How to Publish Linked Data on the Web

Tom Heath, Michael Hausenblas, Chris Bizer, Richard Cyganiak, Olaf Hartig

Half-day Tutorial at ISWC2008
27th October 2008, Karlsruhe, Germany

Objectives

- Introduce the concept of Linked Data
- Highlight why you would want to publish Linked Data on the Web
- Introduce the principles and best practices of publishing Linked Data on the Web
- Provide an in-depth understanding of the technical design decisions required when publishing Linked Data
- Demonstrate the consumption of Linked Data from the Web
- Look ahead to the future
- Answer your burning Linked Data publishing questions

Tutorial Schedule

- 09:00 – 09:10 **Opening**
- 09:10 – 09:40 **Introduction: What and Why**
- 09:40 – 10:30 **Publishing Linked Data on the Web: How**
- 10:30 – 11:00 ***Coffee Break***
- 11:00 – 11:40 **Publishing Linked Data on the Web: How**
- 11:40 – 12:00 **Consuming Linked Data from the Web**
- 12:00 – 12:10 **Conclusions and Outlook**
- 12:10 – 12:30 **Discussion and Linked Data Clinic**

ISWC 2008, Tutorial on
How to Publish Linked Data on the Web

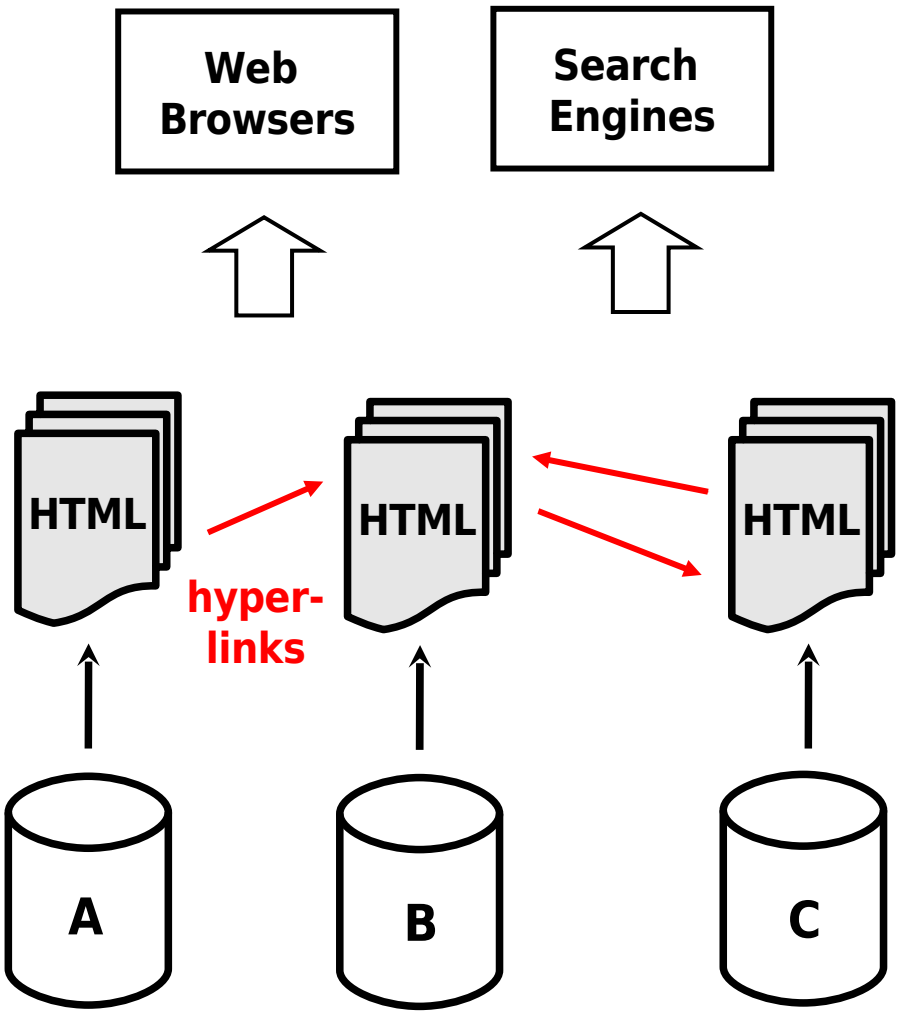
Introduction: What and Why

Christian Bizer
Freie Universität Berlin

Overview

1. **From a Web of Documents to a Web of Data**
   - **Web APIs, Microformats, and Linked Data**
2. **Linked Data Deployment on the Web**
   - **What data is out there?**
3. **Applications**
   - **What is being done with the data?**

The Classic Web



Single global information space

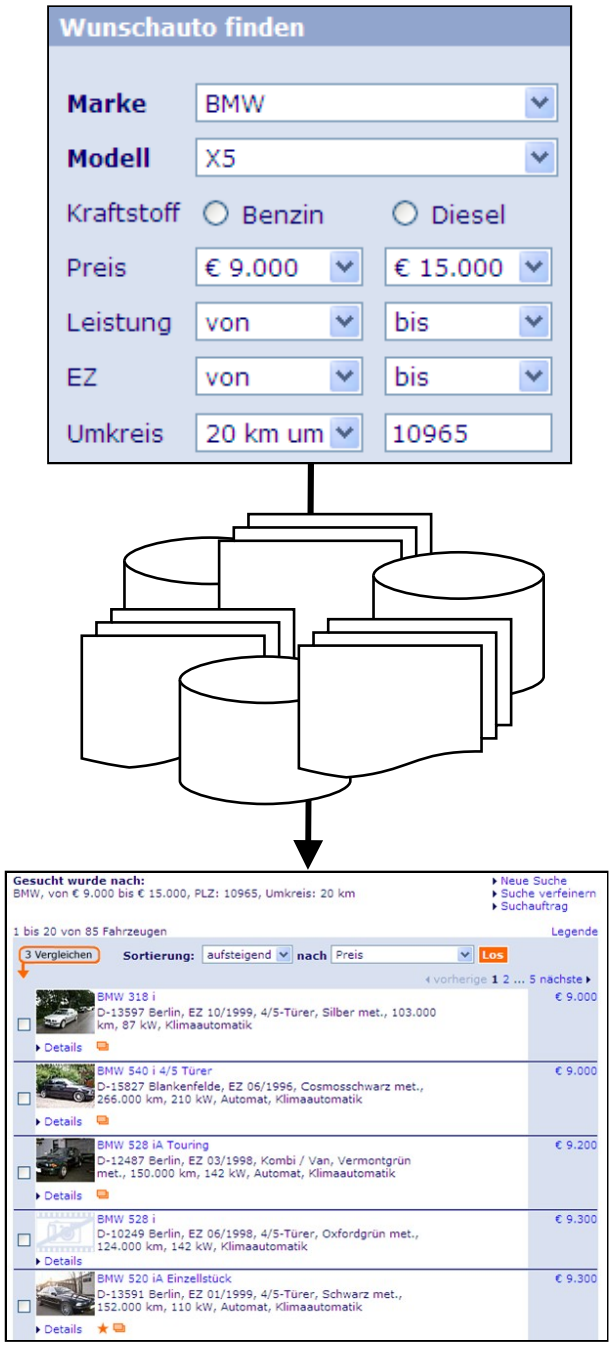
2. URLs as
   - globally unique IDs
   - retrieval mechanism
3. HTML as shared content format
4. Hyperlinks

Shortcomings

- Content is not well structured
- You can not ask expressive queries
- You can not process content within applications

What do we actually want?

Use the Web like a single global database.



Solution

Publish structured data directly on the Web.

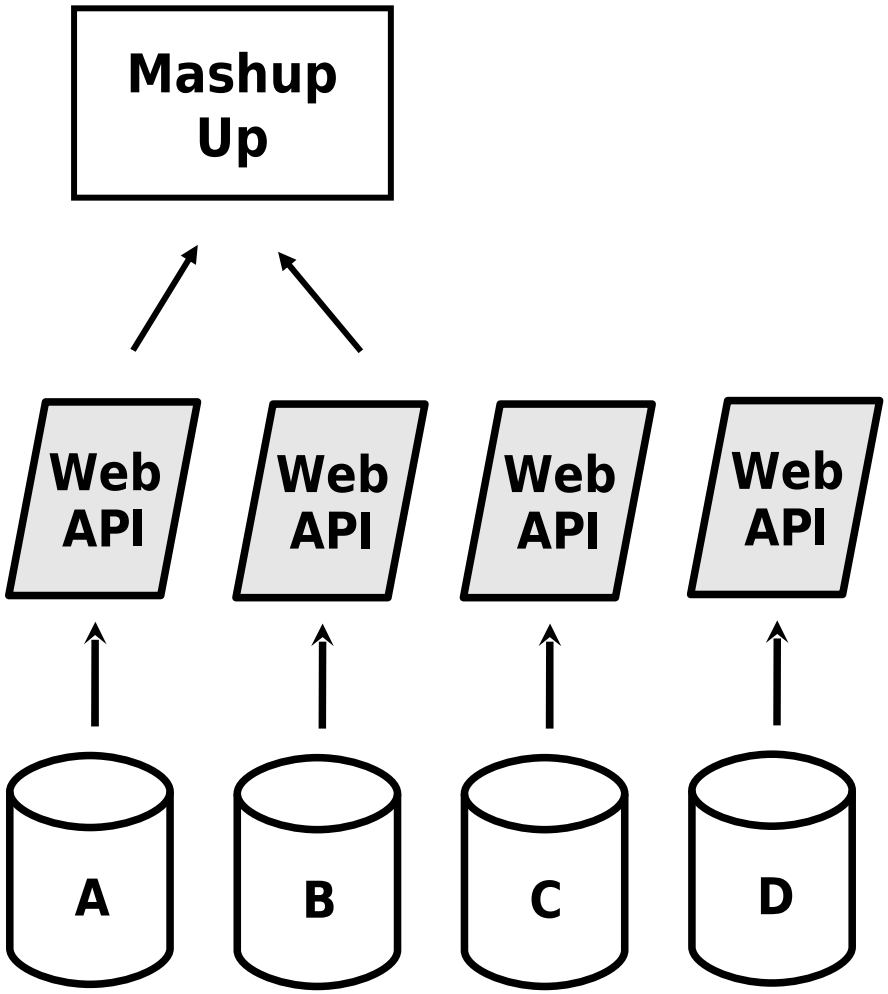
Different Approaches

2. Web APIs
3. Microformats
4. Linked Data

Web APIs

cloudalicious
orkut
WordPress
del.icio.us
social bookmarking
photocase
ourmedia
messenger
talk
30DAYTAGS
fresh tags. free delivery.
FireAnt
YouTube
Gmail
skype
flickr
mercora
MusicSearch
eventful
YAHOO! 360°
last·fm
the social music revolution
Google
MAGNOTO
PicPix

Mashups



Positive

2. APIs expose structured data
3. APIs enable new applications

Negative

6. Proprietary interfaces
7. Mashups are based only on fixed set of sources
8. You can not set hyperlinks between data objects

Web APIs slice the Web into separate data silos

Image: Bob Jagensdorf, http://flickr.com/photos/darwinbell/, CC-BY

Microformats

- Embed structured data into HTML pages.
- hCard, hCalender, hReview, XFN, ...

```
<div class="vevent">
  <span class="summary">bdigital</span>
  <abbr class="dtstart" title="2008-05-20">May 20</abbr> -

  <abbr class="dtend" title="2007-05-22">22</abbr>
</div>
```

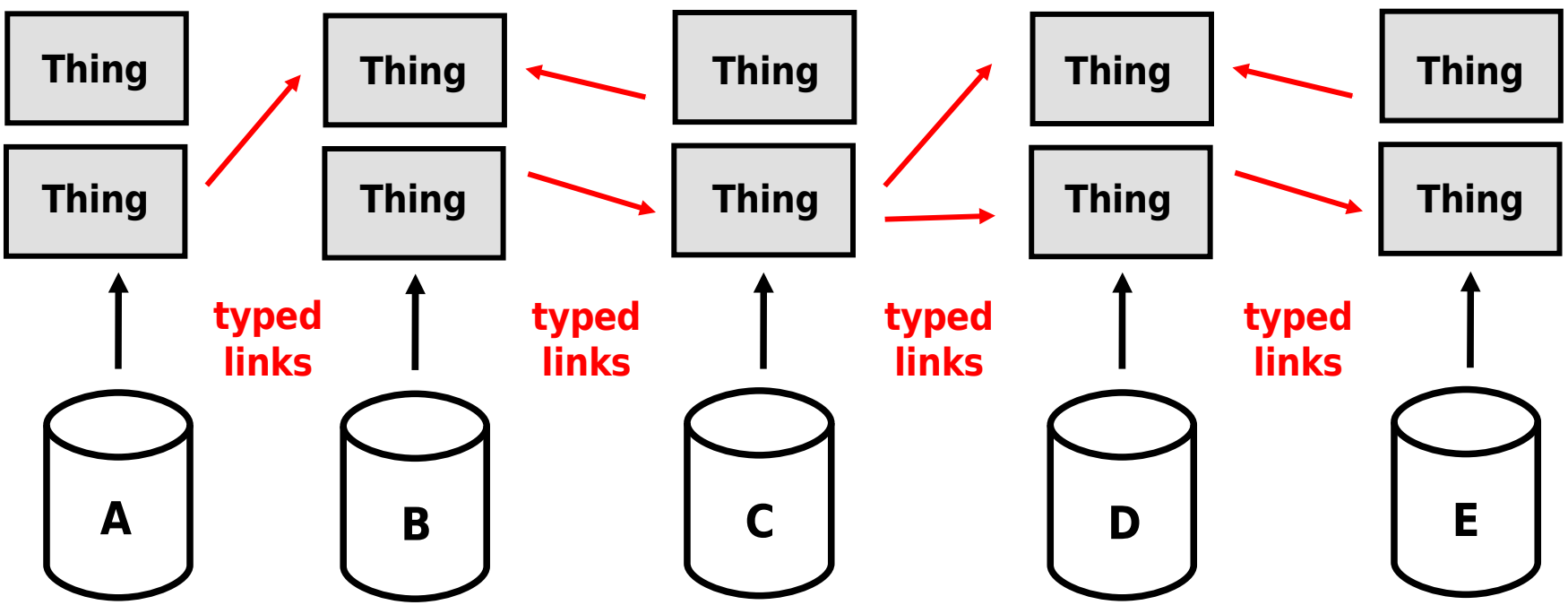
- Compatible with the idea of the Web as single information space.
- Shortcomings
  - Only a fixed set of microformats exist.
  - No way to connect data items.

Linked Data

Use Semantic Web technologies to

2. publish structured data on the Web,

3. set links between data from one data source to data within other data sources.



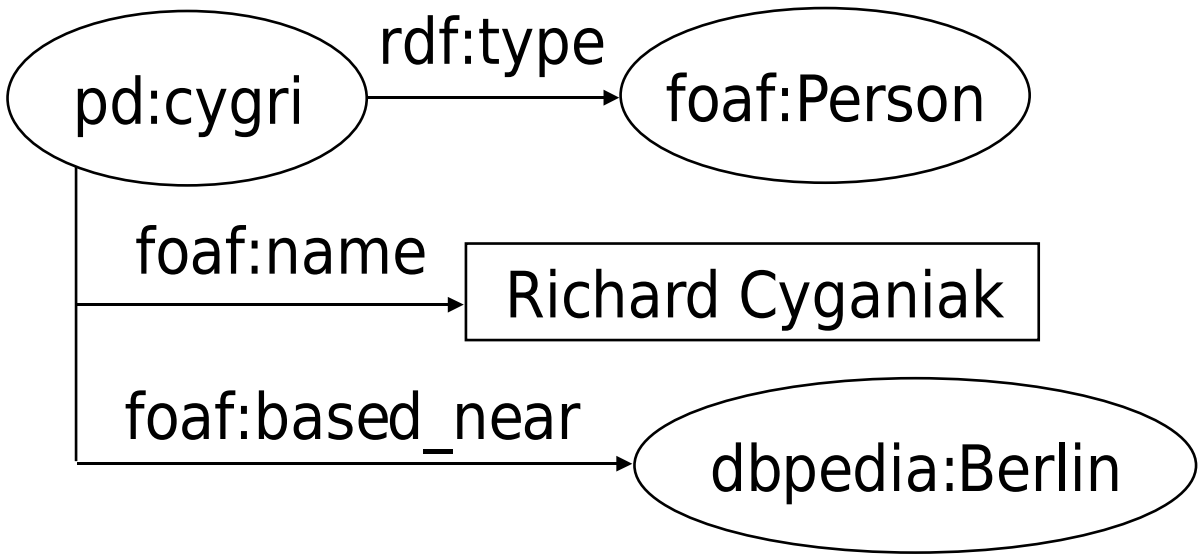
Linked Data Principles

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.

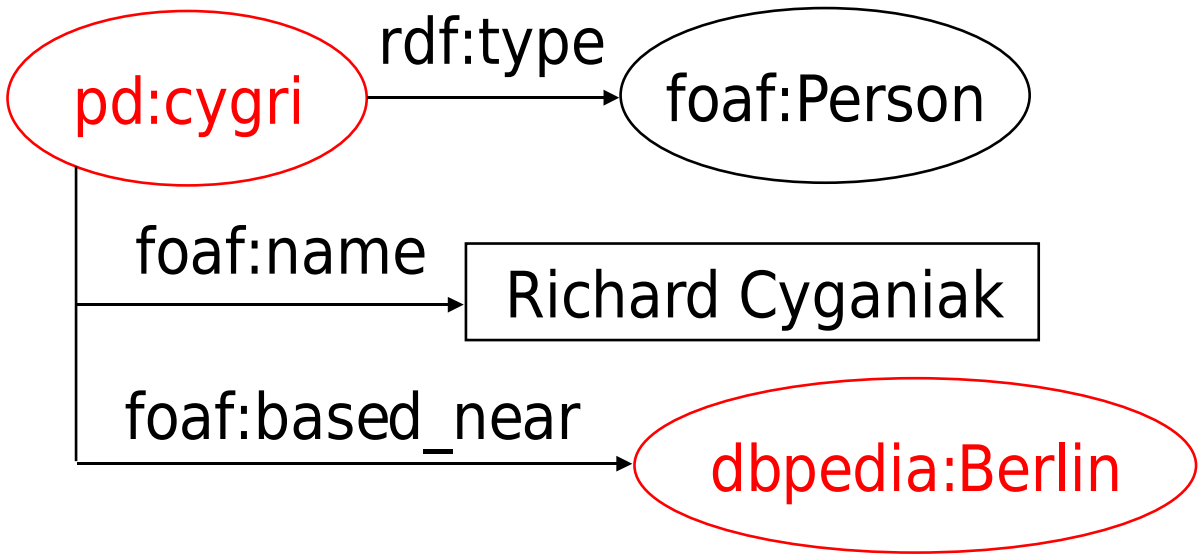
Tim Berners-Lee 2007

http://www.w3.org/DesignIssues/LinkedData.html

The RDF Data Model



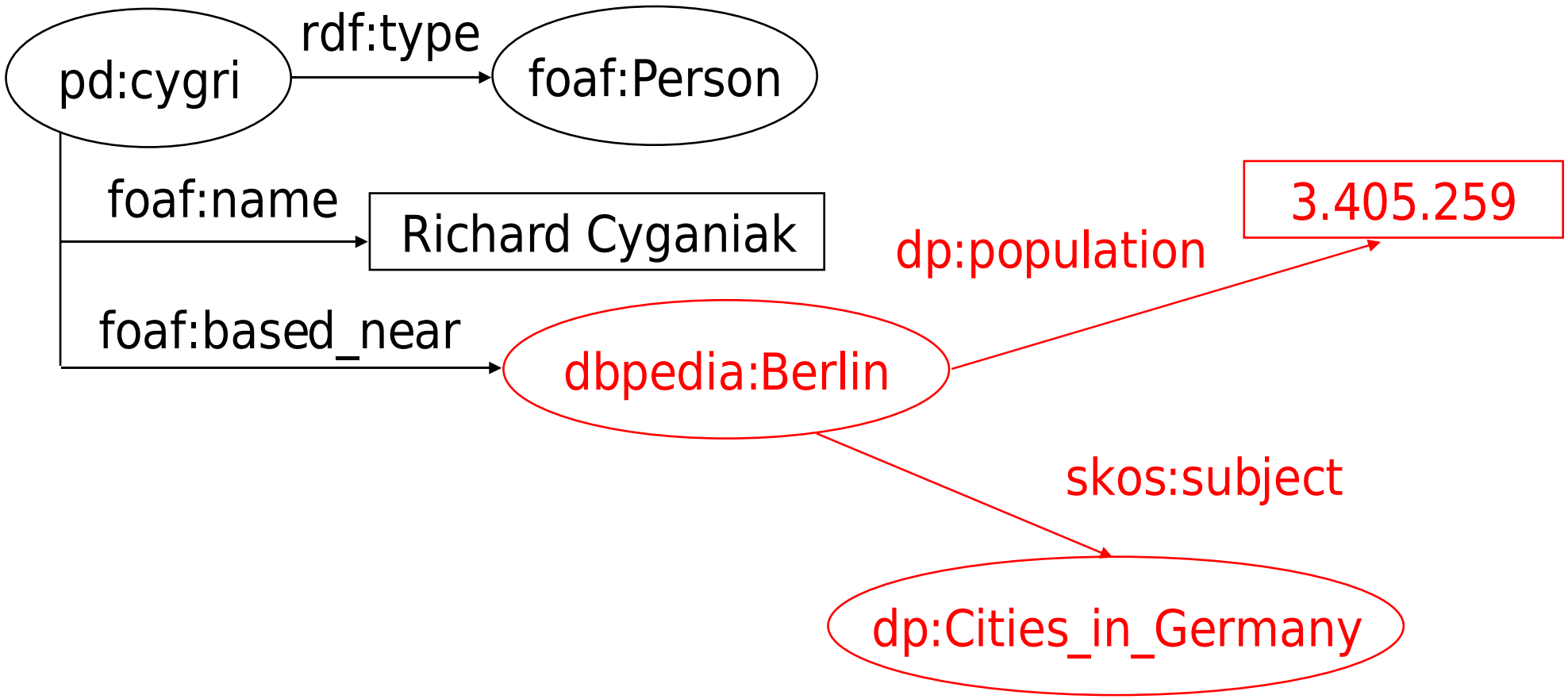
Data objects are identified with HTTP URIs



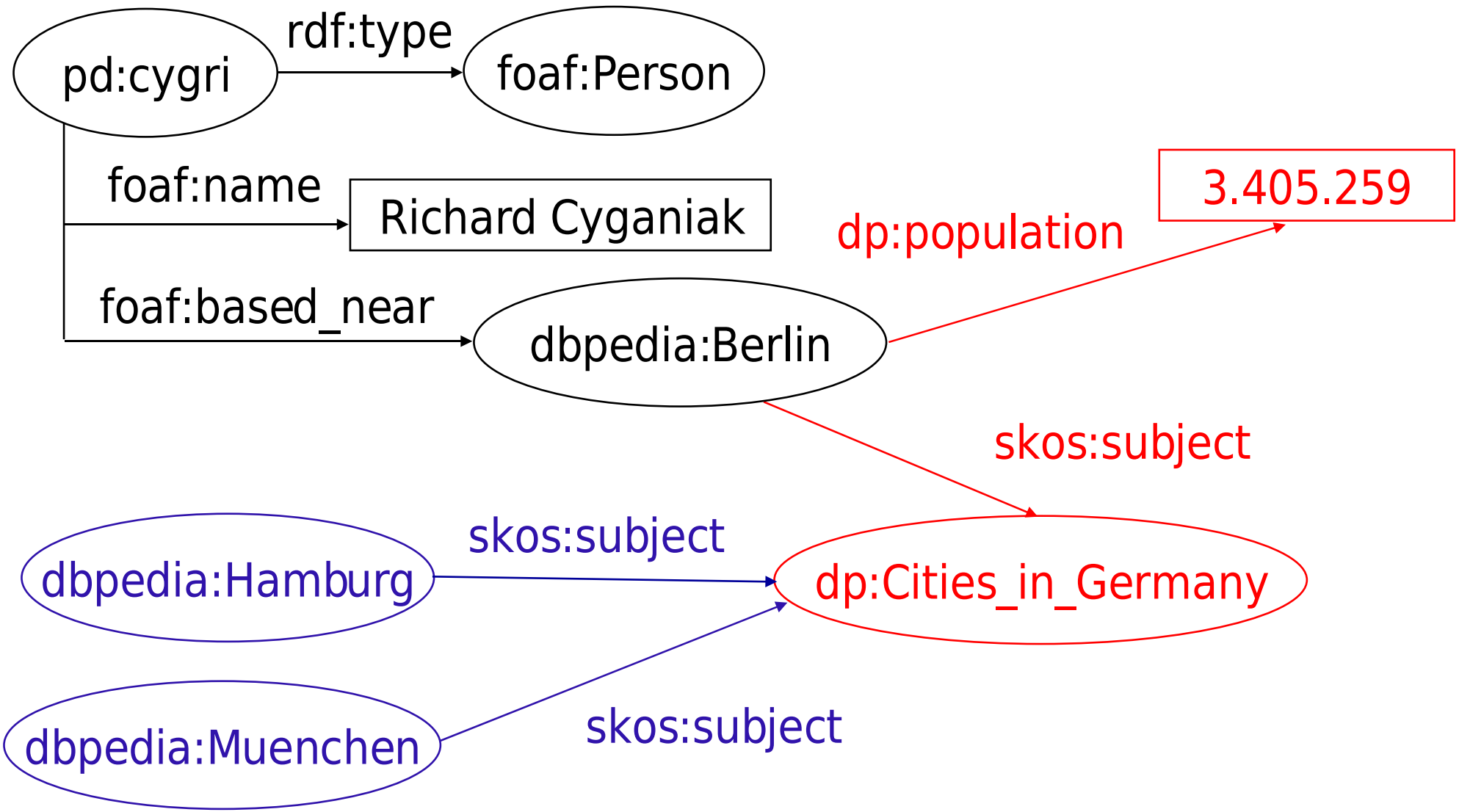
pd:cygri = http://richard.cyganiak.de/foaf.rdf#cygri

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

Dereferencing URIs over the Web



Dereferencing URIs over the Web



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)

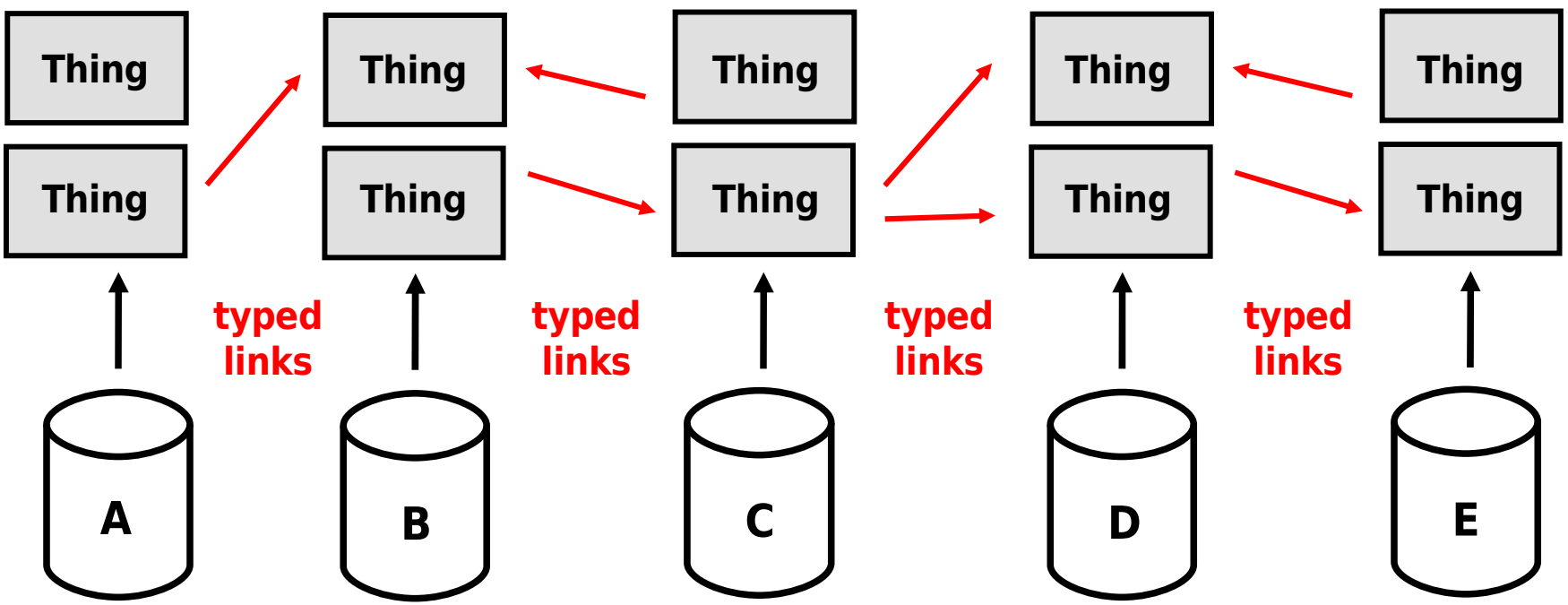
Berlin

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

2. Linked Data Deployment on the Web

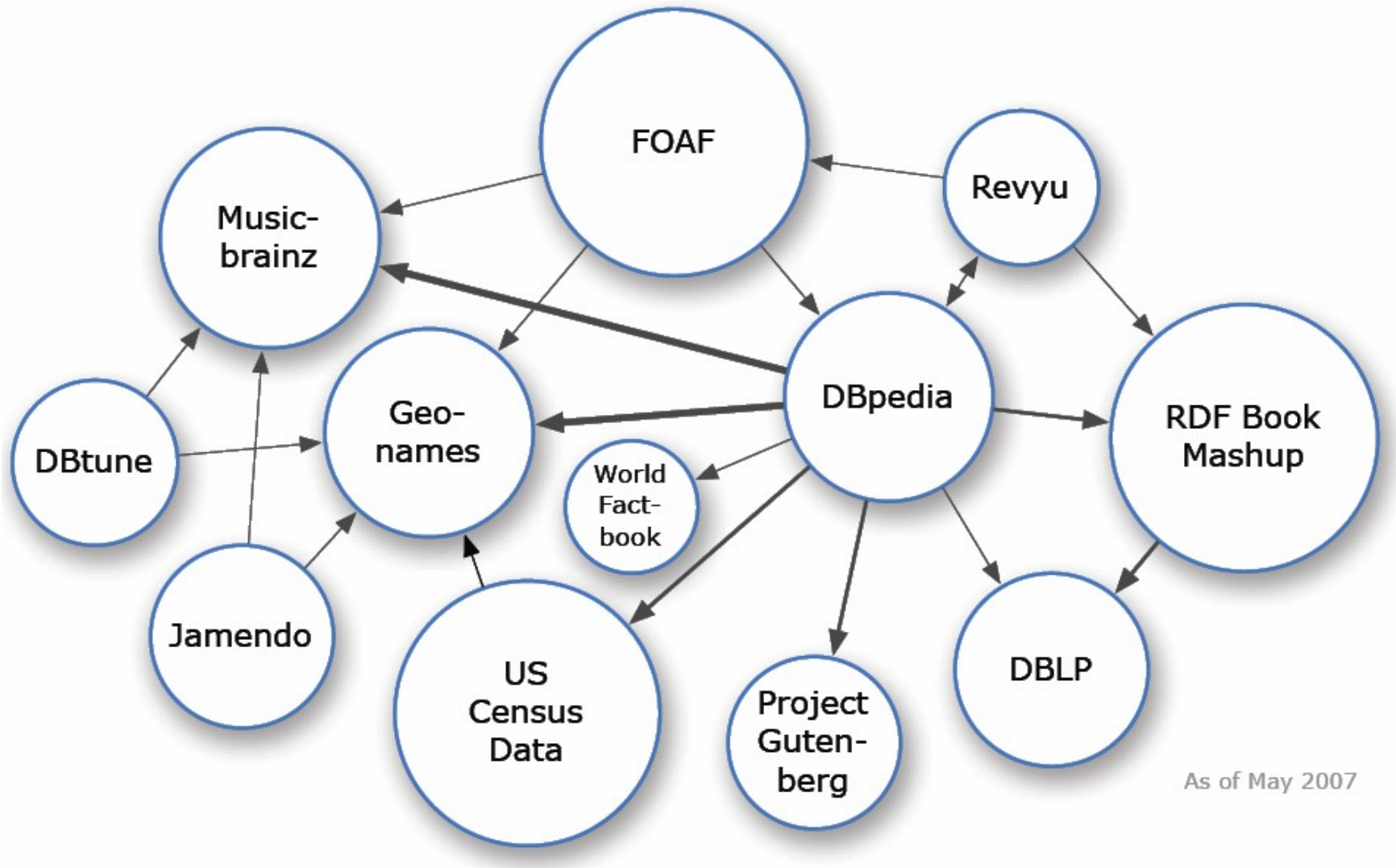
- **Is this real?**



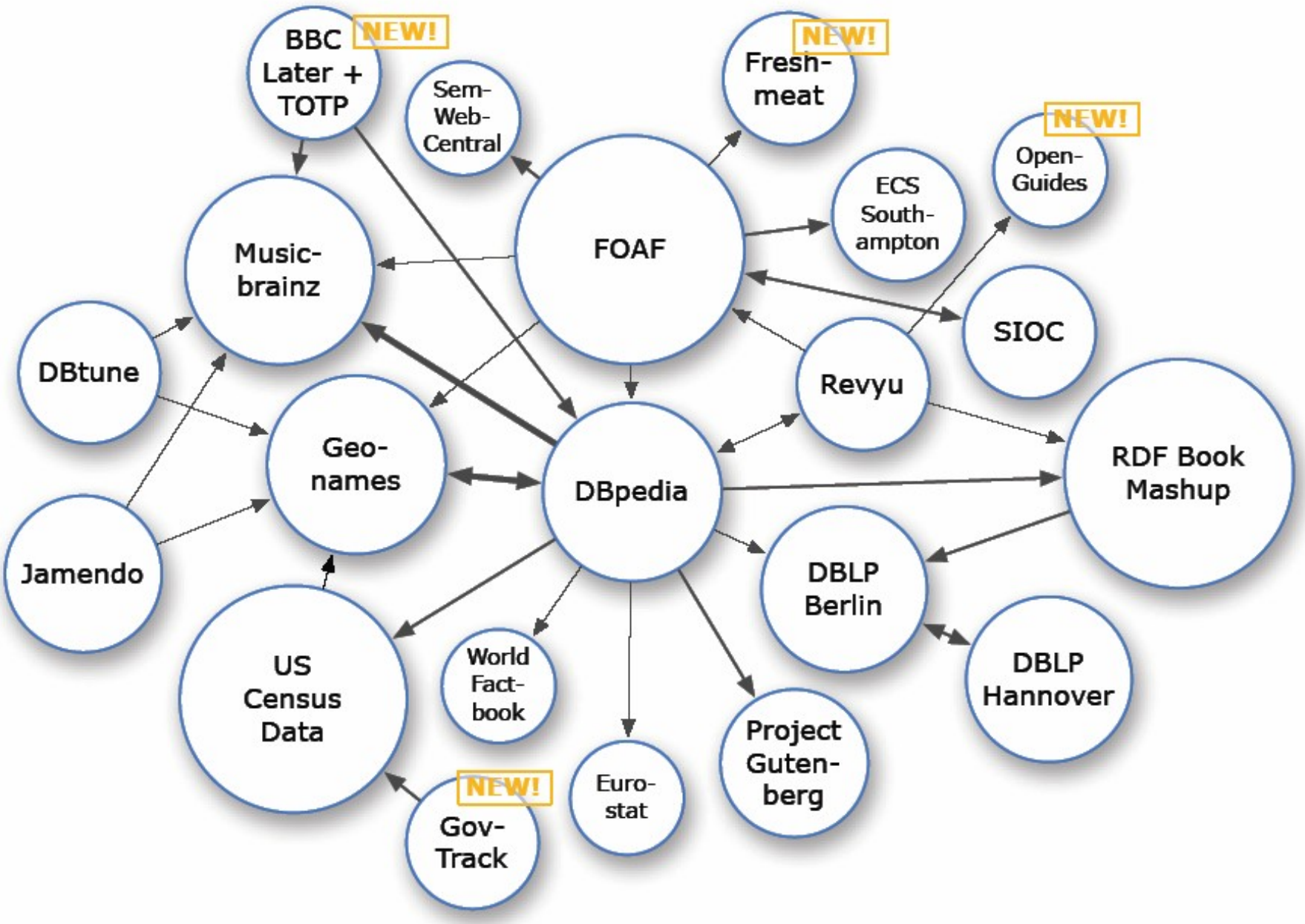
W3C Linking Open Data Project

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

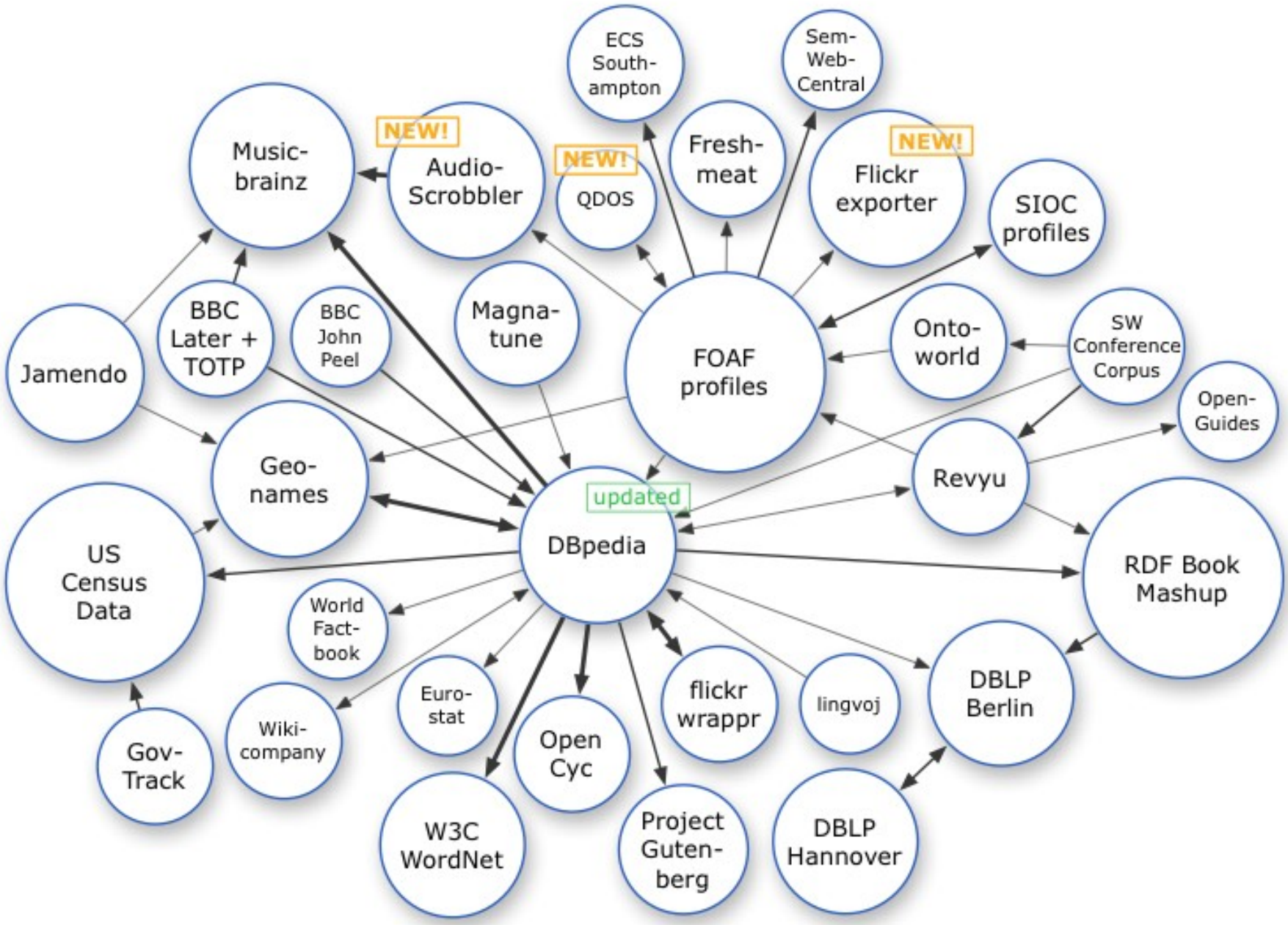
LOD Datasets on the Web: May 2007



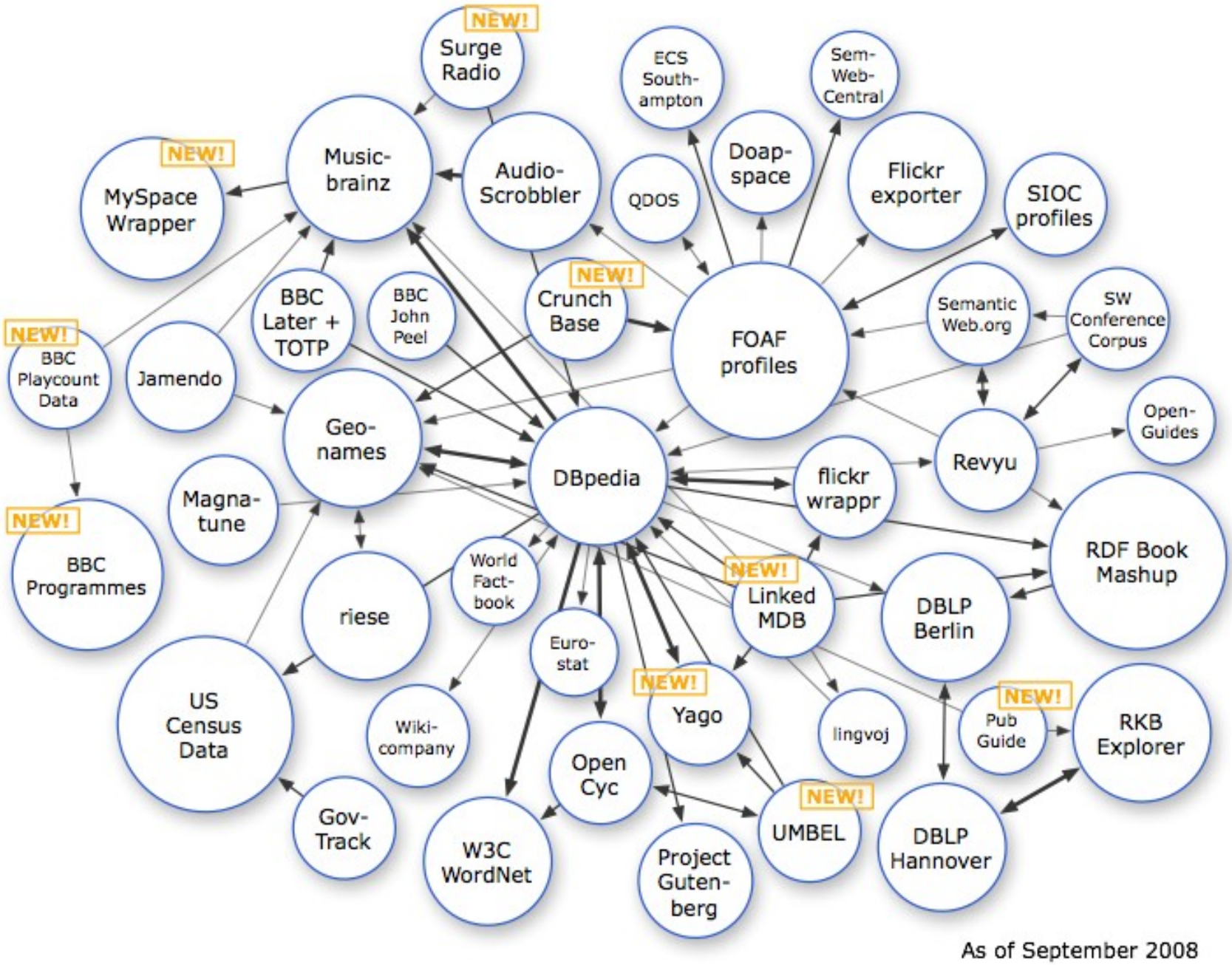
LOD Datasets on the Web: August 2007



LOD Datasets on the Web: February 2008



LOD Datasets on the Web: September 2008



Spotlight: Geonames

- over 8 million geographical locations
- feature hierarchy

GeoNames - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.geonames.org/maps/google_52.373_4.9.html

GeoNames Home | Postal Codes | Download / Webservice | About

search login

Map center : N 52° 22' 13" E 4° 54' 3"

ExploreOurPla.net google earth tagzania mapquest

Map Satellite Hybrid Terrain

GeoNames Wikipedia

features
- city, village,...
- mountain,hill,rock,...
- stream, lake, ...
- country, state, region,...
- parks,area, ...
- road, railroad
- spot, building, farm
- forest,heath,...
- undersea

only 50 objects displayed, zoom in or deselect some features

	Name	country	feature	km to center
1	Amsterdam Centraal Station	Netherlands	railroad station	0.96 km
2	Anne Frank House	Netherlands	museum	1.25 km
3	Jordaan	Netherlands	section of populated place	1.26 km
4	Gemeente Amsterdam	Netherlands	second-order administrative division	1.26 km
5	Het IJ	Netherlands	navigation channel	1.44 km
6	Onze Lieve Vrouwe Gasthuis Locatie Oosterpark	Netherlands	clinic	1.68 km
7	Oosterpark	Netherlands	park	1.73 km
8	Dijksgracht	Netherlands	docking basin	1.79 km
9	Westerkanaal	Netherlands	canal	1.87 km
10	Singelgracht	Netherlands	canal	1.87 km

Done

Spotlight: DBpedia

- extracts structured data from Wikipedia.
- covers over 2.2 million concepts from various domains.

Calgary

Downtown Calgary.

Government	
- Mayor	Dave Bronconnier (Past mayors)
- Governing body	Calgary City Council
- Manager	Owen A. Tobert
Area [1]	
- City	726.50 km² (280.5 sq mi)
- Metro	5,107.43 km² (1,972 sq mi)
Elevation	1,048 m (3,438.3 ft)
Population (2006)[1]	
- City	988,193
- Density	1,360.2/km² (3,522.9/sq mi)
- Metro	1,079,310
- Population rank	3rd
- Metro rank	5th

```
http://en.wikipedia.org/wiki/Calgary
```

```
<http://dbpedia.org/resource/Calgary>
  dbpedia:native_name "Calgary" ;
  dbpedia:altitude "1048" ;
  dbpedia:population_city "988193" ;
  dbpedia:population_metro "1079310" ;
  mayor_name
        dbpedia:Dave_Bronconnier ;
  governing_body
        dbpedia:Calgary_City_Council ;
 ...
```

Example RDF Links

- **RDF links from DBpedia to other data sources**

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

```
<http://dbpedia.org/resource/Tim_Berners-Lee> owl:sameAs
<http://www4.wiwiss.fu-berlin.de/dblp/resource/person/100007> .
```

- **RDF link from a FOAF profile to DBpedia**

```
<http://richard.cyganiak.de/foaf.rdf#cygri> foaf:topic_interest
<http://dbpedia.org/resource/Semantic_Web> .
```

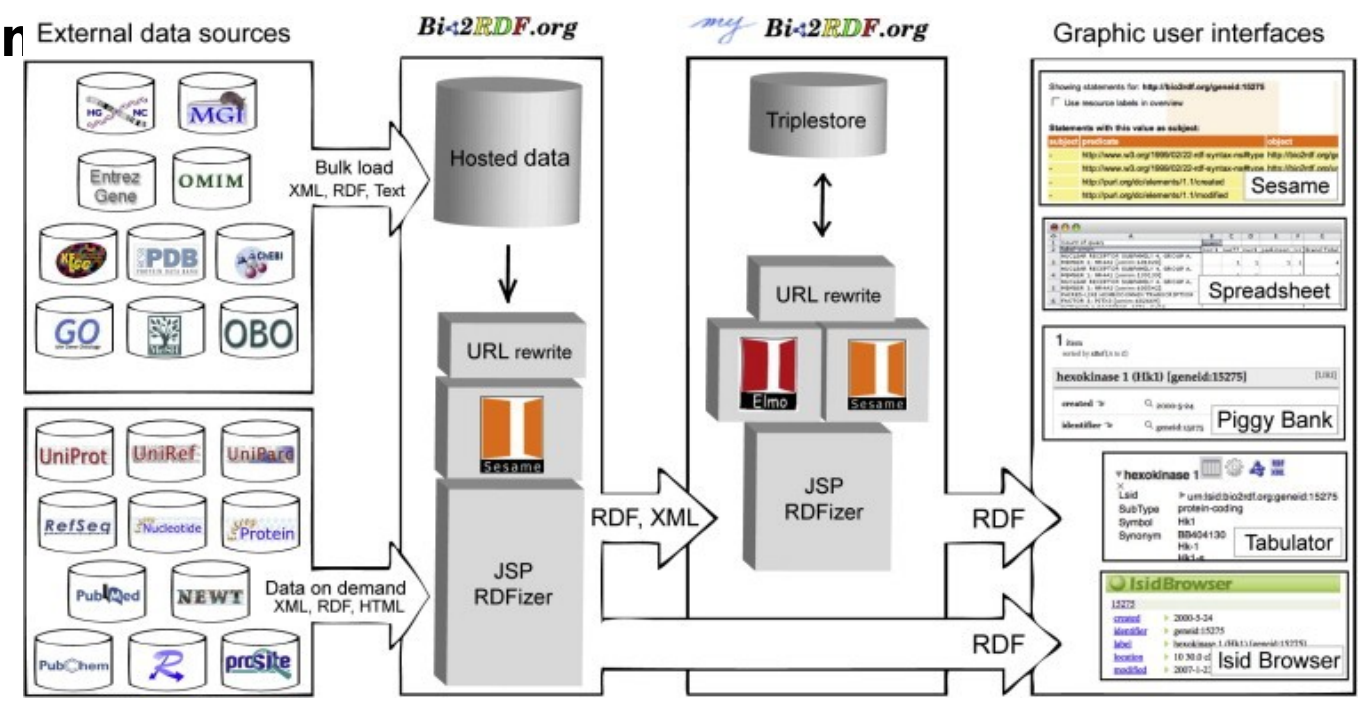
Organizations publishing Linked Data

- **Universities and Research Institutes**
  - Massachusetts Institute of Technology (USA)
  - University of Southampton (UK)
  - Freie Universität Berlin (DE)
  - DERI (IRE)
  - KMi, Open University (UK)
  - University of London (UK)
  - Universität Hannover (DE)
  - University of Pennsylvania (USA)
  - Universität Leipzig (DE)
  - Universität Karlsruhe (DE)
  - Joanneum (AT)
  - University of Toronto (CA)
- **Companies**
  - BBC (UK)
  - OpenLink (UK)
  - Zitgist (USA)
  - Talis (UK)
  - Garlik (UK)
  - Mondeca (FR)
  - Cyc Foundation (USA)

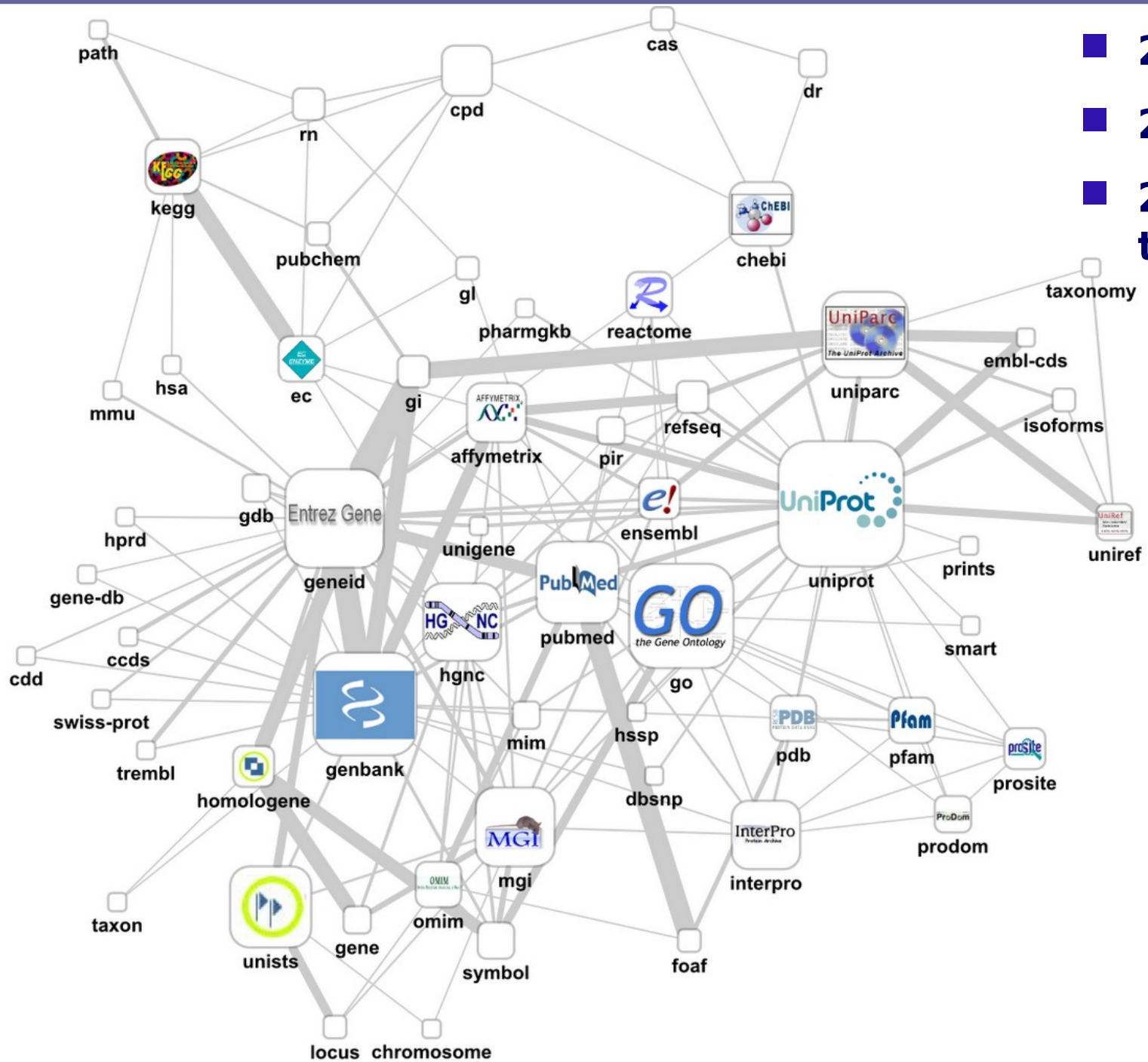
The Bio2RDF Project

- **Goals**
  1. **Make bioinformatics data available in RDF format on the Web.**
  2. **Promote the linked data vision within the bioinformatics community.**
  3. **Answer questions which were not possible or practical to ask before.**

- **Participants**
  - **Université Laval, Canada**
  - **Queen**



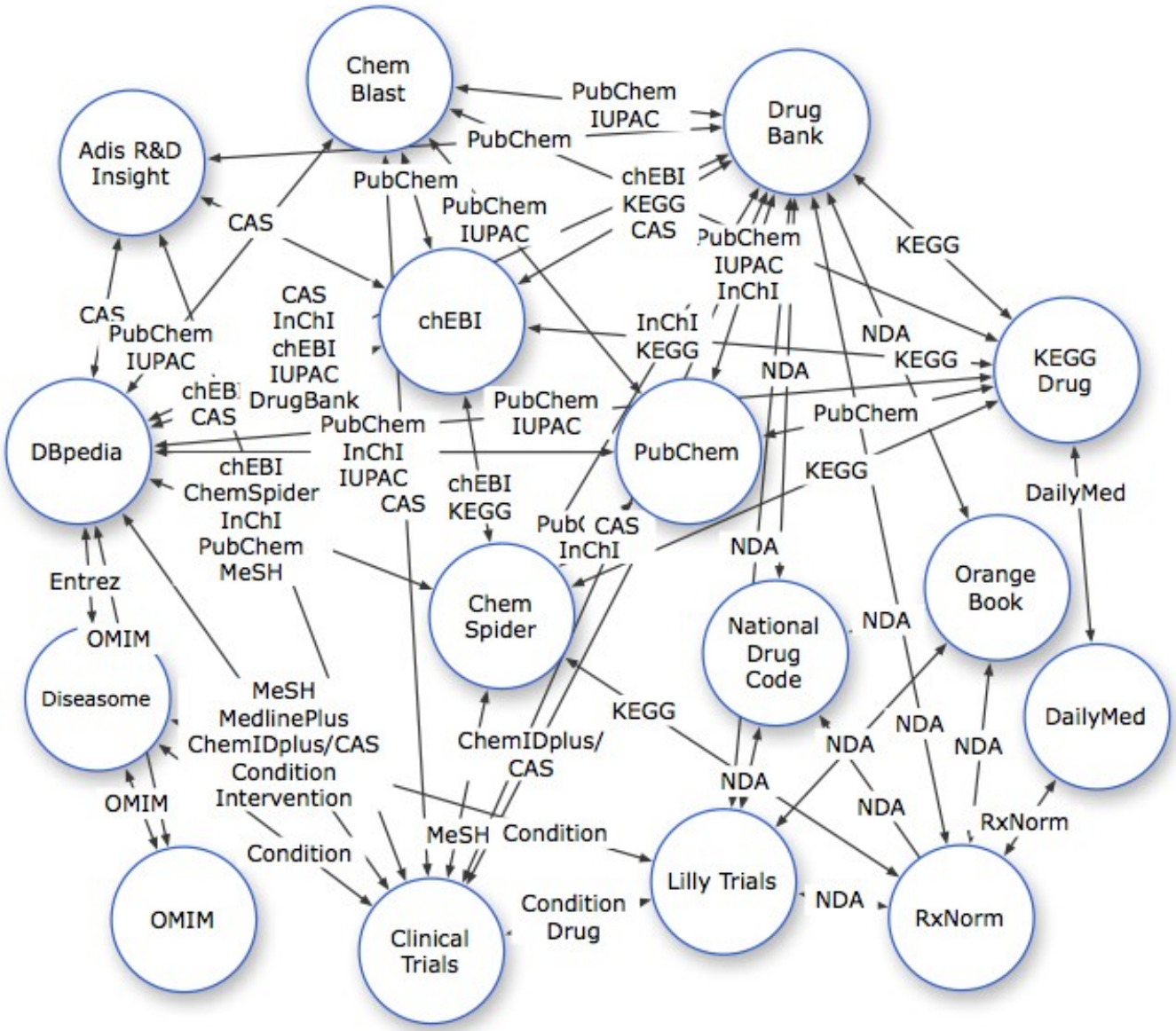
The Bio2RDF Cloud



- 27 data sources
- 260 million records
- 2,7 billion RDF triples

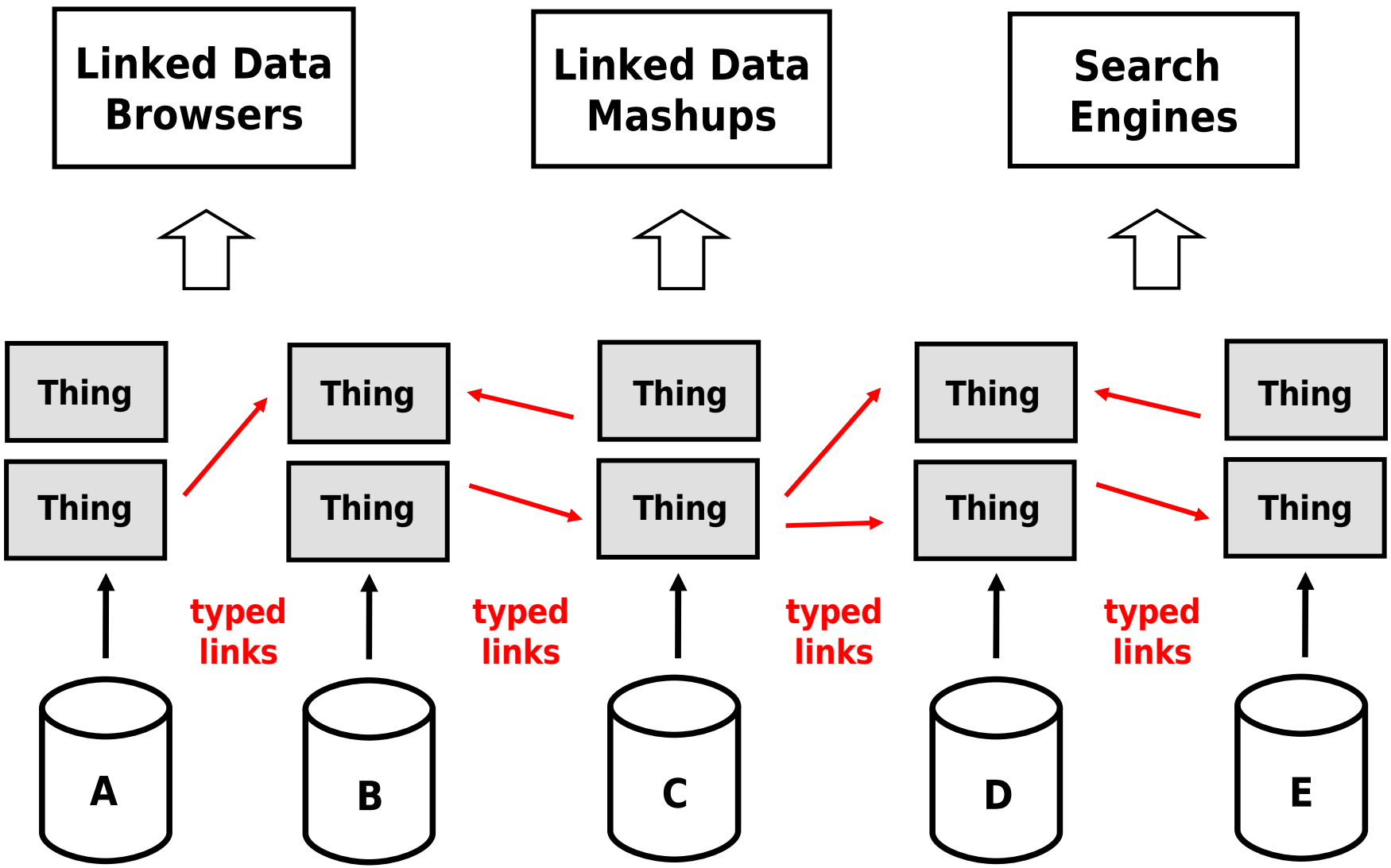
The Linking Open Drug Data Effort

- W3C HCLSIG task started October 1st, 2008
- Goal: Publish and interlink data sets about drugs and clinical trials.



3. Applications

- What can I do with this?



Linked Data Browsers

- Tabulator Browser (MIT, USA)
- Marbles (FU Berlin, DE)
- OpenLink RDF Browser (OpenLink, UK)
- Zitgist RDF Browser (Zitgist, USA)
- Disco Hyperdata Browser (FU Berlin, DE)
- Fenfire (DERI, Irland)

URI: http://www.w3.org/People/Berners-Lee/card#i Add to Outliner Help About

This is release 0.8 of the Tabulator Project. The live development trunk is also available.

Tim Berners-Lee

request	Request for <http://www4.wiwiss.fu-berlin.de/bookmashup/persons/Tim+Berners-Lee>
	Request for <http://www4.wiwiss.fu-berlin.de/dblp/resource/person/100007>
requested by	Tim Berners-Lee's FOAF file
type	Male
	Person
label	Tim Berners-Lee
seeAlso	RDF document describing: Tim Berners-Lee
	RDF Description of Tim Berners-Lee
assistant	Amy van der Hiel
homePage	http://www.w3.org/People/Berners-Lee/
work	...
organization	**World Wide Web Consortium**
	title: W3C (large: includes all specs and groups)
	type: SocialEntity
	Organization
	label: W3C
	seeAlso: W3C Groups and Organizational Structure
	W3C Standards and Technical Reports
	publicHomePage: http://www.w3.org/
	homepage: http://www.w3.org/
	logo:
	name: World Wide Web Consortium
	is organization of: Tim Berners-Lee
publicHomePage	http://www.w3.org/People/Berners-Lee/

http://www.w3.org/People/Berners-Lee/card#i Open

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 www6.wowsa.tu-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	• 985c47c5a70db7407210cef8e4e6f5374a525c5c
workplace homepage	• http://www.w3.org/
nickname	• TimBL
nickname	• TimBL • timbl

Linked Data Mashups

- **Domain-specific applications using Linked Data from the Web**

Revyu

- **Website for rating everything**
- **Uses Linked Data to augment ratings**

Broken Flowers

Links

Homepage: http://www.brokenflowersmovie.com/
See Also: http://en.wikipedia.org/wiki/Broken_flowers

Tags

bill-murray film jessica-lange jim-jarmusch julie-delpy movie sharon-stone

Reviews (1)

★★★★★ by tom on 30 Jan 2007

Broken Flowers provides a fantastic vehicle for a classic deadpan Bill Murray performance. The film centers around his character Don, who one day receives a letter from an ex-girlfriend, telling him he has a teenage son. The letter is unsigned, so (with encouragement from his neighbour) he sets off round the country, visiting each the exes who could be the mother of his son. Predictably they're all different in personality and life situation, giving plenty of raw material for awkward silences and dubious encounters. This is great viewing for any Bill Murray fans, or anyone who likes their humour intelligent and a little bit quirky. The soundtrack is also excellent, and deserves a separate review.

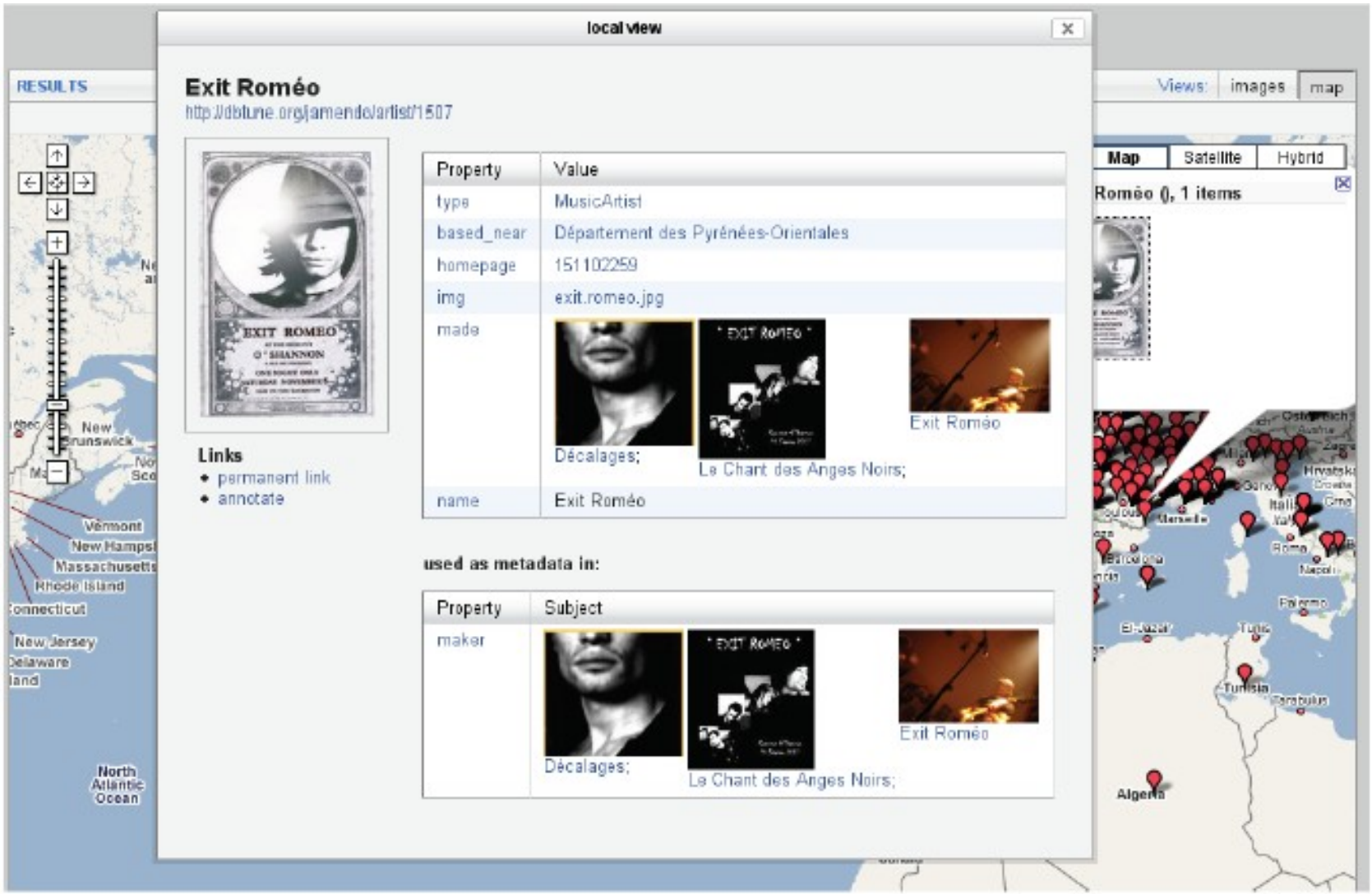
What do you think of **Broken Flowers**? Write Your Own Review...

Broken Flowers

directed by Jim Jarmusch

DBtune Slashfacet

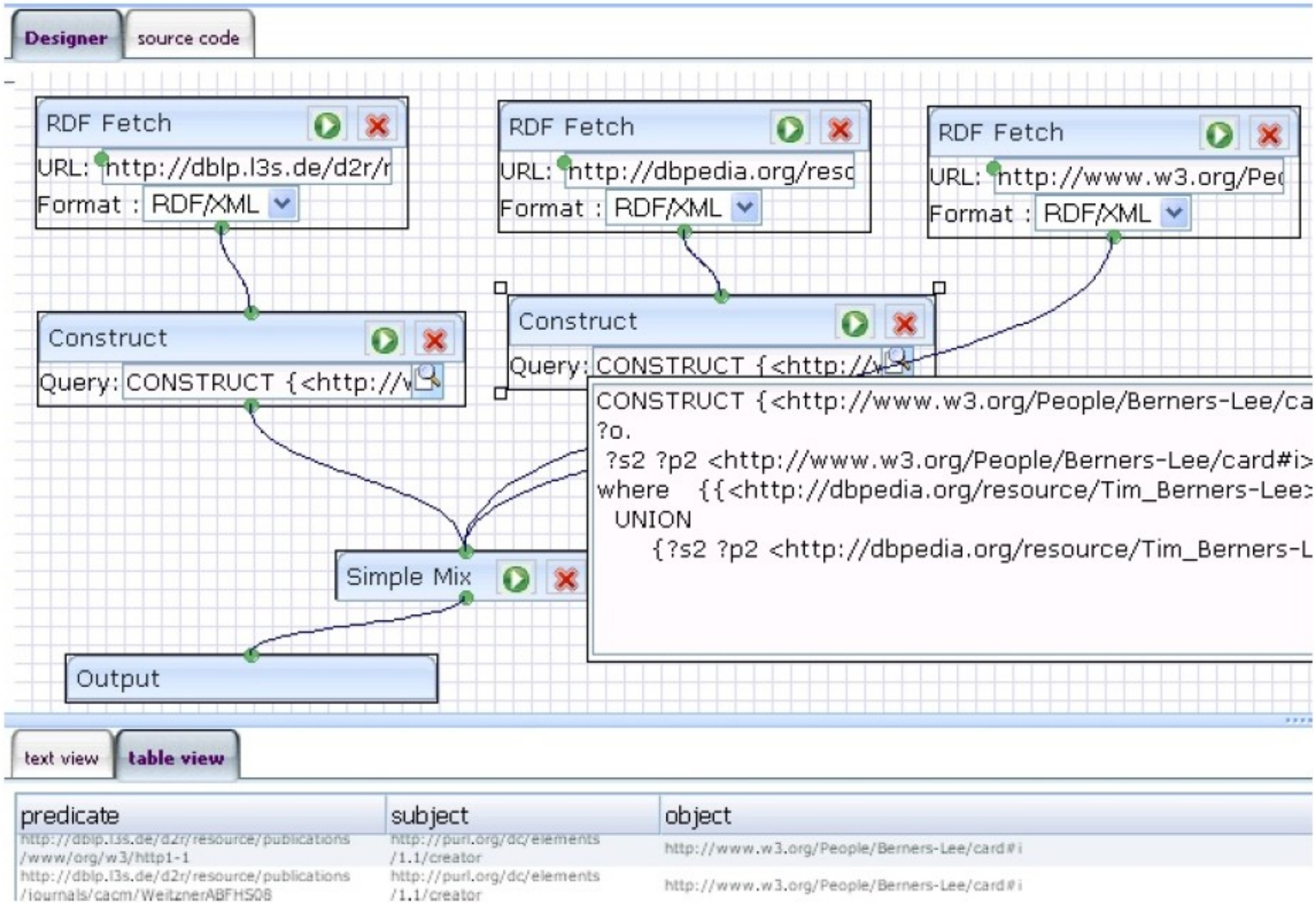
- Visualizes music-related Linked Data
- Uses LastFM, MySpace, and BBC data



DBpedia Mobile

- **Geospatial entry point into the Web of Data**
- **Starts with DBpedia, Revyu and Flickr data**

Semantic Web Pipes



Web of Data Search Engines

- Falcons (IWS, China)
- Sindice (DERI, Ireland)
- MicroSearch (Yahoo, Spain)
- Watson (Open University, UK)
- SWSE (DERI, Ireland)
- Swoogle (UMBC, USA)

Falcons

Object Search Concept Search

Beijing Search Objects

Supports Boolean operators, quotes, and wildcard characters.

All

Artifact Capital City City Document Group
Institution Landmark Location Noun Synset Ontology
Organization Person Publication Subject System

Objects **1** - **10** of **8634** for your search **Beijing** (**1.223** seconds)

Beijing
Types: Capital, City
Labels: 北京" || Pekin || Пекин" || 北京市" || Pequim || Pechino || **Beijing** || Pékin" || Peking || Pekin"
http://dbpedia.org/resource/**Beijing** - Described in 184 documents

Beijing
Types: Subject,
Labels: **Beijing**
http://ontoworld.org/wiki/Special:URIResolver/**Beijing** - Described in 11 documents

Beijing Guoan
Types: Club
Labels: **Beijing** Hyundai || 北京国安" || 北京国安足球俱乐部" || **Beijing** Guoan
http://dbpedia.org/resource/**Beijing**_Guoan - Described in 30 documents

Beijing
The declaration of this URI may be unauthorized.
Types: Capital City
Labels: **Beijing**
http://lonely.org/russia#**Beijing** - Described in 5 documents

Sindice

http://www.w3.org/People/Berners-Lee/card#i

Search

http://www.w3.org/People/Berners-Lee/card#i mentioned in 33 documents:

Tim Berners-Lee's FOAF file [+]
2008-02-03 – 236 triples – 18.9 kb
http://foaf-visualizer.org/?uri=http://www.w3.org/People/... search

http://pipes.deri.org:8080/pipes/Pipes/?id=simplemix [+]
2008-02-24 – 277 triples – 27.3 kb
http://pipes.deri.org:8080/pipes/Pipes/?id=simplemix search

Tim Berners-Lee's FOAF file [+]
2008-04-11 – 233 triples – 18.1 kb
http://www.w3.org/People/Berners-Lee/card.rdf search

Tim Berners-Lee's FOAF file [+]
2008-01-20 – 234 triples – 18.7 kb
http://foaf-visualizer.org/?uri=http://www.w3.org/People/... search

Tim Berners-Lee's FOAF file [+]
2007-11-09 – 613 triples – 17.3 kb
http://www.w3.org/People/Berners-Lee/card search

http://dannyayers.com/me.rdf [+]
2007-12-01 – 164 triples – 9.6 kb
http://dannyayers.com/me.rdf search

http://public.dbin.org/usersaccount/5872362/adamtest.rdf [+]
2007-11-15 – 37 triples – 3.8 kb
http://public.dbin.org/usersaccount/5872362/adamtest.rdf search

Why publish Linked Data on the Web?

- **Linked Data builds on the classic architecture of the Web.**
  - Your data becomes part of a single global data space (the Web of data aka Semantic Web).
  - People can use various data browsers to explore your data.
  - Your data is crawled by Semantic Web search engines and is used by various applications.
  - People start setting links to your data, which might make more people find and use your data.
- **Linked Data is more generic then WebAPIs and Microformats.**
  - Builds on standards in contrast to proprietary Web APIs
  - Enables applications that work against an unbound set of data sources and incorporate new data sources as they become available on the Web.

Publishing Linked Data on the Web

Making a FOAF File into Linked Data

Making a FOAF File into Linked Data

- http://www.ldodds.com/foaf/foaf-a-matic

The Forms

Personal

Some information about you, and how people can contact you.

Title (Mr, Mrs, Dr, etc)
First Name
Last Name (Family/Given)
Nickname
Your Email Address
Homepage
Your Picture
Phone Number

Work

Information about the place you work.

Work Homepage
Page describing what you do at work

School

Where did you go to school?

School Homepage

People You Know

Tell FOAF-a-matic about some people you know. Click "Add Friend" to add space to add more people. If you friend already has a FOAF description, then place a link to it in the 'seeAlso' field.

Friend-- Name Email See Also
Friend-- Name Email See Also
Friend-- Name Email See Also

Making a FOAF File into Linked Data

```
<foaf:knows>
  <foaf:Person>
    <foaf:name>Michael Hausenblas</foaf:name>
    <foaf:mbox_sha1sum>636480acf3cca05e96e612e5e6da6090efd5c71f</foaf:mbox_sha1sum>
    <rdfs:seeAlso rdf:resource="http://www.sw-app.org/foaf/mic.rdf"/>
  </foaf:Person>
</foaf:knows>

<foaf:knows>
  <foaf:Person>
    <foaf:name>Chris Bizer</foaf:name>
    <foaf:mbox_sha1sum>50c02ff93e7d477ace450e3fbddd63d228fb23f3</foaf:mbox_sha1sum>
  </foaf:Person>
</foaf:knows>

<foaf:knows>
  <foaf:Person>
    <foaf:name>Richard Cyganiak</foaf:name>
    <foaf:mbox_sha1sum>39f3c9b7479a83c76596a7c92b61f76dee3f5343</foaf:mbox_sha1sum>
    <rdfs:seeAlso rdf:resource="http://richard.cyganiak.de/foaf.rdf"/>
  </foaf:Person>
</foaf:knows>
```

Making a FOAF File into Linked Data

- Adding URIs for People

Making a FOAF File into Linked Data

- Adding URIs for People

```
<foaf:knows>
  <foaf:Person rdf:about=”http://sw-app.org/foaf/mic.rdf#me”>
    <foaf:name>Michael Hausenblas</foaf:name>
    <foaf:mbox_sha1sum>636480acf3cca05e96e612e5e6da6090ef
    <rdfs:seeAlso rdf:resource="http://sw-app.org/foaf/mic.rdf"/>
  </foaf:Person>
</foaf:knows>
```

Making a FOAF File into Linked Data

- Adding URIs for People

```
<foaf:knows>
  <foaf:Person rdf:about="http://semanticweb.org/id/Chris_Bizer">
    <foaf:name>Chris Bizer</foaf:name>
    <foaf:mbox_sha1sum>50c02ff93e7d477ace450e3fbddd63d228
  </foaf:Person>
</foaf:knows>
```

Making a FOAF File into Linked Data

- Enriching Your Profile

Making a FOAF File into Linked Data

```
<rdf:RDF
      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
      xmlns:foaf="http://xmlns.com/foaf/0.1/"
      xmlns:admin="http://webns.net/mvcb/">

<foaf:PersonalProfileDocument rdf:about="">
  <foaf:maker rdf:resource="#me"/>
  <foaf:primaryTopic rdf:resource="#me"/>
  <admin:generatorAgent rdf:resource="http://www.ldodds.com/foaf/foaf-a-matic"/>
  <admin:errorReportsTo rdf:resource="mailto:leigh@ldodds.com"/>
</foaf:PersonalProfileDocument>

<foaf:Person rdf:ID="me">
<foaf:name>Tom Heath</foaf:name>
<foaf:givenname>Tom</foaf:givenname>
<foaf:family_name>Heath</foaf:family_name>
<foaf:nick>cerealtom</foaf:nick>
<foaf:mbox_sha1sum>1dd0dab717be578c153b8ed70bee284845439706</foaf:mbox_sha1sum>
<foaf:homepage rdf:resource="http://tomheath.com/home/html"/>
<foaf:workplaceHomepage rdf:resource="http://www.talis.com/"/>
```

Making a FOAF File into Linked Data

- Adding Geodata
  - :me foaf:based_near <http://sws.geonames.org/123456>
- Adding Interests
  - :me foaf:topic_interest <http://dbpedia.org/resource/Semantic_Web>
  - :me foaf:topic_interest <http://dbpedia.org/resource/Whisky>
- Adding Your Other Identities
  - :me owl:sameAs <http://data.semanticweb.org/people/tom-heath>
  - :me owl:sameAs <http://kmi.open.ac.uk/people/tom/>

Publishing Linked Data - Process

1. Understand your Data
2. Publish it on the Web as RDF
3. Link it with other Data Sources

Understanding Your Data

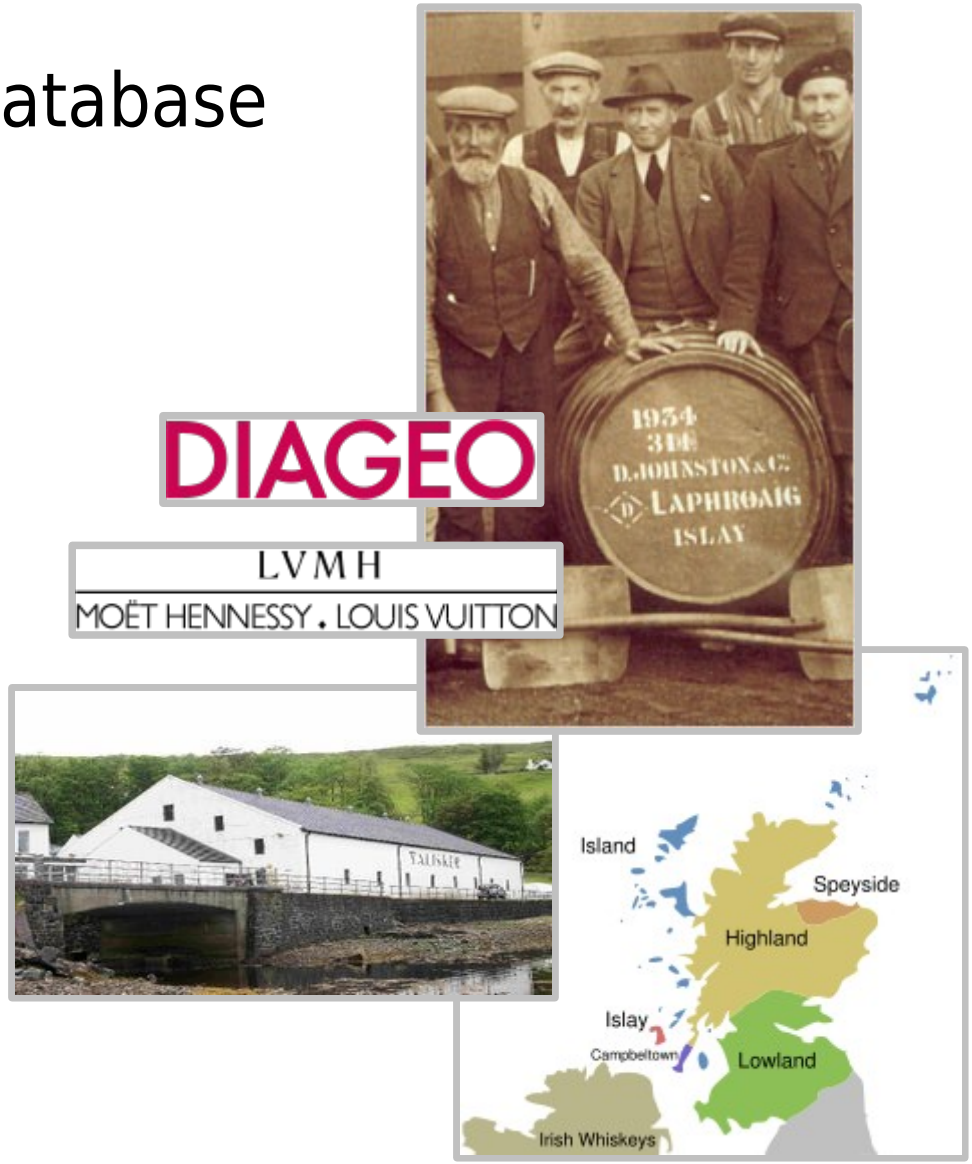
- What are the key entities in the dataset?
- What properties do they have?
- How do they relate to other entities?

The Wiskii.com Scenario

- Online whisky shop: *Wiskii.com*
- New business venture
- For the whisky connoisseur
- Detailed background information from experts
- Contributions from customers
- Custom web app, relational backend
- Simultaneous publication in HTML and RDF

Understanding Your Data

- Things in the *Wiskii.com* database
  - Distilleries
  - Regions and Locations
  - Founders
  - Owners
  - Brands
  - Products
  - Photos
  - Reviews
  - Comments
  - Prices/Offers



Publishing RDF on the Web as Linked Data

Tutorial “How to Publish Linked Data” at ISWC 2008

Richard Cyganiak

Linked Data in 7 Easy Steps

1. Select vocabularies
2. Partition the RDF graph into “data pages”
3. Assign a URI to each data page
4. Create HTML variants of each data page
5. Assign a URI to each entity
6. Add page metadata and link sugar
7. Add a Semantic Sitemap

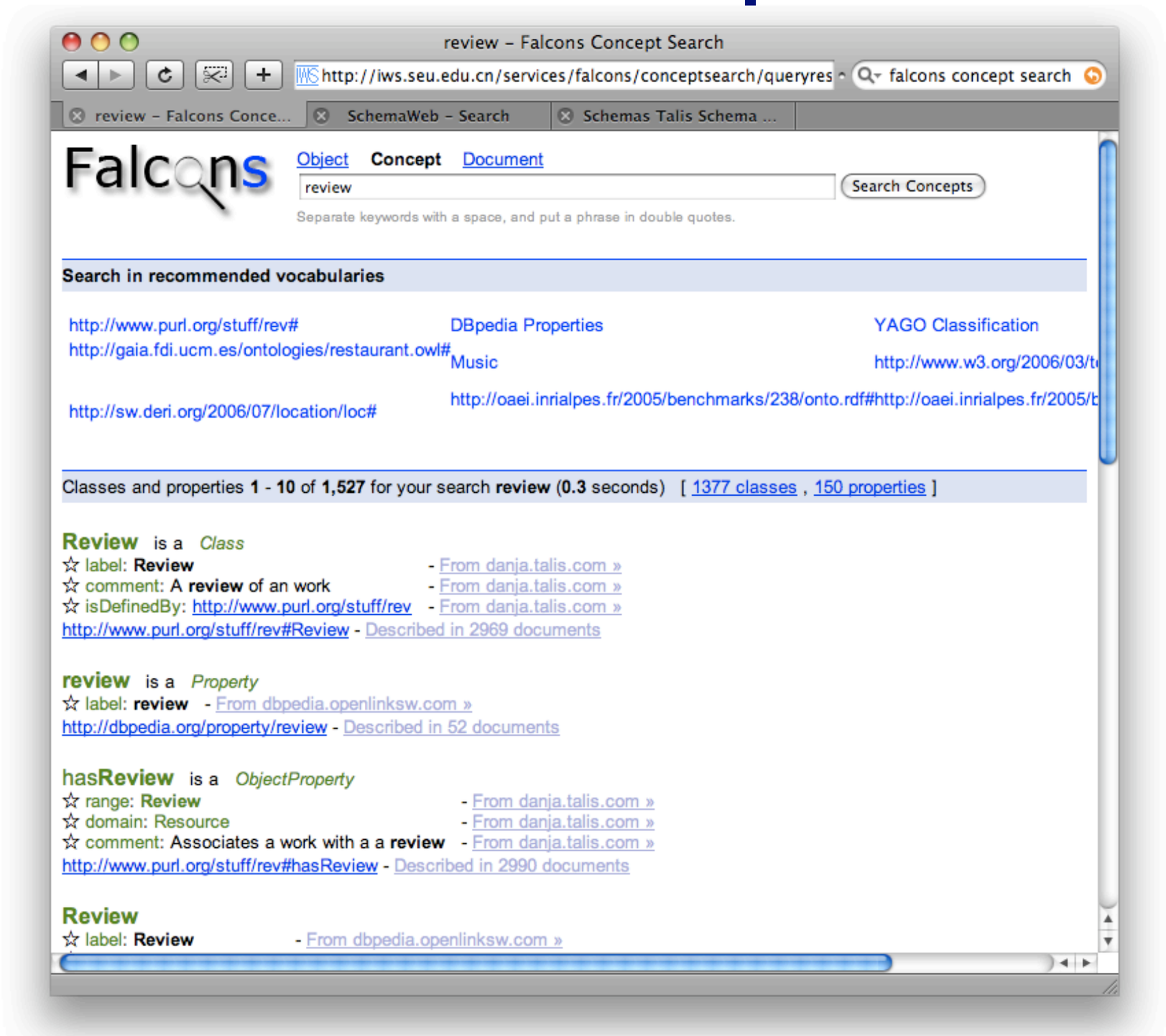
Linked Data in 7 Easy Steps

1. Select vocabularies
2. Partition the RDF graph into “data pages”
3. Assign a URI to each data page
4. Create HTML variants of each data page
5. Assign a URI to each entity
6. Add page metadata and link sugar
7. Add a Semantic Sitemap

Selecting Vocabularies

- To create RDF graph from our data
- Re-use if possible, it makes your data more valuable
- Create your own if re-use not possible
- Be aware of DC, FOAF, SKOS, SIOC
- Expect to mix & match

Falcons Concept Search



SchemaWeb.info

SchemaWeb – Search

http://www.schemaweb.info/search/Search.aspx

schemaweb

review – Falcons Conce... | SchemaWeb – Search | Schemas Talis Schema ...

SchemaWeb

Home | Browse | Search | Query | Web Services | Submit a Schema | Contact | Login

Search SchemaWeb Directory

review [Search]

Search Results

Review Vocabulary
This document provides a vocabulary specification for the purpose of creating sharable reviews and ratings of blogs, CDs, books, software, anything identifiable. The vocabulary is suitable for inclusion in any RDF-compatible language : FOAF, RSS etc. ...

Reviews on the Web
A schema for identifying resources that are reviews of other resources.

Geographic Information Metadata
The metadata ontology was developed from International Standard Organization (ISO) Geographic Information - Metadata publications ISO 19115 . This work is incomplete and continuing towards a more complete representation.

Platform for Privacy Preferences
The Platform for Privacy Preferences Project (P3P) enables Web sites to express their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents. P3P user agents will allow users to be informed of ...

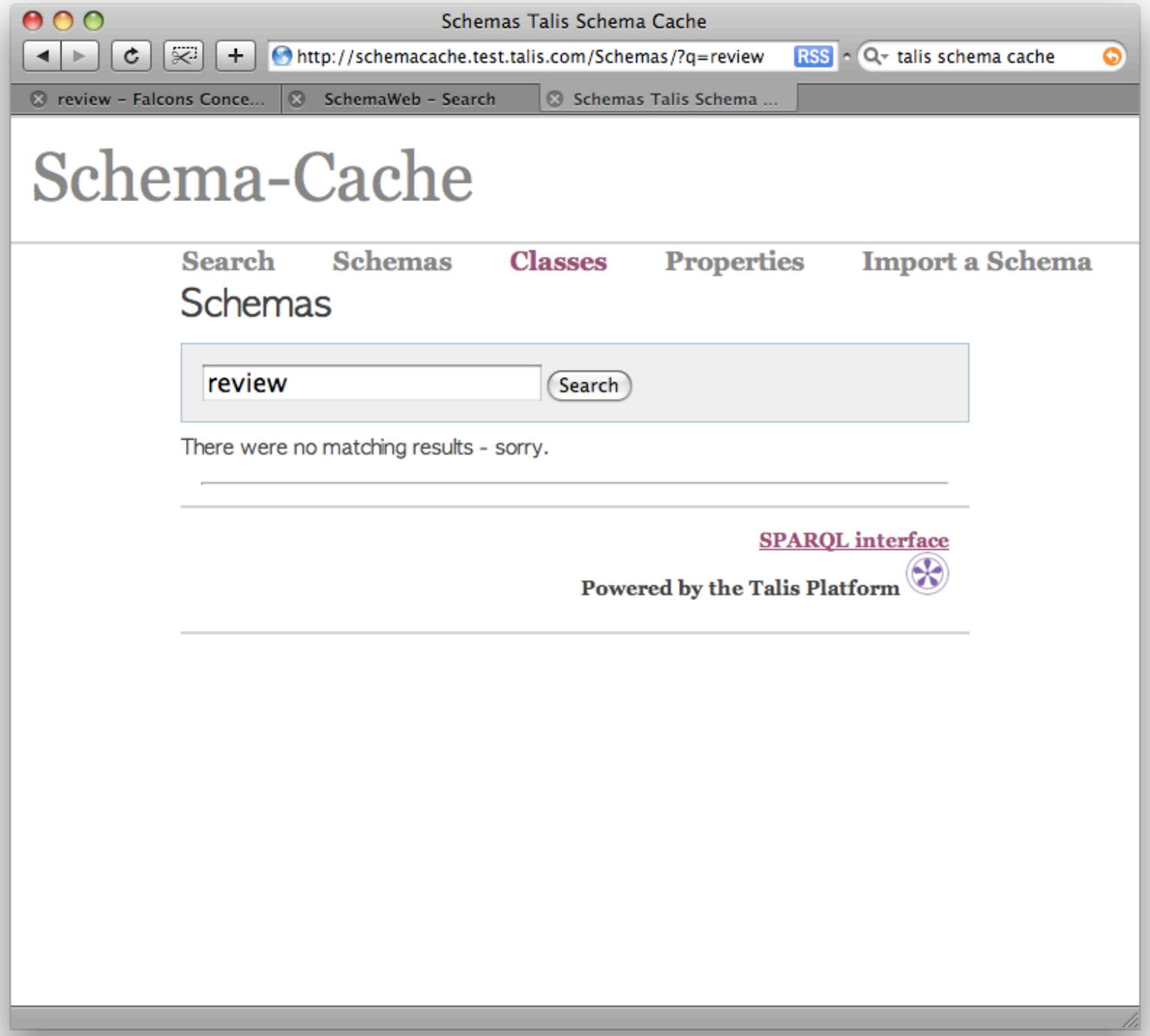
MeNow
The motivation for the MeNow schema is to be able to describe a variety of aspects of the current status of someone, either online or off, in a way that the data can be easily aggregated or retrieved. This schema allows the definition of a variety of ...

GEM Element Set Qualifier Vocabulary v2.0
The GEM Element Set Qualifier Vocabulary is intended to facilitate more precise description of learning objects for purposes of discovery and retrieval than is possible with unqualified GEM. The Gateway to Educational Materials (http://www.geminfo.org/) ...

Band Description Vocabulary
This vocabulary is intended to describe relations between persons, any musical instrument they play and persons who played something together. For the scope of this document singing is equivalent to playing an instrument.

SPASE Ontology
The Space Physics Archive Search and Extract (SPASE) Consortium is an international working group developing a space physics data model for scientific data systems (http://www.spase-group.org). They are working on creating a full dictionary of space physics

Talis Schema-Cache



Spotting good vocabularies

- Existing applications (!)
- Active community
- Good documentation
- Backed by reputable organizations
- Simple
- Few constraints or ontological assumptions

Creating your own

- Stick to what your app needs
- Publish at least an RDFS/OWL file
- Tools: Protégé, Neologism, OpenVocab, …

Linking to existing vocabularies

- rdfs:subClassOf
- rdfs:subPropertyOf
- owl:equivalentClass
- owl:equivalentProperty
- owl:inverseOf

Now we have
an RDF graph
(with blank nodes)

Linked Data in 7 Easy Steps

1. Select vocabularies
2. Partition the RDF graph into “data pages”
3. Assign a URI to each data page
4. Create HTML variants of each data page
5. Assign a URI to each entity
6. Add page metadata and link sugar
7. Add a Semantic Sitemap

Partitioning into “data pages”

- Put the graph online as RDF document(s)
- Huge graph = huge document?
- Hypertext principle: split into sections, interlink them

How to split

- Everything in one document?
- One document per entity?
- Should some entities be grouped together?
- Consider access time, ease of updates, ease of backend access, total # of requests to answer user question

If you already have HTML pages, use the same granularity for the data pages.

Linked Data in 7 Easy Steps

1. Select vocabularies
2. Partition the RDF graph into “data pages”
3. Assign a URI to each data page
4. Create HTML variants of each data page
5. Assign a URI to each entity
6. Add page metadata and link sugar
7. Add a Semantic Sitemap

URIs for data pages

- To put each data page online as RDF doc
- Like web pages, but serve RDF
- E.g. http://wiskii.com/brand/talisker/about.rdf
- “Cool URIs” – stable, no implementation cruft
- ~~http://wiskii.com:2020/demos/cgi-bin/resources.php?id=talisker&output=rdf~~

Linked Data in 7 Easy Steps

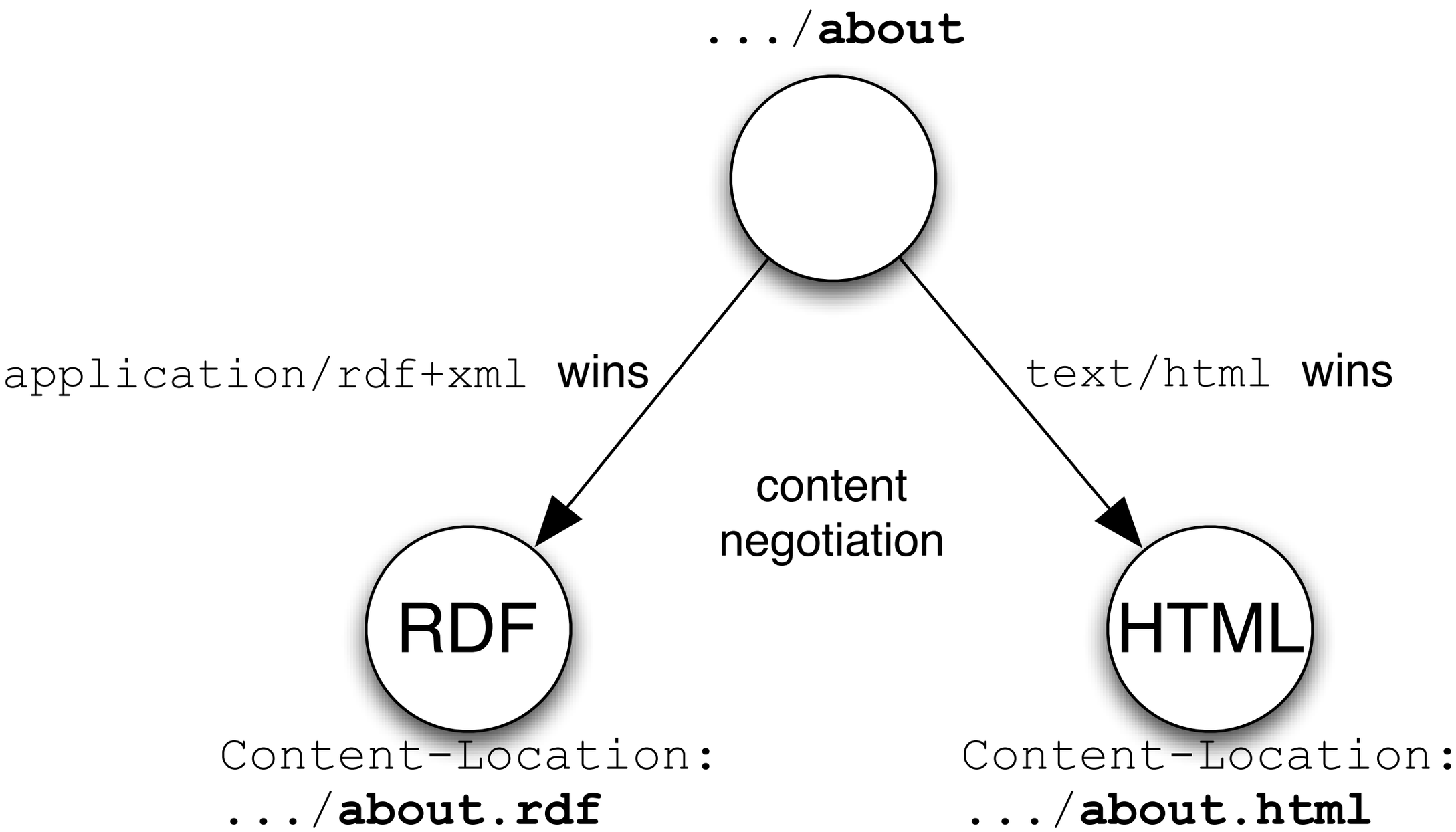
1. Select vocabularies
2. Partition the RDF graph into “data pages”
3. Assign a URI to each data page
4. Create HTML variants of each data page
5. Assign a URI to each entity
6. Add page metadata and link sugar
7. Add a Semantic Sitemap

HTML Variants

- For compatibility with HTML browsers
- HTML rendering of each data page
- *Do we need to add something to the data?*

Content Negotiation

- “generic document” with RDF and HTML variants
- Clients express preferences for formats in *Accept* HTTP header
- Server decides which variant to serve
- Generic document: e.g. .../about
- Format-specific: e.g. .../about.rdf, .../about.html


.../about
application/rdf+xml wins
text/html wins
content
negotiation
RDF
HTML
Content-Location:
.../about.rdf
Content-Location:
.../about.html

HTTP Request/Response

```
GET /brand/talisker/about HTTP/1.0
Host: wiskii.com
Accept: application/rdf+xml
```

```
HTTP/1.0 200 OK
Content-Type: application/rdf+xml
Content-Location:
      http://wiskii.com/brand/talisker/about.rdf

<rdf:RDF xmlns:rdf=....
```

…or put HTML and RDF into one page with RDFa

What we have now

- The RDF graph is online
- In easily digestible chunks
- Chunks can be looked at as RDF or HTML

Linked Data in 7 Easy Steps

1. Select vocabularies
2. Partition the RDF graph into “data pages”
3. Assign a URI to each data page
4. Create HTML variants of each data page
5. Assign a URI to each entity
6. Add page metadata and link sugar
7. Add a Semantic Sitemap

Rules

- Permalinks
- Different URIs for different things
- Can be looked up
- URI ownership – don’t squat URI space

Don’t use these

- http://en.wikipedia.org/wiki/Talisker
- http://wiskii.com/brand/talisker/about.rdf
- http://wiskii.com/brand/talisker/about
- urn:x-wiskii:brand:talisker

Remember, generic document is at

http://wiskii.com/brand/talisker/about

Hash vs. slash

- http://wiskii.com/brand/talisker
  (with HTTP 303 redirect to .../about)
- http://wiskii.com/brand/talisker/about#it
  (#it is removed for lookup)
- Hash is quick and easy
- 303 is future-proof and less cluttered

Linked Data in 7 Easy Steps

1. Select vocabularies
2. Partition the RDF graph into “data pages”
3. Assign a URI to each data page
4. Create HTML variants of each data page
5. Assign a URI to each entity
6. Add page metadata and link sugar
7. Add a Semantic Sitemap

Page metadata

- To help clients understand each data page
- Add some triples to about.rdf
- dc:date, dc:publisher, dc:license
- foaf:primaryTopic, foaf:topic

Link sugar

- Add a bit of information about other entities mentioned in the page
- To support rendering and navigation
- Clients need to make less HTTP requests
- rdfs:label, rdf:type, …
- Redundancy is okay

Linked Data in 7 Easy Steps

1. Select vocabularies
2. Partition the RDF graph into “data pages”
3. Assign a URI to each data page
4. Create HTML variants of each data page
5. Assign a URI to each entity
6. Add page metadata and link sugar
7. Add a Semantic Sitemap

Semantic Sitemaps

- If you publish Linked Data *and* SPARQL endpoint or RDF dump
- Allows crawlers to find dumps and endpoints
- Add a line to `robots.txt`:
  `Sitemap: sitemap.xml`
- Add a file `sitemap.xml`

```
<urlset>
   <sc:dataset>
      <sc:datasetLabel>
            The Wiskii.com dataset
      </sc:datasetLabel>
      <sc:linkedDataPrefix>
            http://wiskii.com/
      </sc:linkedDataPrefix>
      <sc:dataDumpLocation>
            http://downloads.wiskii.com/dump.nt.gz
      </sc:dataDumpLocation>
      <sc:sparqlEndpointLocation>
            http://wiskii.com/sparql
      </sc:sparqlEndpointLocation>
      <changefreq>daily</changefreq>
   </sc:dataset>
</urlset>
```

Publishing Tools

Pubby

- When your data is already in RDF
- Java server in front of SPARQL store

D2R Server

- When your data is in a relational database
- Java server
- Mapping language for describing database-to-RDF mappings
- Provides SPARQL endpoint too

Triplify

- For LAMP applications
- Simple PHP script
- Specify some SQL queries and how the results should be rendered as RDF

Roll your own?

- Build normal HTML site
- Add content negotiation
- Add RDF version of all pages

Linked Data in 7 Easy Steps

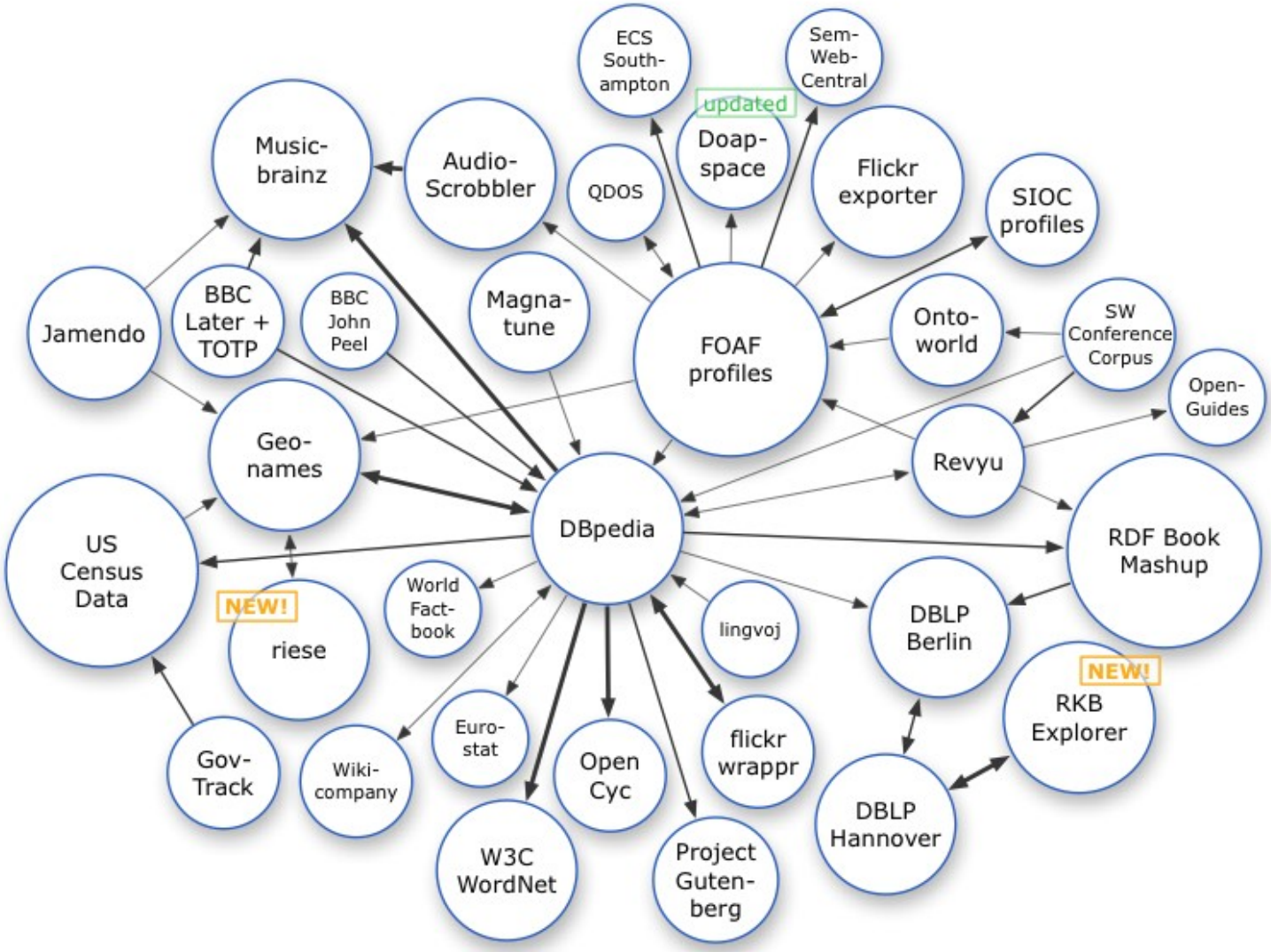
1. Select vocabularies
2. Partition the RDF graph into “data pages”
3. Assign a URI to each data page
4. Create HTML variants of each data page
5. Assign a URI to each entity
6. Add page metadata and link sugar
7. Add a Semantic Sitemap

Revisiting the Principles

- Use URIs as names for things
- Use HTTP URIs
- Provide useful information in RDF
- Include RDF links to other URIs

Linking

Other Available Data Sets



Link to other Data Sets

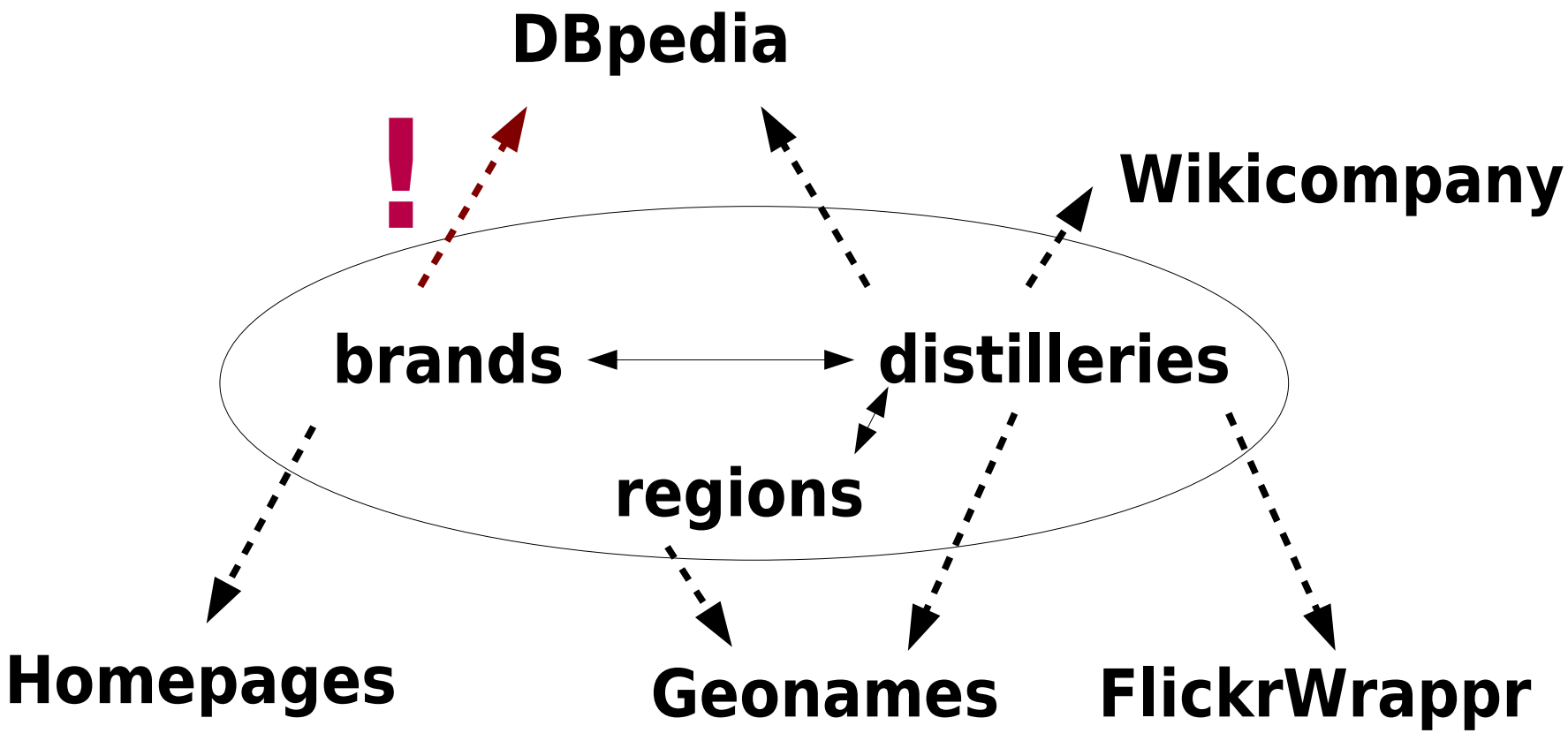
- Popular Predicates for Linking
  - owl:sameAs
  - foaf:homepage
  - foaf:topic
  - foaf:based_near
  - foaf:maker/foaf:made
  - foaf:depiction
  - foaf:page
  - foaf:primaryTopic
  - rdfs:seeAlso

Link to other Data Sets

- Linking Opportunities for *Wiskii.com* Data
  - Distilleries
    - Link to their DBpedia entries
    - Link to their Parent Companies in DBpedia and WikiCompany
    - Link their Locations to Towns in Geonames/DBpedia
    - Link to Photos internally and externally (via FlickrWrapper)
  - Regions
    - Link to DBpedia and Geonames
  - Brands
    - Link to DBpedia entries? *Caution!*
    - Link to external Brand Homepages
  - Reviews
    - Link to Brands and Products internally

Link to other Data Sets

- Linking Opportunities for *Wiskii.com* Data



Link to other Data Sets

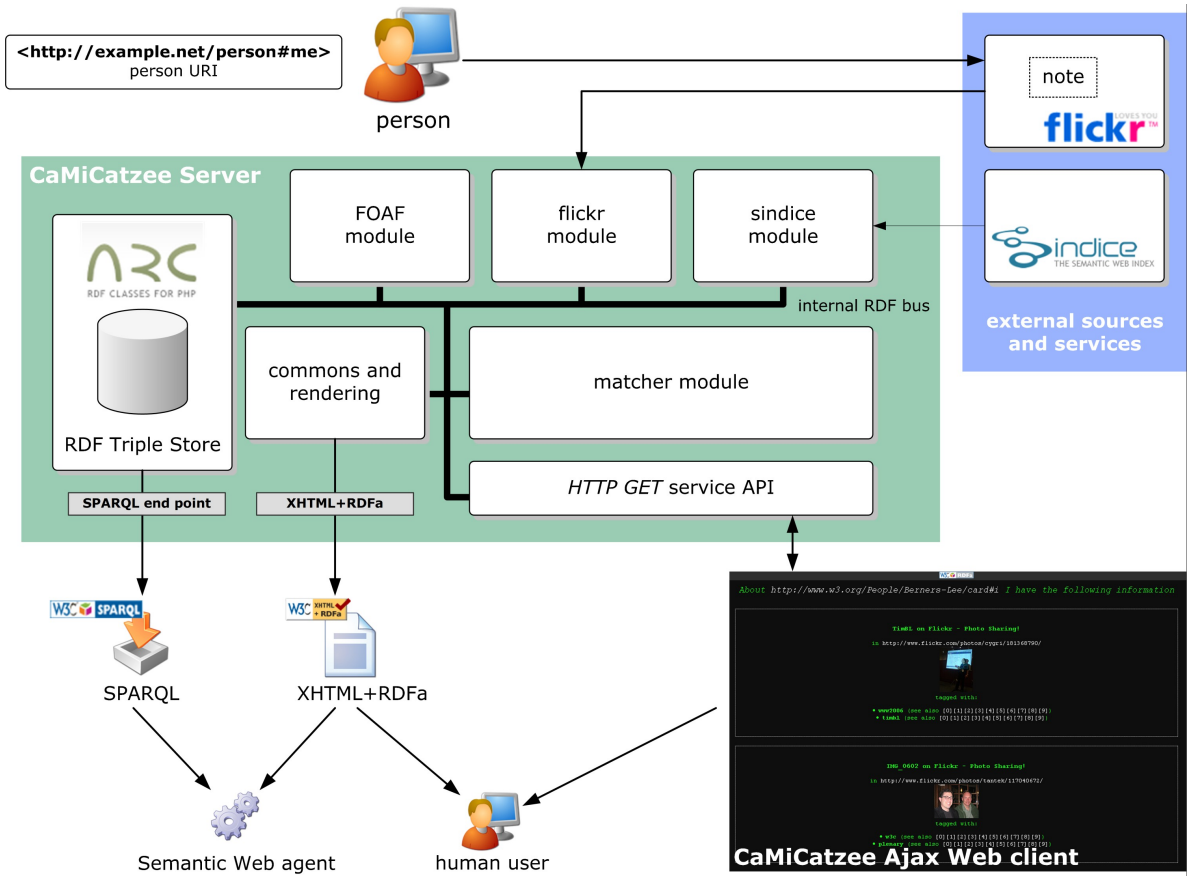
- Linking Algorithms
  - String Matching
  - Common Key Matching
    - e.g. ISBN, Musicbrainz IDs
  - Property-based Matching
    - Do these two things have the same label, type and coordinates

Manual Linking

- just as with Wikis, Tags, GWAPs, etc.: humans are good and willing to contribute high-quality content (semantic links, in our case)
- certain use cases and/or resource types (e.g. multimedia assets with fine-grained spatio-temporal annotations) are good candidates for manual interlinking

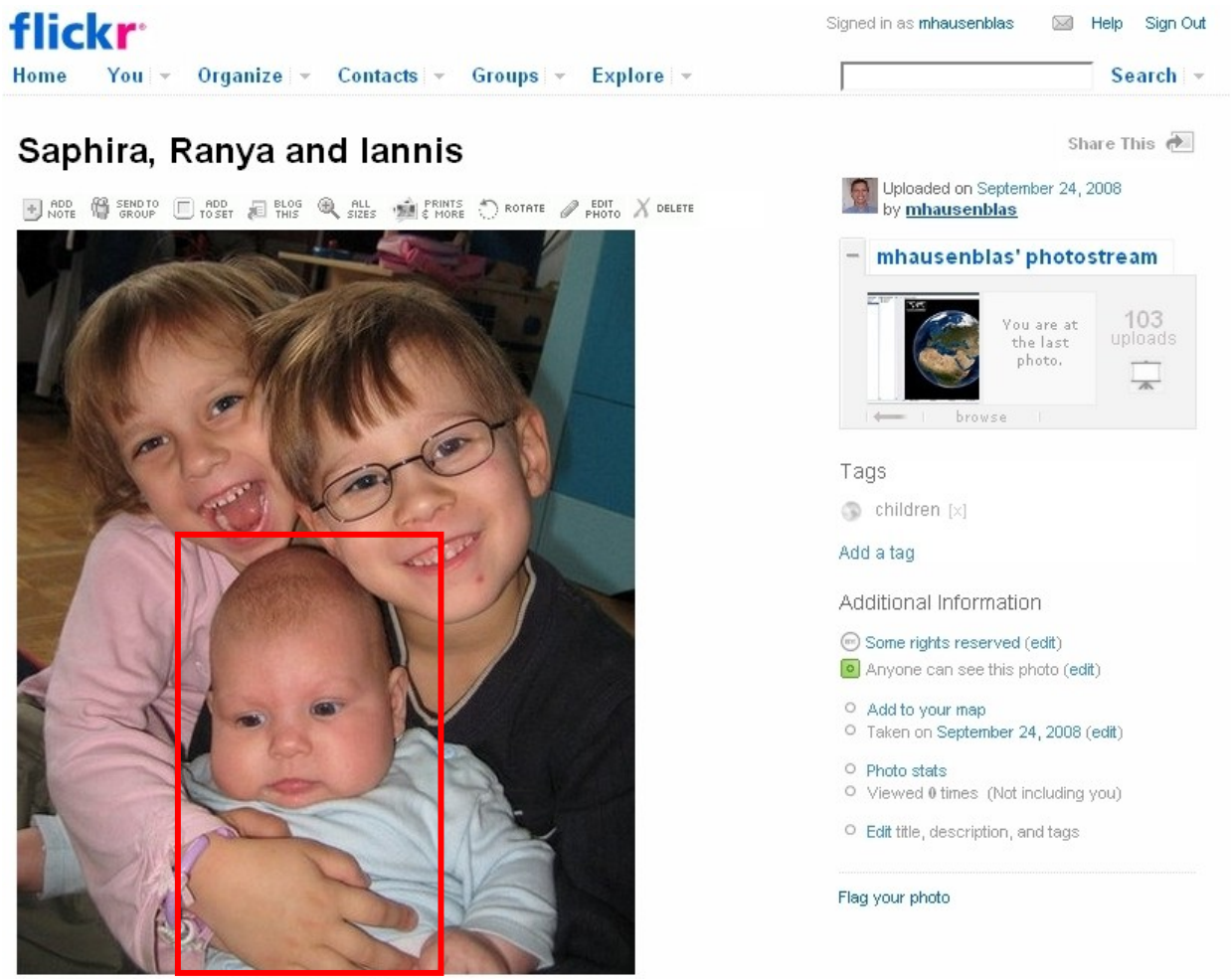
Manual Linking

CaMiCatzee [1], a concept demonstrator allowing the **FOAF-based** search for **person depictions** on flickr photos



[1] http://sw.joanneum.at/CaMiCatzee

Manual Linking



foaf:depicts <http://saphira.blogr.com/#me>

Manual Linking

- quite new linking paradigm, not much experience/research available, yet
- issues
  - exposing link generation vs. hiding it
  - provenance, trust & privacy
  - motivation for end-user

The
Semantic Web Client Library

Consuming Linked Data in Your Applications

http://www4.wiwiss.fu-berlin.de/bizer/ng4j/semwebclient/

Overview

- Introduction
- How does the library work?
- Using the command line tool
- Using the library in applications

Example

- **What's the interests of the people Tom knows?**

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?i WHERE {
   <http://kmi.open.ac.uk/people/tom/> foaf:knows ?p .
   ?p foaf:interest ?i
}
```

- **Answer:**
  - "travel"
  - <http://tyne.shef.ac.uk/t-rex/>
  - <http://www.3worlds.org/>

  ...
  - <http://www.dcs.shef.ac.uk/%7Efabio/X-Media/>
  - "new things"
  - "climbing"
- **29 RDF documents retrieved**

Main Features

- **Enables to query the whole Web**
  - SPARQL queries
  - find(SPO) queries
- **Retrieves relevant RDF documents from the Web dynamically**
  - dereferences HTTP URIs
  - follows rdfs:seeAlso links
  - follows alternate or meta links in HTML headers
  - queries Sindice
- **Stores retrieved RDF documents as Named Graphs**
- **Supports GRDDL**

Query Processing

- **Splitting SPARQL queries into triple patterns**

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?i WHERE {
  <http://kmi.open.ac.uk/people/tom/> foaf:knows ?p .
  ?p foaf:interest ?i
}
```

1

2

- **Executing a directed-browsing algorithm for each triple pattern**
  - retrieves relevant RDF graphs iteratively
  - finds matching triples (i.e. solutions) in retrieved graphs

Directed-Browsing Algorithm

<http://kmi.open.ac.uk/people/tom/> foaf:knows ?p

1

- **Step 1: Look up URIs in the triple pattern**
  - Look up of Tom's URI

    → GET http://kmi.open.ac.uk/people/tom/
    Accept: application/rdf+xml;q=1, text/html;q=0.5

    ← Response: 303 See Other (http://kmi.open.ac.uk/people/tom/html)

    → GET http://kmi.open.ac.uk/people/tom/html

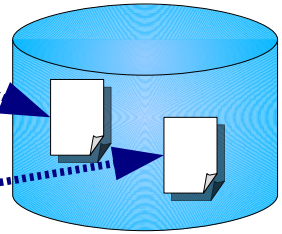
    ← Response: HTML document with

    ```
    <link rel="meta" type="application/rdf+xml"
          title="FOAF" href="/people/tom/rdf"/>
    ```

    → GET http://kmi.open.ac.uk/people/tom/rdf

    ← Response: RDF document

  - Look up of http://xmlns.com/foaf/0.1/knows
    - similar procedure
    - RDF document from http://xmlns.com/foaf/spec/



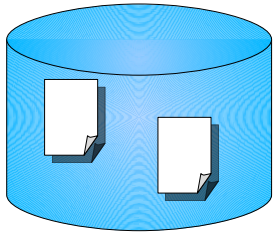
Directed-Browsing Algorithm

<http://kmi.open.ac.uk/people/tom/> foaf:knows ?p (1)

- **Step 2: Follow rdfs:seeAlso links**

FOR EACH triple (*a*, rdfs:seeAlso, *b*) in the local graph set
where *a* is a URI in the current triple pattern **DO**
Look up *b*

- For (1) we have:
  (<http://kmi.open.ac.uk/people/tom/> , rdfs:seeAlso , ?t)
  and
  (foaf:knows , rdfs:seeAlso , ?k)
- Look up each URI that matches ?t or ?k

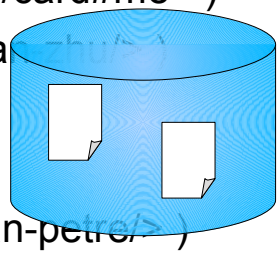


Directed-Browsing Algorithm

<http://kmi.open.ac.uk/people/tom/> foaf:knows ?p (1)

- **Step 3: Match the triple pattern against all graphs in the local graph set**
  - For (1) we get:

(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , -11454bb1:11d1409ca3c:-7ff0)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , <http://identifiers.kmi.open.ac.uk/people/enrico-motta/>)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , <http://danbri.org/foaf.rdf#danbri>)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , <http://www.dcs.shef.ac.uk/~sam/foaf.rdf#samchapman>)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , -11454bb1:11d1409ca3c:-7ff4)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , <http://semanticweb.org/id/Richard_Cyganiak>)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , <http://semanticweb.org/id/Chris_Bizer>)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , -11454bb1:11d1409ca3c:-7ff8)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , <http://identifiers.kmi.open.ac.uk/people/michele-pasin/>)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , <http://www.semantic-web.at/people/blumauer/card#me>)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , <http://identifiers.kmi.open.ac.uk/people/jianha
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , -11454bb1:11d1409ca3c:-7ff1)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , -11454bb1:11d1409ca3c:-7fee)
(<http://kmi.open.ac.uk/people/tom/> , foaf:knows , <http://identifiers.kmi.open.ac.uk/people/marian-petre/>)
...



Directed-Browsing Algorithm

<http://kmi.open.ac.uk/people/tom/> foaf:knows ?p

1

- **Step 4: For each matching triple:**

  e.g. (<http://kmi.open.ac.uk/people/tom/> , foaf:knows , <http://semanticweb.org/id/Richard_Cyganiak>)

  1. Look up all new URIs in the triple

  - For <http://semanticweb.org/id/Richard_Cyganiak> we retrieve a new RDF document from:

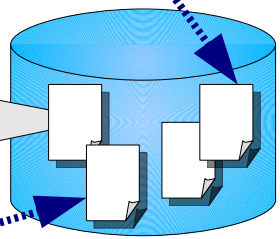
    http://semanticweb.org/index.php?title=Special:ExportRDF/Richard_Cyganiak&xmlmime=rdf

  2. Follow rdfs:seeAlso links for all new URIs in the triple

  - For <http://semanticweb.org/id/Richard_Cyganiak> we find

Tom's FOAF document

```
...
<http://semanticweb.org/id/Richard_Cyganiak>
                    rdfs:seeAlso <http://richard.cyganiak.de/foaf.rdf> .
...
```



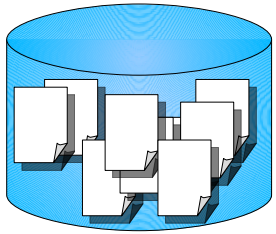
  - New document http://richard.cyganiak.de/foaf.rdf

Directed-Browsing Algorithm

<http://kmi.open.ac.uk/people/tom/> foaf:knows ?p (1)

- **Step 4: For each matching triple ...**
  - overall 21 new graphs in local graph set
- **Step 5: Match the triple pattern against all newly retrieved graphs**
  - nothing new for (1)
- **Step 6: Repeat steps 4 and 5 alternately until**
  - no new matching triples in step 5,
  - maximum number of retrieval steps reached, or
  - timeout reached

- **Another query:**
  - triple pattern: `?p1 foaf:knows ?p2`
  - seeded with Tom's FOAF document
  - after 1min: 9812 matching triples, 372 graphs



Sindice Support

- **Sindice**
  - index of documents with structured data
  - provides an API to find documents
  - URI-based search: finds documents that mention the given URI
- **URI look up triggers a query to the Sindice service**
- **More complete results:**
  - triple pattern: `?prop rdfs:range foaf:Person`
  - 1 matching triple without Sindice look up
  - 19 matching triples with Sindice look up
- **Beware: number of discovered graphs may grow significantly**
  - 2 graphs vs. 134 graphs

Implementation

- **Implemented in Java**
- **Based on the Jena framework**
- **BSD license**
- **Part of the NG4J (Named Graphs API for Jena)**
  - extends Jena with methods the parse, manipulate, query, and serialize sets of Named Graphs
- **Multi-threaded for faster retrieval**

Command Line Tool

- **Execute SPARQL or find(SPO) queries**

```
./bin/semwebquery -retrieveduris -sindice
    -find "ANY rdfs:range <http://xmlns.com/foaf/0.1/Person>"
```

- **Parameters (selection):**
  - `-find <Filename>` – executes a find(SPO) query (use `ANY` as wildcard)
  - `-sparql <Query>` – executes the given SPARQL query
  - `-sparqlfile <Filename>` – executes the SPARQL query in the specified file
  - `-load <URI>` – loads graph from the given URI into local cache before execution
  - `-sindice` – enables Sindice-based URI search during execution
  - `-maxsteps <Integer>` – sets maximum number of retrieval steps
  - `-timeout <Integer>` – sets timeout of the query in seconds

Using the Library

- **Main interface: `SemanticWebClient` class**
  - implements the NamedGraphSet interface defined by NG4J
  - methods (selection):

`read(url,lang)` – reads a Named Graph into the local graph set

`addRemoteGraph(uri)` – issues a URI look up

`find(pattern)` – executes find(SPO) query and returns iterator over all matching triples

`asJenaModel(nameOfDfltGraph)` – returns a jena model view on the local graph set

Using the Library

```
import com.hp.hpl.jena.query.*;
import de.fuberlin.wiwiss.ng4j.semwebclient.SemanticWebClient;

SemanticWebClient semweb = new SemanticWebClient();

String queryString = ...  // Specify the query

// Execute the query and obtain the results
Query query = QueryFactory.create( queryString );
QueryExecution qe = QueryExecutionFactory.create( query,
                                        semweb.asJenaModel("default") );
ResultSet results = qe.execSelect();

// Consume the results
while ( results.hasNext() )
{
  QuerySolution s = results.nextSolution();
  ...
}
```

Using the Library

- **Methods of `SemanticWebClient` for custom control:**

  `reloadRemoteGraph(uri)` – refresh local copy

  `clear()` – clears the local graph set

  `requestDereferencing(uri,step,listener)` – initiates URI look up

  `requestDereferencingWithSearch(uri,step, derefListener, searchListener)`

  `setConfig(option,value)` – sets configuration option
  - CONFIG_MAXSTEPS
  - CONFIG_TIMEOUT
  - CONFIG_DEREF_CONNECT_TIMEOUT
  - CONFIG_DEREF_READ_TIMEOUT
  - CONFIG_ENABLE_SINDICE
  - ...

Provenance Information

- **`SemWebTriple` class**
  - provided by `find(pattern)` method of `SemanticWebClient`
  - method `getSource()` returns URI of the containing graph
- **Provenance graph**
  - always in the local graph set
  - contains statements about the source URL and retrieval time for each retrieved graph
- **Dereferenced URIs**
  - lists of successfully and unsuccessfully retrieved URIs

    ```
    successfullyDereferencedURIs()
    unsuccessfullyDereferencedURIs()
    ```

  - information about redirected URIs

    ```
    redirectedURIs()
    getRedirectURI(uri)
    ```

Conclusion

- **The Semantic Web Client Library**
  - enables queries over the whole Web
  - dynamically retrieves RDF data during query execution
  - is available from:
    http://www4.wiwiss.fu-berlin.de/bizer/ng4j/semwebclient/

- **Future work:**
  - smart caching and replacement of retrieved graphs

Conclusions and Outlook

Summary

- Linked Data is a generic approach for publishing structured data on the Web.
  - Builds on standards in contrast to proprietary Web APIs
- Linked Data builds on the classic architecture of the Web.
  - Links allow you to discover unexpected things
- The Web of Linked Data is growing rapidly.
- There is an increasing number of application prototypes that consume Linked Data from the Web.

Linked Data Prospects in 2009

Even More Data!

1. **Conversion of further open license datasets into RDF**
2. **Wrappers around existing applications**

- Growing number of tools available
  - D2R Server
  - Triplify
  - Pubby
- Growing number of wrappers for existing systems
  - Drupal
  - Wordpress
  - osCommerce

Research Directions and Challenges

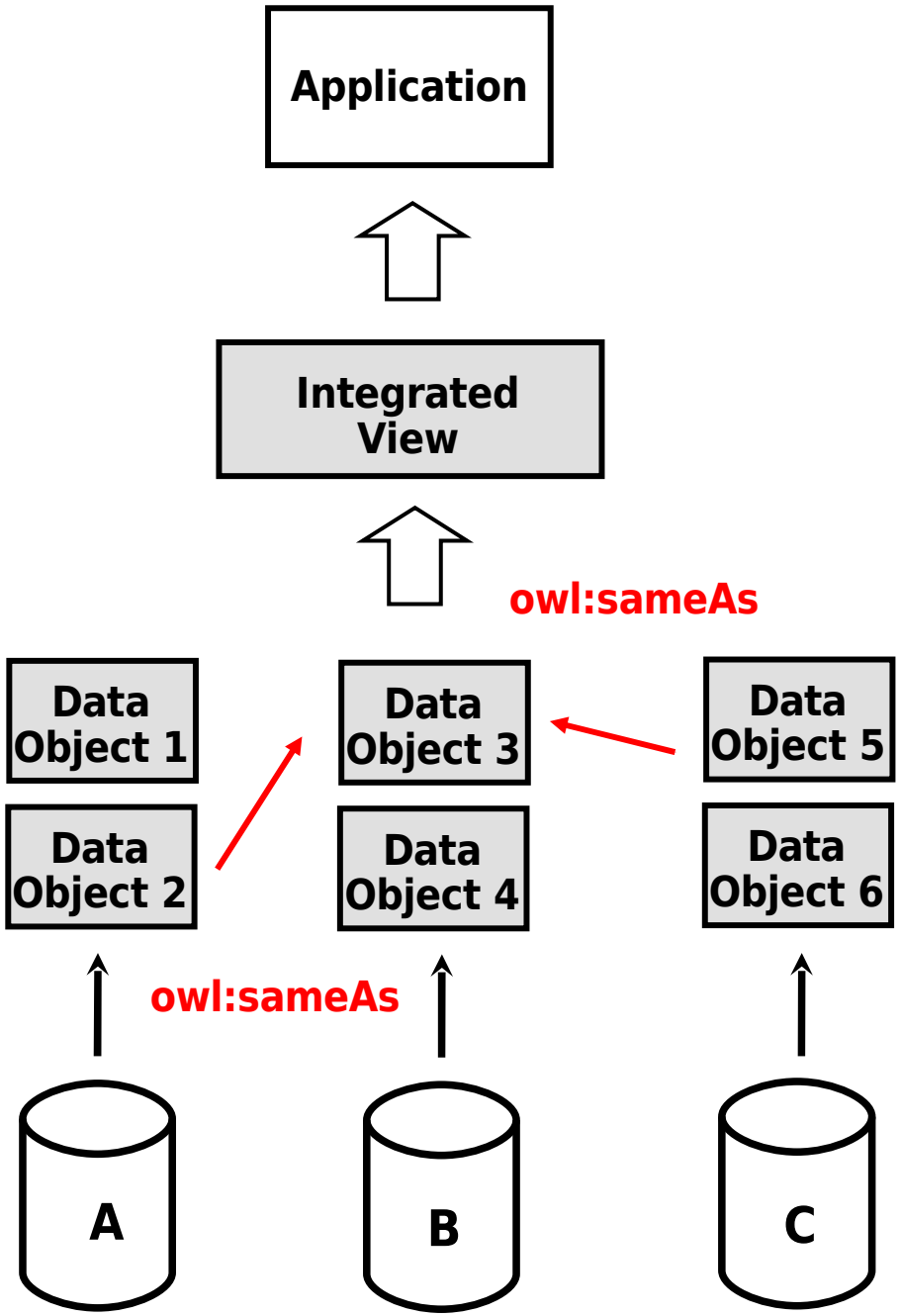
Linking

1. **Increase the amount of links between datasets**
2. **Increase the quality of these links**

- Today: Simple pattern- and graph-matching based techniques used for automated interlinking.

- There is lots of existing work in database and knowledge representation communities on identity resolution to be used.

Data Fusion



Users want an integrated view on all data that is available about an object!

Raises well known but hard problems:

- Schema mapping
- Inconsistency resolution
- Trust / information quality

Licensing

In order to do anything serious with data from the Web, its license terms have to be clear.

- Need for
  - proper licensing vocabularies for dedicating data to the public domain
  - best practices on how to annotate data with licensing meta-data
- Can build on
  - Open Data Commons Public Domain Dedication & Licence (PDDL) (see LDOW2008 paper)
  - Creative Commons Licensing Framework

Browsers and Search Engines for the END USER

Need for real tools, not only proof of concept prototypes!

- End user friendly views on the data
  - ordering and merging of properties
  - dealing with information overflow
- More advanced data analysis features
  - aggregation, drill down
  - calculations, Web-Excel
- Explanations about data provenance and trustworthiness
- Interesting work happening around Freebase

IJSWIS Special Issue on Linked Data

- Special Issue of International Journal on Semantic Web and Information Systems
- Editor-in-Chief: Amit Sheth
- Guest Editors: Chris Bizer, Tom Heath, Martin Hepp
- Submission deadline in January 2009

Getting Involved

- Wiki Page
  - http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
- Mailing List
  - public-lod@w3.org
  - http://lists.w3.org/Archives/Public/public-lod/
- Participating in the project
  - Put your name on the Wiki page
  - Subscribe to the mailing list
  - Do something useful
- Tutorial: How to Publish Linked Data on the Web
  - http://linkeddata.org/docs/how-to-publish

Discussion and Linked Data Clinic