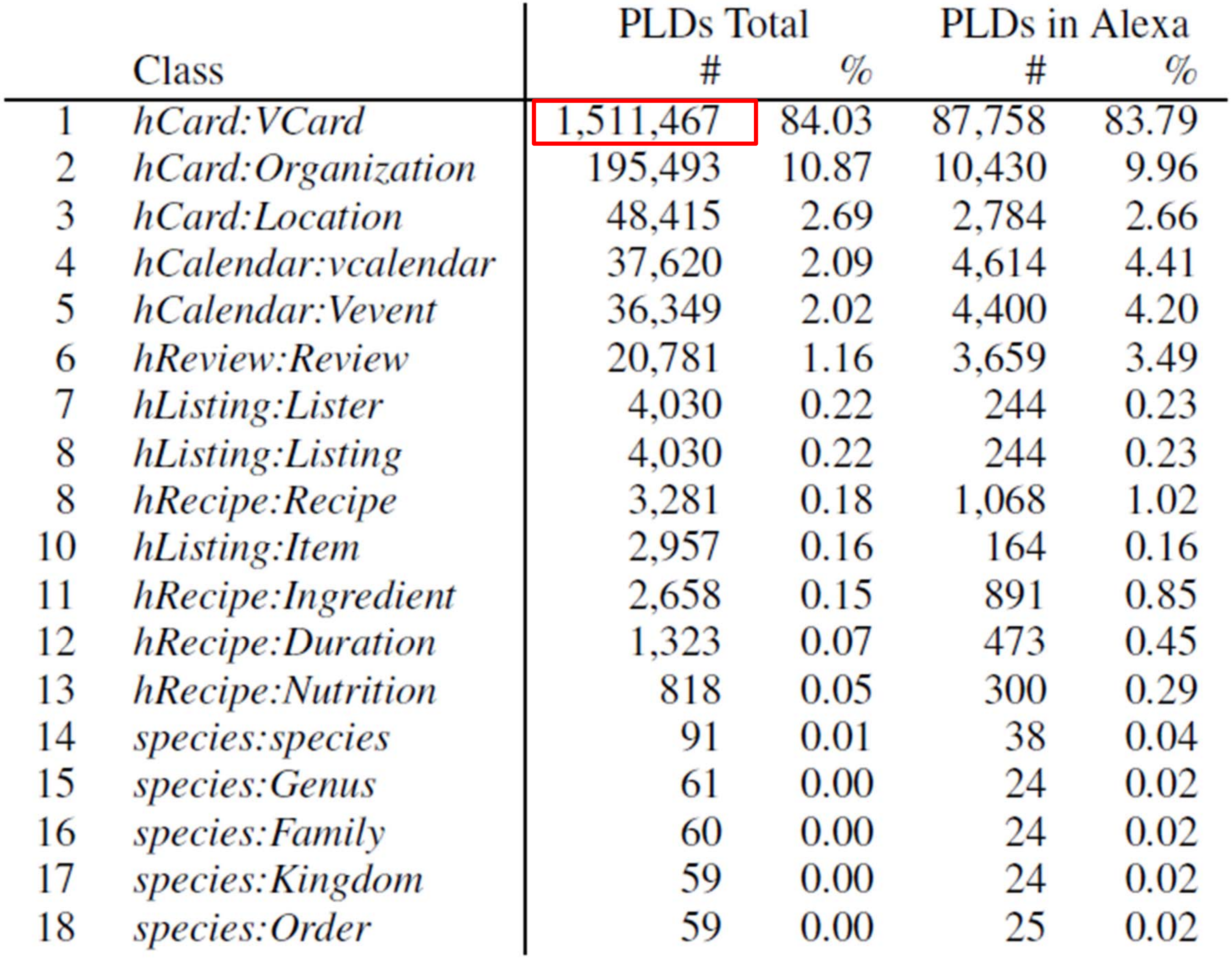


	Class	PLDs Total #	PLDs Total %	PLDs in Alexa #	PLDs in Alexa %
1	*schema:BlogPosting*	25,235	17.98	1,502	6.63
2	*datavoc:Breadcrumb*	21,729	15.49	5,244	23.13
3	*schema:PostalAddress*	19,592	13.96	1,404	6.19
4	*schema:Product*	16,612	11.84	3,038	13.40
5	*schema:LocalBusiness*	16,383	11.68	845	3.73
6	*schema:Article*	15,718	11.20	3,025	13.35
7	*datavoc:Review-aggregate*	8,517	6.07	2,376	10.48
8	*schema:Offer*	8,456	6.03	1,474	6.50
9	*datavoc:Rating*	7,711	5.50	1,726	7.61
10	*schema:AggregateRating*	7,029	5.01	1,791	7.90
11	*schema:Organization*	7,011	5.00	1,270	5.60
12	*datavoc:Product*	6,770	4.82	1,156	5.10
13	*schema:WebPage*	6,678	4.76	2,112	9.32
14	*datavoc:Organization*	5,853	4.17	654	2.89
15	*datavoc:Address*	5,559	3.96	654	2.89
16	*schema:Person*	5,237	3.73	890	3.93
17	*schema:GeoCoordinates*	4,677	3.33	312	1.38
18	*schema:Place*	4,131	2.94	488	2.15
19	*schema:Event*	4,102	2.92	659	2.91
20	*datavoc:Person*	2,877	2.05	523	2.31
21	*datavoc:Review*	2,816	2.01	783	3.45

datavoc = Google‘s
Rich Snippet Vocabulary
schema = Schema.org

Microformats

- Top Classes:
- Topics
  - Persons
  - Organisations
  - Events
  - Listings and Reviews
  - Recipes

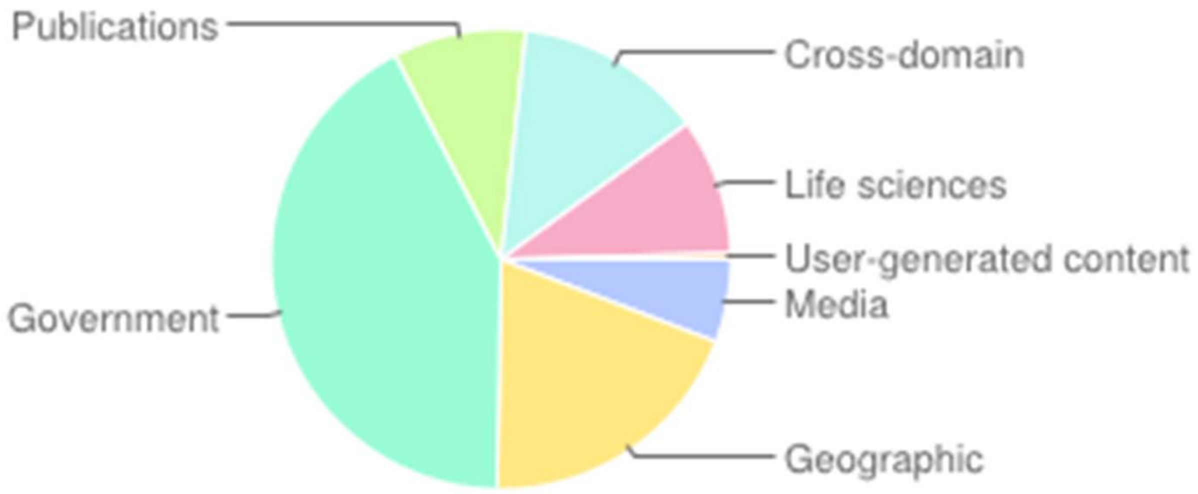


		PLDs Total		PLDs in Alexa	
	Class	#	%	#	%
1	*hCard:VCard*	1,511,467	84.03	87,758	83.79
2	*hCard:Organization*	195,493	10.87	10,430	9.96
3	*hCard:Location*	48,415	2.69	2,784	2.66
4	*hCalendar:vcalendar*	37,620	2.09	4,614	4.41
5	*hCalendar:Vevent*	36,349	2.02	4,400	4.20
6	*hReview:Review*	20,781	1.16	3,659	3.49
7	*hListing:Lister*	4,030	0.22	244	0.23
8	*hListing:Listing*	4,030	0.22	244	0.23
8	*hRecipe:Recipe*	3,281	0.18	1,068	1.02
10	*hListing:Item*	2,957	0.16	164	0.16
11	*hRecipe:Ingredient*	2,658	0.15	891	0.85
12	*hRecipe:Duration*	1,323	0.07	473	0.45
13	*hRecipe:Nutrition*	818	0.05	300	0.29
14	*species:species*	91	0.01	38	0.04
15	*species:Genus*	61	0.00	24	0.02
16	*species:Family*	60	0.00	24	0.02
17	*species:Kingdom*	59	0.00	24	0.02
18	*species:Order*	59	0.00	25	0.02

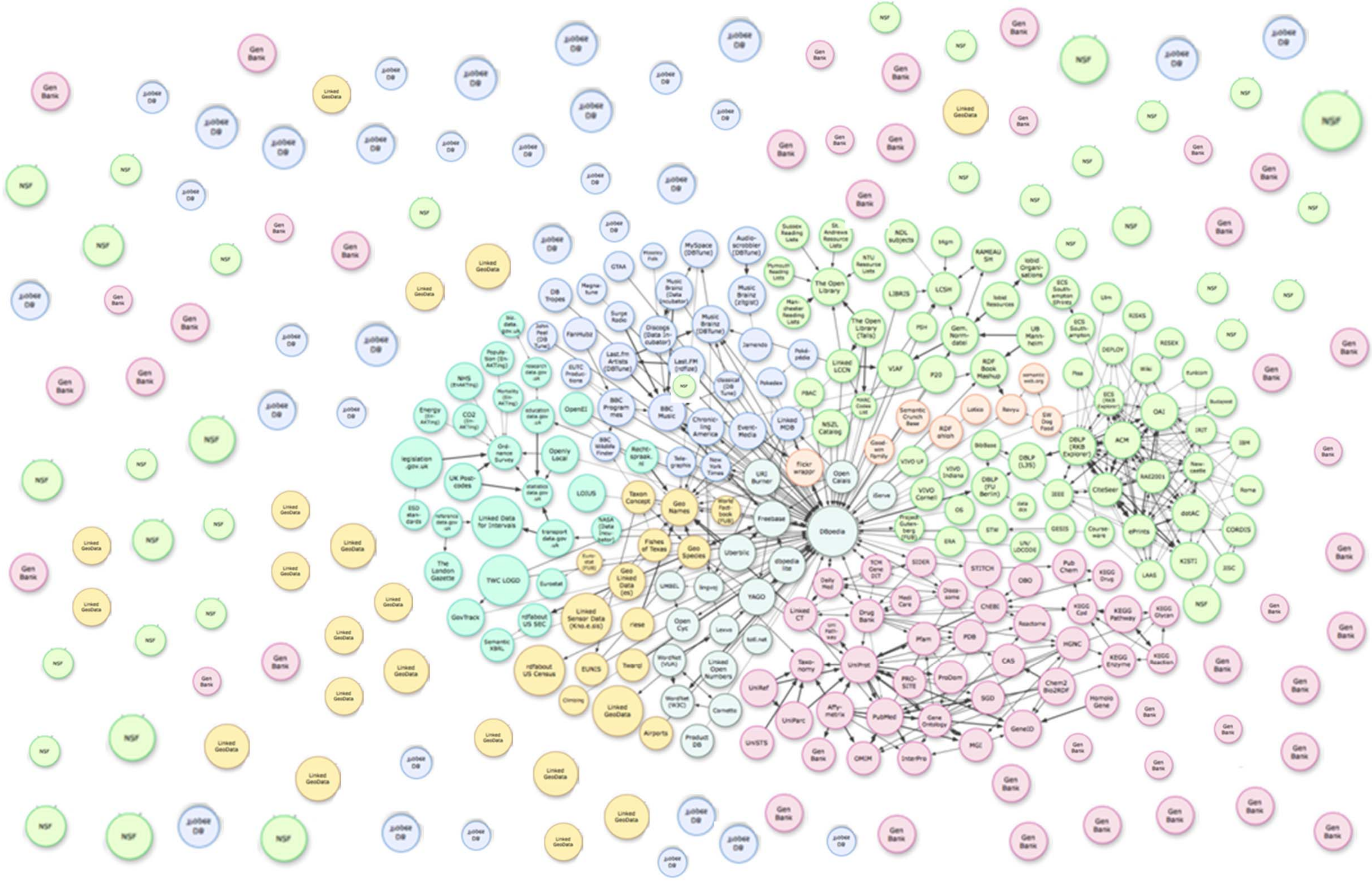
Linked Data vs. HTML-embeded Data

Compared to Microformats, Microdata, RDFa

- **the LOD Cloud covers a wider range of topics**
- **the LOD Cloud contains more complex data structures**
- **Emphasis on setting RDF Links between sources**

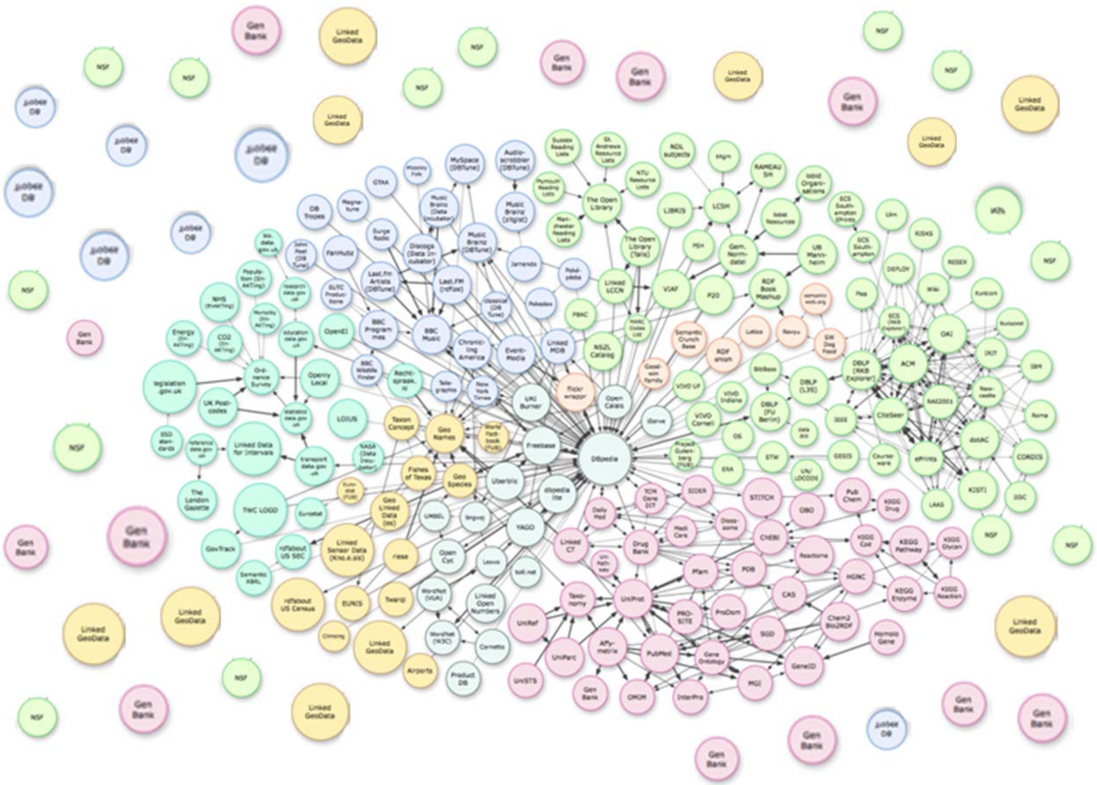


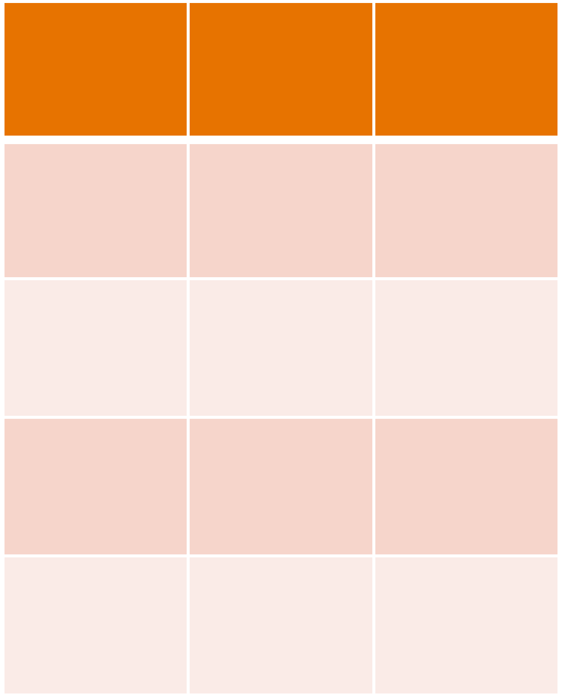
Overall Topology of the Web of Data



5. Challenges involved in using Web Data

Applications hate heterogeneity and uncertain data quality!





The wild wild west

My little world

Current Research Challenges

1. **More research on data space profiling is needed.**
   - What is in the data space and how does the content change over time?
2. **More research on data quality assessment and SPAM detection is needed.**
3. **More research on learning mappings and identity resolution heuristics within the Web context.**
   - Identity links make it easier to learn vocabulary links.
   - Vocabulary links make it easier to learn identity links.
4. **More research on pay-as-you-go data integration is needed.**
   - How do human, community and machine contributions play together over time?

Conclusion

- **The Web of Data is growing rapidly**
  - Active deployment communities exist in various domains
  - Value-able resource of background knowledge for many applications
- **Web search is evolving into query answering**
  - Search engines increasingly rely on structured data from the Web
- **Next step: Linked Data within Enterprises**
  - alternative to data warehouses and EAI middleware
  - advantages: schema-less data model, pay-as-you go data integration
- **You are looking for a topic for your PhD thesis?**
  - There are many exciting research challenges around consuming Web Data
  - Examples: Web-scale data integration, data quality assessment

Thanks!

References

- Christian Bizer, Tom Heath, Tim Berners-Lee: Linked Data – The Story So Far
  http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf
- Tom Heath, Christian Bizer: Linked Data – Evolving the Web into a global data space.
  http://linkeddatabook.com/editions/1.0/
- 4ht Workshop on Consuming Linked Data at ISWC 2013
  http://db.uwaterloo.ca/cold2013/
- 6th Linked Data on the Web Workshop at WWW 2013
  http://events.linkeddata.org/ldow2013/