

I-KNOW / I-SEMANTICS '09

2-4 September 2009, Graz, Austria

The Emerging Web of Linked Data

Christian Bizer, Freie Universität Berlin



1. From a Web of Documents to a Web of Data

- Web APIs and Linked Data

2. Linked Data Deployment on the Web

- What data is out there?

3. Applications

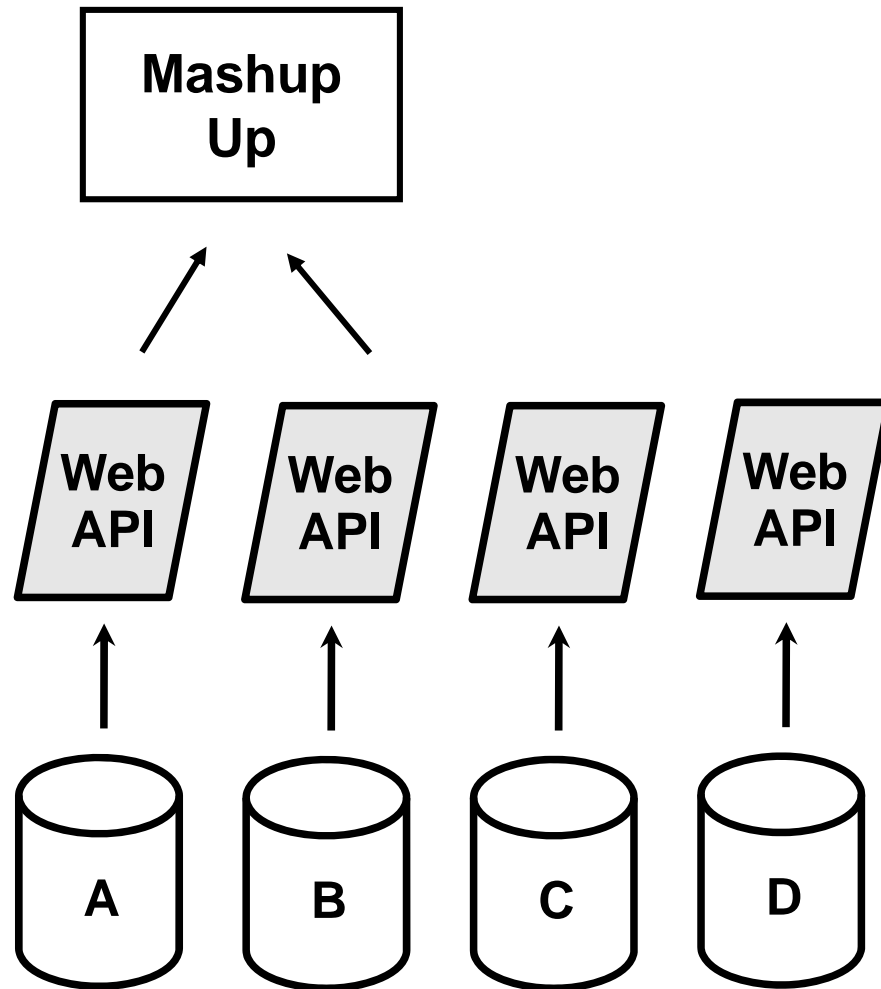
- What is being done with the data?

4. Research Issues

- What is still missing?

Web APIs and Mashups

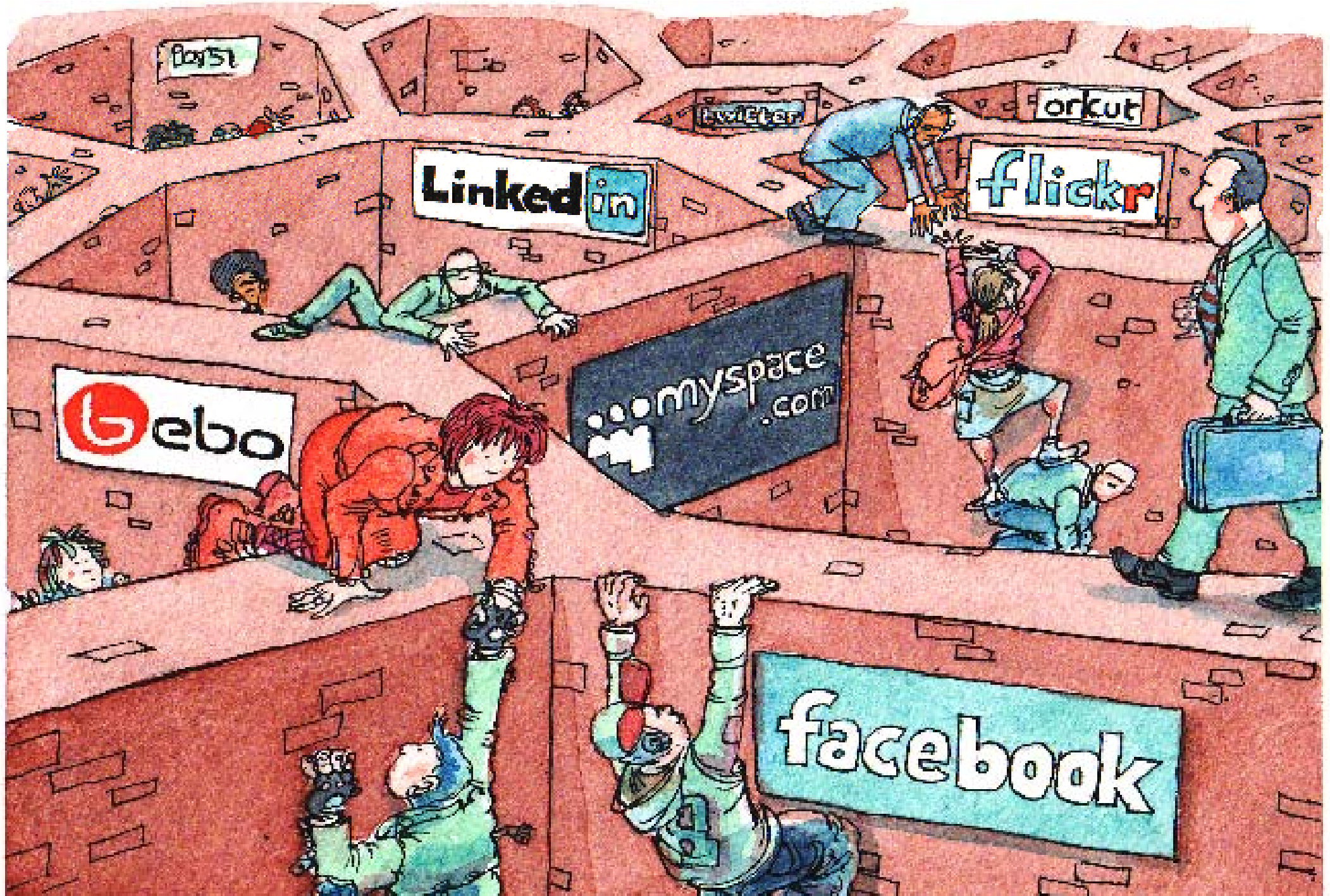




Shortcomings

1. APIs provide proprietary interfaces
2. No global identification mechanism for data objects
3. You can not set hyperlinks between data objects
4. Mashups are implemented against a fixed set of data sources

Web APIs slice the Web into Walled Gardens

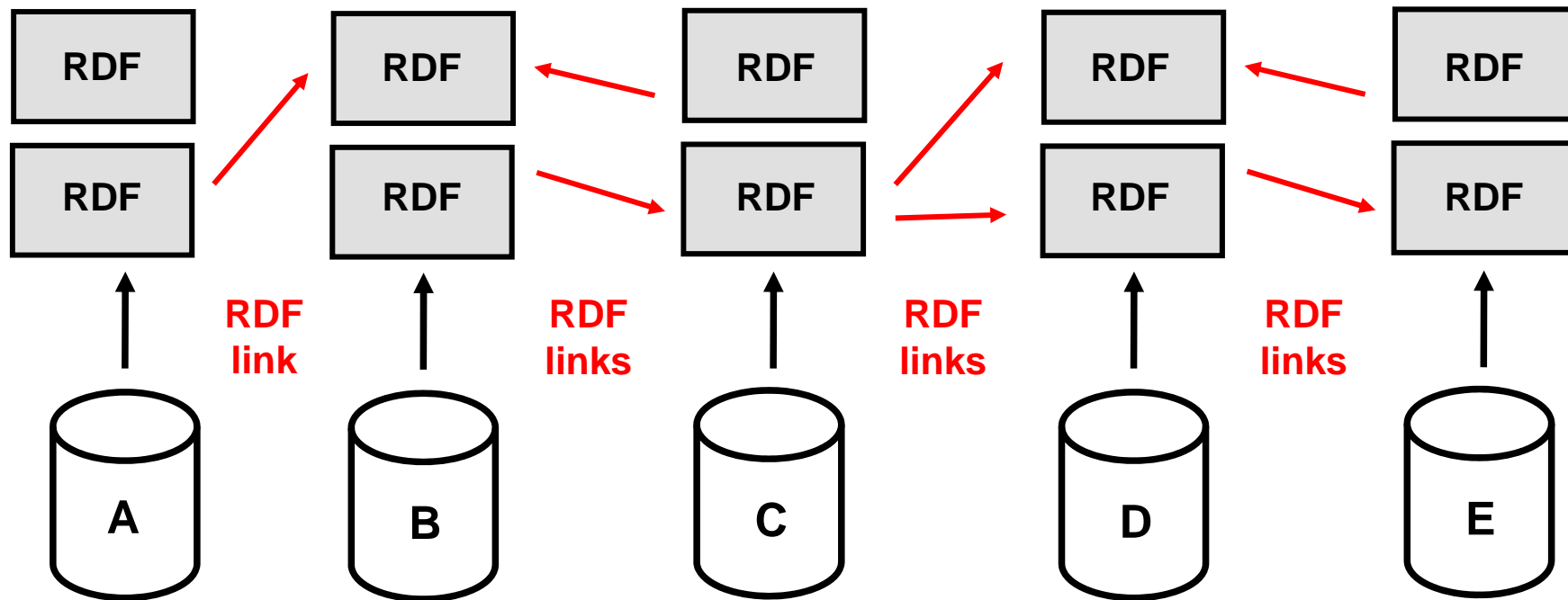


Linked Data

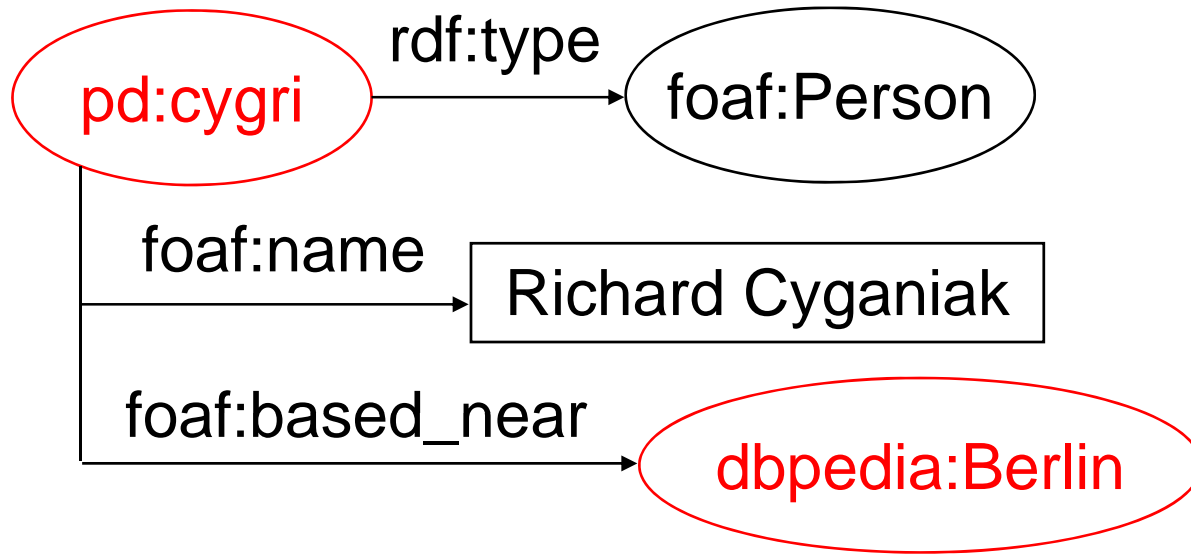


Use Semantic Web technologies to

1. publish structured data on the Web,
2. set links between data from one data source to data within other data sources.



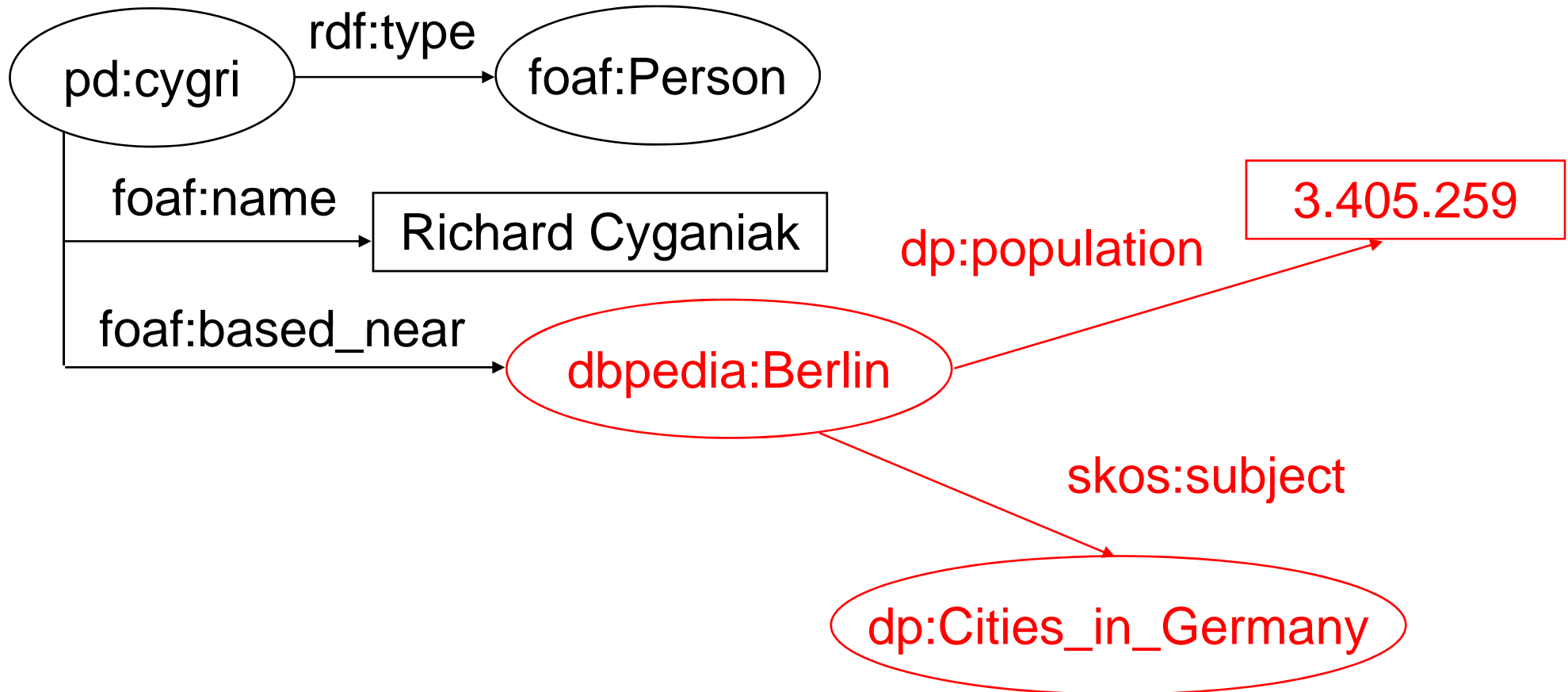
Data items are identified with HTTP URIs



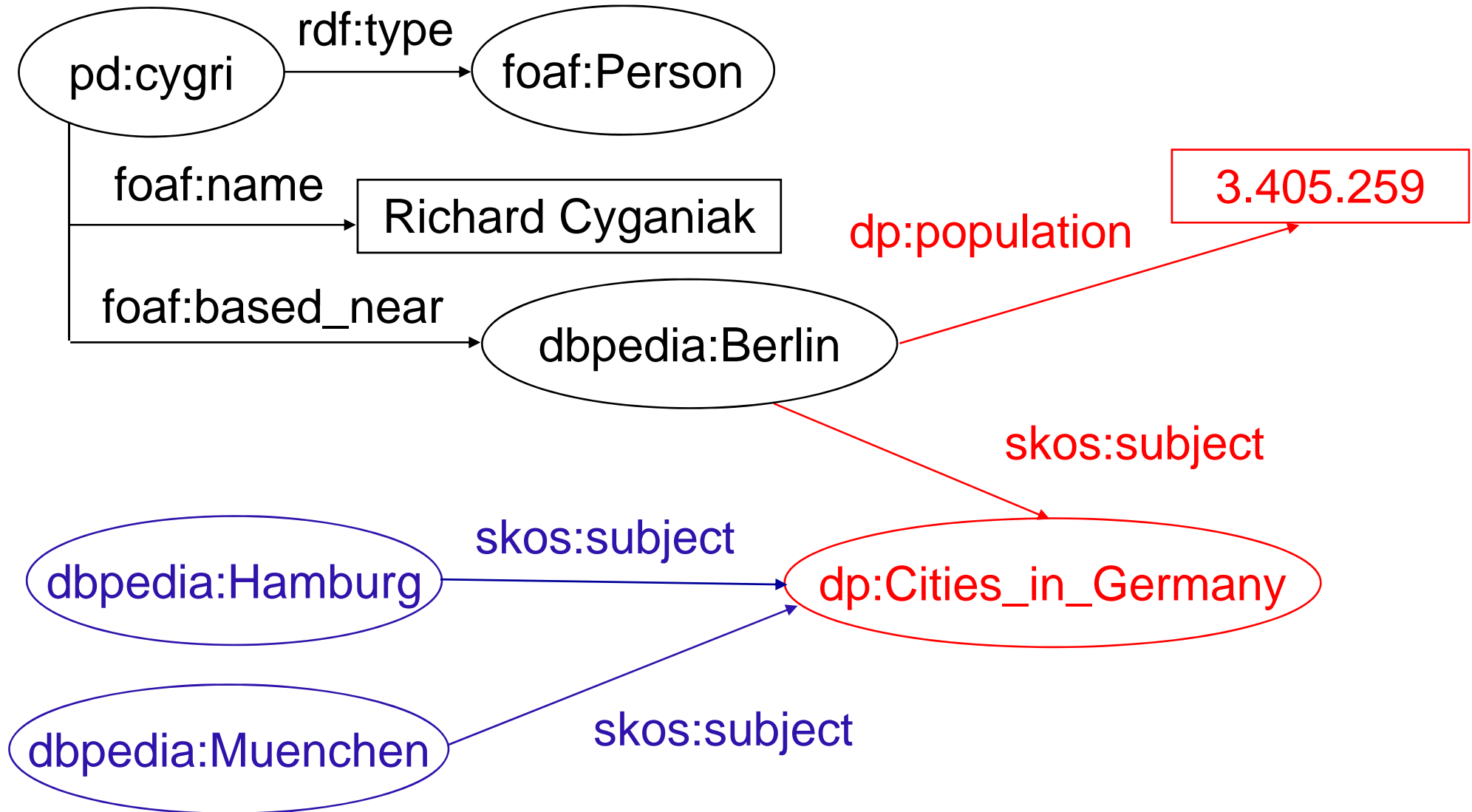
pd:cygri = <http://richard.cyganiak.de/foaf.rdf#cygri>

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

Resolving URIs over the Web



Resolving URIs over the Web



The Web of Linked Data

■ Is a global data space

- Anyone can publish data to this data space
- Entities in the dataspace are connected by links

■ Data is self-describing

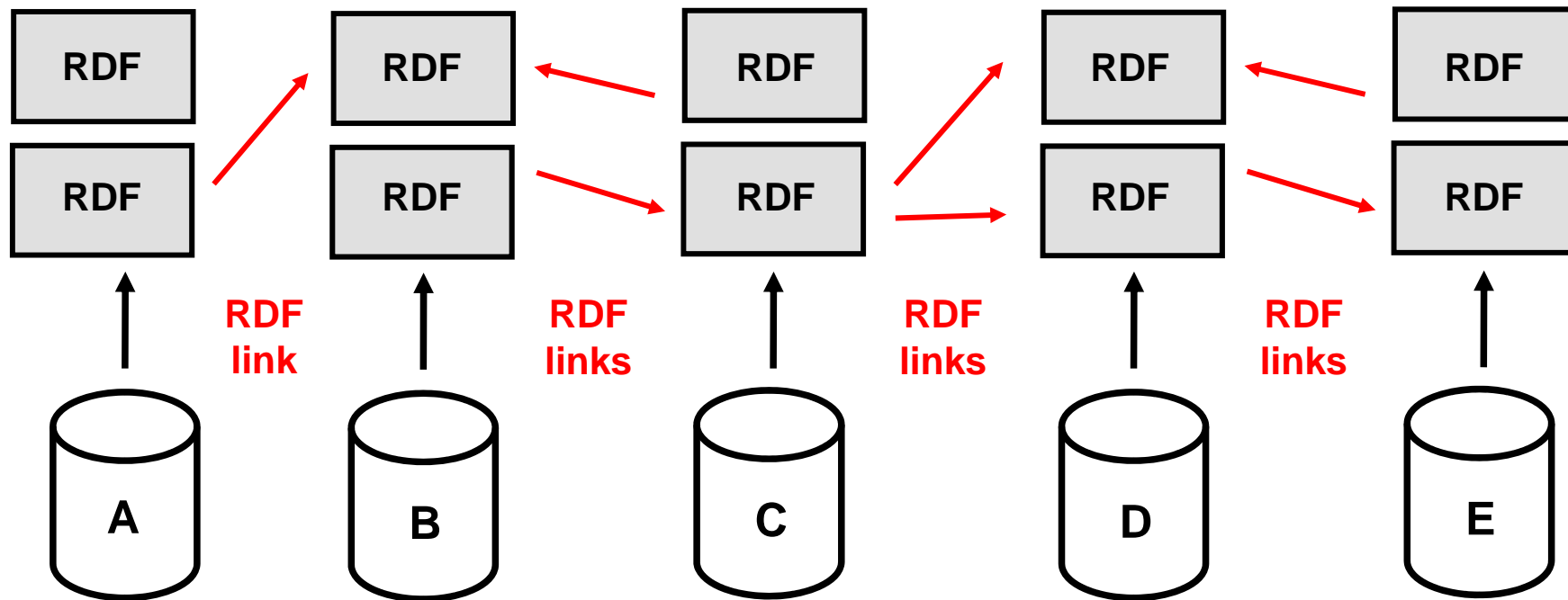
- If an application encounters data represented using an unfamiliar vocabulary/ontology, the application can resolve the URIs that identify vocabulary terms in order to find their RDFS or OWL definition.

■ The Web of Data is open

- meaning that applications can discover new data sources at run-time by following links.

2. Linked Data Deployment on the Web

■ Is this real?



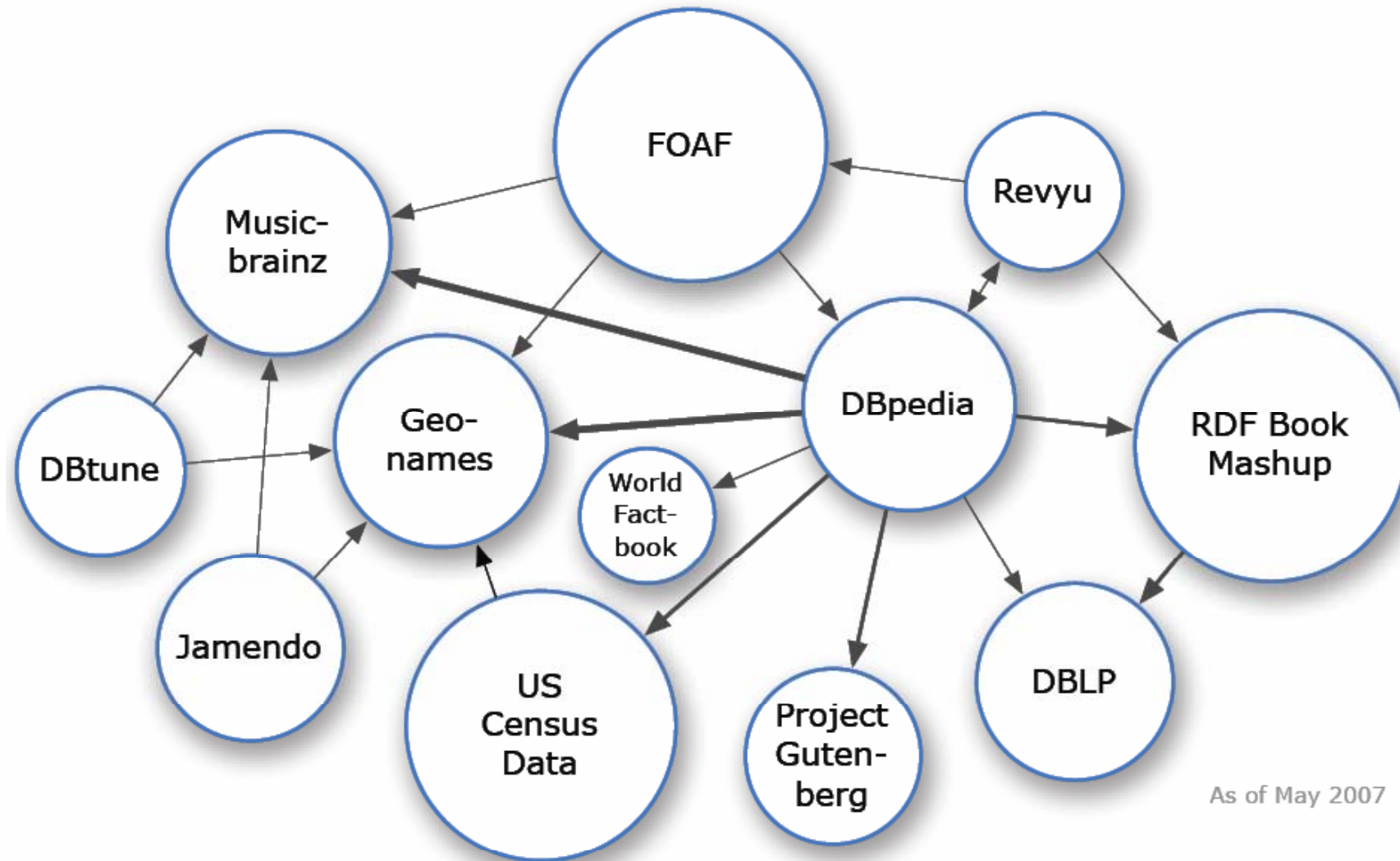
W3C Linking Open Data Project



■ Grassroots community effort to

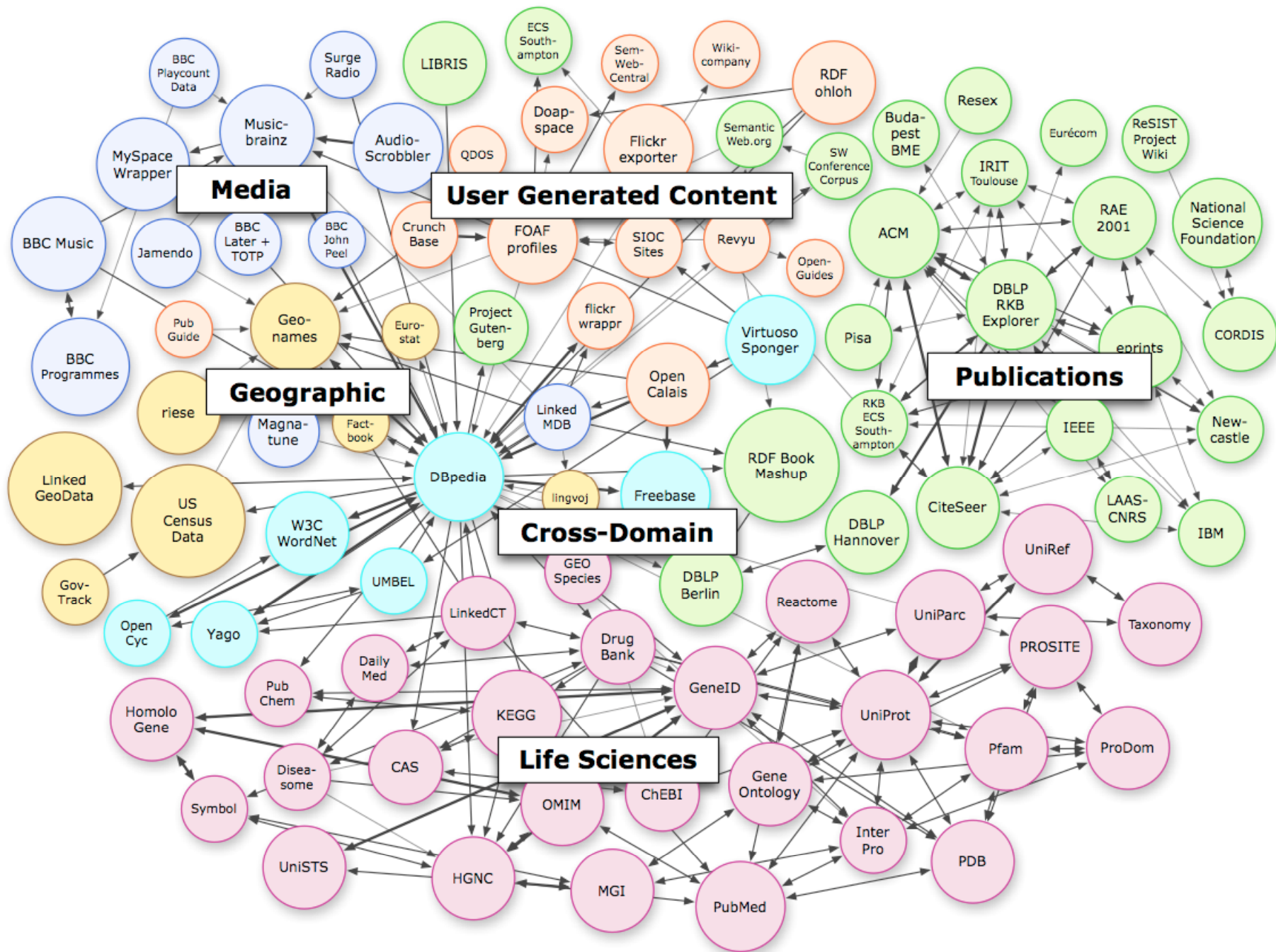
- publish existing open license datasets as Linked Data on the Web
- interlink things between different data sources

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: July 2009



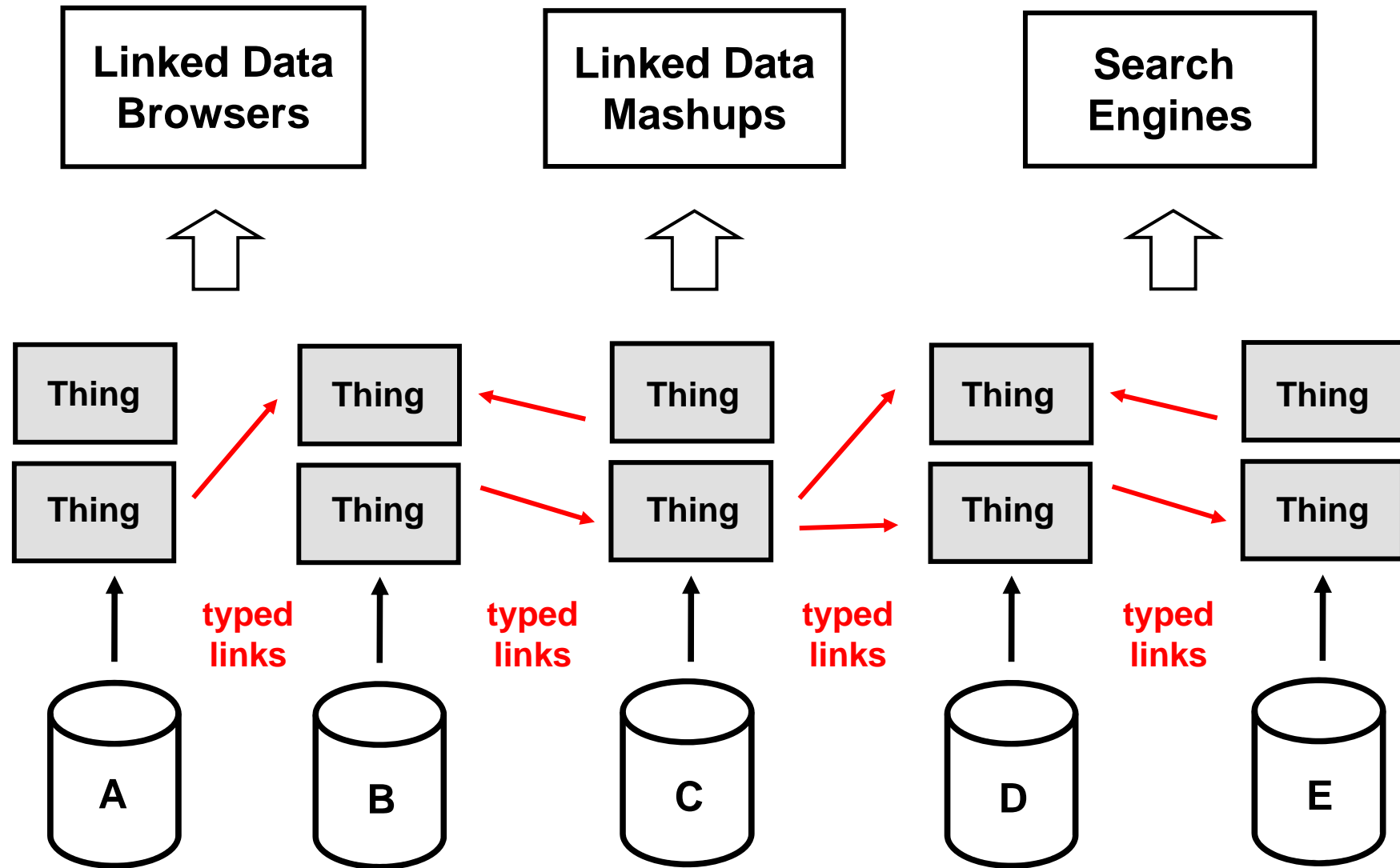
LOD data set statistics as of July 2009

<i>Domain</i>	<i>No of Triples</i>	<i>% of Cloud</i>	<i>No of Links</i>	<i>% of Links</i>
Media	698.000.000	10,4%	1.238.000	0,8%
Publications	212.000.000	3,2%	4.922.000	3,3%
Life Sciences	2.429.000.000	36,1%	133.199.000	89,4%
Geographic Data	3.097.000.000	46,0%	4.038.000	2,7%
User Generate Content	76.000.000	1,1%	1.559.000	1,0%
Cross-Domain	214.000.000	3,2%	3.992.000	2,7%
Total	6.726.000.000		148.948.000	

+ 2 billion triples from Data.gov published recently.

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)**
- **Humboldt (HP Labs, UK)**
- **Disco Hyperdata Browser (FU Berlin, DE)**
- **Fenfire (DERI, Irland)**

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 \(also at www4.wiwiw.fu-berlin.de\)](#)

[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](#)

Web of Data Search Engines

- **Falcons (IWS, China)**
- **Sig.ma (DERI, Ireland)**
- **Swoogle (UMBC, USA)**
- **VisiNav (DERI, Ireland)**
- **Watson (Open University, UK)**

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](#)

Tim Berners-Lee


Knows

weblog

New Search

Ok

[Detail View](#) [List View](#) [Table View](#) [Timeline View](#)  [RSS](#)

[next](#)  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, merrag?

What are the big players doing?

- **Yahoo! and Google have started to crawl Linked Data in its RDFa serialization as well as Microformats.**
- **Yahoo!**
 - provides access to crawled data through the Yahoo BOSS API
 - is using the data within Yahoo Search Monkey to make search results more useful and visually appealing.
- **Google**
 - uses crawled RDF data for its Social Graph API
 - is uses crawled data to enhance search results snippets for reviews and people.

Yahoo! Search Monkey

Yahoo! My Yahoo! Mail Welcome, Guest [Sign In] Help

Web | Images | Video | Local | Shopping | more ▾

Dr. Seuss Horton Hears a Who Movie Options ▾

1 - 10 of about 17,600,000 for Dr. Seuss' Horton Hears a Who Movie (About this page) - 0.33 sec

Dr Seuss' Horton Hears A Who
www.hortonmovie.com/splash.html - [Cached](#)

Dr. Seuss' Horton Hears a Who (2008) Movie - Acme Movies
Dr. Seuss' Horton Hears a Who (2008) Movie - Acme Movies - how tickets and maps. Connect with friends and share reviews.
acmemovies.com/hortonhearsawho

Dr. Seuss' Horton Hears a Who (2008) Movie - Acme Movies
Find Dr. Seuss' Horton Hears a Who (2008) Movie on Fandango.com. ... Dr. Seuss' Horton Hears a Who (2008) Movie - Acme Movies
www.fandango.com/dr-seuss-horton-hears-a-who

Dr. Seuss' Horton Hears a Who (2008) Movie - Acme Movies
Dr. Seuss' Horton Hears a Who (2008) Movie - Acme Movies
THEATERS / DR. SEUSS' HORTON HEARS A WHO
www.rottentomatoes.com/m/hortonhearsawho

Dr. Seuss' Horton Hears a Who (2008) Movie - Acme Movies
Read the Dr. Seuss' Horton Hears a Who! movie overview. Learn more about this Animated, Family movie, buy tickets, find showtimes, and read reviews at Fandango.com.
[www.fandango.com/dr-seuss-horton-hears-a-who_102891/movieoverview?date="](http://www.fandango.com/dr-seuss-horton-hears-a-who_102891/movieoverview?date=) - 57k - [Cached](#)

SPONSOR RESULTS

Hottest Movies of 2008
Access to All the Info You Need On Horton Hears a Who.
www.RottenTomatoes.com

Dr. Seuss' Horton Hears a Who (2008) Movie - Acme Movies
[Movie Details](#) | [Showtimes & Tickets](#) | [Trailers & Clips](#) | [Reviews](#)

- Reviews: ★★★★★ (173)
- MPAA Rating: G
- Running Time: 1 hr. 28 min.
- Release Date: March 14th, 2008

 acmemovies.com/hortonhearsawho - [Cached](#)

Connecting the classic Web and Linked Data

■ Annotate Web documents with Linked Data URIs

```
<http://data.semanticweb.org/conference/eswc/2007/paper-69>  
dc:subject <http://dbpedia.org/resource/Machine_learning> .
```

■ (Semi-) Automated Annotation Services using Named Entity Recognition

- Open Calais (Thomsons Reuters) for news
- Zemanta (startup) for blog posts

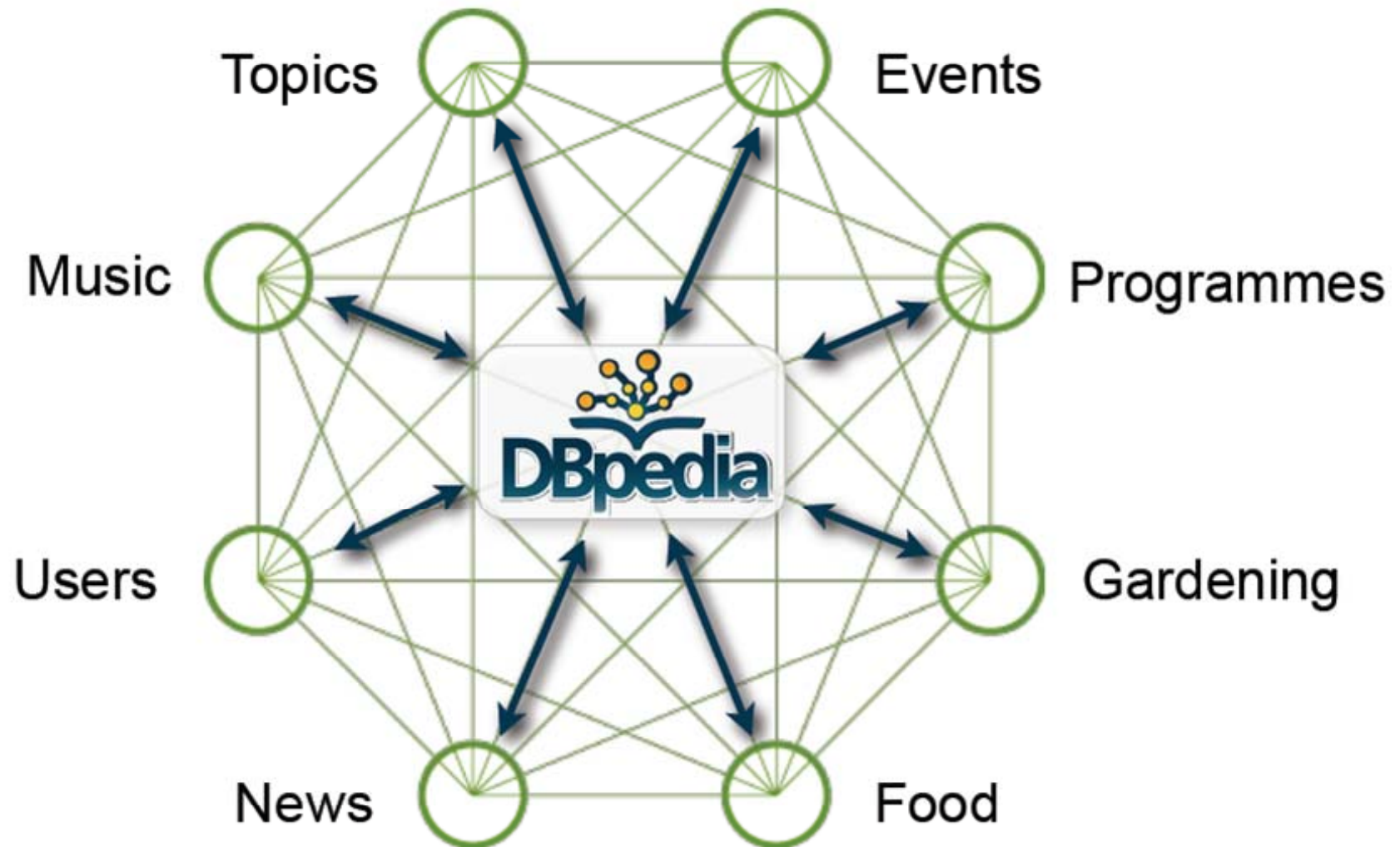


■ Goals

- Improve search by using Linked Data as background knowledge.
- Display Web of Data content as info boxes next to news, blog posts.

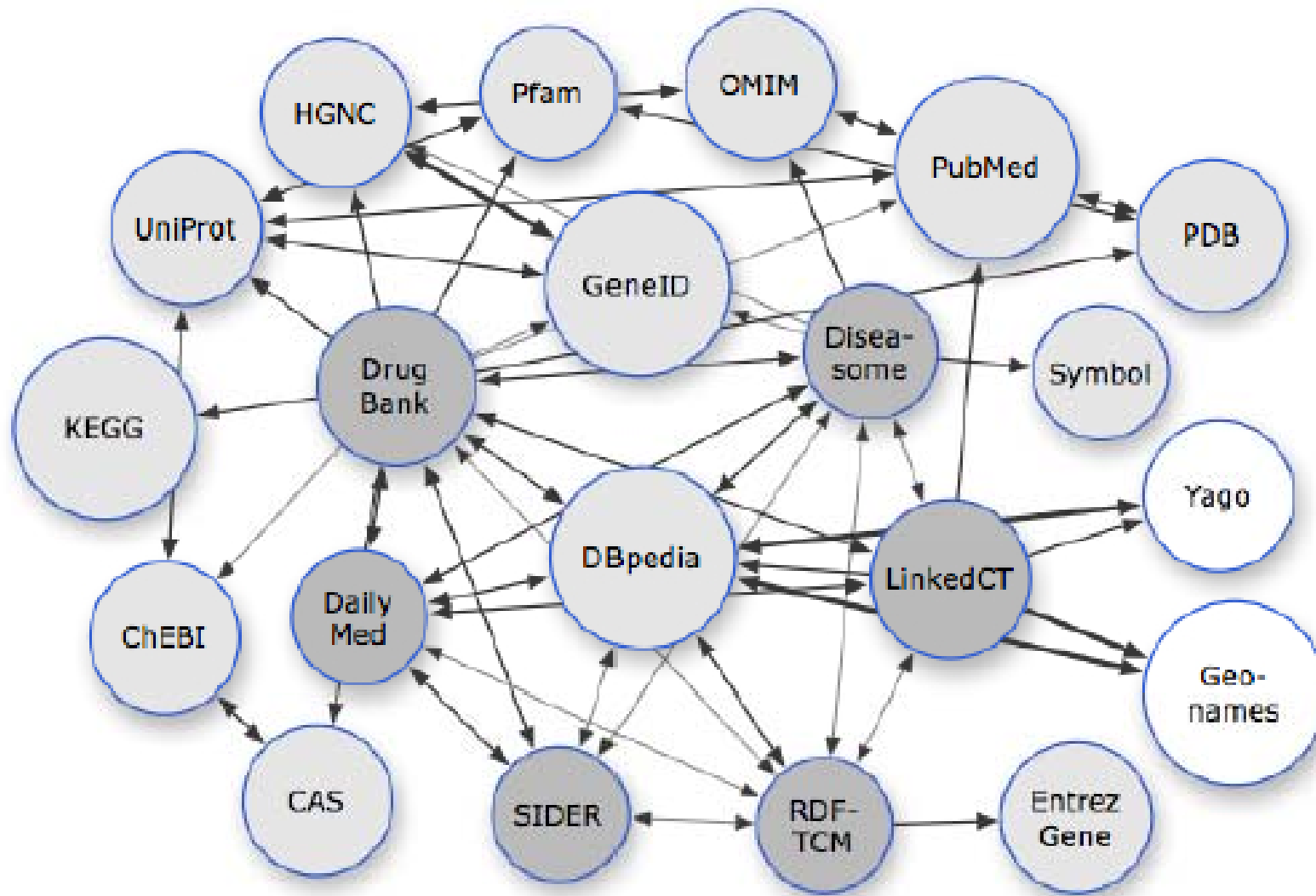
Linked Data inside the BBC

- The BBC uses Linked Data to connect content residing in various repositories.



W3C Linking Open Drug Data

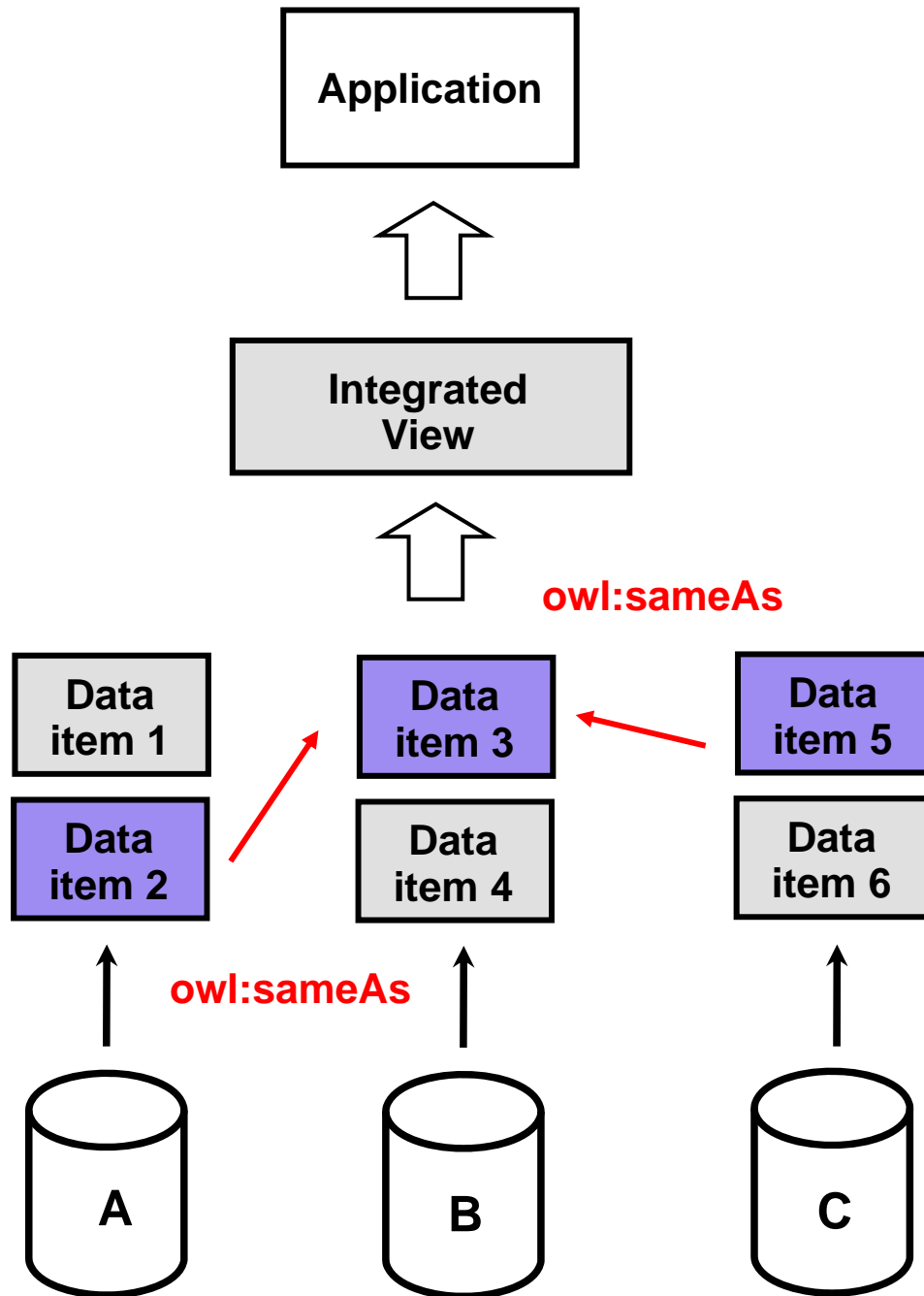
- Eli Lilly, Johnson & Johnson, Astra Zeneca and several research institutes cooperate on interlinking drug discovery data.



4. Research Issues

- What is still missing?

Linked Data Fusion



Applications want an integrated view on all data that is available about an real-world entity!

Linked Data Fusion - Requirements

- 1. Map data into a single schema**
 - so that data can be rendered and queried properly.
- 2. Smush data from all sources about a single real-world entity**
 - while keeping track of information provenance.
- 3. Resolve inconsistencies in the data**
 - by applying different trust heuristics.

Pay As You Go Data Integration

- **There is a pay as you go data integration paradigm emerging on the Web of Data**
 - Publish data first using different schemata
 - Maybe use common vocabularies
 - Publish mappings to the Web afterwards

- **How to derive best-effort answers based on heterogeneous Web data and partial mappings?**
 - Alon Halevy, et al.: Web-scale Data Integration: You can only afford to Pay As You Go

User Interfaces and Interaction Paradigms

- **How do we build interfaces that operate over such large amounts of data?**
 - How to aggregate the data in a meaningful way?
- **What will be their interaction paradigm?**
 - Will the browser be something like a Web-Excel including drill-down?
- **What will Google and Yahoo do with the data?**
 - Will search engines turn into answer engines?

Hands on: How to publish Linked Data?

■ Read the “How to Publish Linked Data on the Web” tutorial

- <http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>

■ Publishing Tools

- D2R Server: Publishes relational data bases as Linked Data and via SPARQL
- Pubby: Linked Data wrapper that can be used together with any RDF store

■ Link Generation Tools

- Silk – Link Discovery Framework
- LinQuer

■ Join the W3C Linking Open Data community

- Wiki: <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
- Mailing list: public-lod@w3.org

Thanks!

References

- Overview Article
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>
- Linking Open Data Project Wiki
<http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
- Tutorial on How to Publish Linked Data on the Web
<http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>