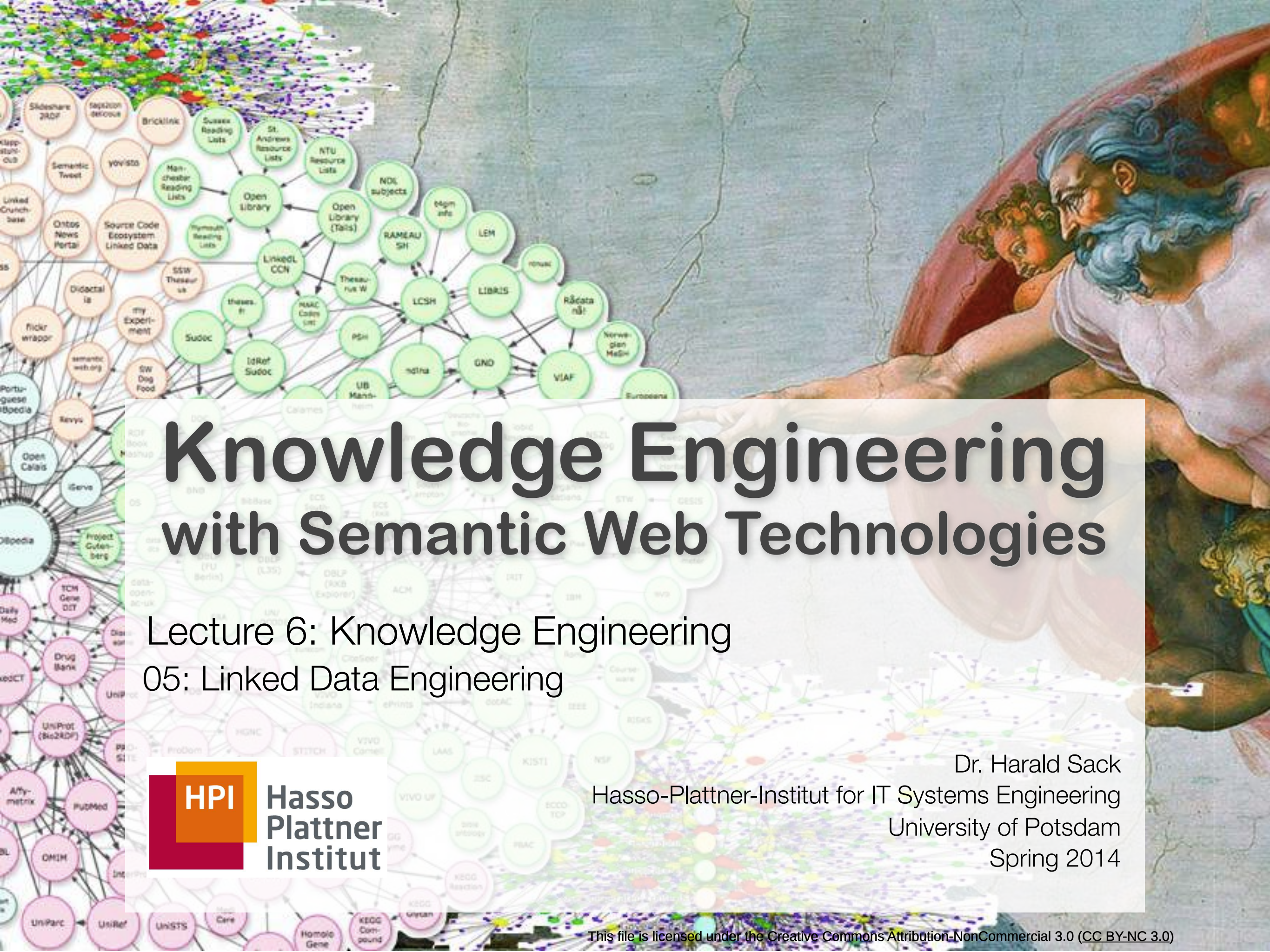




Knowledge Engineering with Semantic Web Technologies

Lecture 6: Knowledge Engineering 05: Linked Data Engineering

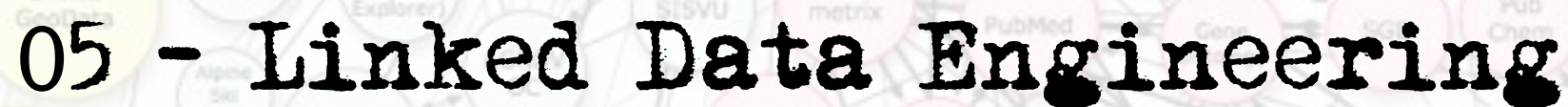


Dr. Harald Sack
Hasso-Plattner-Institut for IT Systems Engineering
University of Potsdam
Spring 2014



Lecture 6: Knowledge Engineering

Open HPI Course: Knowledge Engineering with Semantic Web Technologies

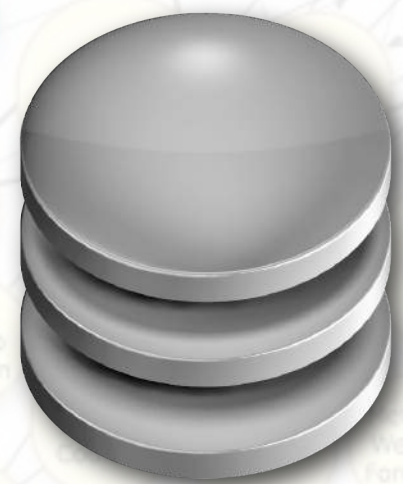


Open HPI - Course: Knowledge Engineering with Semantic Web Technologies

Lecture 6: Knowledge Engineering

How to get Data from the Web?

- Data can only be found on the Web, if it is available at some website

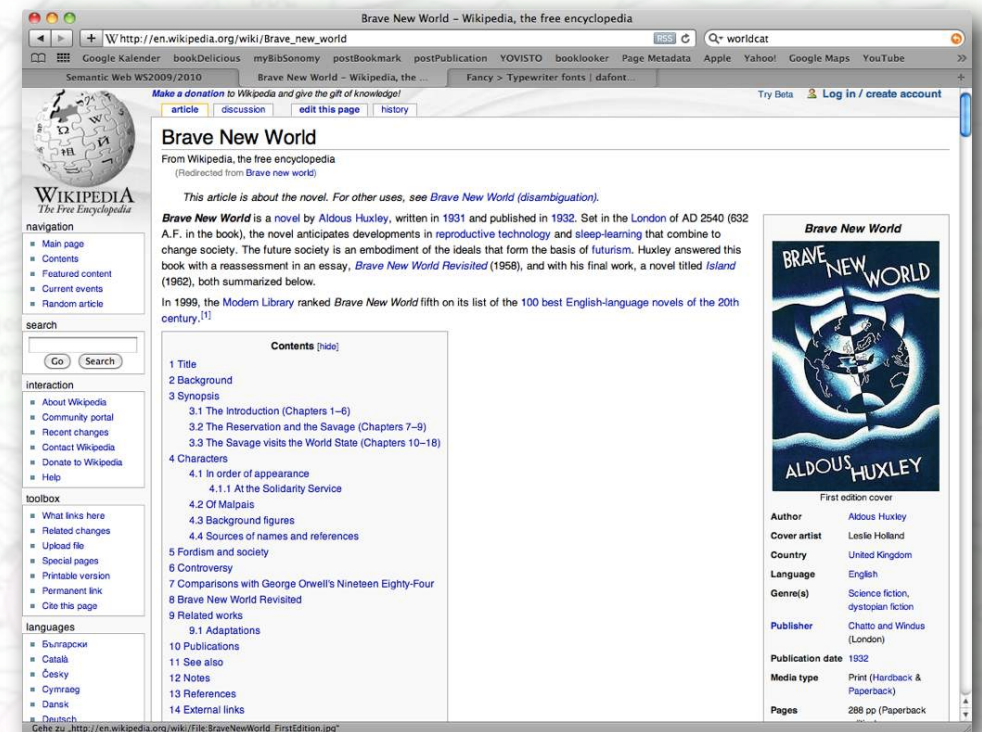


JDBC



Web-Server

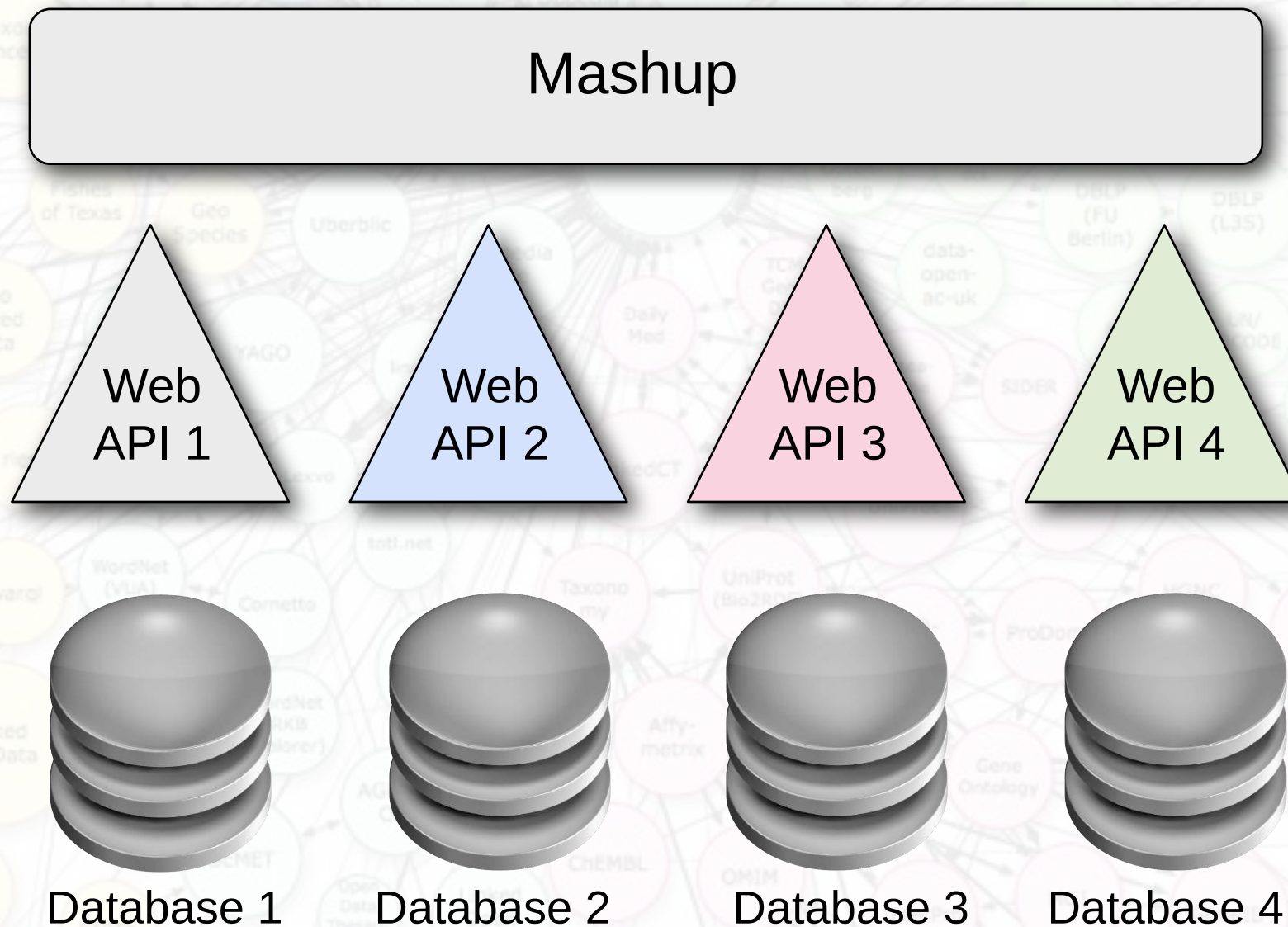
HTTP
HTML



Browser

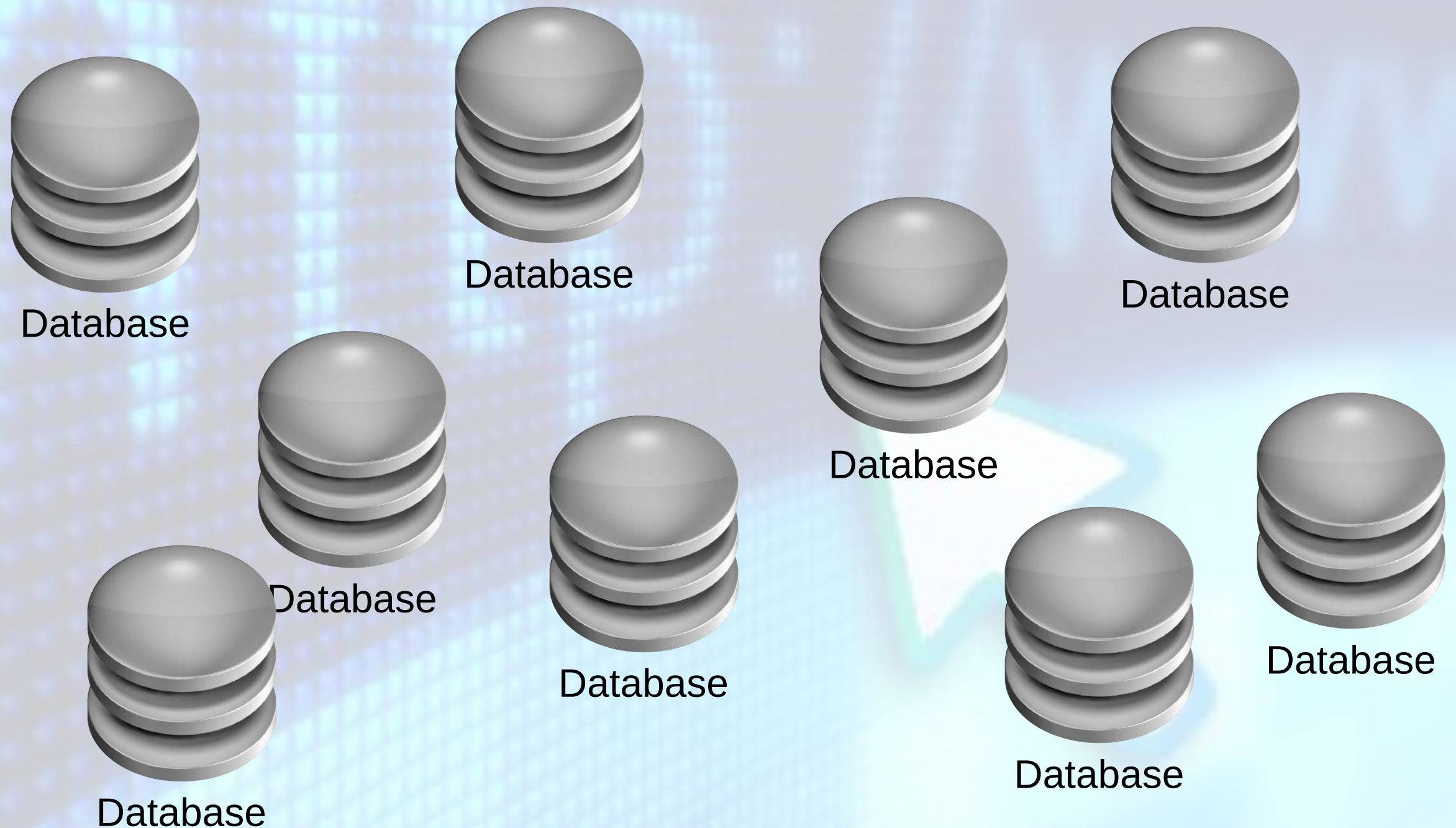
How to get Data from the Web?

- There is a number of different (proprietary) **Web APIs**, data exchange formats and **Mashups** on top of that



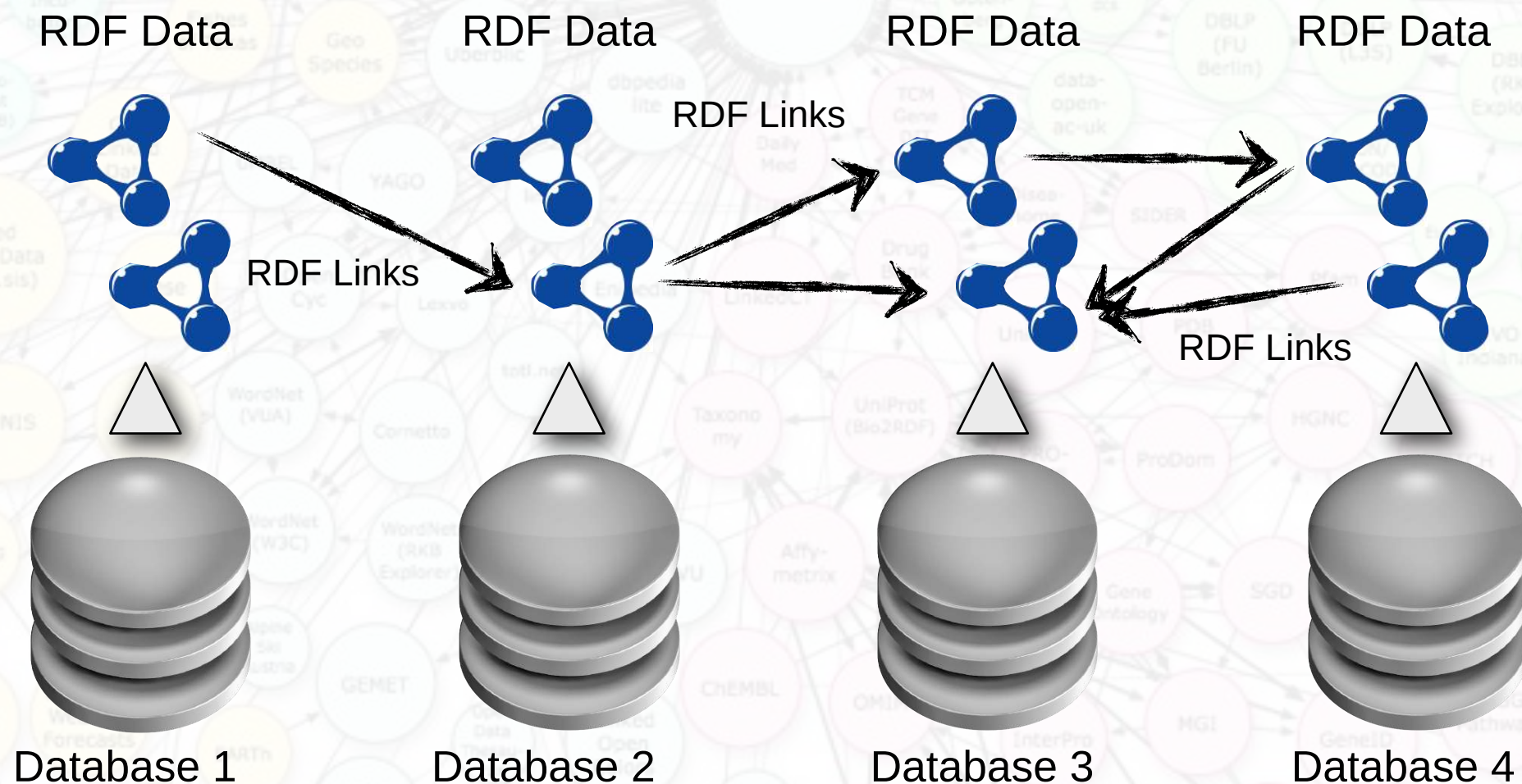
In the Web today...

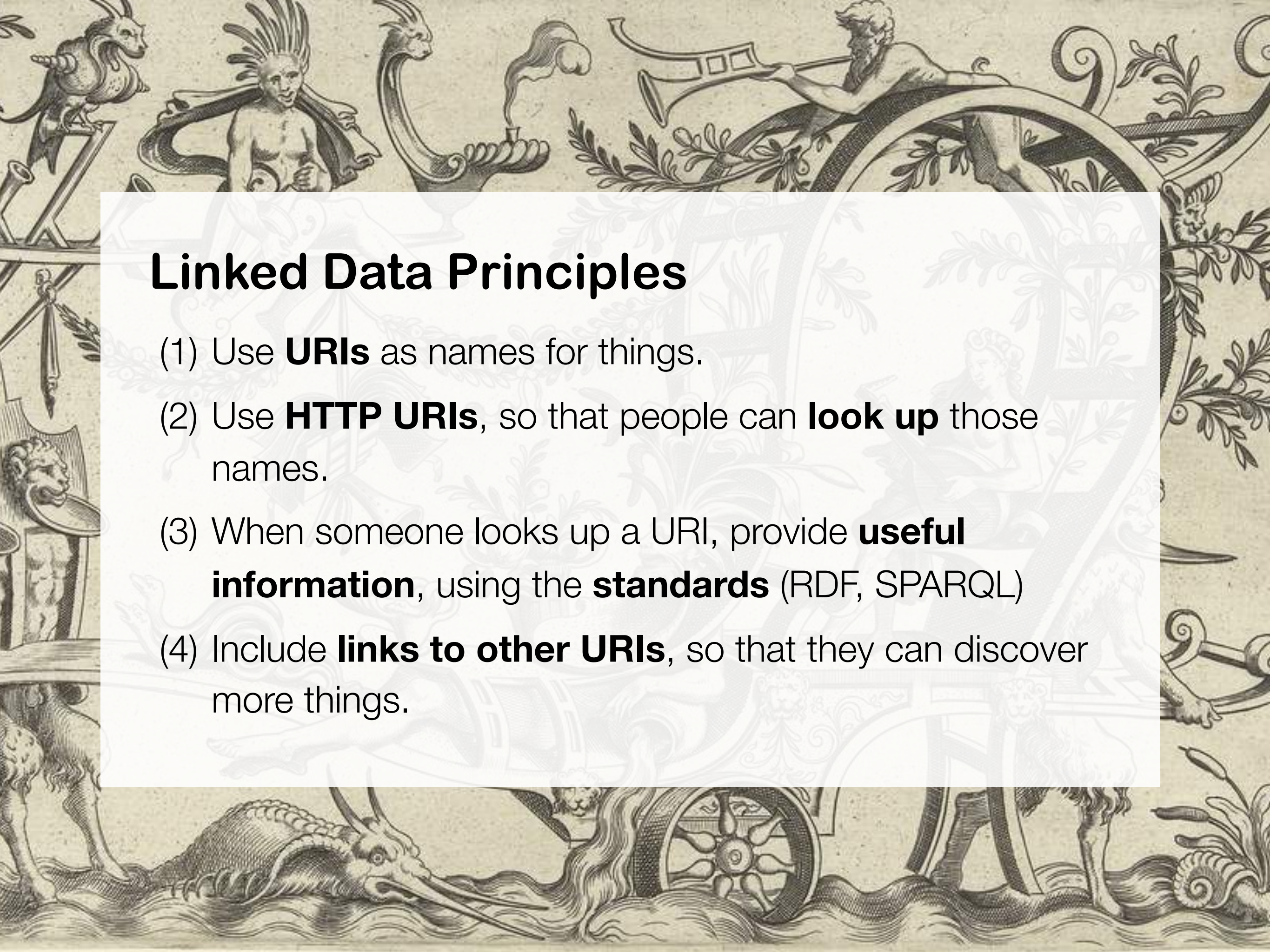
- Data is locked up in small data islands
- Other applications usually cannot access this data...



How to get of Closed Data Islands?

- **Apply semantic technologies:**
 - to publish structured data on the web
 - to draw connections from one data source to data from other data sources

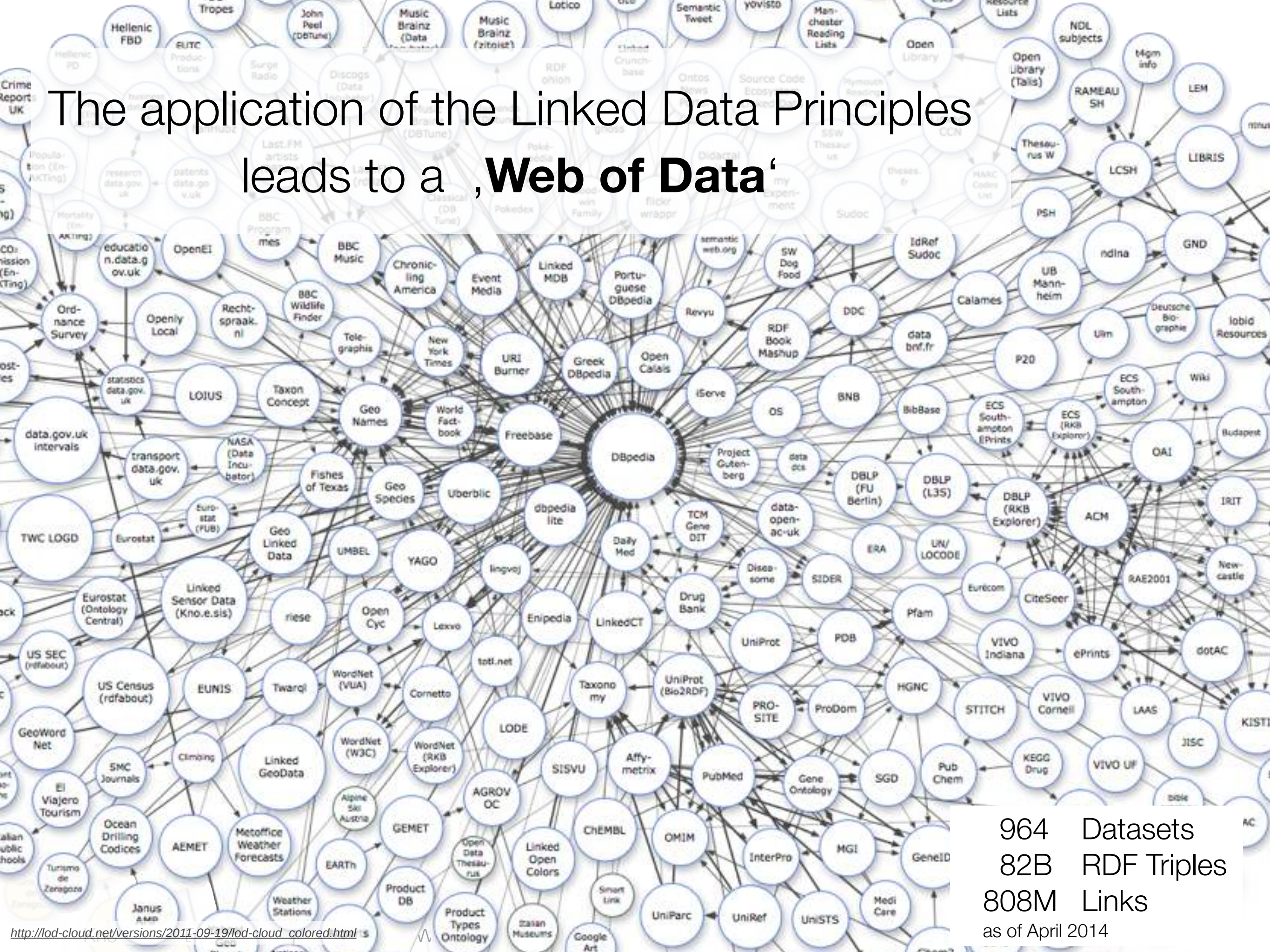




Linked Data Principles

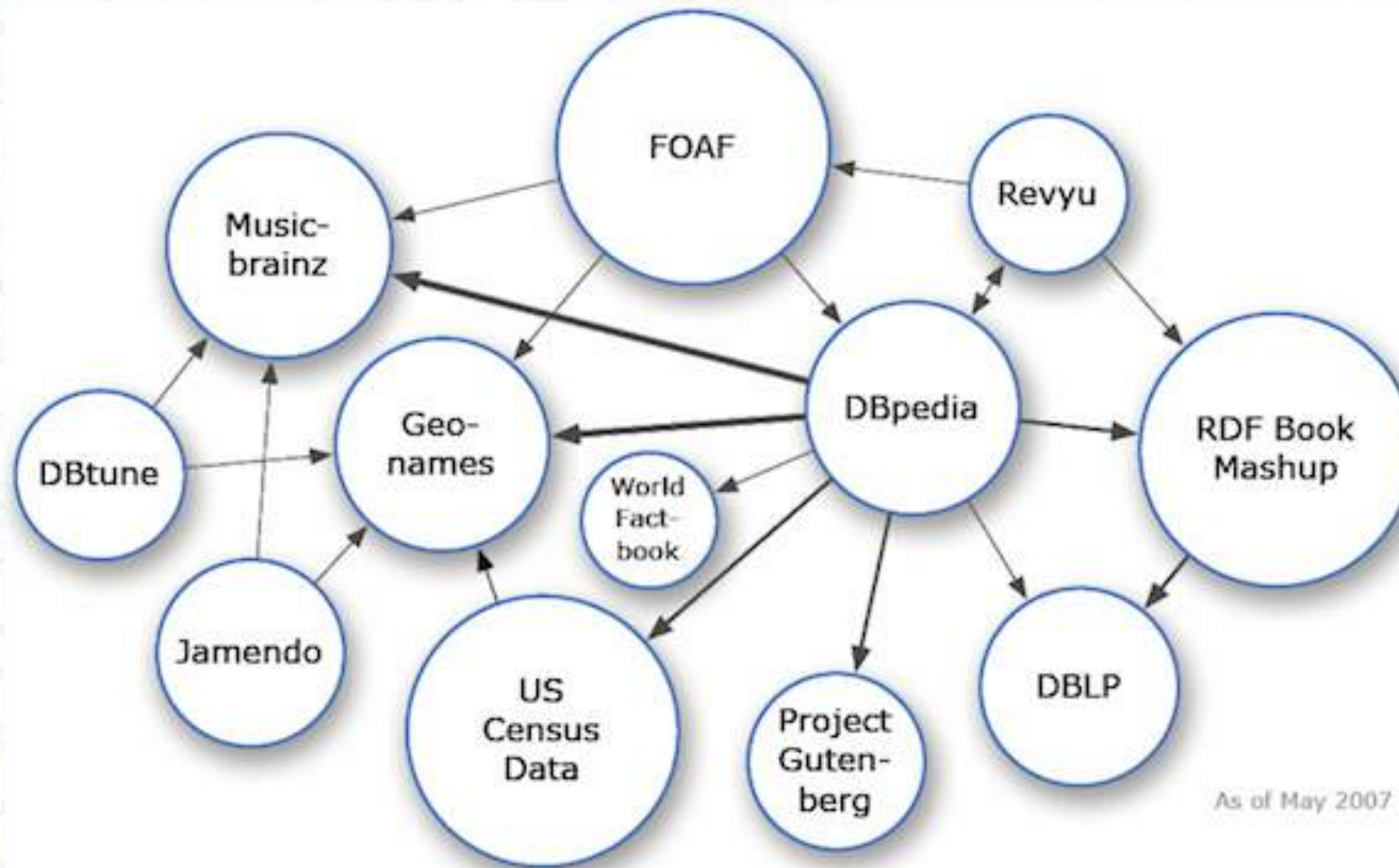
- (1) Use **URIs** as names for things.
- (2) Use **HTTP URIs**, so that people can **look up** those names.
- (3) When someone looks up a URI, provide **useful information**, using the **standards** (RDF, SPARQL)
- (4) Include **links to other URIs**, so that they can discover more things.

The application of the Linked Data Principles leads to a **Web of Data**



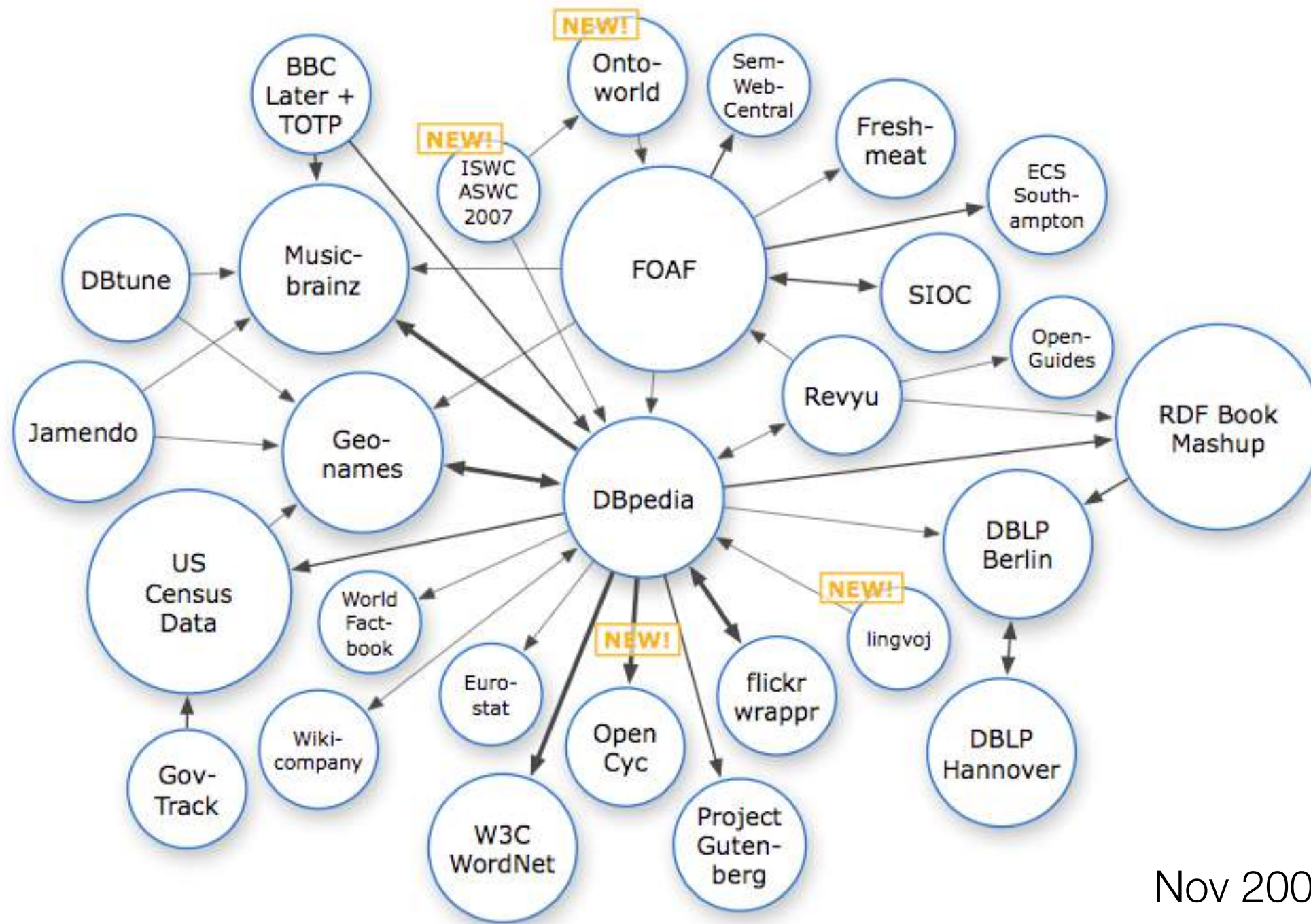
964 Datasets
82B RDF Triples
808M Links
as of April 2014

Development of the Web of Data



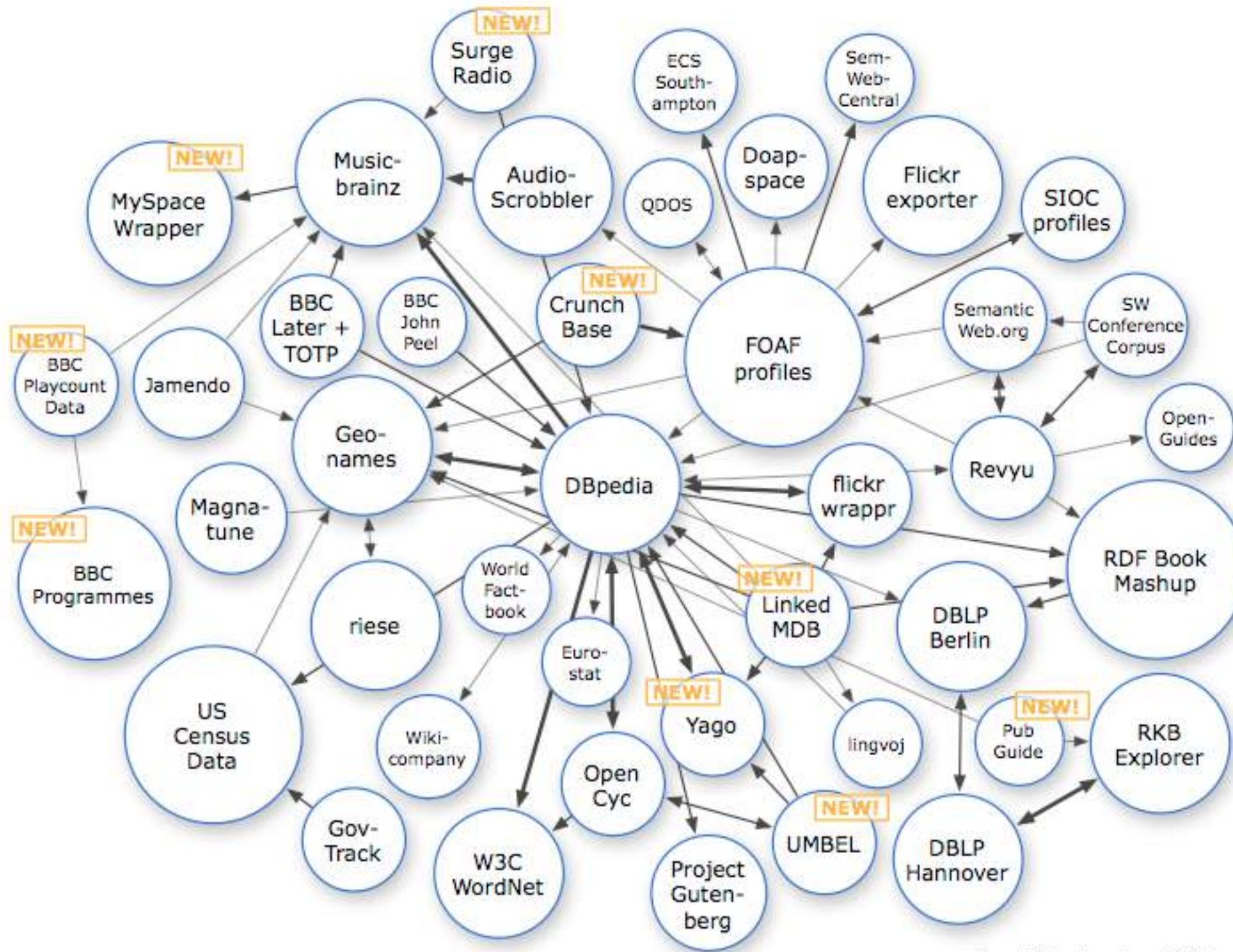
May 2007

Development of the Web of Data



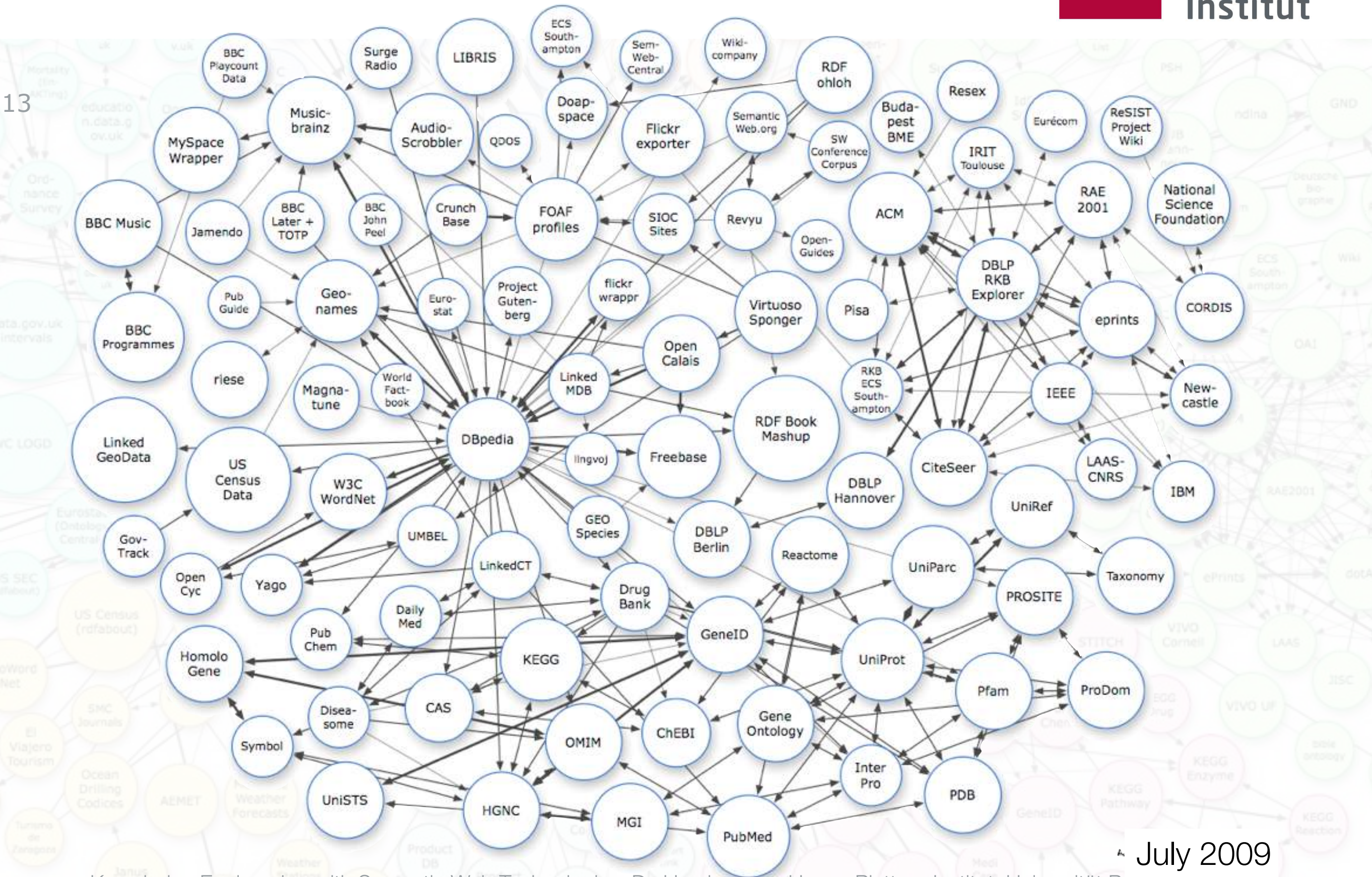
Nov 2007

Development of the Web of Data

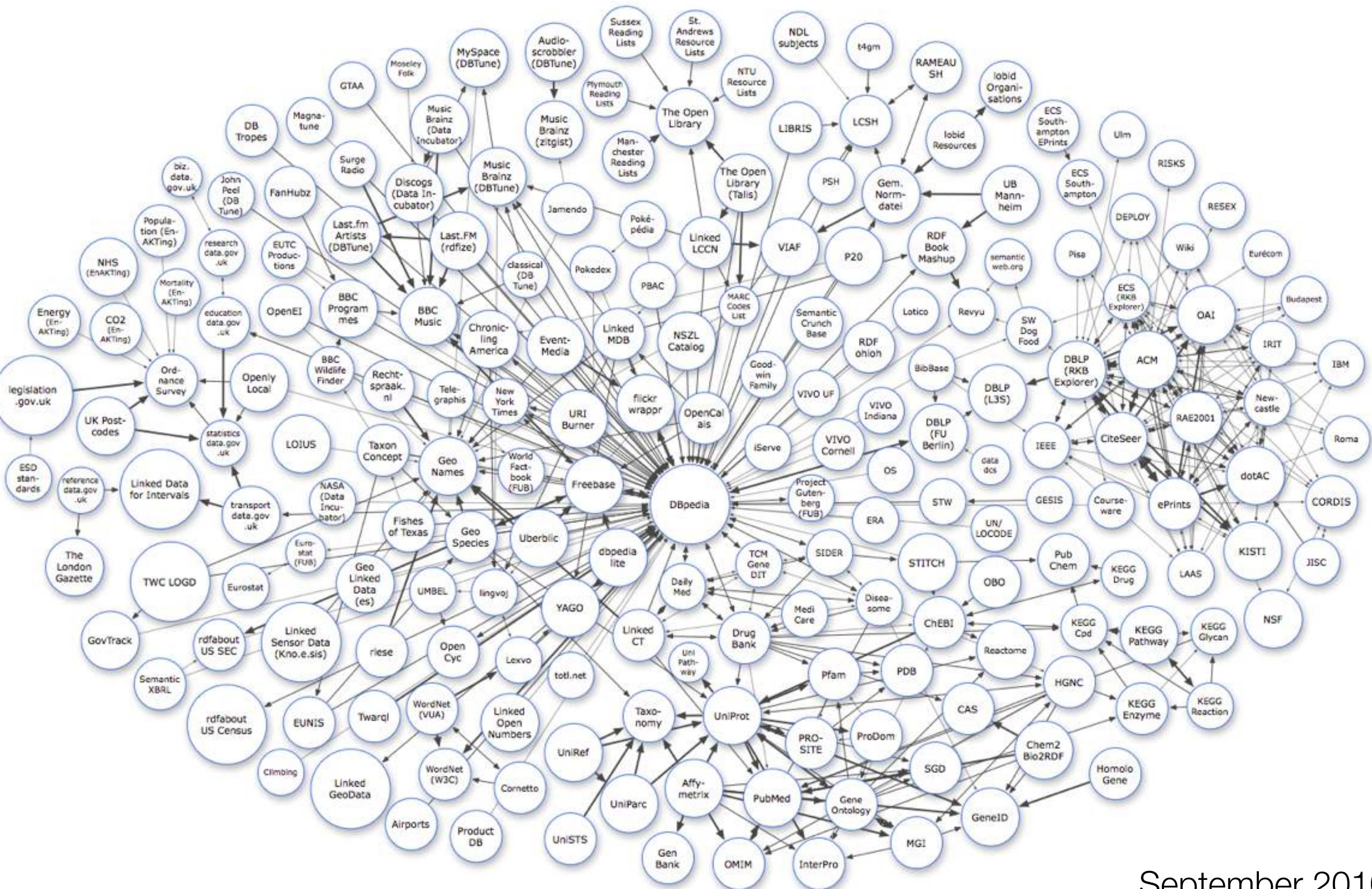


As of September 2008

Development of the Web of Data

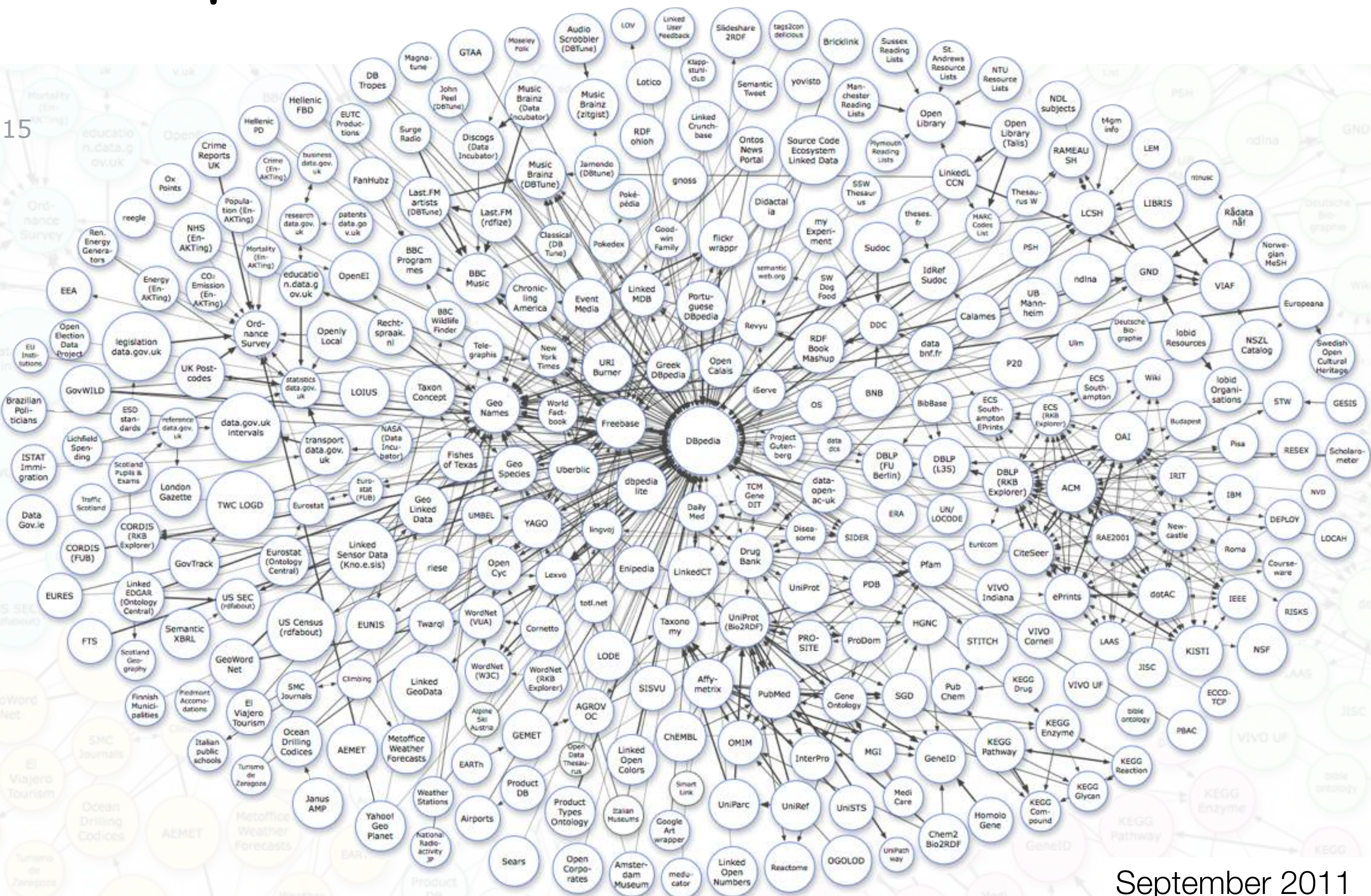


Development of the Web of Data



September 2010

Development of the Web of Data



September 2011

Linked Open Data

- Public Linked Data resources in the Web, licensed as „Creative Common CC-BY“
- Tim Berners-Lee's 5-Star Criteria for Linked Open Data

★ Available on the web (whatever format) but with an **open licence**, to be Open Data

★ ★ Available as **machine-readable structured data** (e.g. excel instead of image scan of a table)

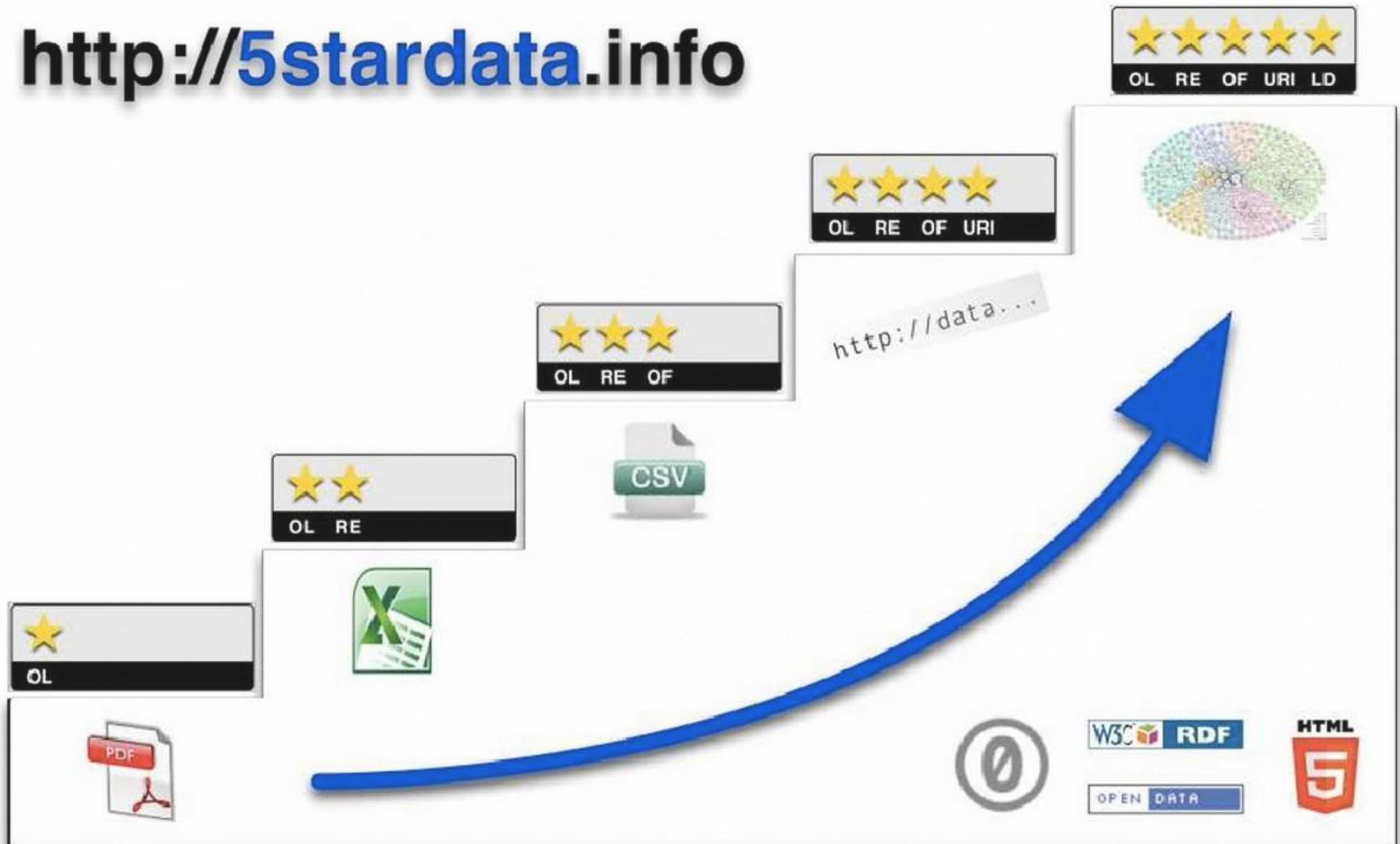
★ ★ ★ as (2) plus **non-proprietary format** (e.g. CSV instead of excel)

★ ★ ★ ★ All the above plus: use **open standards from W3C** (RDF and SPARQL) to identify things, so that people can point at your stuff

★ ★ ★ ★ ★ All the above, plus: **link your data to other people's data** to provide context

Linked Open Data

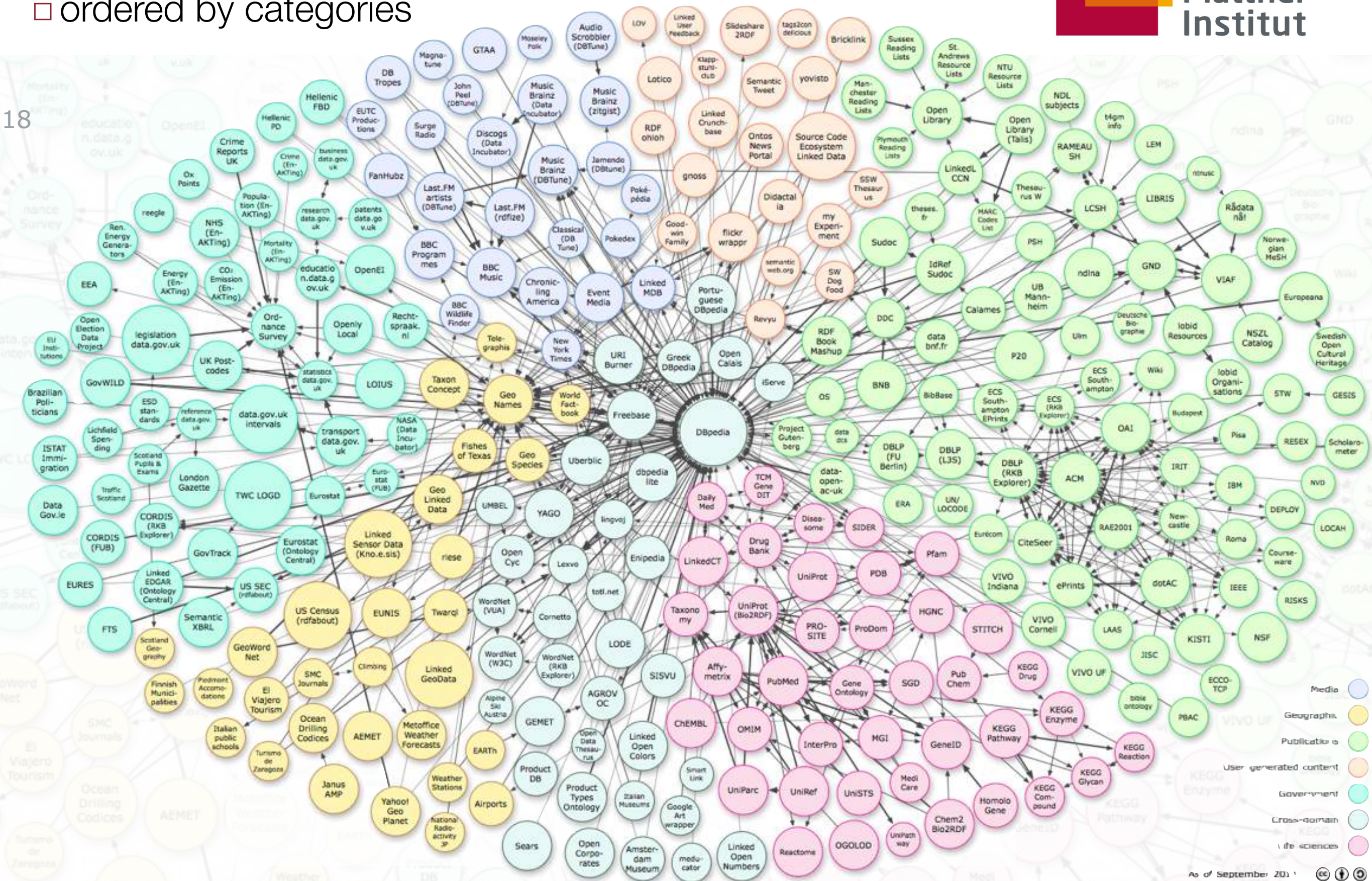
17 <http://5stardata.info>



Linked Data - A Guided Tour...

□ ordered by categories

18

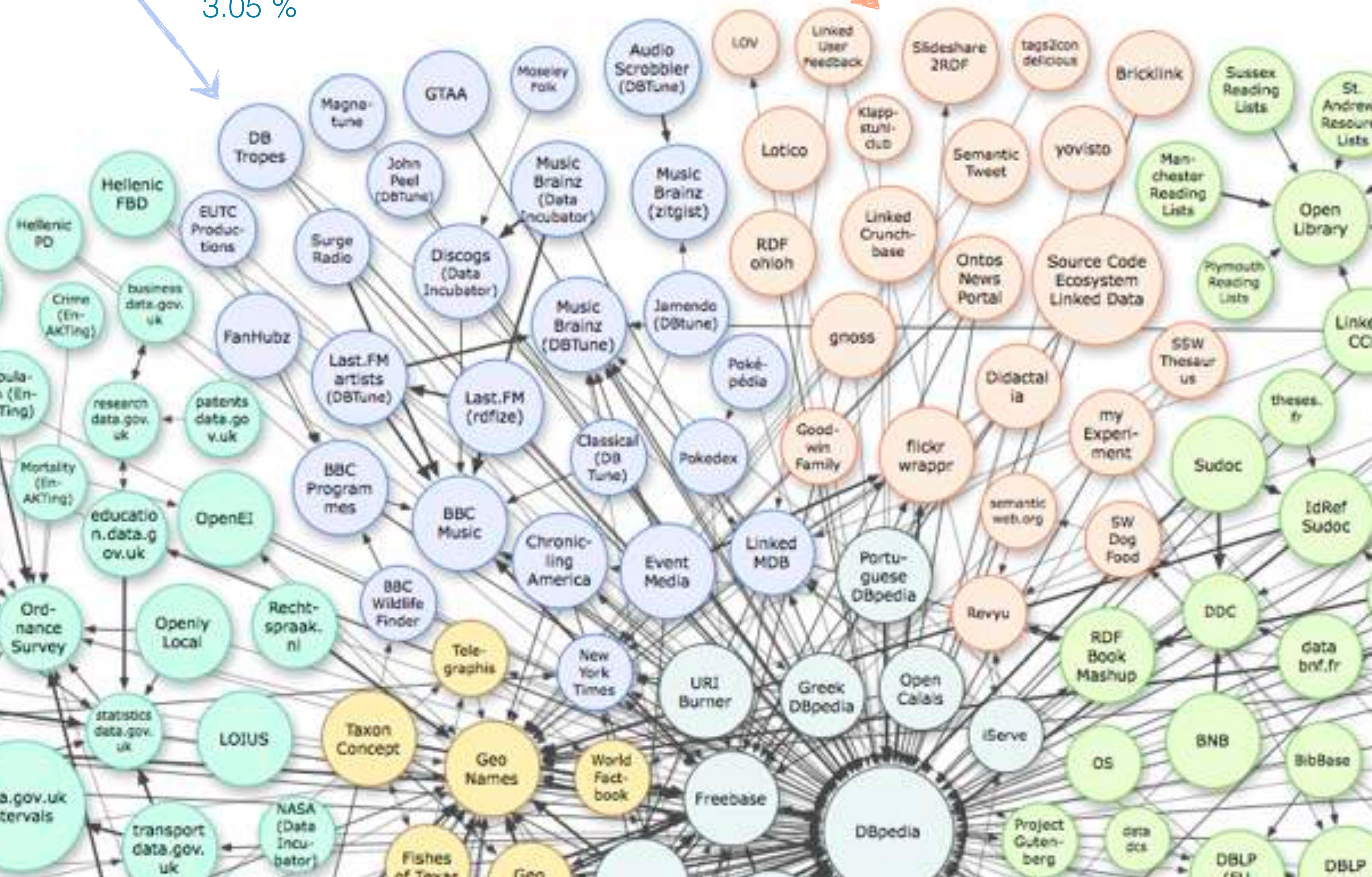


Media

41 data sets
2.568.964.515 RDF Triples
3.05 %

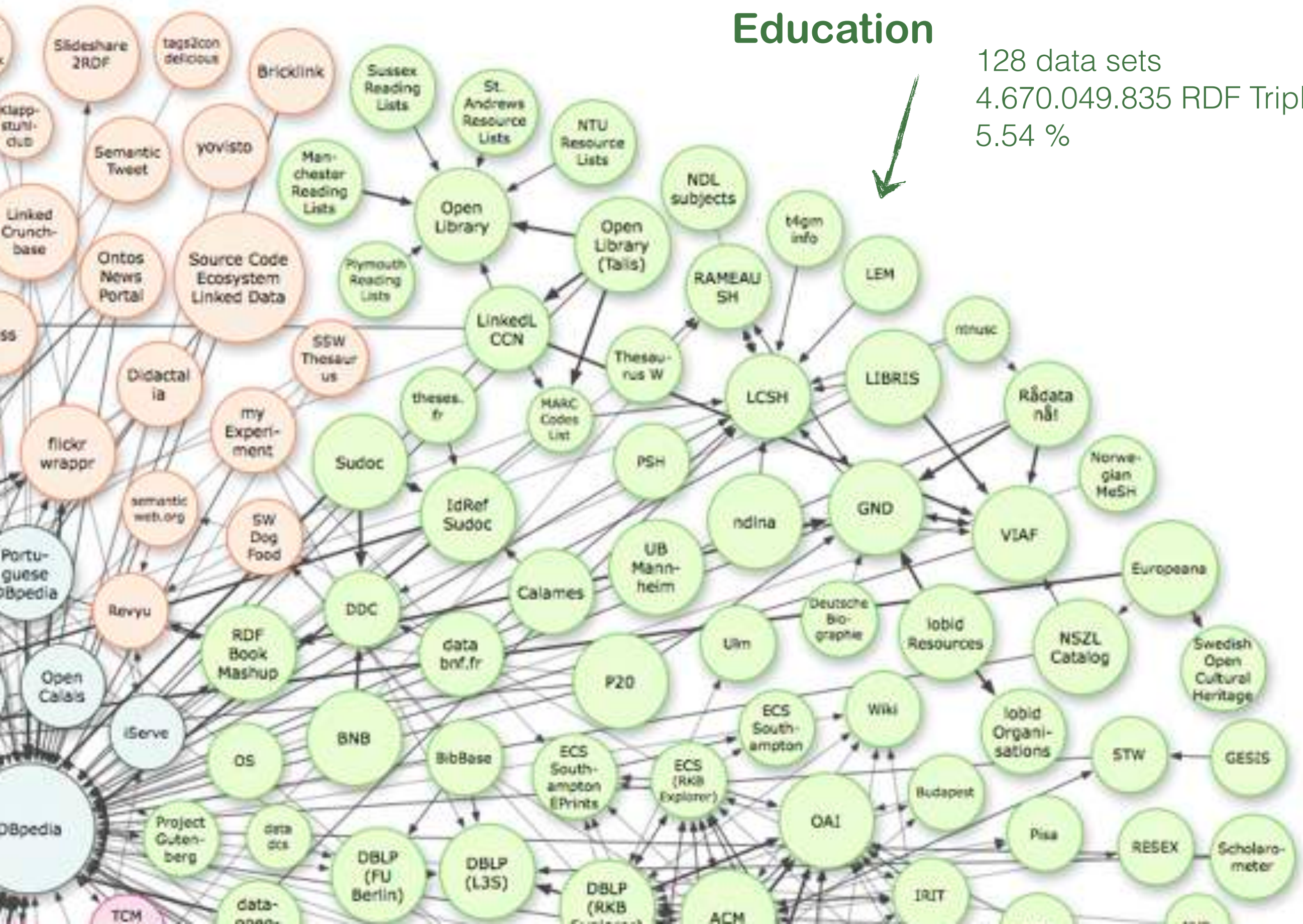
User Generated Content

49 data sets
2.283.111.223 RDF Triples
2.71 %



Libraries and Education

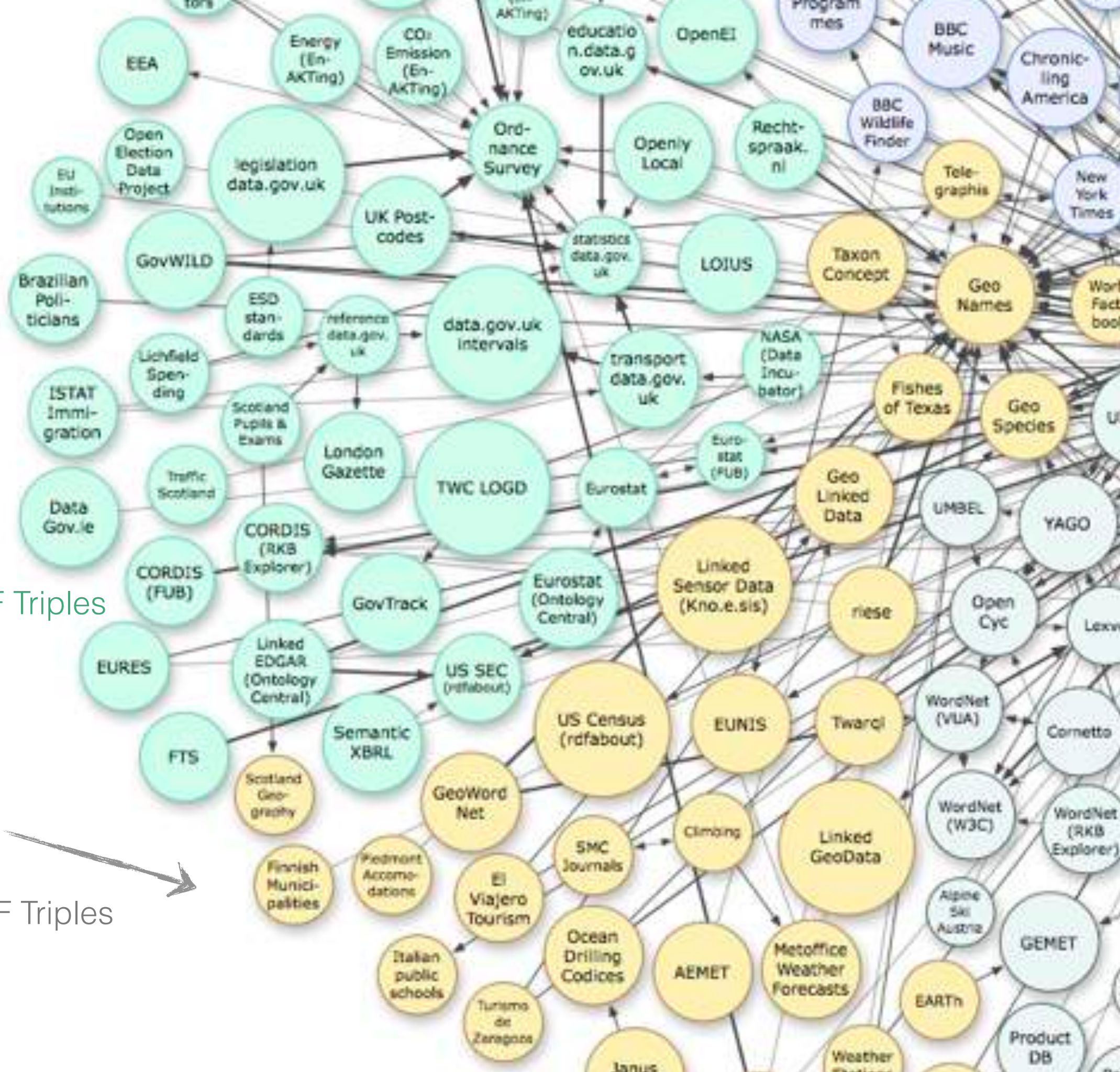
128 data sets
4.670.049.835 RDF Triples
5.54 %



Government



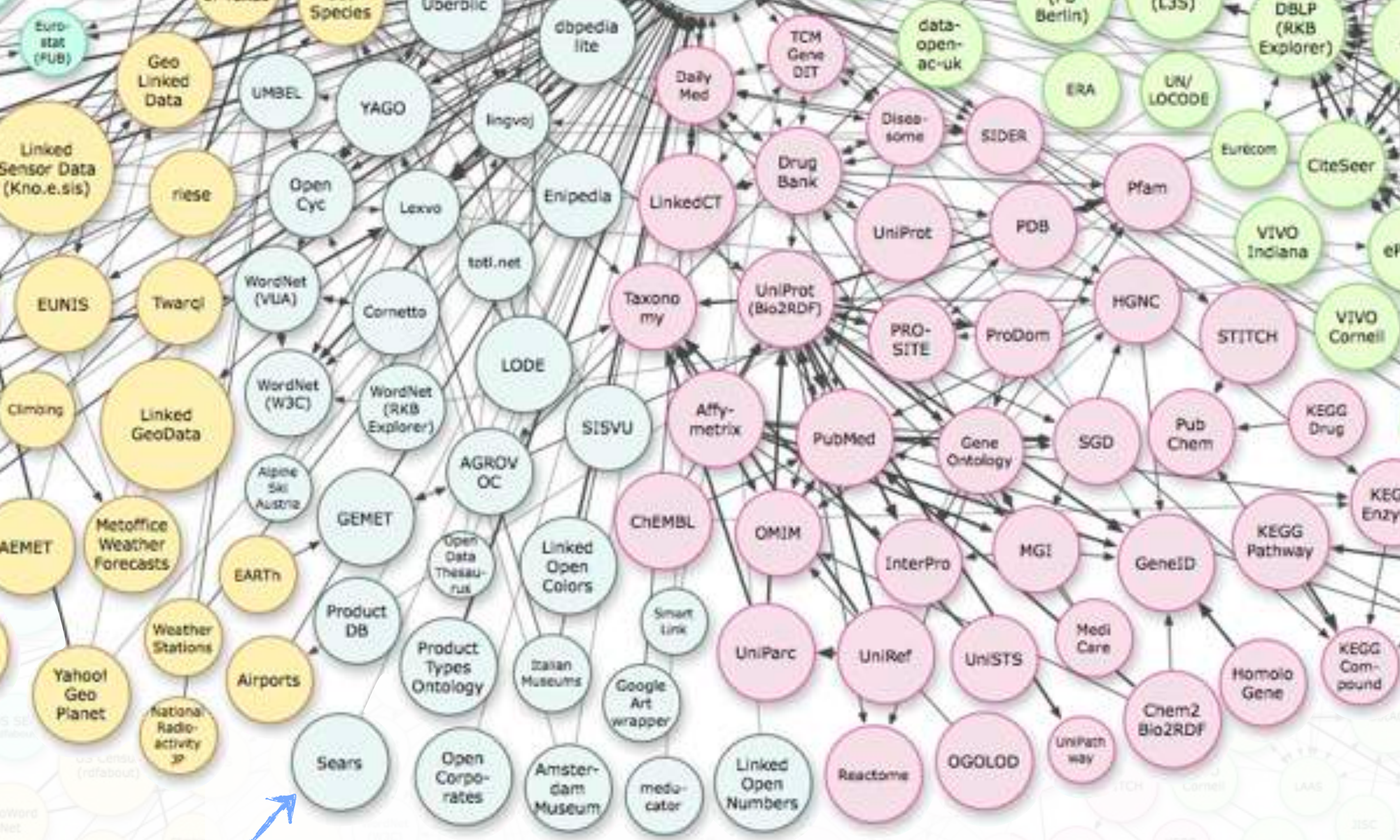
139 data sets
38.246.327.678 RDF Triples
45,5 %



Geographic



46 data sets
4.830.360.509 RDF Triples
5,73 %



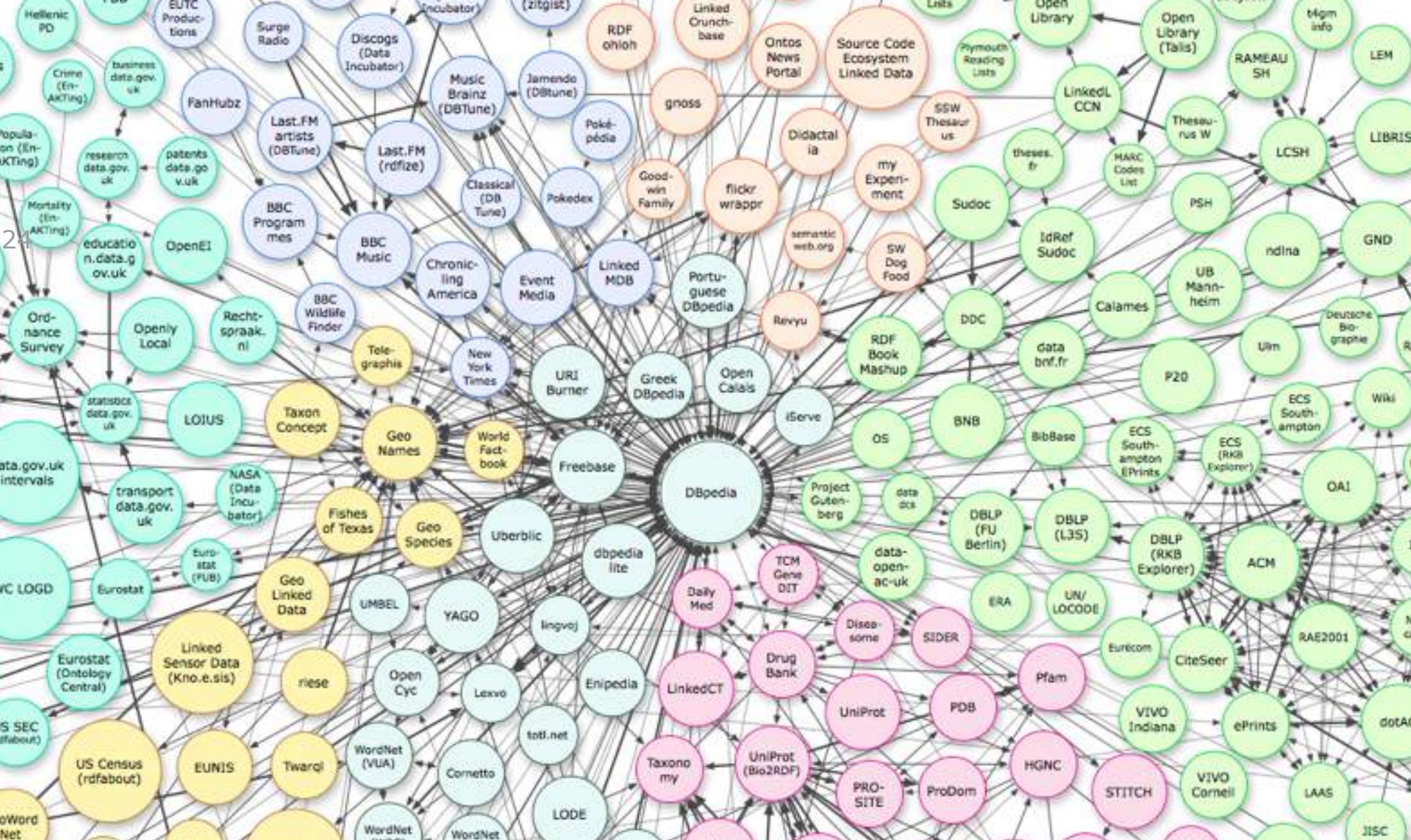
Cross-Domain

84 data sets
9.161.202.967 RDF Triples
10,87 %

Life Sciences

92 data sets
7.370.024.776 RDF Triples
8,75 %

- **Semantic Mashups** are applications that use linked RDF data from various data sources.
- in difference to *interfaces* and *exchange formats* or *ordinary Web APIs*, **Linked Data** offers the following benefits:
 - a flexible and standardized data format (**RDF**)
 - standardized access mechanism (**http**)
 - possibility to put links (**RDF-Links**) among different data sources
 - » enables navigation
 - » is supported by search engines (crawler)
 - » enables expressive search facilities over the crawled data and beyond



06 - Linked Data Engineering - Pt.2

Open HPI - Course: Knowledge Engineering with Semantic Web Technologies
Lecture 6: Knowledge Engineering