Semantic Web Technologies I

Lehrveranstaltung im WS11/12

Dr. Elena Simperl
PD Dr. Sebastian Rudolph
M. Sc. Anees ul Mehdi

Linked Data

- Dr. Elena Simperl

Semantic Web Architecture

entspricht dem ersten Teil von Kapitel 3 des Buches „Semantic Web - Grundlagen" - siehe auch http://www.semantic-web-grundlagen.de/index.php/Kapitel_3

http://sild.cs.vu.nl/
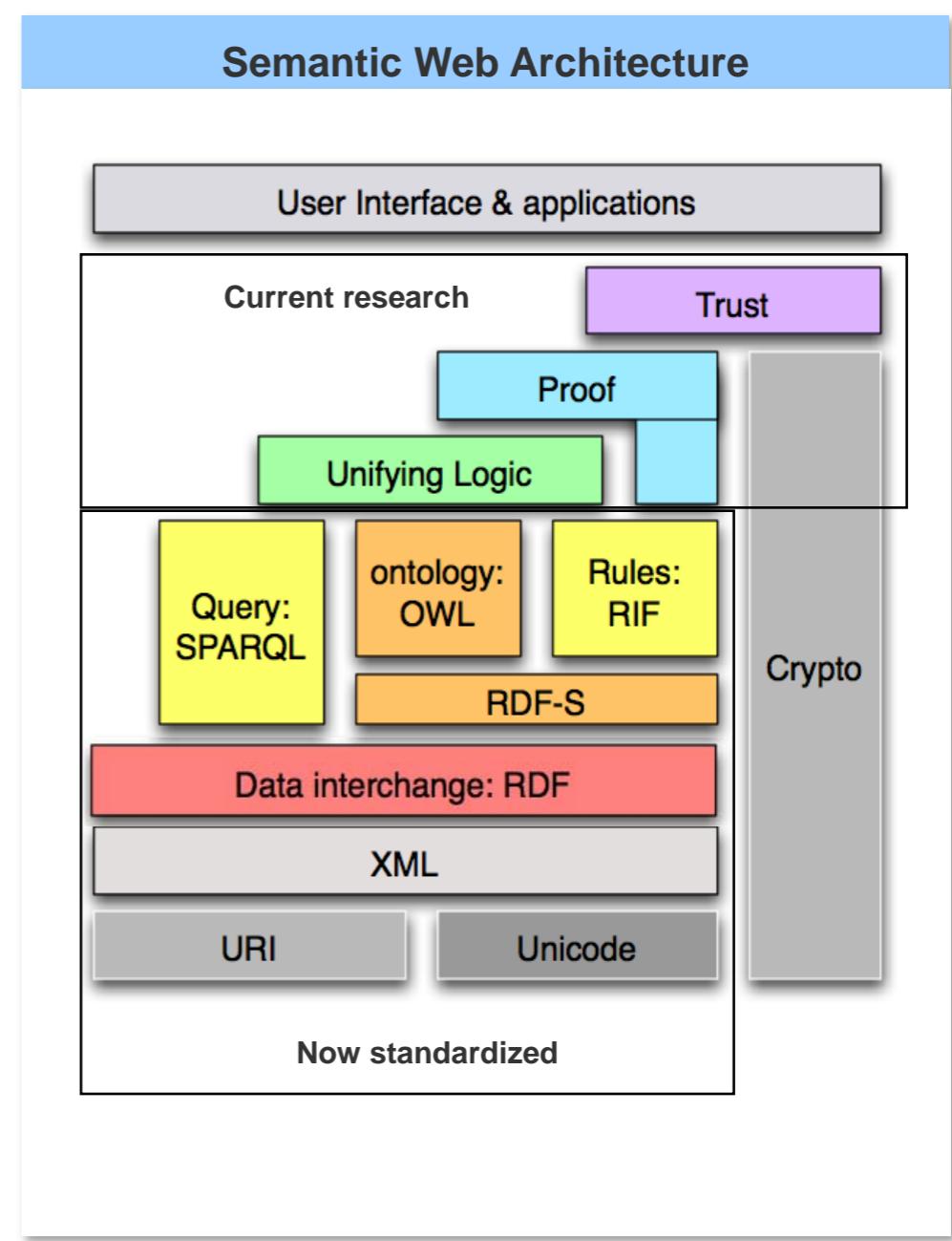
SLIDES DUE TO HARTH, HOGAN, KOTOULAS AND URBANI – SCALABLE INTEGRATION OF LINKED DATA @ ISWC2011

INTRODUCTION

Motivation

- With increased use of computers more and more data is being stored

  - Organisations rely on data for business decisions

  - Data drives policy decisions in government

  - Individuals rely on data from the Web for information and communication

- Data volumes explode

  - More and more data available on the Web is represented in Semantic Web standards

  - Linking Open Data (LOD) initiative

- Semantic Web technologies facilitate the integration of data from multiple sources

- Combining data from multiple sources enables insights

Linked Data Now!
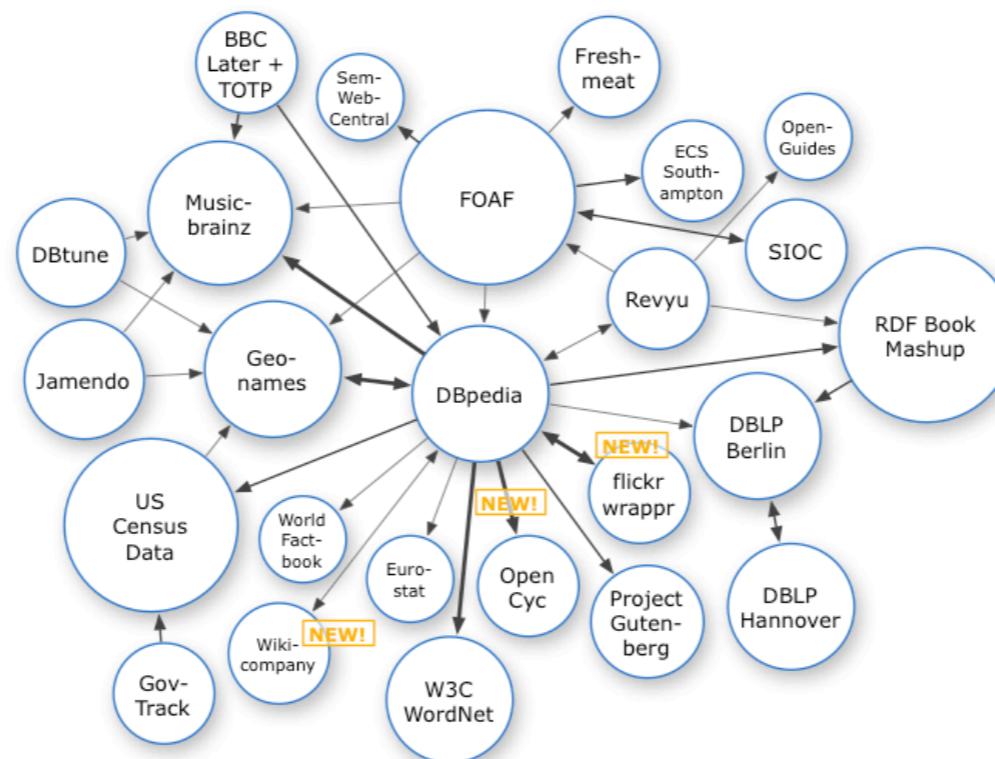
http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html

# Linked Data on the Web
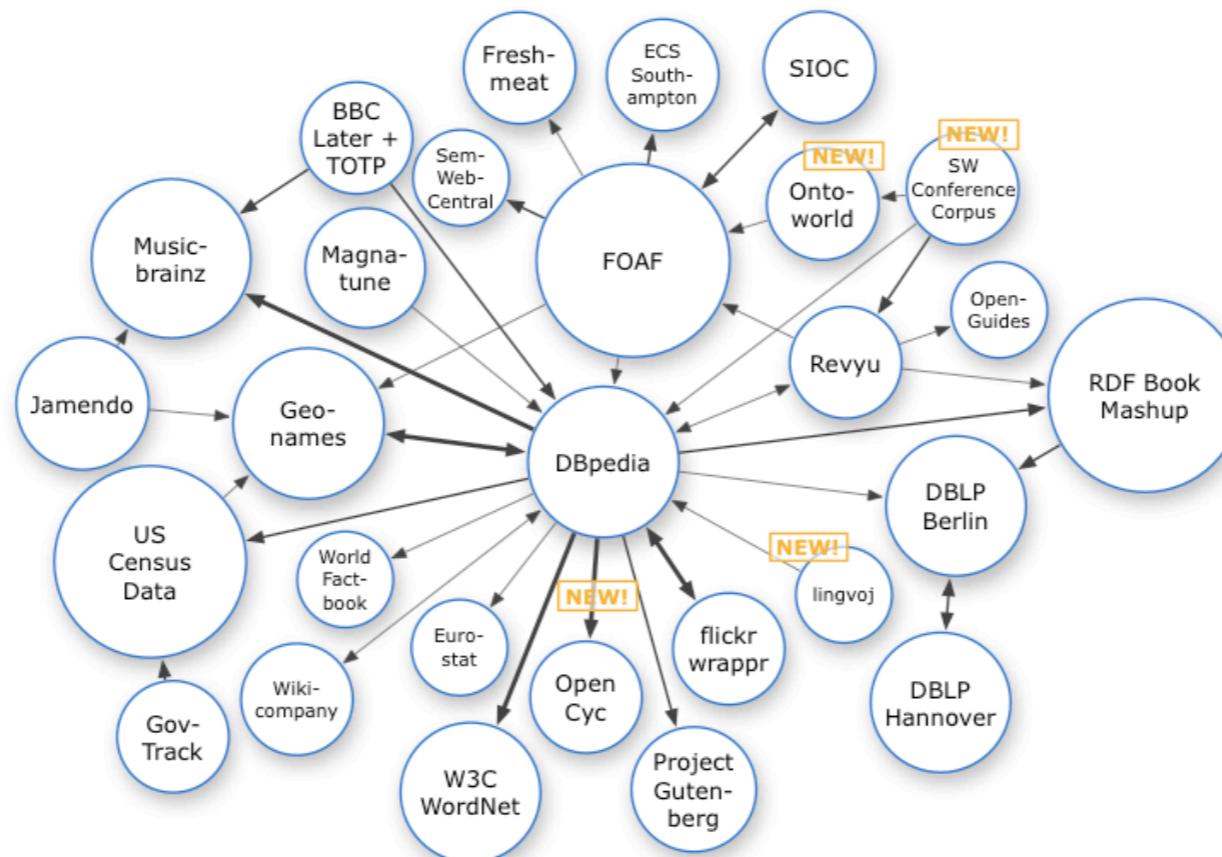
2007-10

Linked Data on the Web
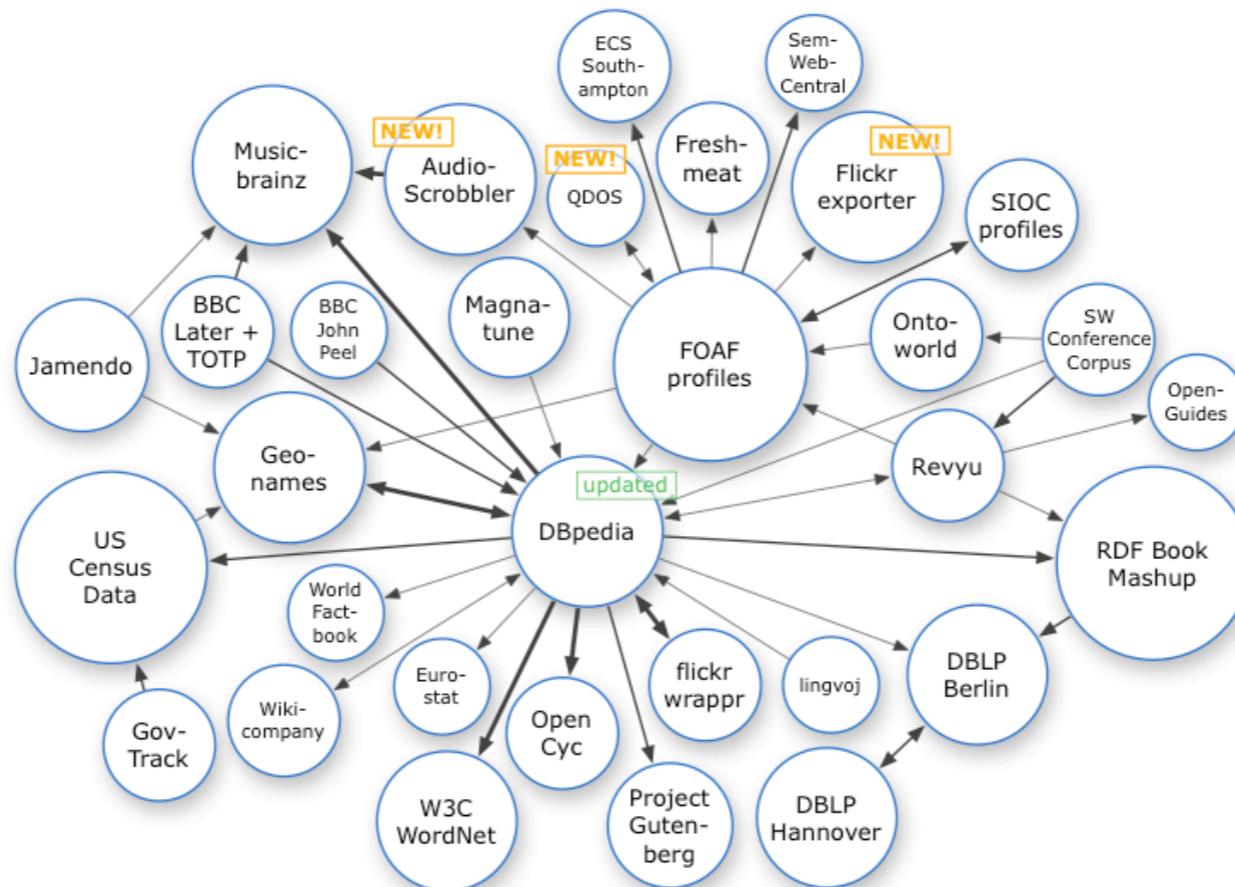
AIFB

2007-11

Linked Data on the Web
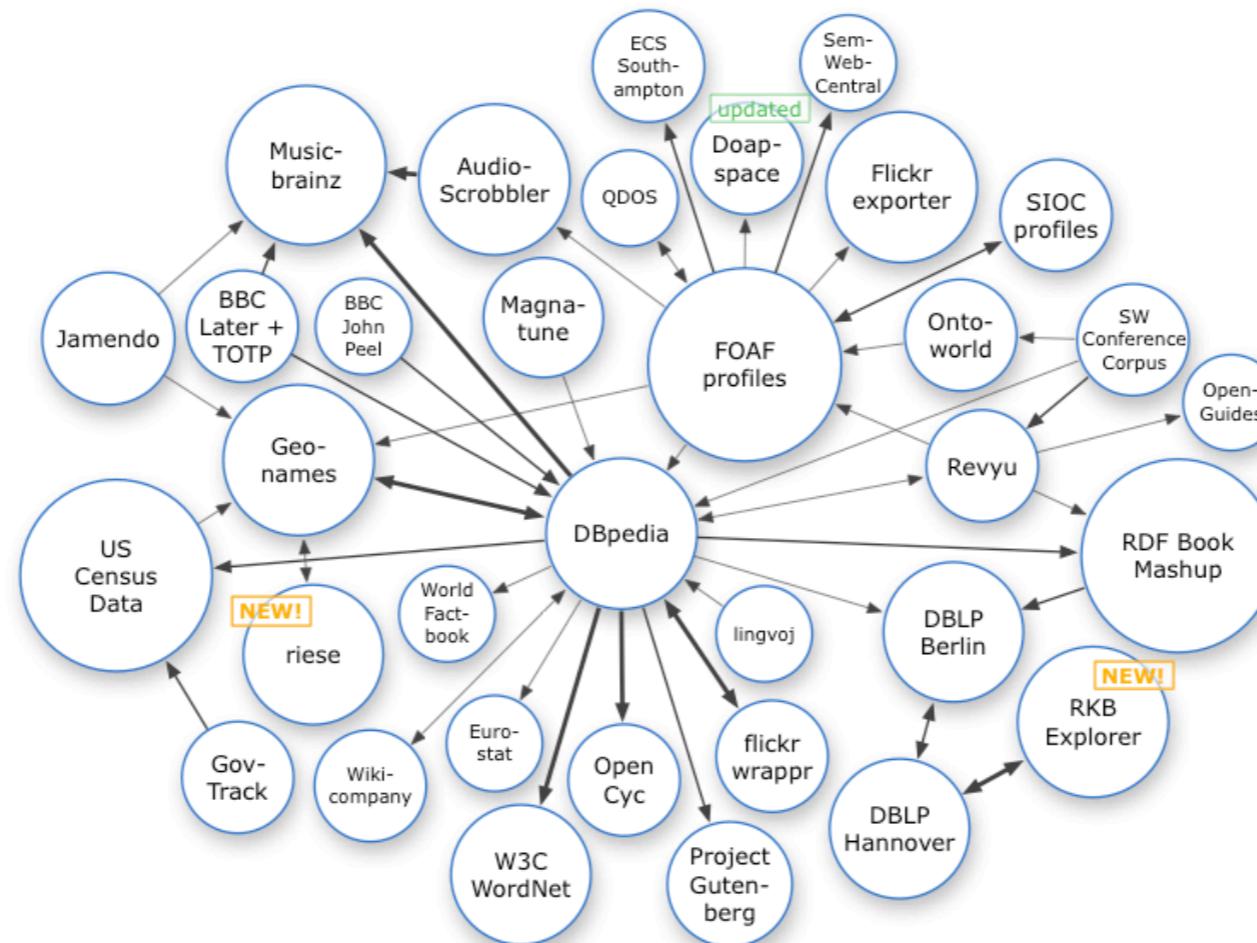
AIFB ◼

2008-02

Linked Data on the Web
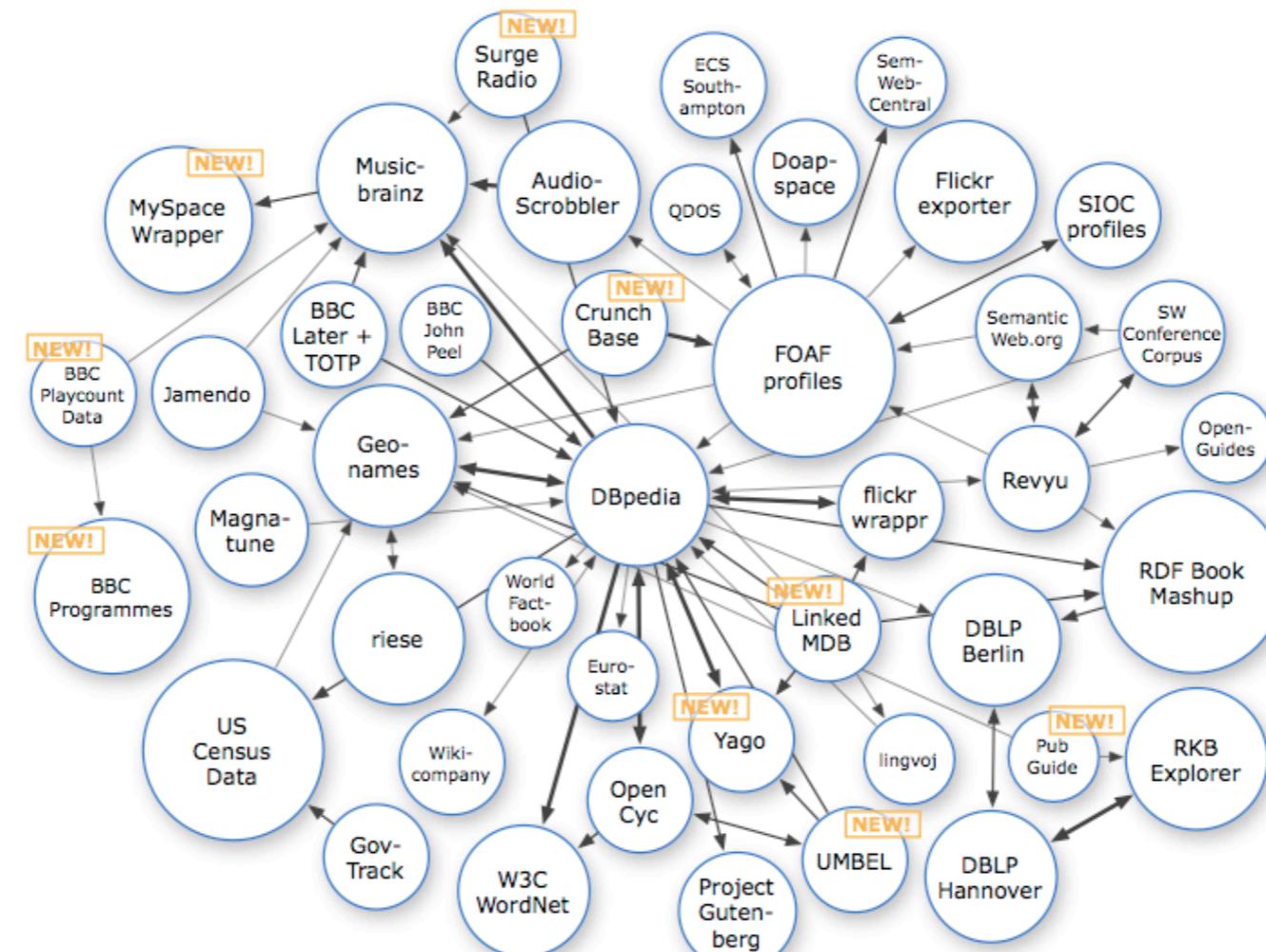
AIFB

2008-03

Linked Data on the Web

2008-09
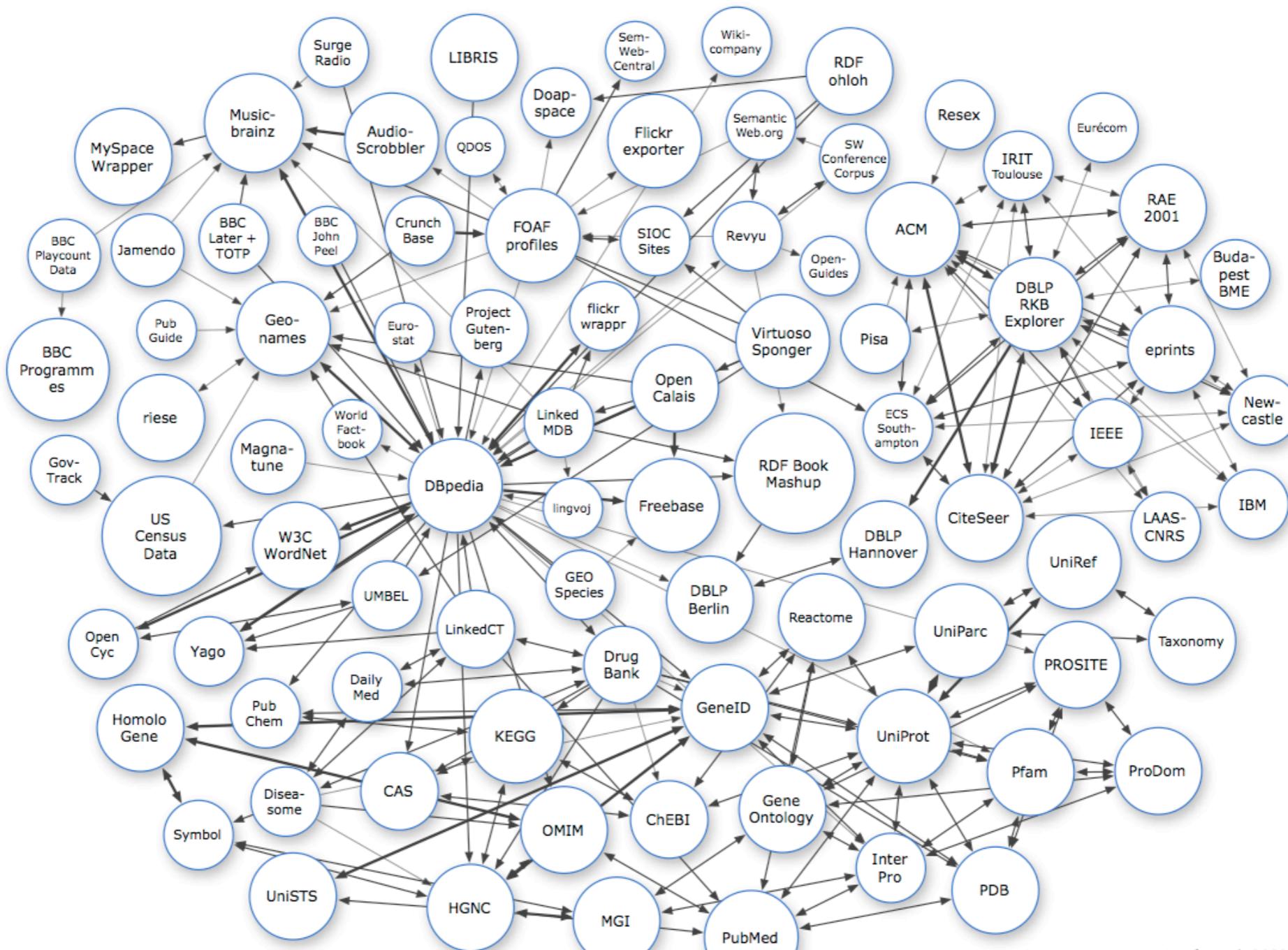
As of September 2008

Linked Data on the Web

AIFB ⬡

2009-03
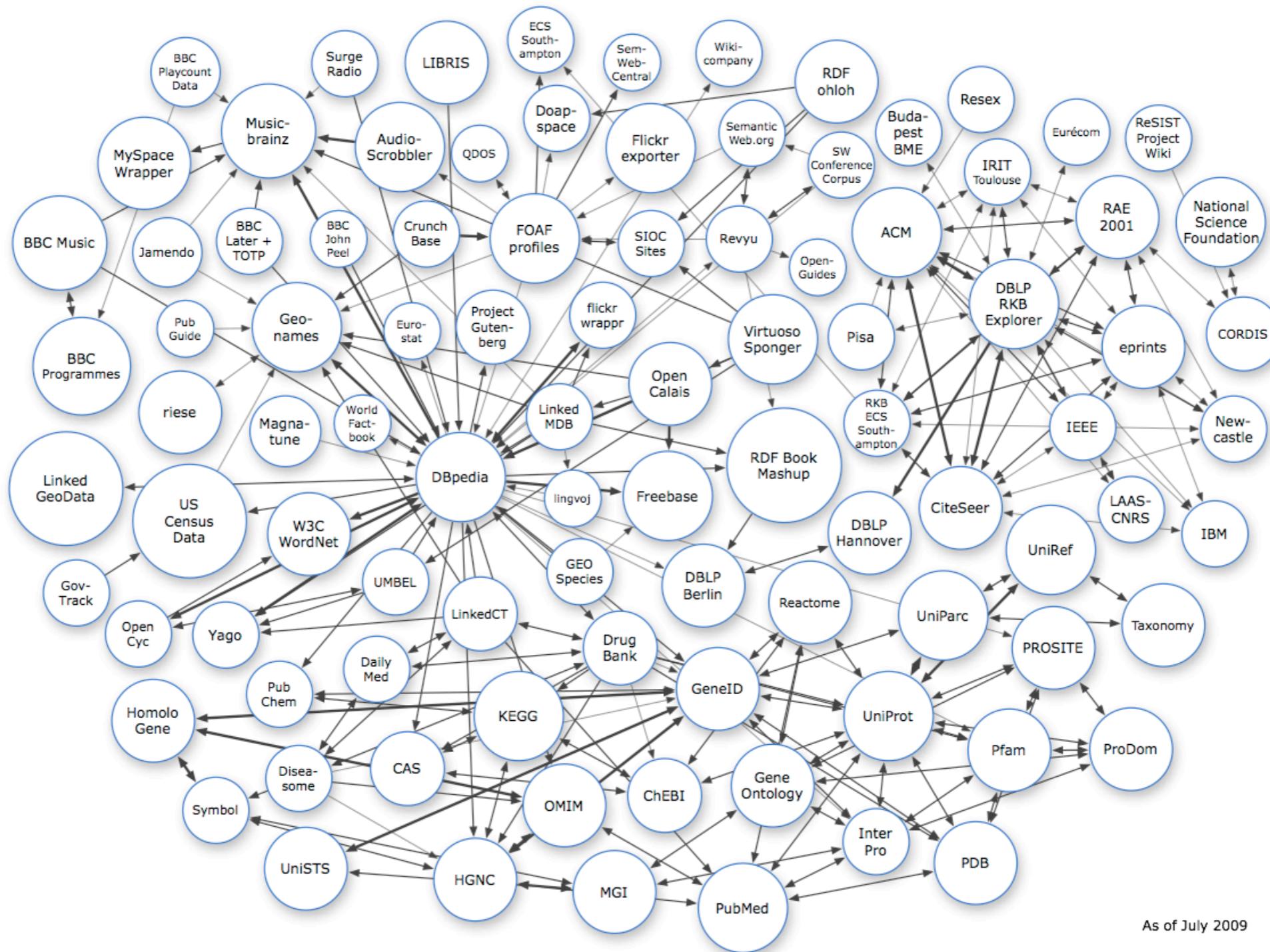
As of March 2009

Linked Data on the Web

AIFB

2009-07

As of July 2009

Linked Data on the Web

AIFB

2009-09

As of September 2010
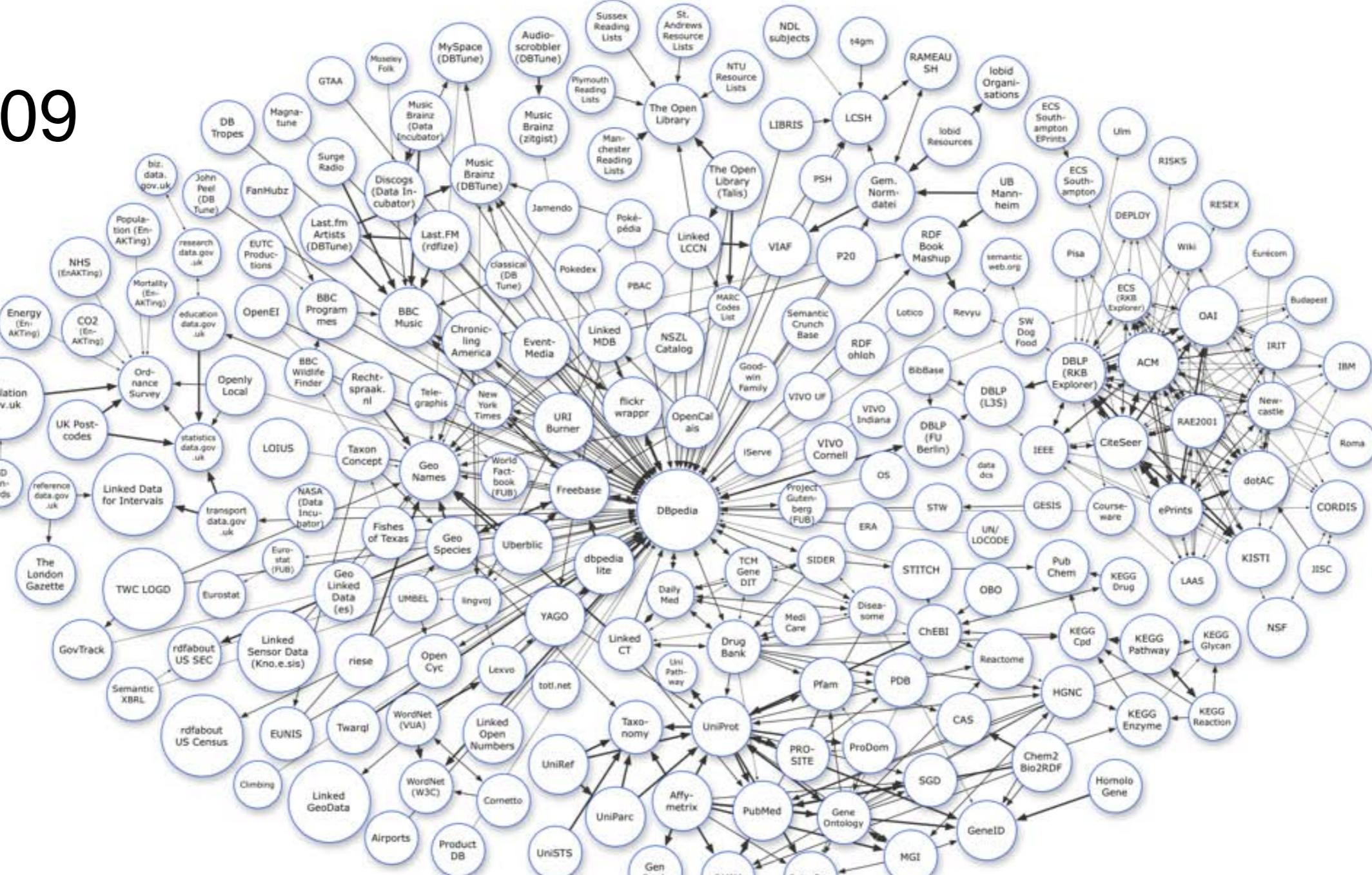
Linked Data on the Web
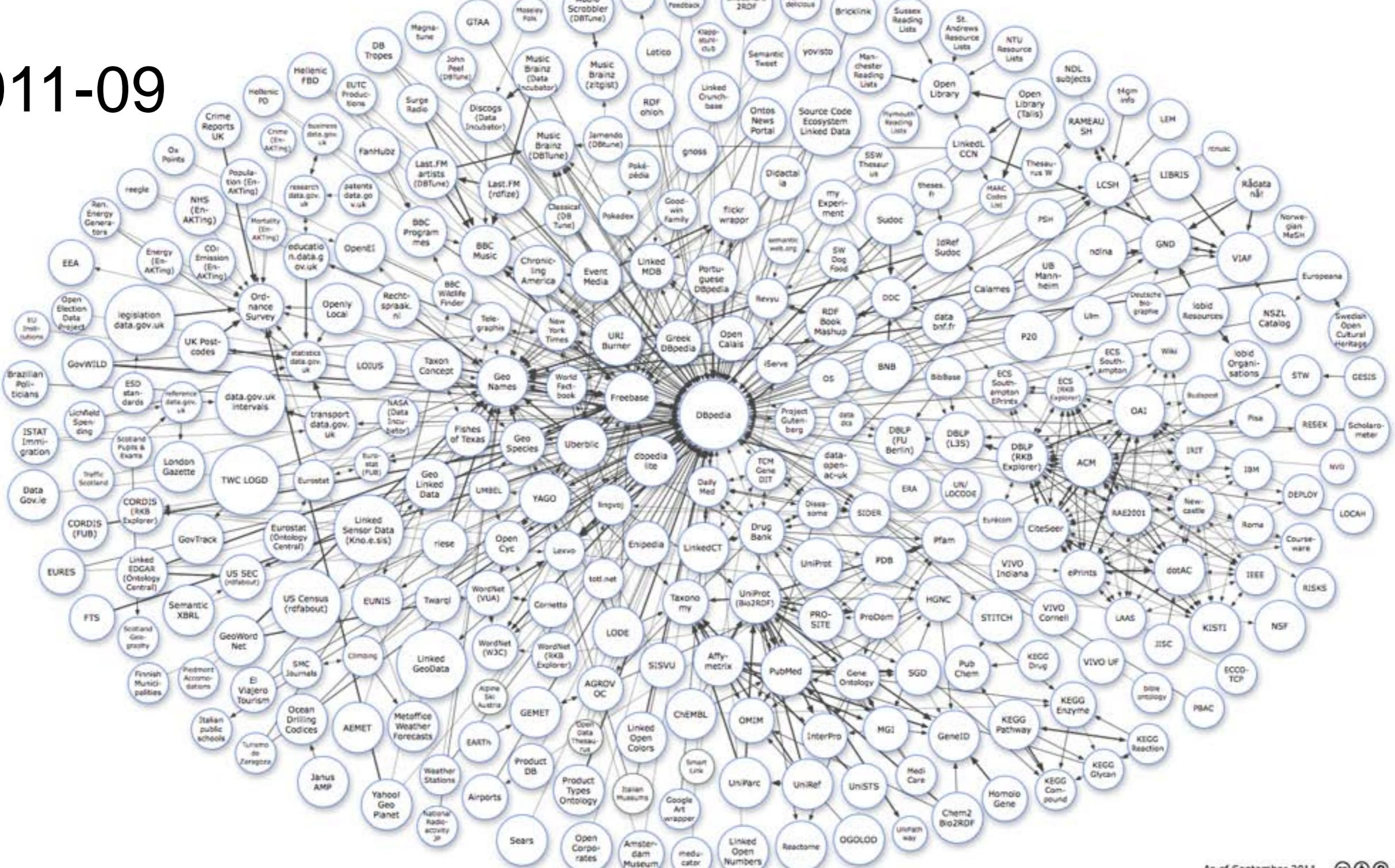
Semantic Web TECHNOLOGIES
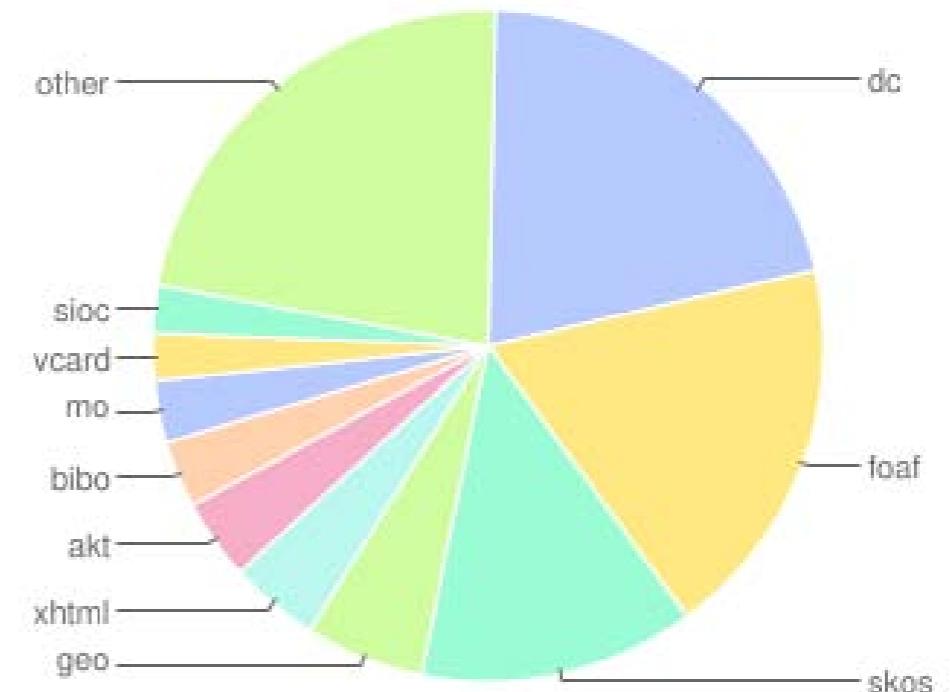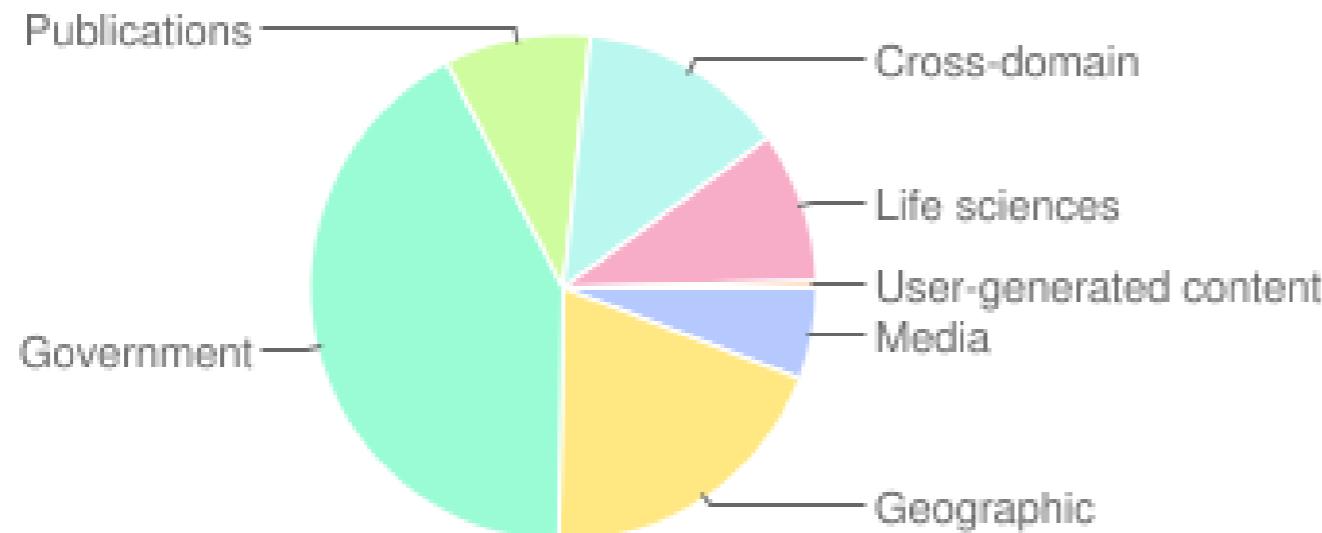
AIFB

2011-09

As of September 2011

Data and ontologies in the Linked Open Data Cloud

AIFB

Publications — Cross-domain
Life sciences
User-generated content
Media
Government
Geographic
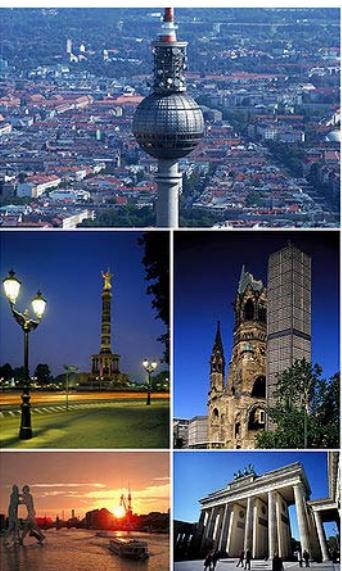
http://www4.wiwiss.fu-berlin.de/lodcloud/state/
(September 2011)

Scenario overview
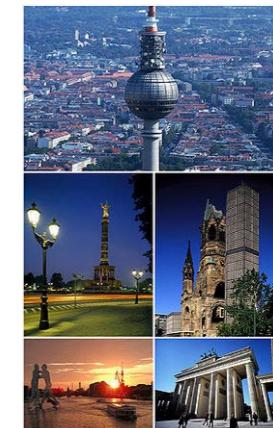
- Semantic technologies facilitate access to data

  - Q: data about Berlin?

  - Q: famous people that died in Berlin?

  - Q: data about Hegel?

  - Q: Hegel's publications?

  - Q: data about Marlene Dietrich?

  - Q: Dietrich's songs?

DBpedia

- Linked Data version of Wikipedia

- Scripts that extract data (text, links, infoboxes) from Wikipedia

- Published as Linked Data

- Interlinking hub in the Linked Data web

- Berlin

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

- Hegel

  - http://dbpedia.org/resource/Georg_Wilhelm_Friedrich_Hegel

- Marlene Dietrich

  - http://dbpedia.org/resource/Marlene_Dietrich

BBC Music

- Data about BBC (radio) programmes, artists, songs…

- Combination of BBC-internal data (playlists), MusicBrainz (artists, albums), Wikipedia (artists)

- Underpinning the BBC Music website

- Data published according to Linked Data principles

- Marlene Dietrich

  - http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5

Virtual International Authority File (VIAF)

- Joint project of national libraries and related organisations

  - 21 institutions, among them the Deutsche Nationalbibliothek

- Provide access to "authority files"

- Matching and interlinking collections from participating institutions

- Hegel

  - http://viaf.org/viaf/89774942

- Marlene Dietrich

  - http://viaf.org/viaf/97773925
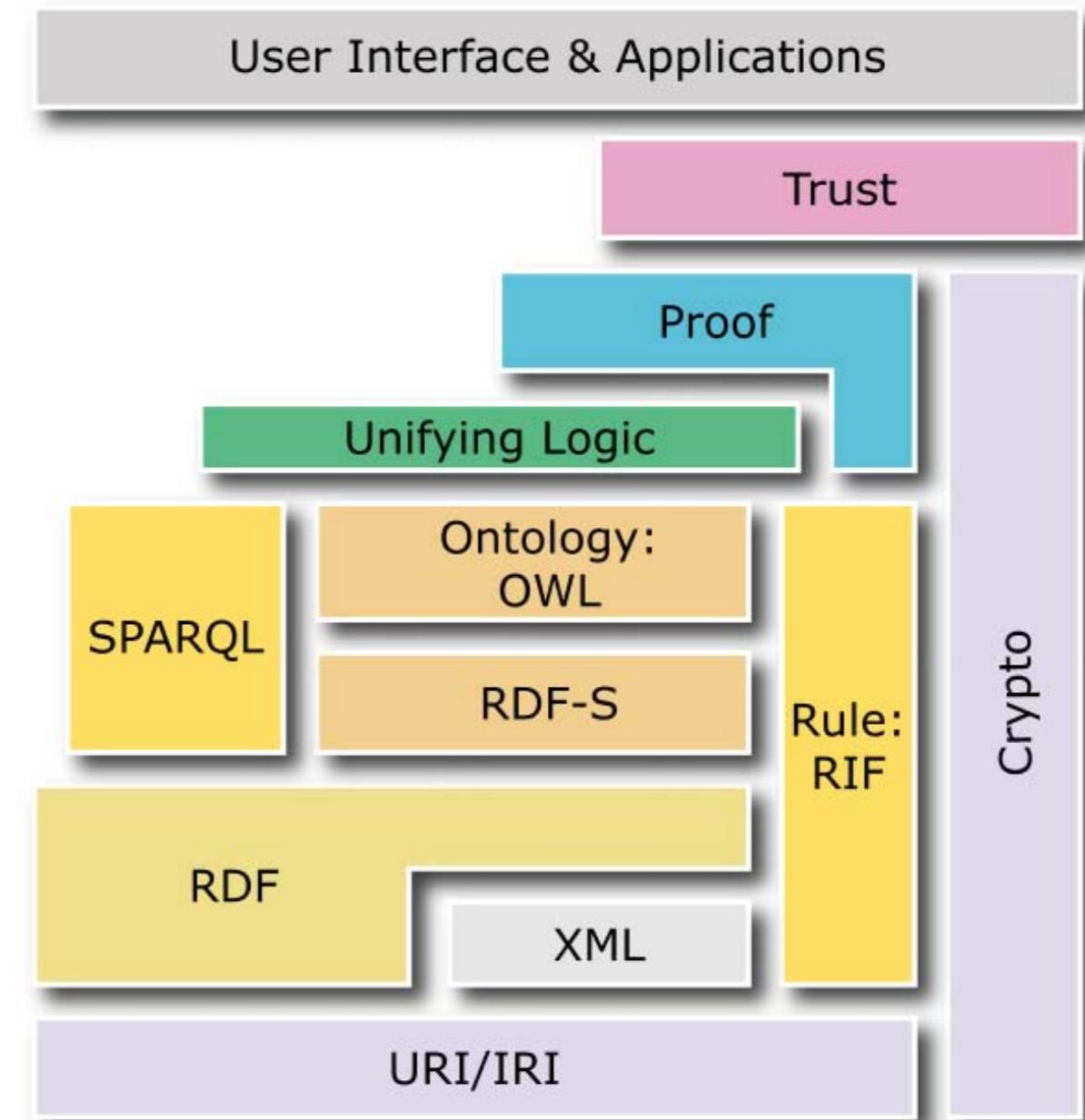
LINKED DATA PRINCIPLES

Semantic technologies

AIFB

- Semantic Web technologies, standardised by the W3C, are mature

  - RDF recommendation in 1999, update in 2004

  - RDFa (RDF in HTML) note in 2008

  - RDFS recommendation in 2004

  - SPARQL recommendation in 2008

  - OWL recommendation in 2004, update in 2009

- Linked Data is a subset of the Semantic Web stack, including web architecture

  - IRI (IETF RFC 3987, 2005)

  - HTTP (IETF RFC 2616, 1999)

User Interface & Applications

Trust

Proof

Unifying Logic

SPARQL

Ontology: OWL

RDF-S

Rule: RIF

Crypto

RDF

XML

URI/IRI

Linked Data principles
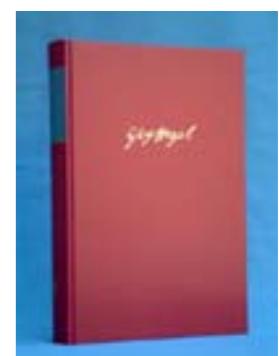
1. Use URIs as names for things

2. Use HTTP URIs so that people can look up those names.

3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)

4. Include links to other URIs. so that they can discover more things.

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

1. Use URIs as names for things

- Use a unique identifier to denote things

- URIs are defined in RFC 2396

- Hegel, Georg Wilhelm Friedrich

  - http://dbpedia.org/resource/Georg_Wilhelm_Friedrich_Hegel

  - http://viaf.org/viaf/89774942

  - …

- Hegel, Georg Wilhelm Friedrich: Gesammelte Werke / Vorlesungen über die Logik

  - urn:isbn:978-3-7873-1964-0

Names for things

"Now! *That* should clear up a few things around here!"

2. Use HTTP URIs

- Enables "lookup" of URIs

- Via Hypertext Transfer Protocol (HTTP)

- Piggy-backs on hierarchical Domain Name System to guarantee uniqueness of identifiers

- Uses established HTTP infrastructure

- Connects logical level (thing) with physical level (source)

- Important: distinction between name/"thing URI" and location/"source URI" („other resource"/„non-information resource" vs. „information resource")

Semantic Web

TECHNOLOGIES

AIFB ☐

Marlene Dietrich, the person
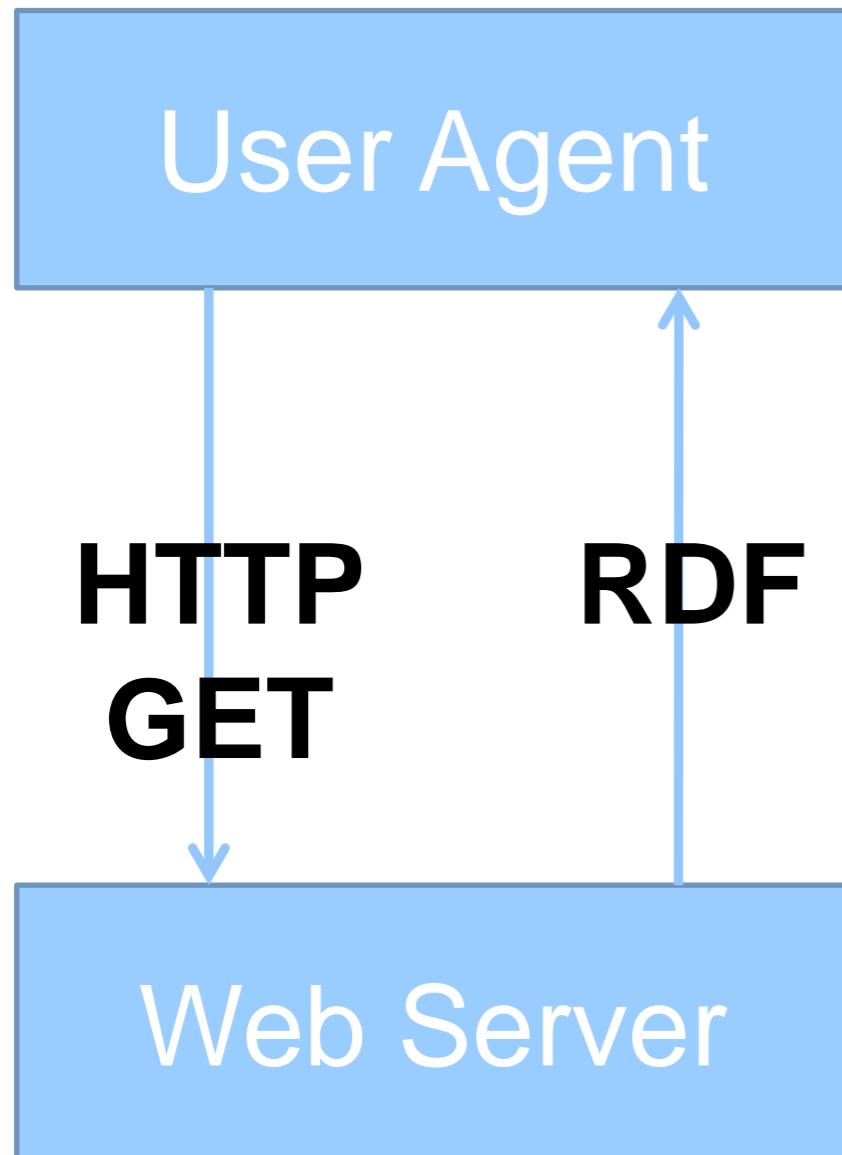
Name?
Creator?
Birth date?
Last change date?
License?
Copyright?
...

File containing data about Marlene Dietrich

Correspondence between thing-URI and source-URI („hash URIs")

User Agent

HTTP GET

RDF

Web Server

http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5#artist

http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5.rdf

Hypertext Transfer Protocol (HTTP)

```
$ curl -H "Accept: application/rdf+xml" -v
     http://www.w3.org/People/Berners-Lee/card#i
```

REQUEST
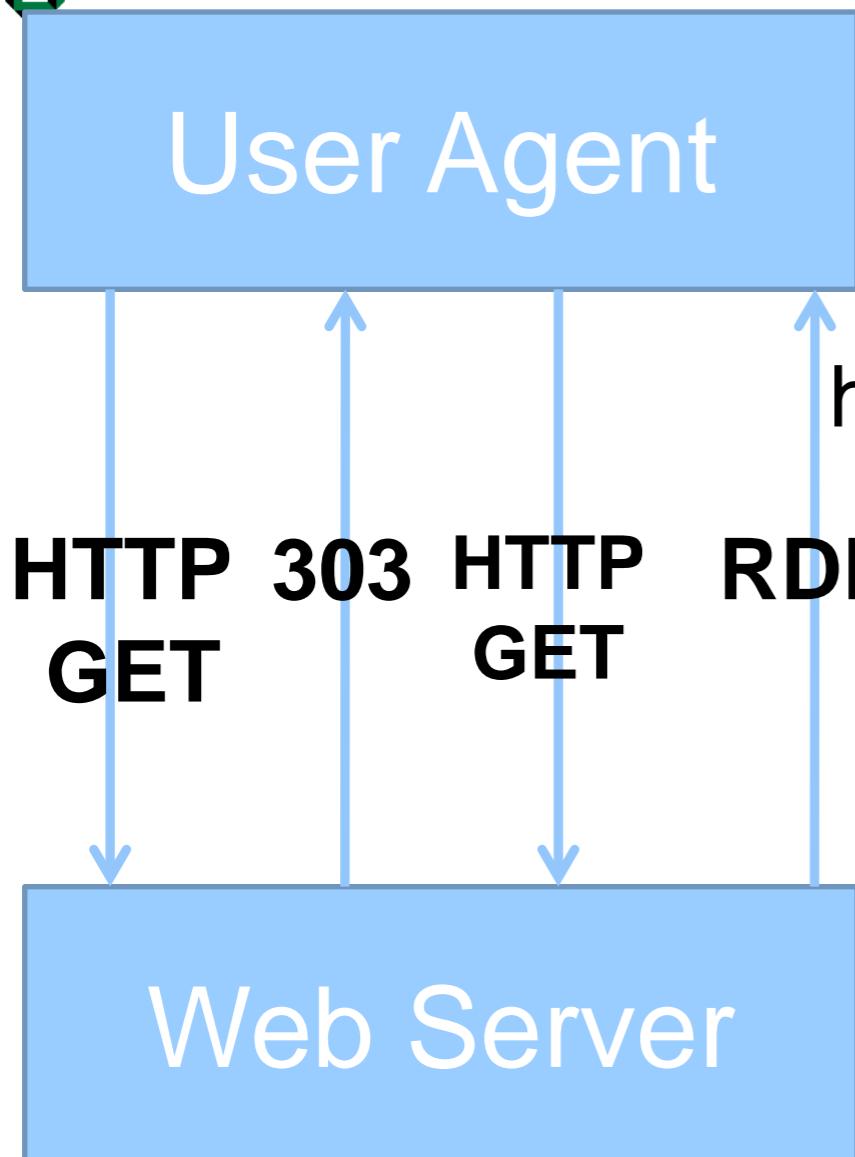
```
> GET /People/Berners-Lee/card HTTP/1.1

> User-Agent: curl/7.21.0

> Host: www.w3.org

> Accept: application/rdf+xml
```

RESPONSE

```
< HTTP/1.1 200 OK

< Date: Mon, 28 Mar 2011 17:16:30 GMT

< Server: Apache/2

< Content-Location: card.rdf
```

Correspondence between thing-URI and source-URI („slash URIs")

User Agent

http://dbpedia.org/resource/Marlene_Dietrich

HTTP GET **303** **HTTP GET** **RDF**

RDF

http://dbpedia.org/data/Marlene_Dietrich

Web Server

http://dbpedia.org/page/Marlene_Dietrich
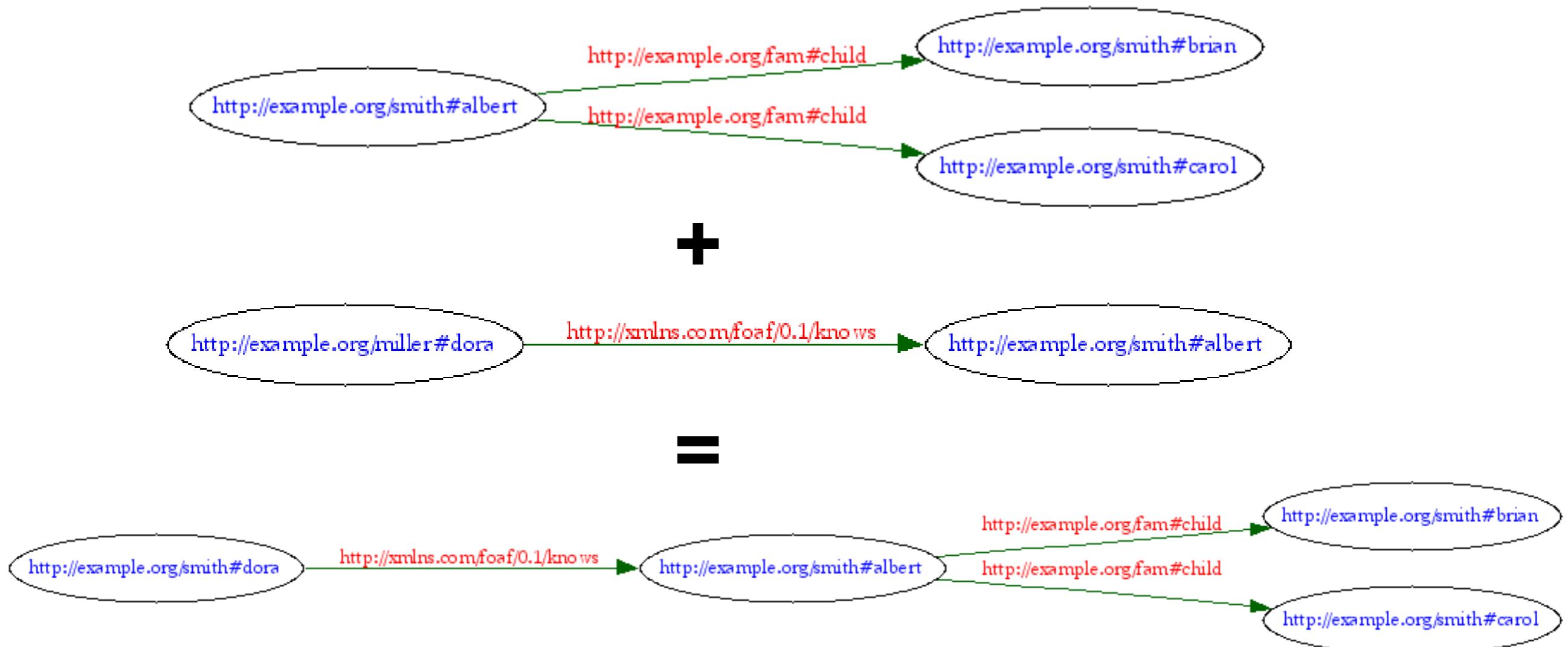
3. Provide useful information

AIFB

- When somebody looks up a URI, return data using the standards (RDF*, SPARQL)

Merging Data with RDF

4. Link to other URIs

- Enable people (and machines) to jump from server to server

- External links vs. internal links (for any predicate)

- Using external vocabularies enables linking

- Vocabularies might be interlinked, too

- Special owl:sameAs links to denote equivalence of identifiers (useful for data merging)

http://viaf.org/viaf/89774942

- http://dbpedia.org/resource/Georg_Wilhelm_Friedrich_Hegel

- http://www.idref.fr/026917467/id

- http://libris.kb.se/resource/auth/190350

- http://d-nb.info/gnd/118547739

http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5#artist

- http://dbpedia.org/resource/Marlene_Dietrich

http://viaf.org/viaf/97773925

- http://dbpedia.org/resource/Marlene_Dietrich .

- http://d-nb.info/gnd/118525565

- http://libris.kb.se/resource/auth/238817

- http://www.idref.fr/027561844/id

http://dbpedia.org/resource/Berlin

- http://mpii.de/yago/resource/Berlin

- http://data.nytimes.com/N50987186835223032381 - Berlin (Germany)

- http://www4.wiwiss.fu-berlin.de/flickrwrappr/photos/Berlin
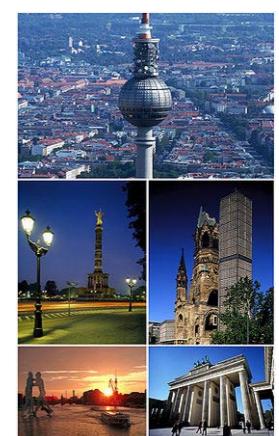
- http://data.nytimes.com/16057429728088573361 - Gaspe Peninsula (Quebec) (?)

Benefits of Linked Data

- Explicit, simple data representation

  - Common data representation (Resource Description Framework, RDF) hides underlying technologies and systems

- Distributed system

  - Decentralized distributed ownership and control facilitates adoption and scalability

- Cross-referencing

  - Allows for linking and referencing of existing data, via reuse of URIs

- Loose coupling with common language layer

  - Large scale systems require loose coupling, via HTTP as common access protocol

- Ease of publishing and consumption

  - Simple and easy-to-use systems and technologies to facilitate uptake

- Incremental data integration

  - Start with merged RDF graphs and provide mappings as you go

Challenges

- Ramp-up cost for data conversion

  - May be alleviated by semi-automatic mappings and adequate tool support for manual conversion

- Integrated data may be messy at first

  - But can be refined as need arises

- Distributed creation and loose coordination may result in inconsistencies

  - Can be detected, diagnosed, and fixed with appropriate tools

Data.gov & public sector information

- Many data sets useful for business intelligence

BBC & Media

- Value of content increased by Linked Data

Semantic technologies at BBC

AIFB

- Various micro-sites built and maintained manually

- No integration across sites in terms of content and metadata

- Use cases

  - Find and explore content on specific (and related) topics

  - Maintain and re-organize sites

  - Leverage external resources

- Ontology: One page per thing, reusing DBpedia and MusicBrainz IDs, different labels…

„Design for a world where Google is your homepage, Wikipedia is your CMS, and humans, software developers and machines are your users"

http://www.slideshare.net/reduxd/beyond-the-polar-bear

Semantic technologies at BBC(2)