Linked Data
Towards a Web of Data

Anja Jentzsch, Freie Universität Berlin

Outline

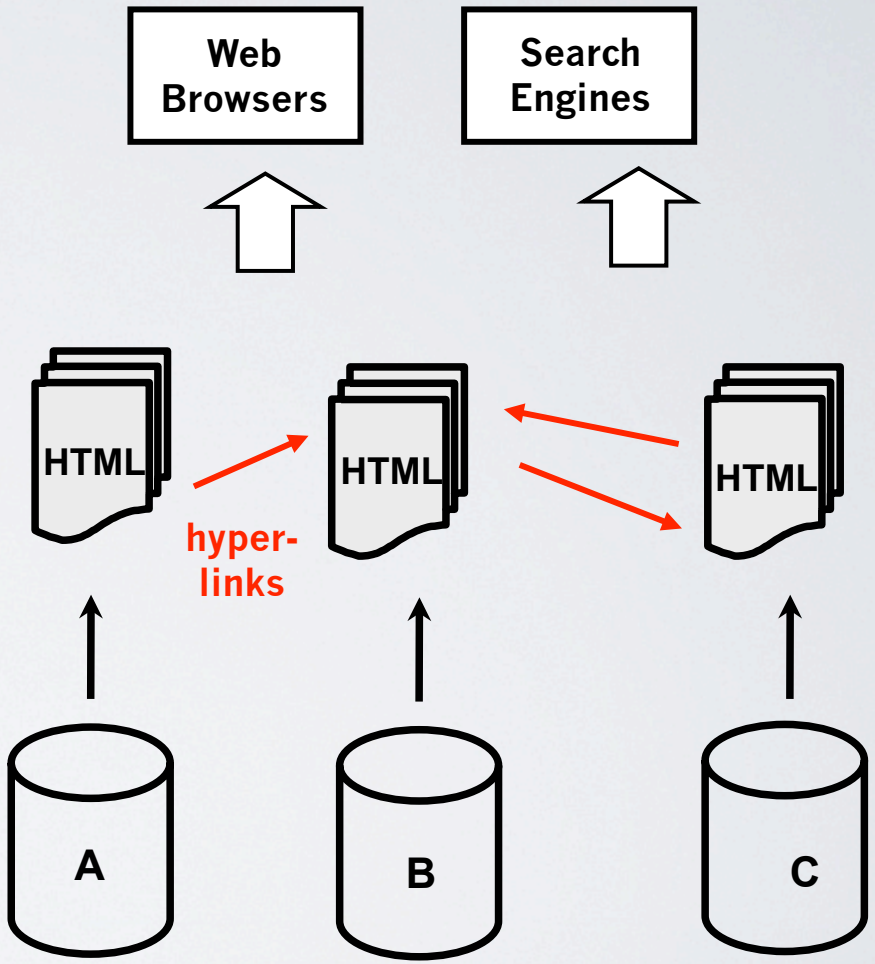
1. Foundations of Linked Data
   - What is the vision and goal?
2. The Web of Linked Data
   - What data is out there?
   - What is being done with the data?
3. How to publish Linked Data?
4. How to consume Linked Data?

Architecture of the classic Web

Single global dataspace

Small set of simple standards

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

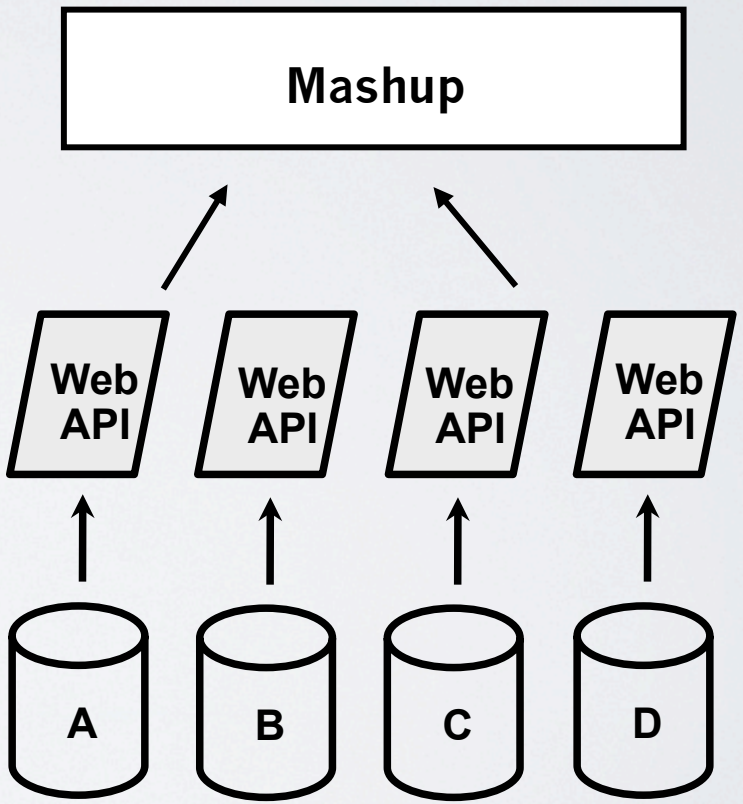


Web 2.0 APIs and Mashups

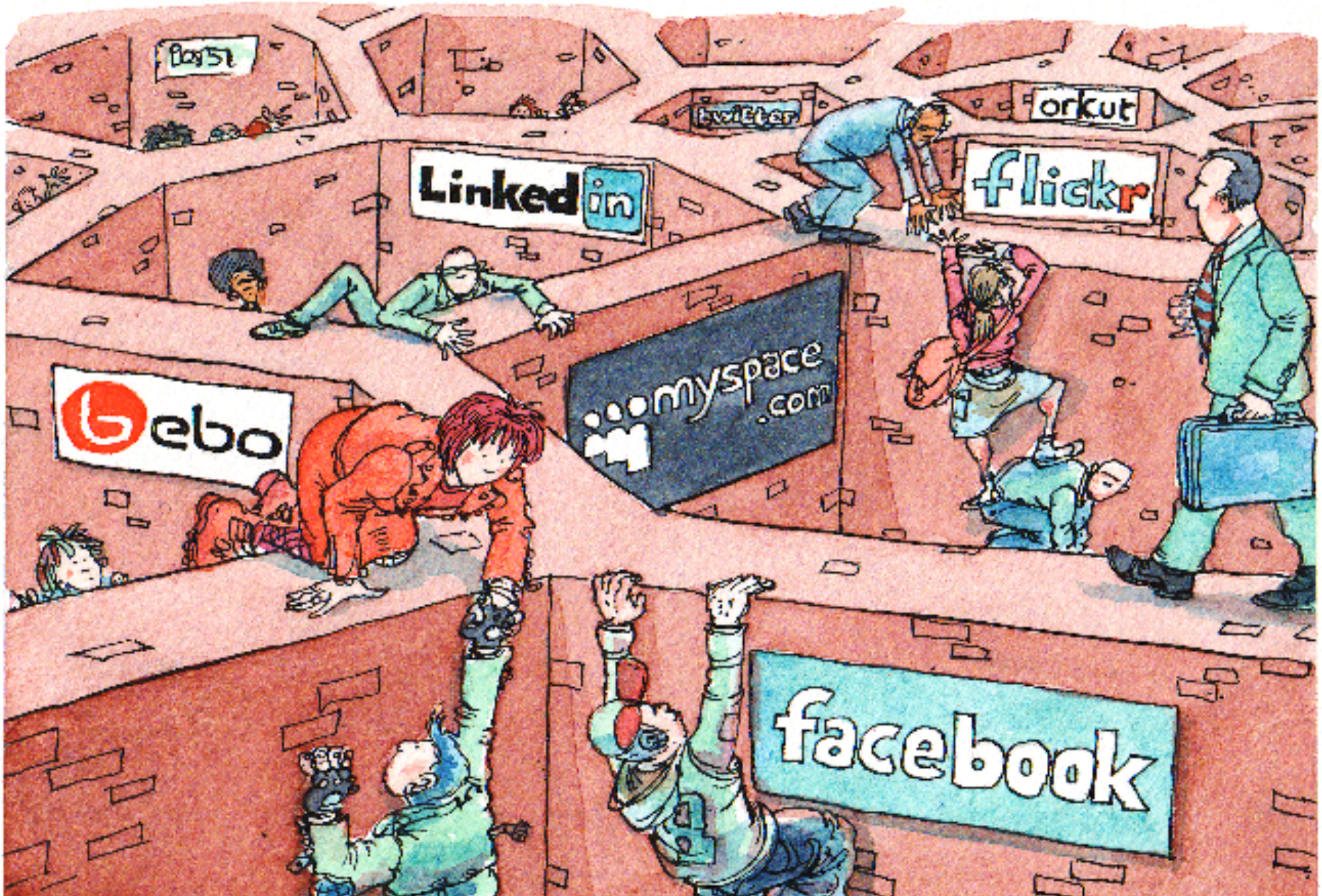
No single global dataspace

Shortcomings

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



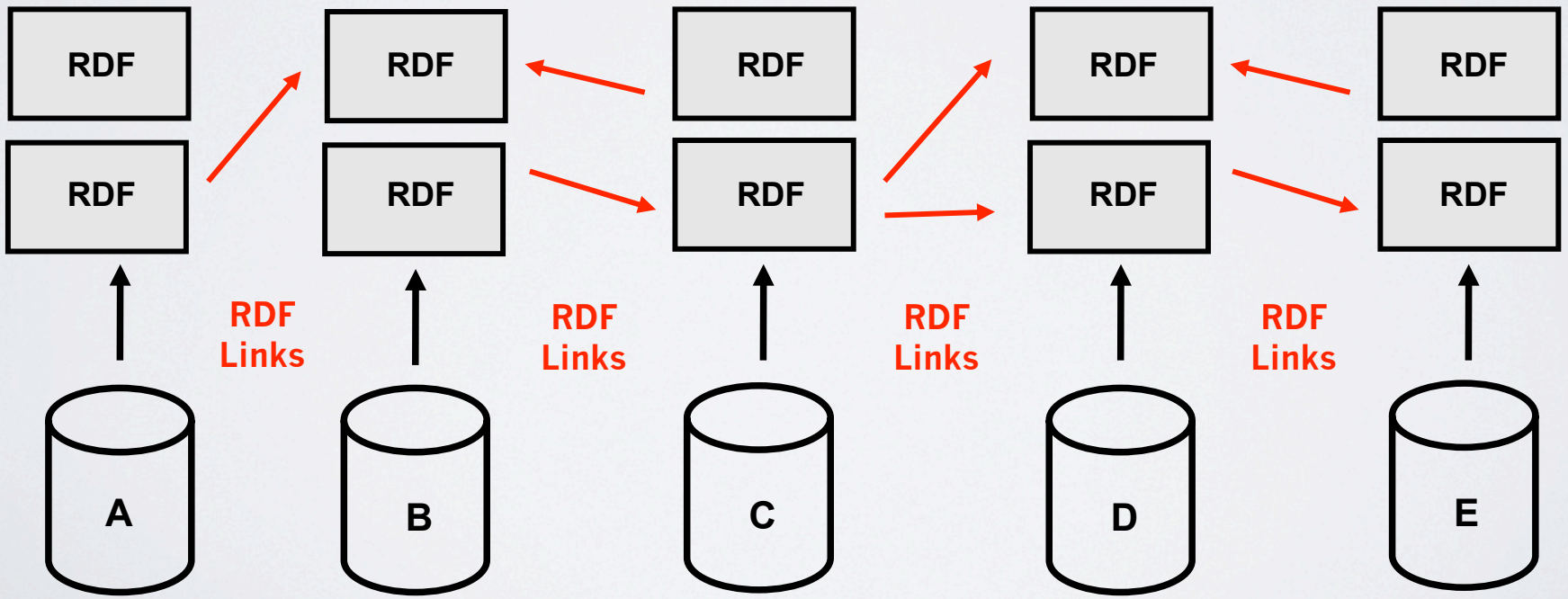
Web APIs slice the Web into Walled Gardens



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 Principles

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

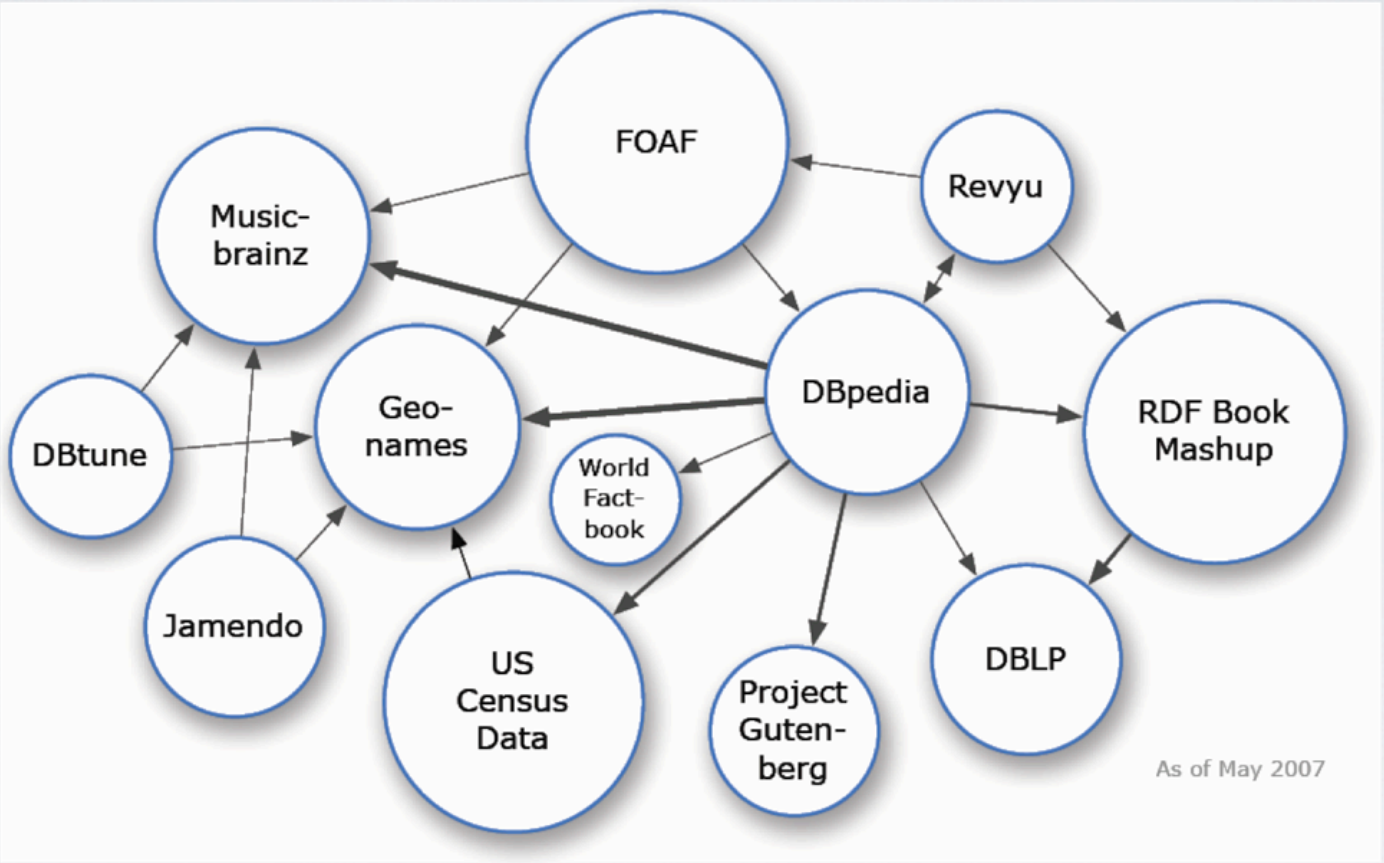
Properties of the Web of Linked Data

- 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
  - Everybody can use the vocabularies/schema that they like

W3C Linking Open Data Project

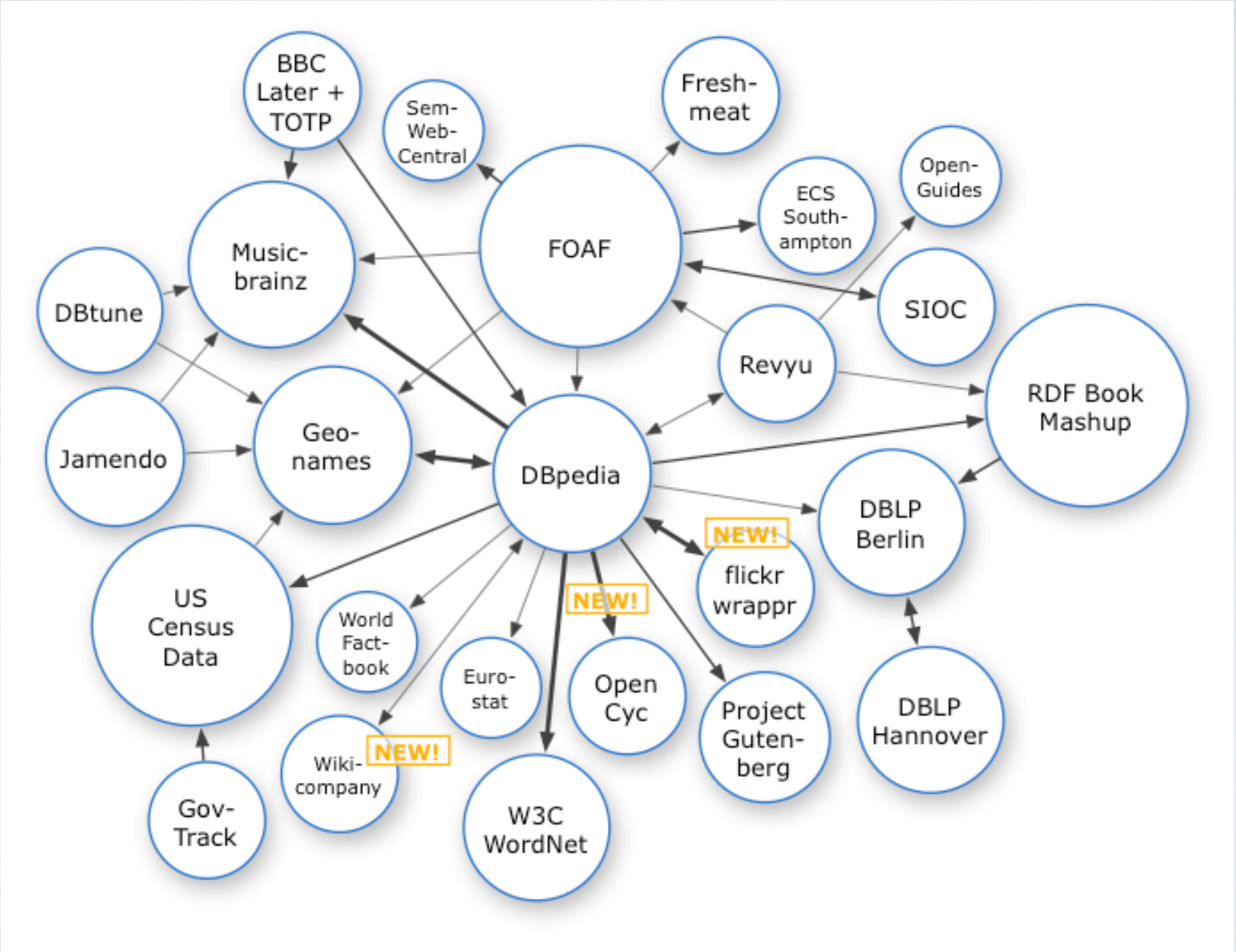
- Grassroots community effort to
  - publish existing open license datasets as Linked Data on the Web
  - interlink things between different data sources

LOD Data Sets on the Web: May 2007

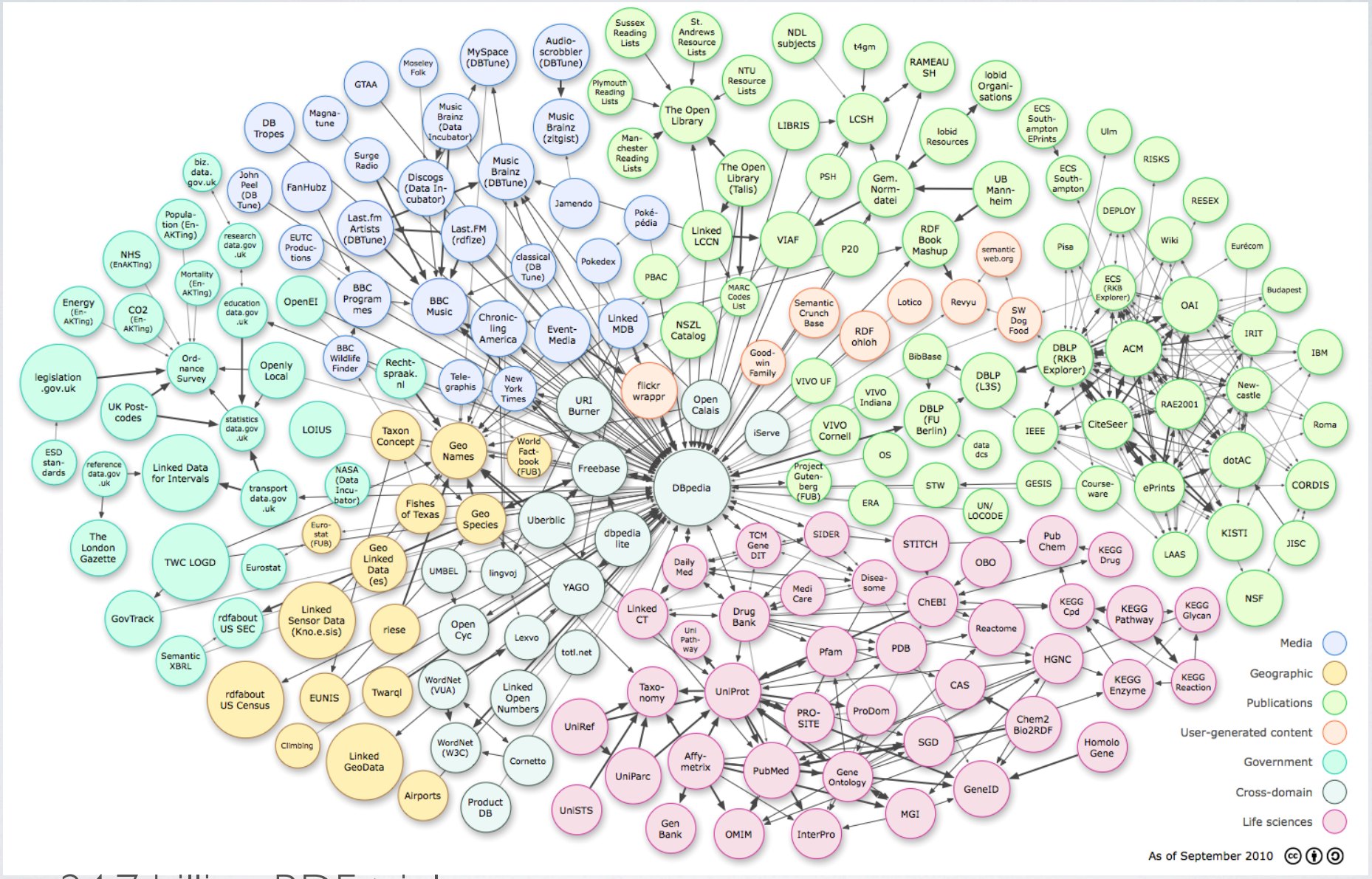


- Over 500 million RDF triples
- Around 120,000 RDF links between data sources

LOD Data Sets on the Web: September 2008



LOD Data Sets on the Web: September 2010



- Over 24,7 billion RDF triples
- Over 436 million RDF links between data sources

LOD Data Set statistics as of Nov 2010

Domain	Data Sets	Triples	Percent	RDF Links	Percent
Cross-domain	20	1,999,085,950	7.42	29,105,638	7.36
Geographic	16	5,904,980,833	21.93	16,589,086	4.19
Government	25	11,613,525,437	43.12	17,658,869	4.46
Media	26	2,453,898,811	9.11	50,374,304	12.74
Libraries	67	2,237,435,732	8.31	77,951,898	19.71
Life sciences	42	2,664,119,184	9.89	200,417,873	50.67
User Content	7	57,463,756	0.21	3,402,228	0.86
	203	**26,930,509,703**		**395,499,896**	

LOD Cloud Data Catalog on CKAN

- http://www.ckan.net/group/lodcloud

More statistics

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

DBpedia – The Hub on the Web of Data

- DBpedia is a joint project with the following goals
  - extracting structured information from Wikipedia
  - publish this information under an open license on the Web
  - setting links to other data sources

- Partners
  - Freie Universität Berlin (Germany)
  - Universität Leipzig (Germany)
  - OpenLink Software (UK)
  - neofonie (Germany)

Extracting structured data from Wikipedia

Log in / create account

WIKIPEDIA
The Free Encyclopedia

Article Discussion Read Edit View history Search

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Toolbox
What links here
Related changes
Upload file
Special pages
Permanent link
Cite this page

Print/export

Languages
Afrikaans
Alemannisch
Ænglisc

Geneva

From Wikipedia, the free encyclopedia

Coordinates: 46°12′N 6°09′E

For other uses, see Geneva (disambiguation).

Geneva (/dʒɨˈnɪvə/; French: *Genève*, IPA: [ʒənɛv]; Arpitan: *Genèva*, IPA: [d͡zənɛva])[note 1] is the second-most-populous city in Switzerland (after Zürich) and is the most populous city of Romandie, the French-speaking part of Switzerland. Situated where the Rhone exits Lake Geneva, it is the capital of the Republic and Canton of Geneva. While the municipality itself *(city centre ville de Genève)* has a population (as of December 2010) of 191,415, the canton of Geneva (*République et Canton de Genève*, which includes the city) has 457,628 residents (as of December 2009).[2] The agglomeration has 1,240,000 residents.

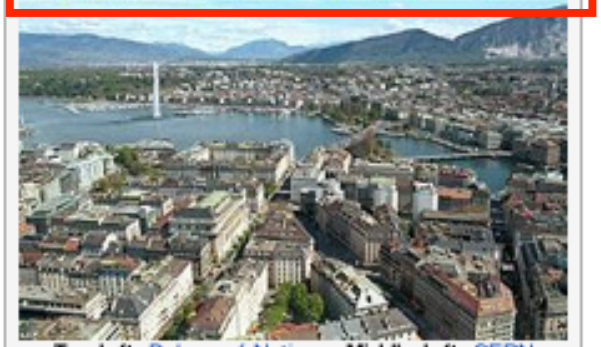
Geneva is a global city, a financial centre, and a worldwide centre for diplomacy and the most important UN international co-operation centre with New York thanks to the presence of numerous international organizations, including the headquarters of many of the agencies of the United Nations[3] and the Red Cross.[4] It is also the place where the Geneva Conventions were signed, which chiefly concern the treatment of wartime non-combatants and prisoners of war.

Geneva has been described as the third European financial centre after London and Zurich,[5] and the world's eighth most important financial centre by the Global Financial Centres Index, ahead of Frankfurt and Sydney,[6] and a 2009 survey by Mercer found Geneva to have the third-highest quality of life of any city in the world (narrowly outranked by Zürich).[7] The city has been referred to as the world's most compact metropolis and the "Peace Capital."[8] In 2009, Geneva was ranked as the fourth most expensive city in the world.[9]

Contents [hide]

1 Etymology
2 History
2.1 Antiquity

Geneva



Top left: Palace of Nations, **Middle left:** CERN Laboratory, **Right:** Jet d'Eau, **Bottom:** View over Geneva and the lake.

Country	Switzerland
Canton	Geneva
District	N/A

Extracting structured data from Wikipedia

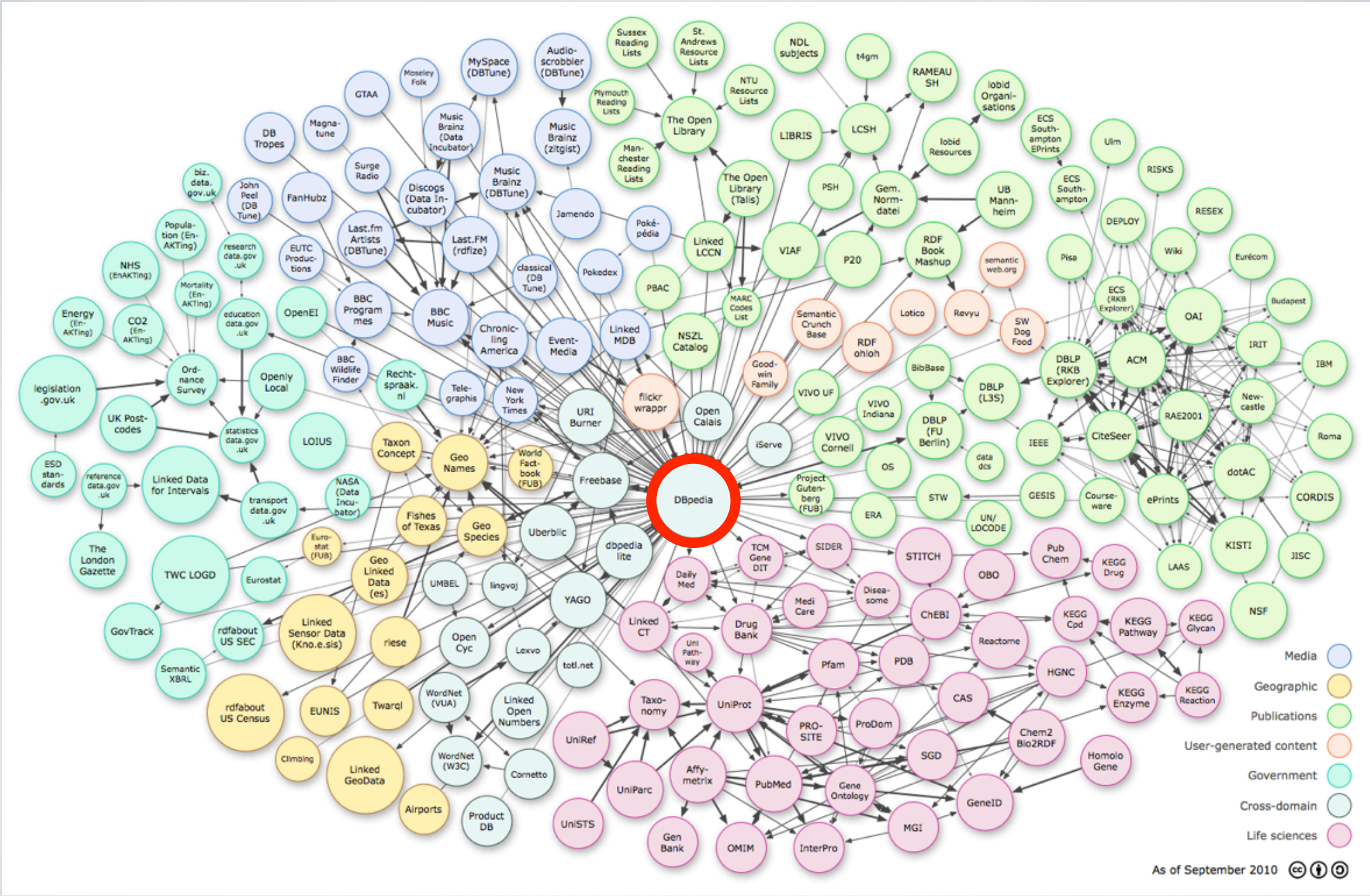
```
dbpedia:Geneva  rdf:type  dbpedia-owl:Place ,
      dbpedia-owl:PopulatedPlace ,
      dbpedia-owl:Settlement ;
   rdfs:label "Geneva"@en ,
      "Genf"@de ;
   dbpedia-owl:population 191237 ;
   dbpedia-owl:postalCode "1200"@en .

dbpedia:CERN dbpedia-owl:headquarter dbpedia:Geneva .
```

- access to DBpedia data:
  - dumps
  - SPARQL endpoint
  - Linked Data interface

The DBpedia Data Set

- Information on more than 3.5 million "things"
  - 364,000 persons
  - 148,000 organizations
  - 462,000 places
  - 99,000 music albums
  - 54,000 movies
  - 169,000 species
- overall more than half a billion RDF triples
  - title and abstract in 97 different languages
  - 1,850,000 links to images
  - 5,900,000 links to external websites
  - 6,500,000 links to other Linked Data sets


DBpedia
Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences
As of September 2010

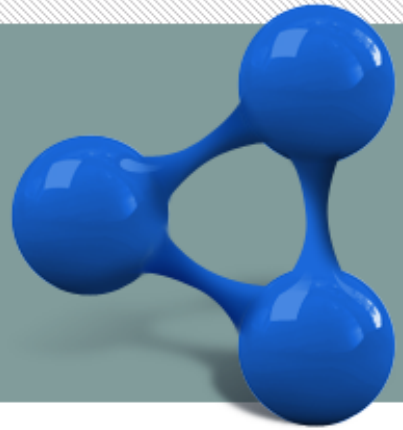
HM Government

data.gov.uk

Home | Blog | Data | SPARQL | Apps | Ideas | Forum | Wiki | Resources | About

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) Search

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

Powered by: CKAN

Browse for Data

List all datasets

By Public Body

Common tags

Subscribe by RSS

Community
Log in / Sign up

Local Data Panel

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.

Read more

Digital Engagement Twitter stream

DATA.GOV
EMPOWERING PEOPLE

HOME | DATA | TOOLS | COMMUNITY | METRICS | DIALOGUE

LINKING OPEN GOVERNMENT DATA

VIEW MORE

Most Popular Datasets

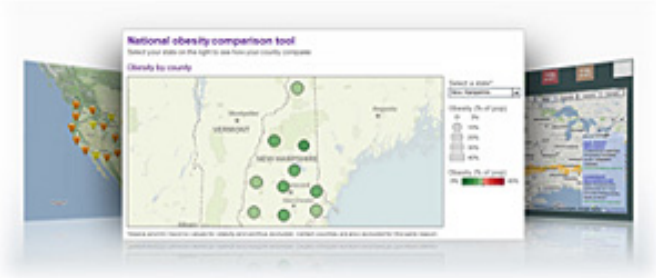
1. Worldwide M1+ Earthquakes, Past 7 Days
2. U.S. Overseas Loans and Grants (Greenbook)
3. Latest Volumes of Foreign Relations of the...
4. OSHA Data Initiative - Establishment...
5. IT Dashboard - Federal IT Spending (major...

SEARCH OUR CATALOGS

Search our catalogs..

SEARCH

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

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:

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

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 project just released data about 4 million artifacts
- W3C Library Linked Data Incubator Group
- OKFN Working Group on Bibliographic Data
- Goals:
  - Integrate Library Catalogs on global scale
  - Interconnect resources between repositories (by topic, by location, by historical period, by ...)

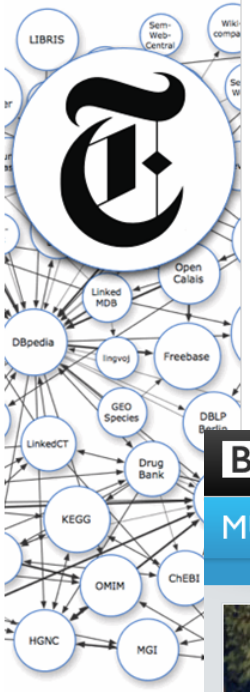
Uptake in the Media Industry

- Publish data as RDF/XML or RDFa
- Goal: Drive traffic to websites via search engines

The New York Times

Linked Open Data BETA

Search data.nytimes.com



data.nytimes.com

For the last 150 years, The New York Times has maintained one of the most authoritative news vocabularies ever developed. In 2009, we began to publish this vocabulary as linked open data.

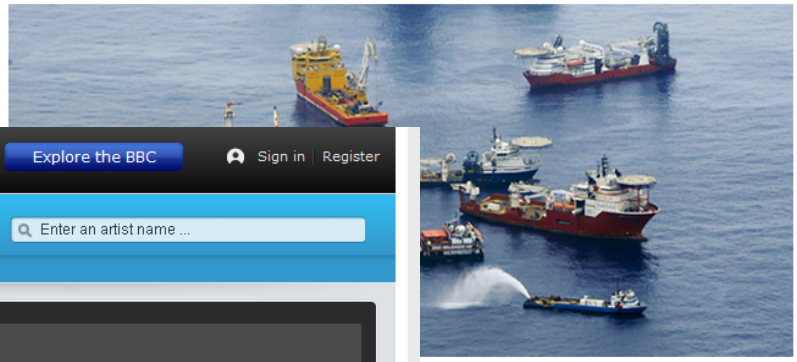
The Data

As of 13 January 2010, The New York Times has published approximately ,10,000 subject headings as linked open data under a CC BY license. We provide both RDF documents and a human-friendly HTML versions. The table below gives a breakdown of the various tag types and mapping strategies on data.nytimes.com.

Type	Manually Mapped Tags	Automatically Mapped Tags
People	4,978	0
Organizations	1,489	1,592
Locations	1,910	0

North Korea's Exit Strategy | Facebook: Friend or Foe? | BP Oil Spill Timeline

Subscribe msnbc.com



tter Plan?

e scenario if they wanted to drill the gulf. But
ıt approved—is alarmingly short on

Newsweek
What matters most
STORIES TOPICS AUTHORS SEARCH

FedEx Ground
Faster to more locations than UPS Ground.®
UPS is a registered trademark of United Parcel Service
SAVE NOW

BBC Text only | Help Search Explore the BBC Sign in | Register

MUSIC BETA GENRES ARTISTS REVIEWS NEWS BLOG QUICK FIND Enter an artist name ...

BBC TWO
BBC TWO
Later Live
With Mumford & Sons, Hole, Joanna Newsom, Ian Hunter and Angelique Kidjo, plus a chat with Iggy Pop.

MORE MUSIC ON THE BBC 1-3 OF 20

LATER LIVE

GORILLAZ - LIVE IN THE FLESH ON LATER...

KATIE MELUA - WEEKEND WOGAN

BROWSE BY GENRE • Classic Pop & Rock • Classical • Country • Dance & Electronica • Desi • Easy Listening, Soundtracks & Musicals • Folk • Hip Hop, RnB & Dancehall • Jazz & Blues • Pop & Chart • Rock & Indie • Soul & Reggae • World

Most Played Artists On The BBC

3-9 MAY 2010

eek NOW

Subscribe now and save up to 78%

ere. We're Queer. We're Retiring.

To Meet The Demands Of America's Aging Gay Population, Developers Are Now Targeting The Lgbt Market With Everything From Active Adult Rental Apartments To Retirement Communities That Promise Life-Long Care. More ›

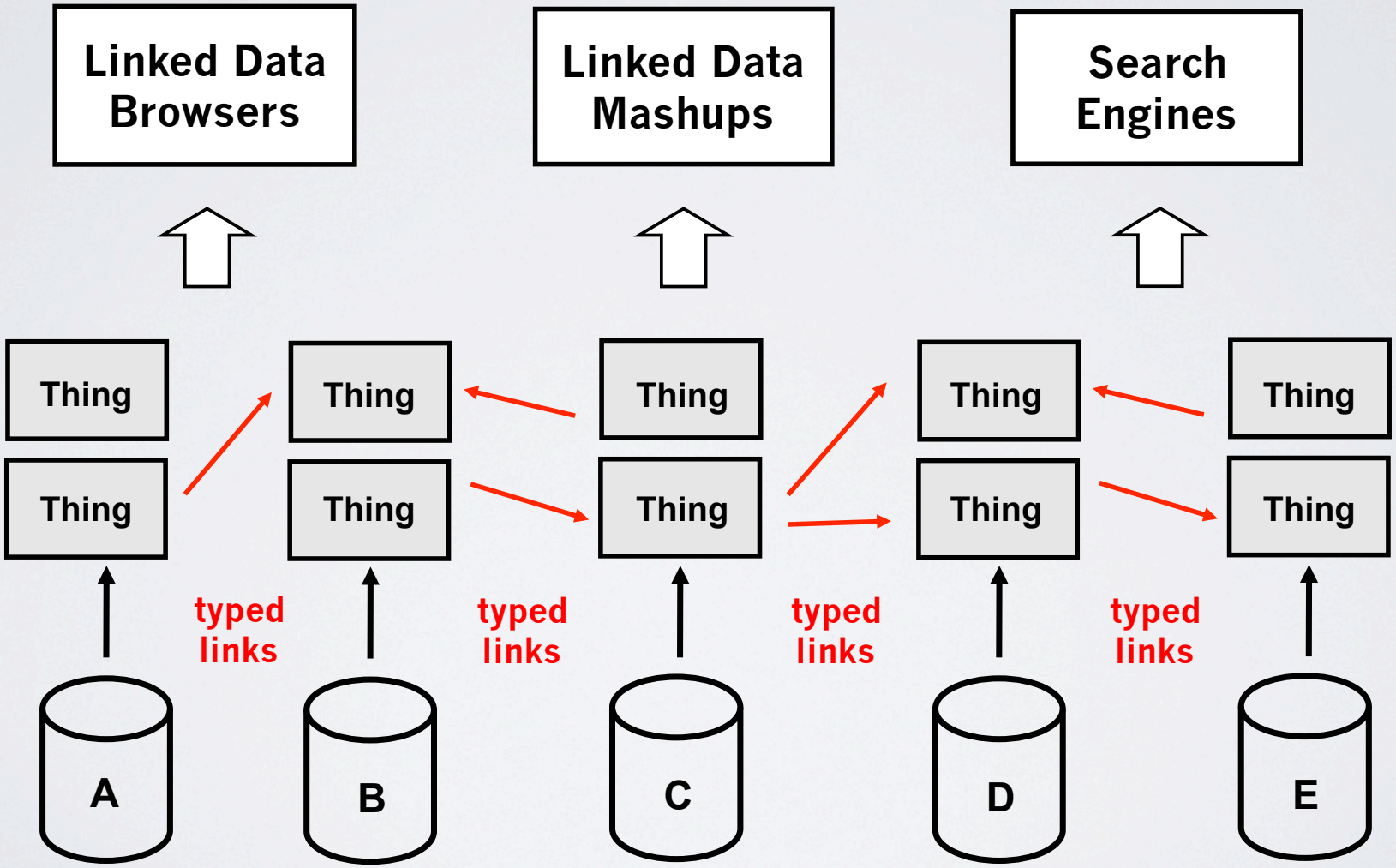
Business

by Linda Stern

May 26, 2010

43

Linked Data Applications



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 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/

SIG.MA
SEMANTIC INFORI
MASHUP

Help | About | Forum

Chris Bizer | Add More Info | Start New | Options | Order | Permalink

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]

hide value | just this value | which sources | reject sources [6]

RAP: RDF API for PHP [16]

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

NG4J - Named Graphs API for Jena [16]

Sources (20) | Approved (0)

1 Chris Bizer - Free Unive... 3
http://videolectures.net/chris_bizer/

2 Chris Bizer - semanticwe... 3
http://ontoworld.org/wiki/Chris_Bizer

3 Untitled document 6 facts | 20
BOSS http://www.facebook.com/pe

4 Chris Bizer - semanticwe... 3
http://semanticweb.org/wiki/Chris_Bizer

5 Chris Bizer - LinkedIn 9 facts
BOSS http://www.linkedin.com/in/c

6 Chris Bizer 10 facts | 2008-05-2
http://data.semanticweb.org/person/chr

7 Chris Bizer - semanticwe... 3
http://semanticweb.org/index.php?title=

8 Flickr: Chris Bizer's Ph... 3 fa
BOSS http://flickr.com/photos/2618

9 Untitled document 8 facts | 20
http://data.semanticweb.org/conference

10 Chris Bizer 6 facts | 2009-06-0
BOSS http://ebiquity.umbc.edu/pers

<- | 1 | 2 | ->

http://example.loc/document.

BBC | Text only | Help

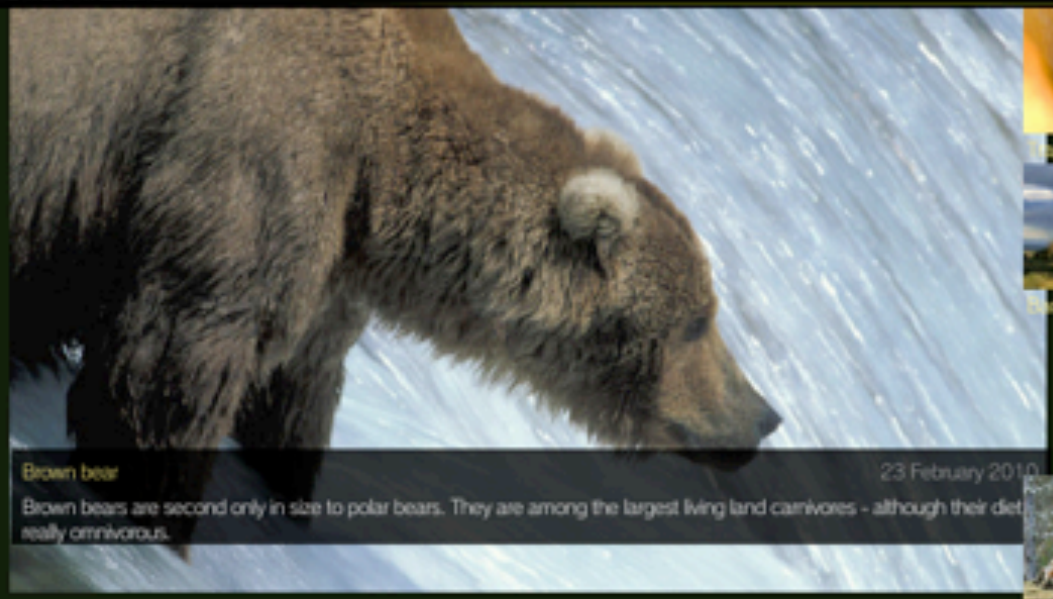
Search | Explore the BBC

WILDLIFEFINDER

WATCH THE MOST AMAZING ANIMALS IN THE WORLD

BBC Nature

LATEST WILDLIFE



Brown bear — 23 February 2010

Brown bears are second only in size to polar bears. They are among the largest living land carnivores - although their diet is really omnivorous.

Trapping predator

Detritus recycler

Common wasp

Baobabs

Goshawks and sparrowhawks

Black grouse

Northern goshawk

Eurasian grouses

Arctic fox

COLLECTIONS

Collections offer a new perspective on the vast catalogue of clips showcasing presenter favourites, aspects of wildlife film -making, and gathering content together in illuminating ways.

Year of the Tiger

2010 is the Year of the Tiger in the Chinese calendar, a zodiac sign associated with power, passion and courage, yet these majestic animals remain under threat from human activity.

What on Earth...? 2009

Watch the year's highlights from the BBC's exploration of the planet's hidden corners and rarest creatures: from the turquoise seas of the South Pacific to the Lost Land of the Volcano.

EDITOR'SCHOICE

In celebration of St David's day this week, meet the real dragons that roam the globe today.

Brown basilisk 1 video

Thorny devil 1 video

Leafy sea dragon 1 video

Weedy sea dragon 1 video

Komodo dragon 3 videos

Marine iguana 3 videos

WELCOMETO WILDLIFE FINDER

Explore a wealth of video, sounds, stories and breaking news to find out more about your favourite animals, how they live and where they live.

Wildlife Finder gives you access to an ever growing catalogue of BBC natural history programmes, with video clips from series such as: Planet Earth, Blue Planet, Life on Earth, Natural World, Life and many more.

WHAT'SNEW?

1 March 2010
Featuring: 11 new clips from this week's Natural World film about forest elephants.

26 February 2010
Three ways to catch your food - new clips and pages on trapping, ambush and pack-hunting.

18 February 2010
Meet the Banded Brothers and watch clips from the new series.

How to publish Linked Data ?

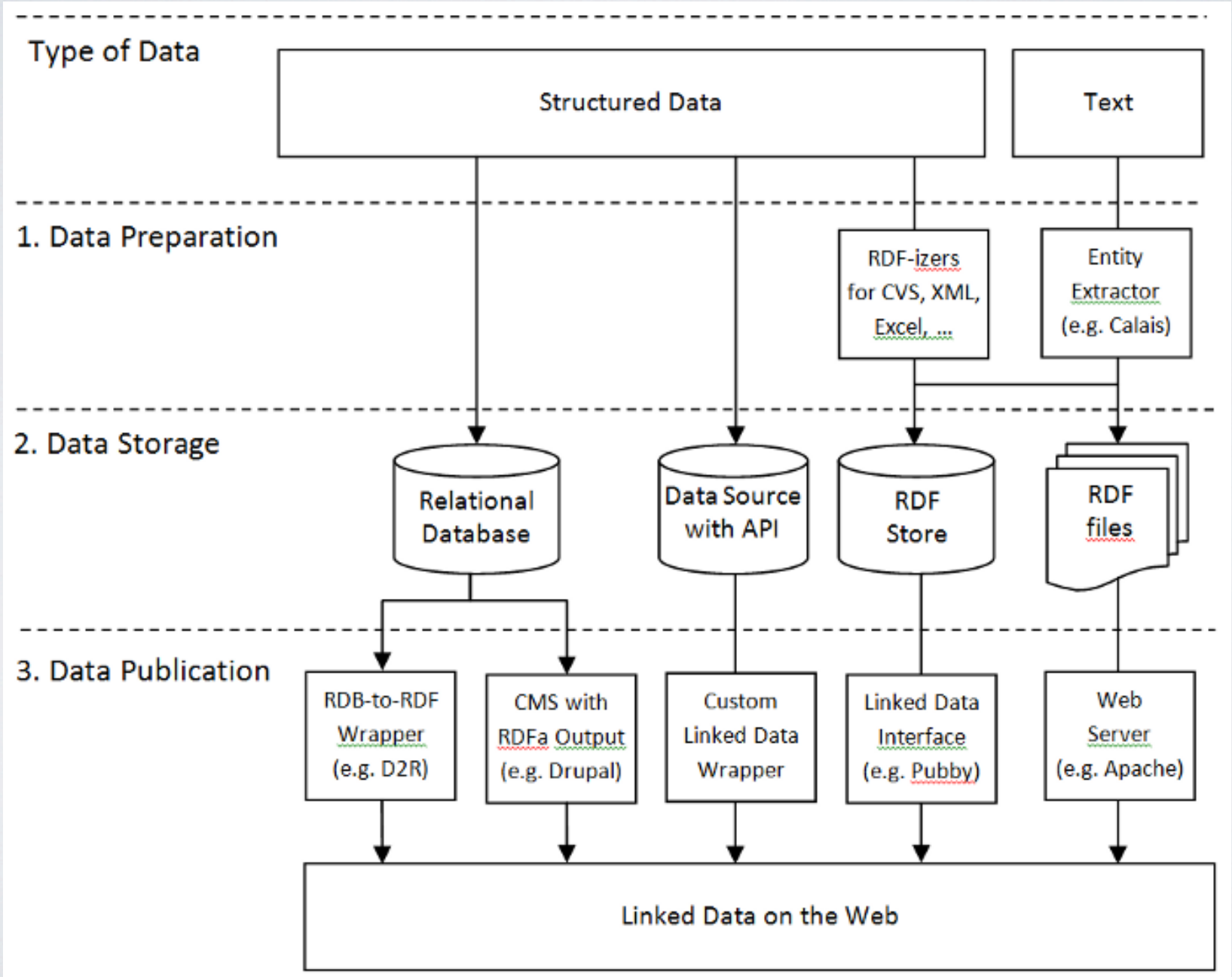
Tasks:

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

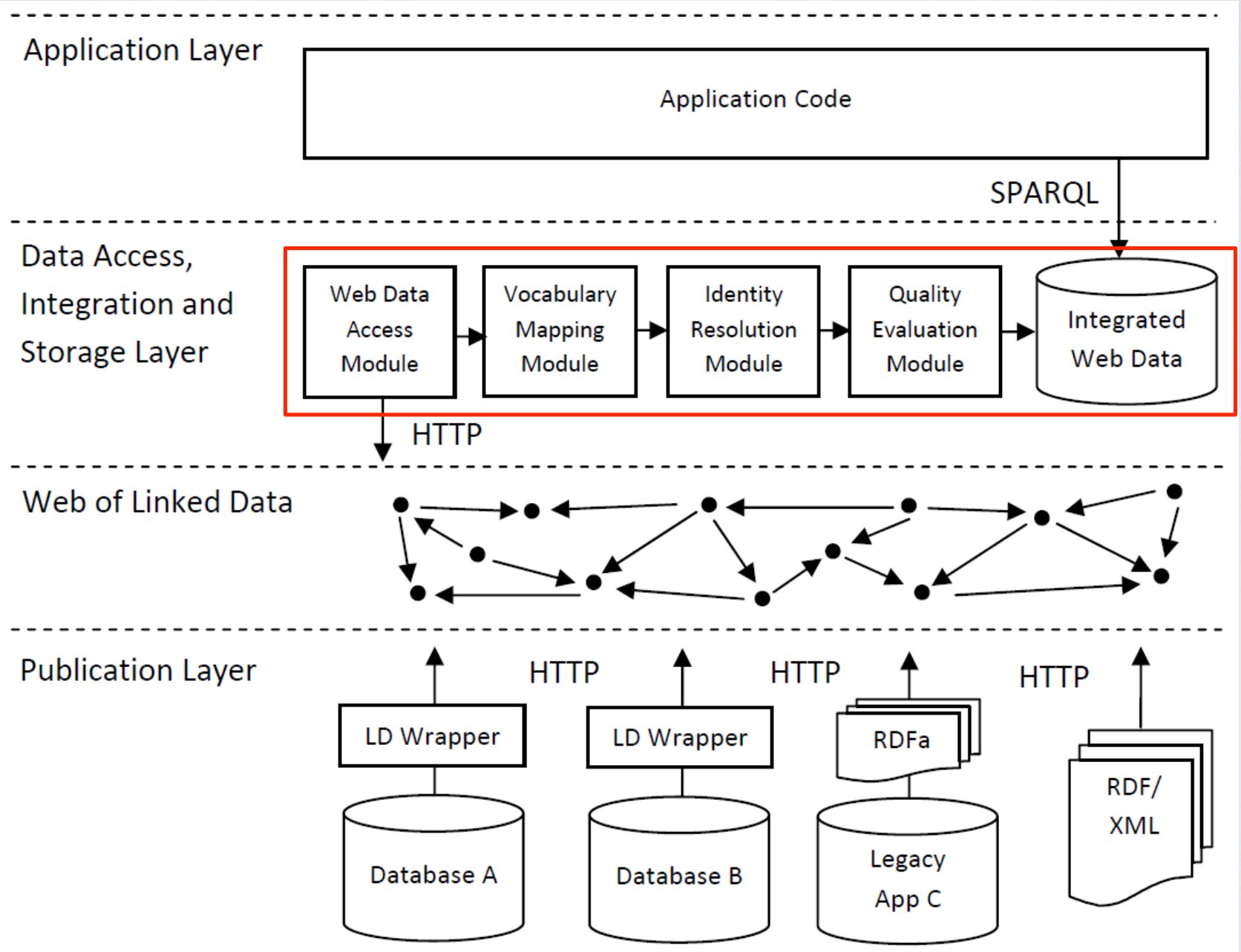
Tom Heath, Christian Bizer: Linked Data: Evolving the Web into a Global Data Space

http://linkeddatabook.com/

Linked Data Publishing Patterns



How to consume Linked Data ?



Conclusion

- The Web of Data is growing fast
- Linked Data provides a standardized data access interface
- Linked Data allows for the development of a variety of tools to integrate, enhance and and view the data
- Web search is evolving into query answering
  - Search engines will 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

Thanks

References

- Textbook: Tom Heath, Christian Bizer: Linked Data: Evolving the Web into a Global Data Space
  http://linkeddatabook.com/
- 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
- 4th Linked Data on the Web Workshop at WWW 2011
  http://events.linkeddata.org/ldow2011/
- 1st Workshop on Consuming Linked Data at ISWC 2010
  http://people.aifb.kit.edu/aha/2010/cold/