Linked Data - Dados Abertos Conectados

Henrique Muniz

Linked Data

- Is a method of publishing structured data so that it can be interlinked and become more useful through semantic queries
- Tim Berners-Lee, director of the World Wide Web Consortium (W3C), coined the term in a 2006
- Semantic Web is a Web of Data

Linked Data

- HTTP
- RDF
- URI

Requirements

- Huge amount of data on the Web available in a standard format, reachable and manageable
- Relationships among data should be made available

Principles

1. Use URIs to name (identify) things.
2. Use HTTP URIs so that these things can be looked up
3. Provide useful information about what a name identifies when it's looked up
4. Refer to other things using their HTTP URI-based names when publishing data on the Web.

Example

Subject - http://data.linkedmdb.org/resource/film/77

Predicate - http://www.w3.org/2002/07/owl#sameAs

Object - http://dbpedia.org/resource/Pulp_Fiction_%28film%29

Web of Data

• The Web of Data is generic and can contain any type of data.

• Anyone can publish data to the Web of Data.

• Data publishers are not constrained in choice of vocabularies with which to

represent data.

• Entities are connected by RDF links, creating a global data graph that spans
data sources and enables the discovery of new data sources.

Characteristics

- Data is strictly separated from formatting and presentational aspects.
- Data is self-describing. If an application consuming Linked Data encounters data described with an unfamiliar vocabulary, the application can dereference the URIs that identify vocabulary terms in order to find their definition.
- The use of HTTP as a standardized data access mechanism and RDF as a standardized data model simplifies data access compared to Web APIs, which rely on heterogeneous data models and access interfaces.
- The Web of Data is open, meaning that applications do not have to be implemented against a fixed set of data sources, but can discover new data sources at run-timeby following RDF links.

Publishing Linked Data on the Web

- Assign URIs to the entities described by the data set and provide for dereferencing these URIs over the HTTP protocol into RDF representations.
- Set RDF links to other data sources on the Web, so that clients can navigate the Web of Data as a whole by following RDF links.
- Provide metadata about published data, so that clients can assess the quality of published data and choose between different means of access.

Linked open data

Linked Open Data (LOD) is Linked Data which is released under an open licence, which does not impede its reuse for free. - Tim Berners-Lee

The term "linked open data" has been in use since at least February 2007, when the "Linking Open Data" mailing list was created.

The mailing list was initially hosted by the SIMILE project at the Massachusetts Institute of Technology.

SIMILE (Semantic Interoperability of Metadata and Information in unLike Environments), funded by the Andrew W. Mellon Foundation

Linked open data

The goal of the W3C Semantic Web Education is to extend the Web with a data commons by publishing various open datasets as RDF on the Web and by setting RDF links between data items from different data sources

In October 2007, datasets consisted of over two billion RDF triples, which were interlinked by over two million RDF links.

By September 2011 this had grown to 31 billion RDF triples, interlinked by around 504 million RDF links.

Datasets

DBpedia: a dataset containing extracted data from Wikipedia, contains about 3.4 million concepts described by 1 billion triples, including abstracts in 11 different languages

FOAF: a dataset describing persons, their properties and relationships

GeoNames: provides RDF descriptions of more than 7,500,000 geographical features worldwide.
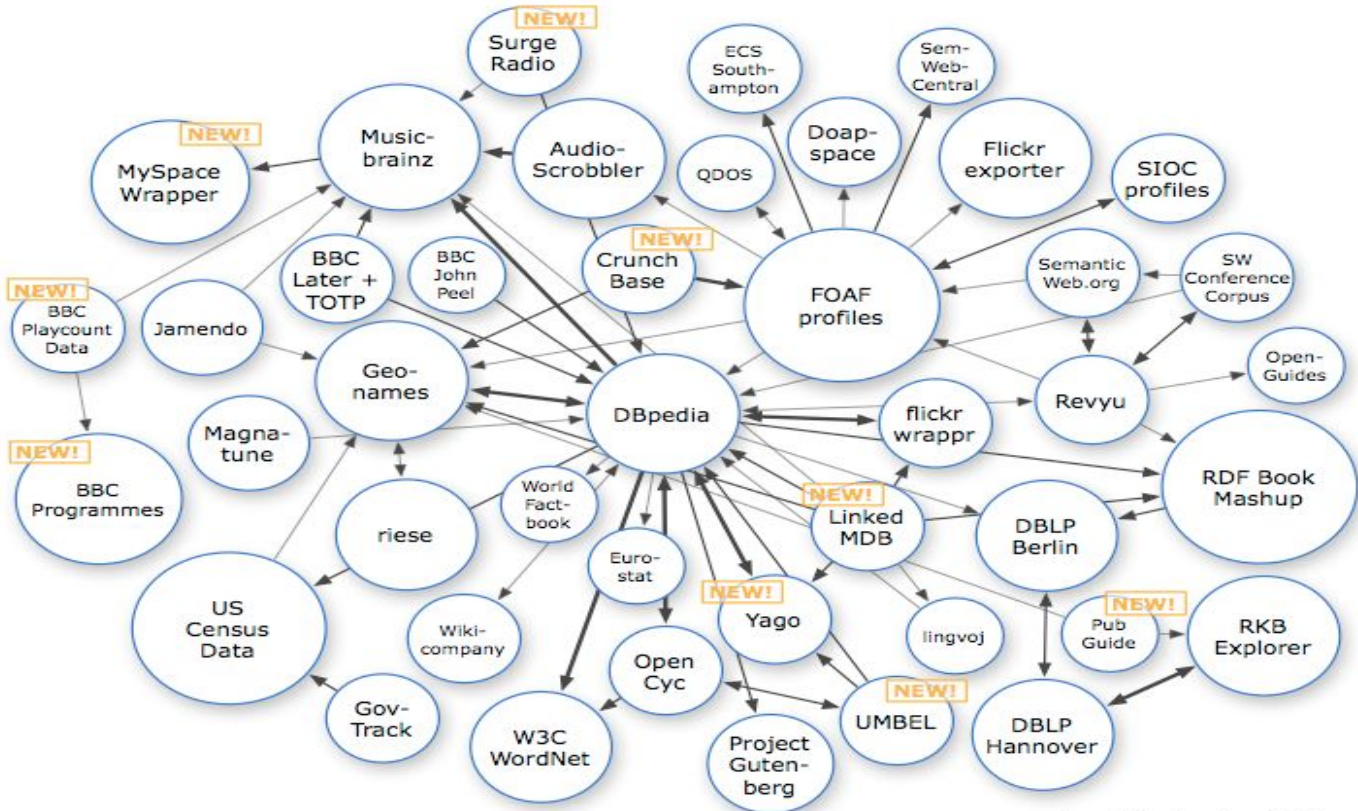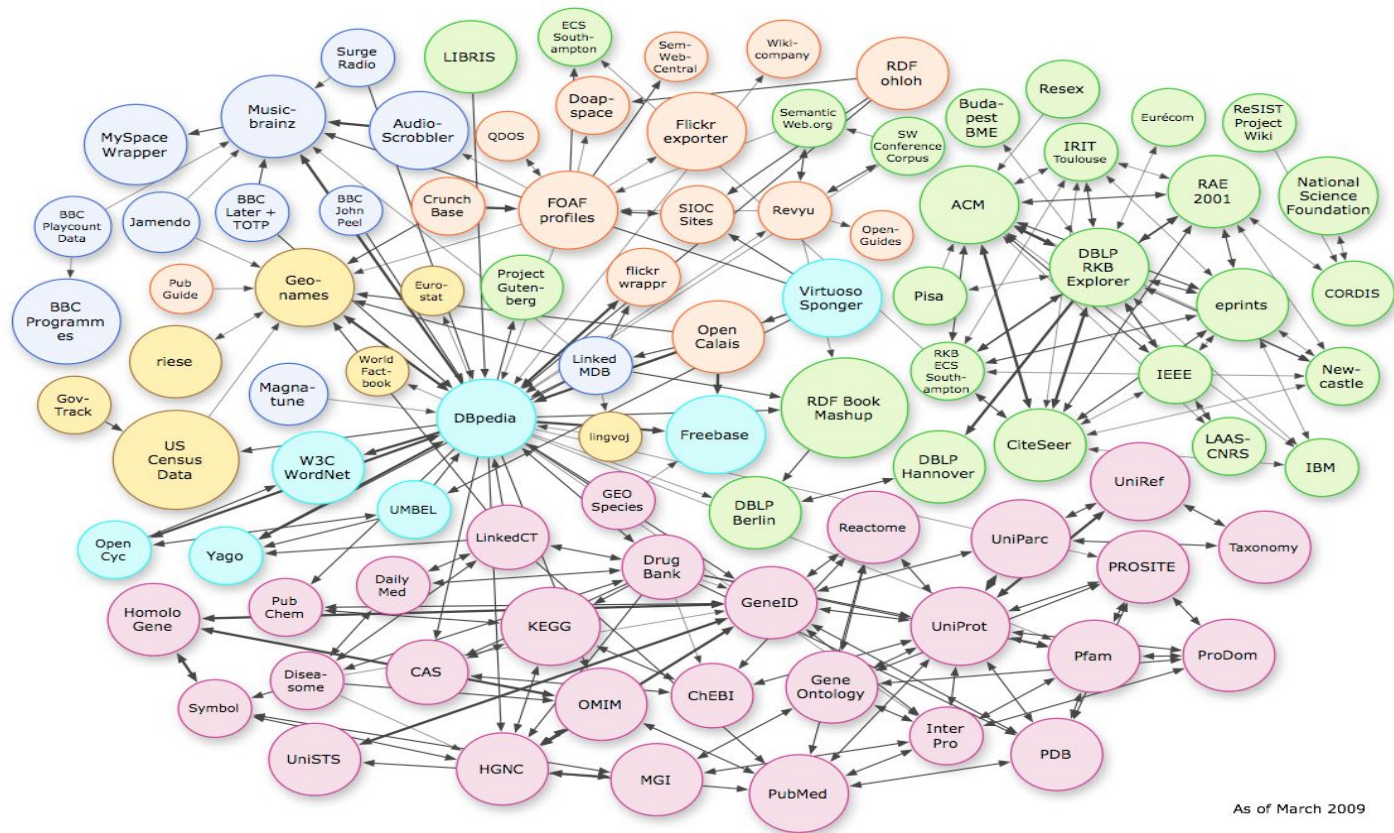
Growth of LOD-Cloud

2007-11-10
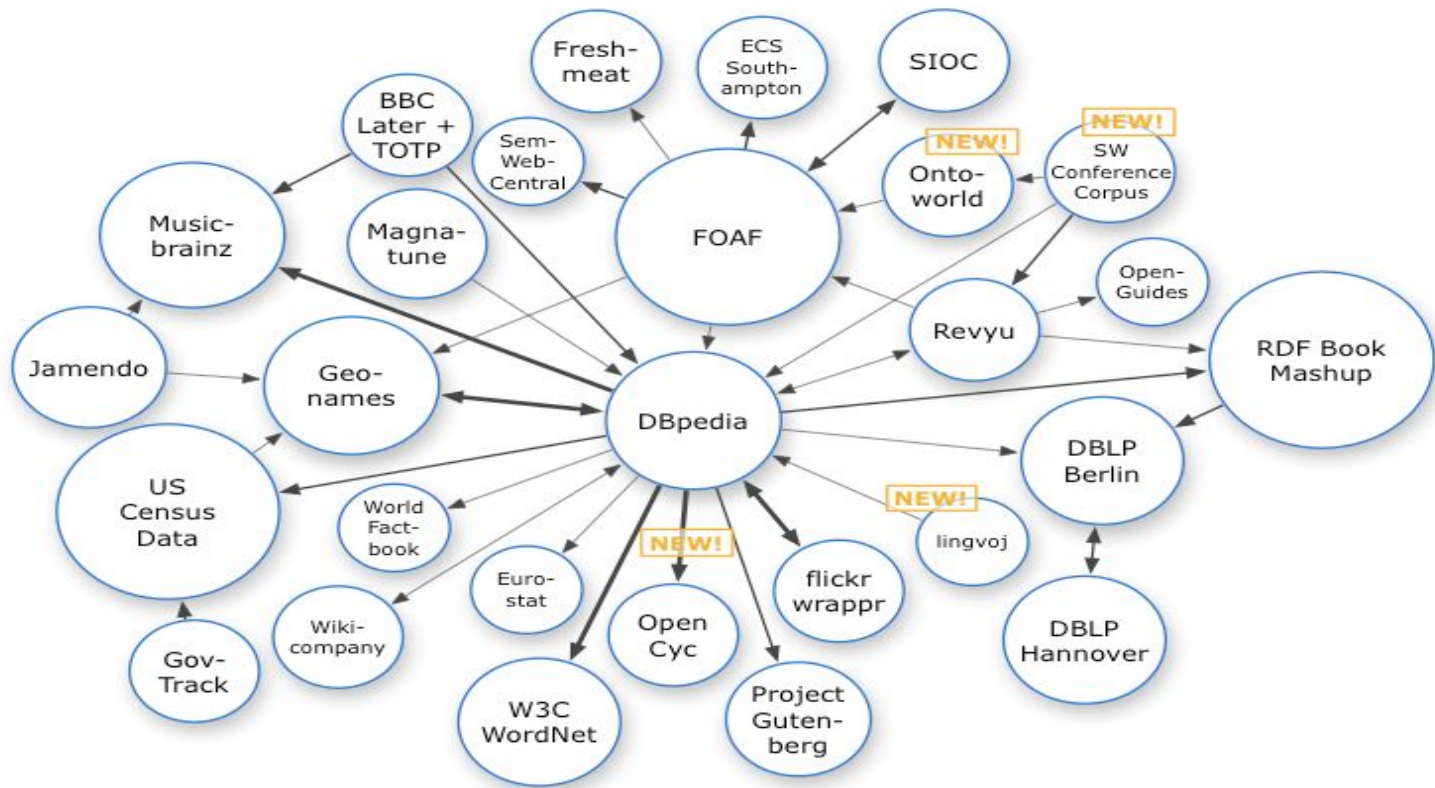
2007-11-10

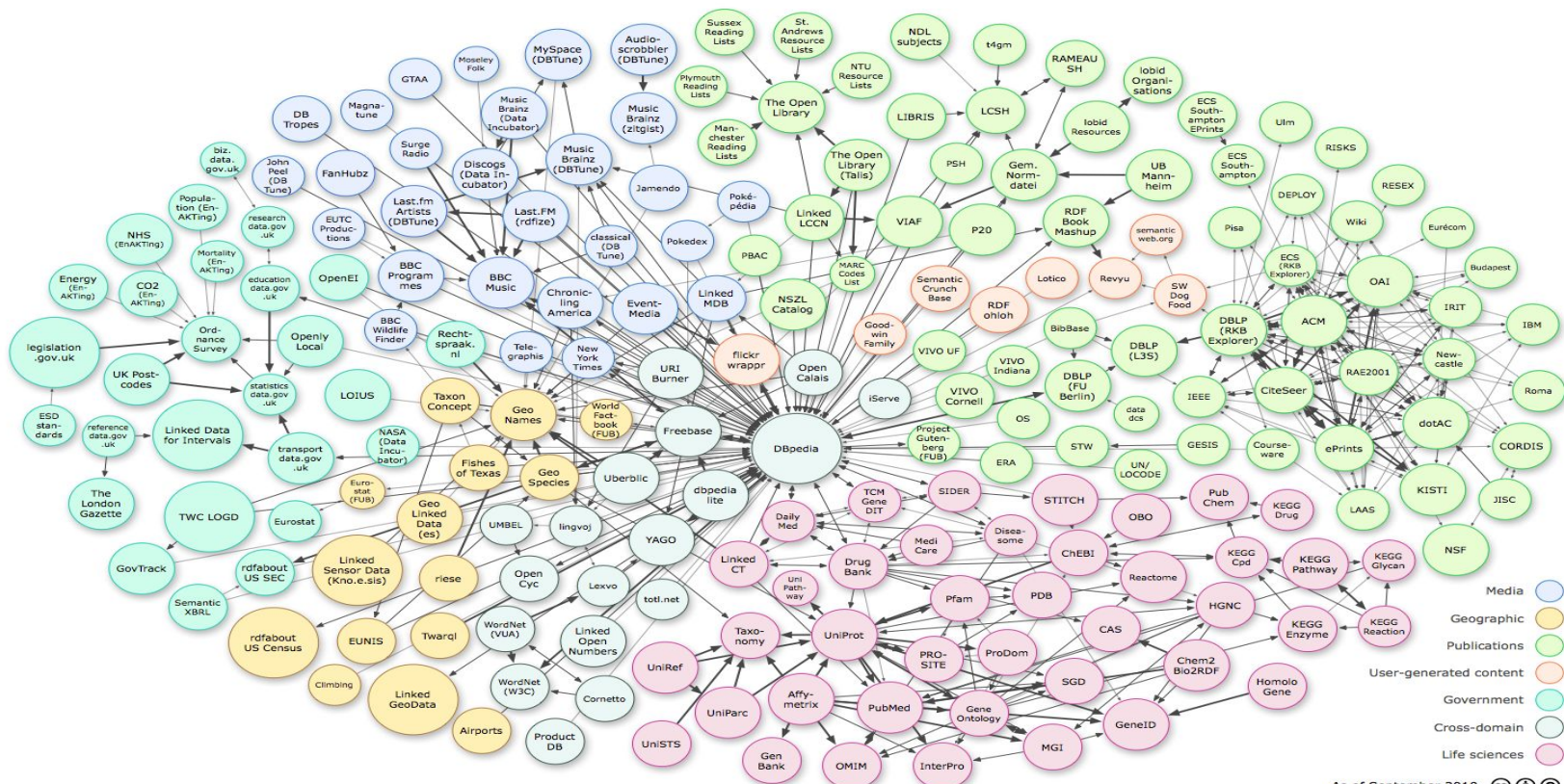- Two billion RDF triples
- Two million RDF links

2008-09-18

As of September 2008

2009-03-27

As of March 2009

2010-09-22

As of September 2010

Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences

2011-09-19

As of September 2011

Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences
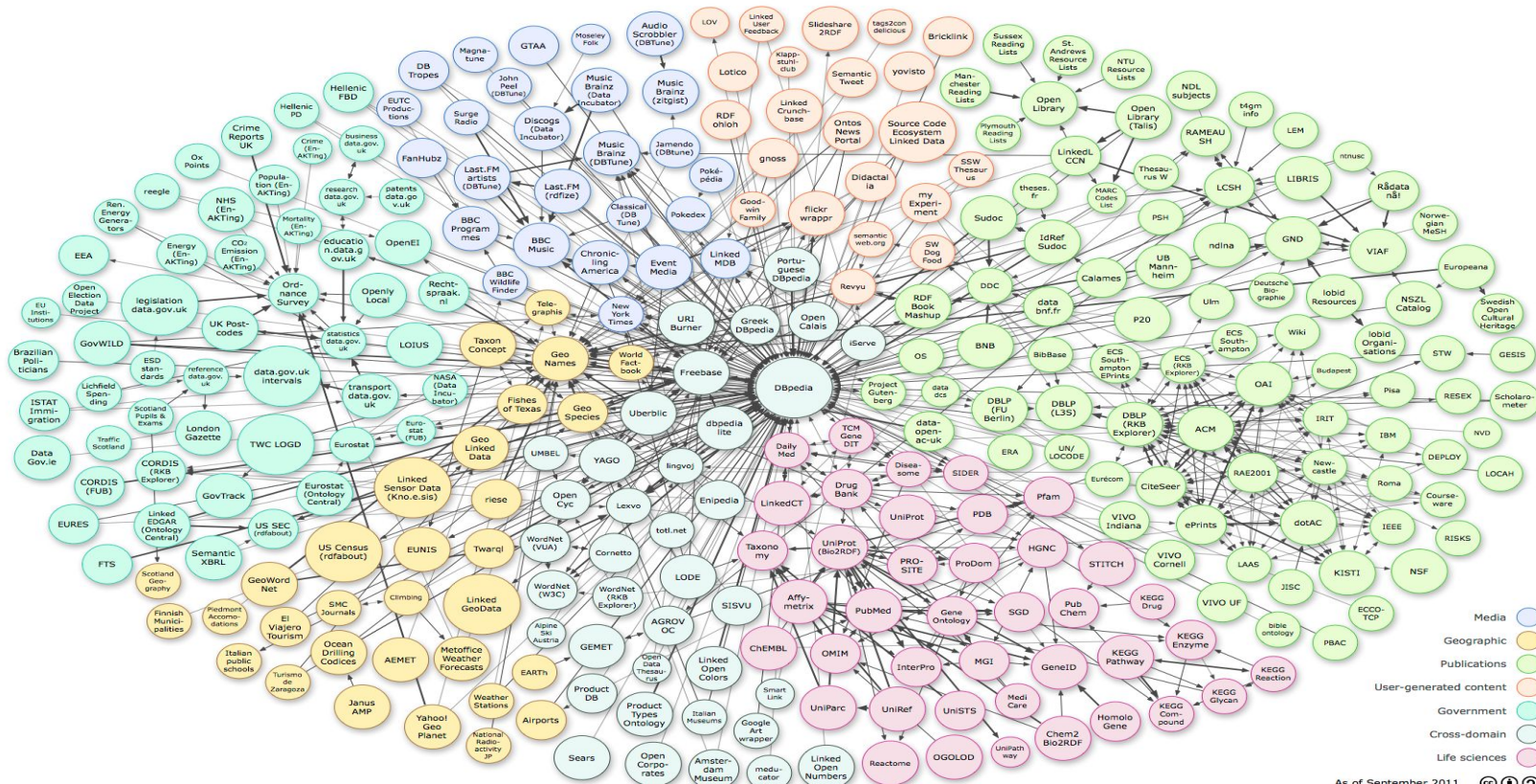
2011-09-19

- 31 billion RDF triples
- 504 million RDF links

2014-08-30

Linked Datasets as of August 2014

Publications
Life Sciences
Cross-Domain
Social Networking
Geographic
Government
Media
User-Generated Content
Linguistics
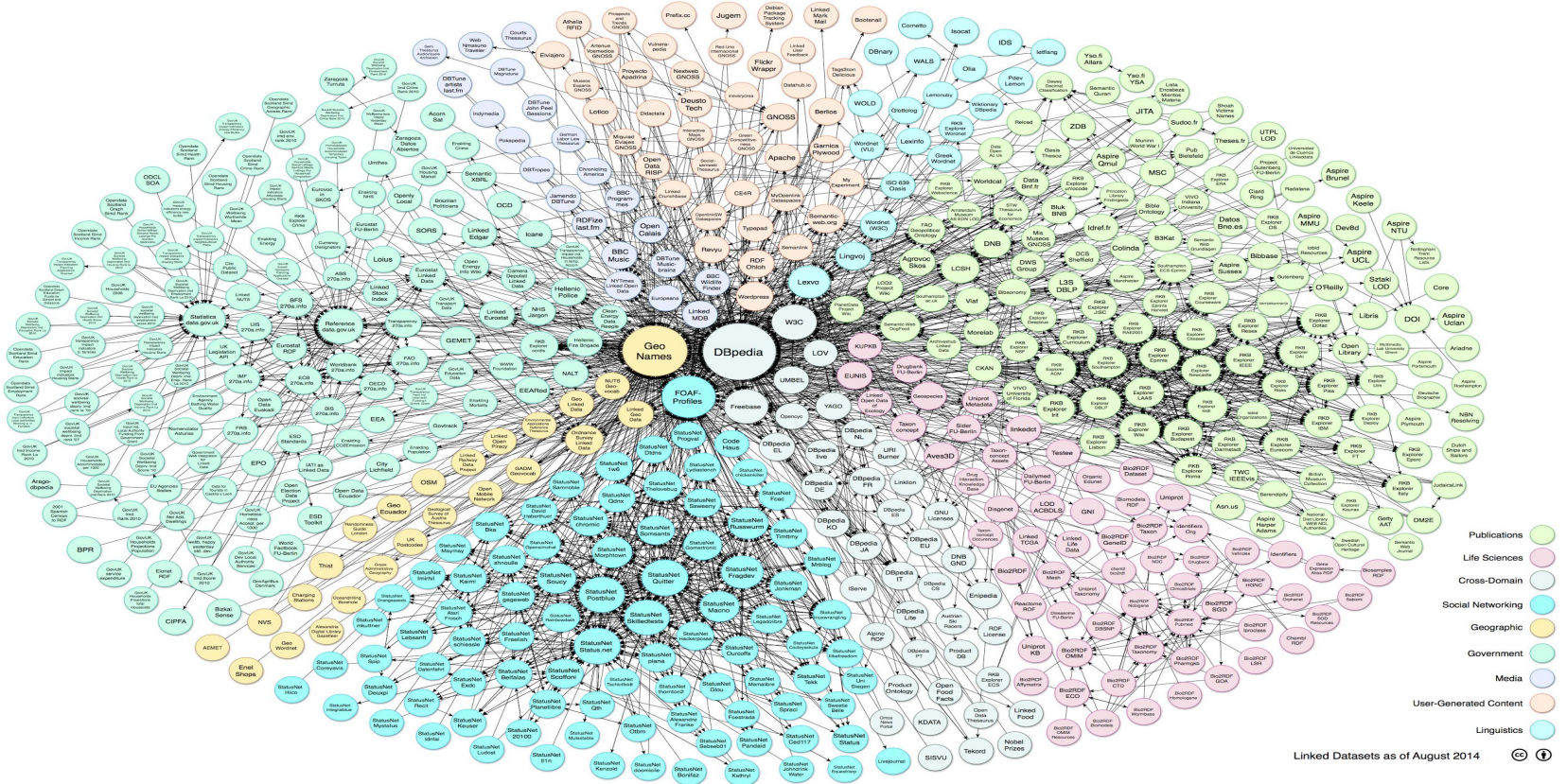
State of the LOD Cloud 2014

http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/

https://upload.wikimedia.org/wikipedia/commons/0/08/LOD_Cloud_2014.svg

Demonstration

SPARQL (SPARQL Protocol and RDF Query Language)

References

https://en.wikipedia.org/wiki/Linked_data

http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/

https://www.w3.org/standards/semanticweb/data

http://eprints.soton.ac.uk/271285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf