

Hasso
Plattner
Institut

IT Systems Engineering | Universität Potsdam

Semantic Web Technologies

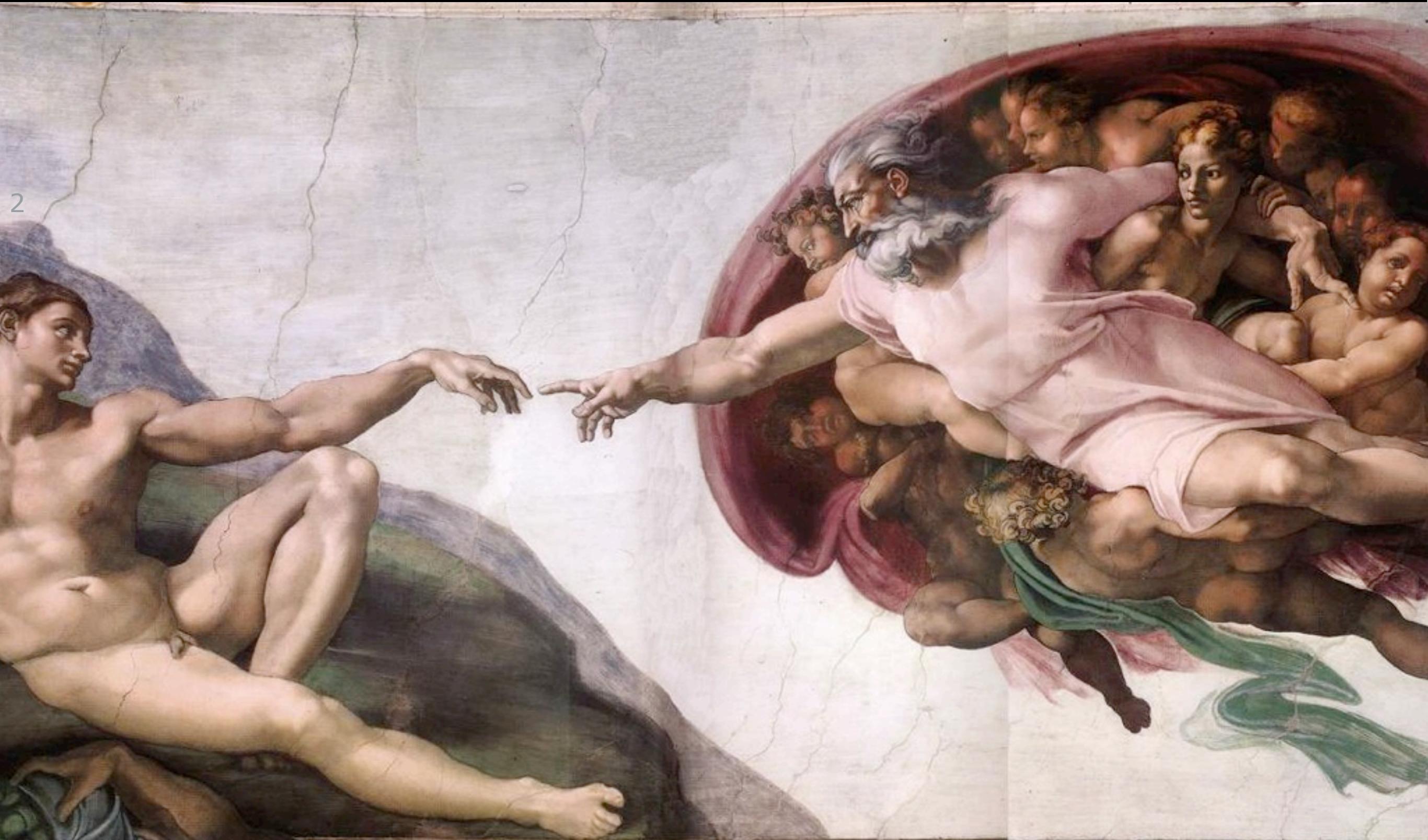
Lecture 6: Applications in the Web of Data
05: Linked Data Engineering (Part 2)

Dr. Harald Sack

Hasso Plattner Institute for IT Systems Engineering

University of Potsdam

Spring 2013



Lecture 6: Applications in the Web of Data

Open HPI - Course: Semantic Web Technologies

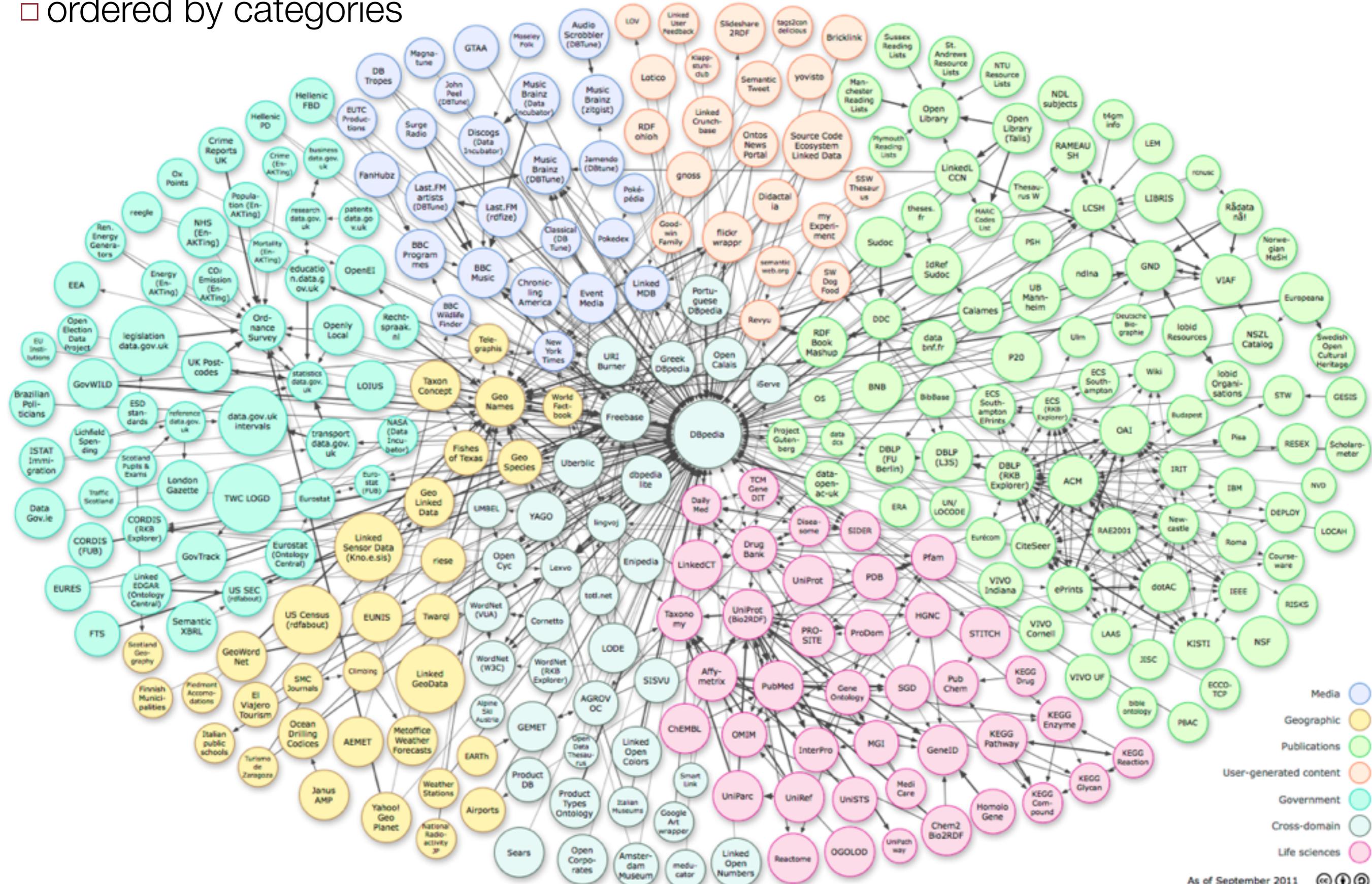
05 - Linked Data Engineering (Part 2)

Open HPI - Course: Semantic Web Technologies - Lecture 6: Applications in the Web of Data

Linked Open Data

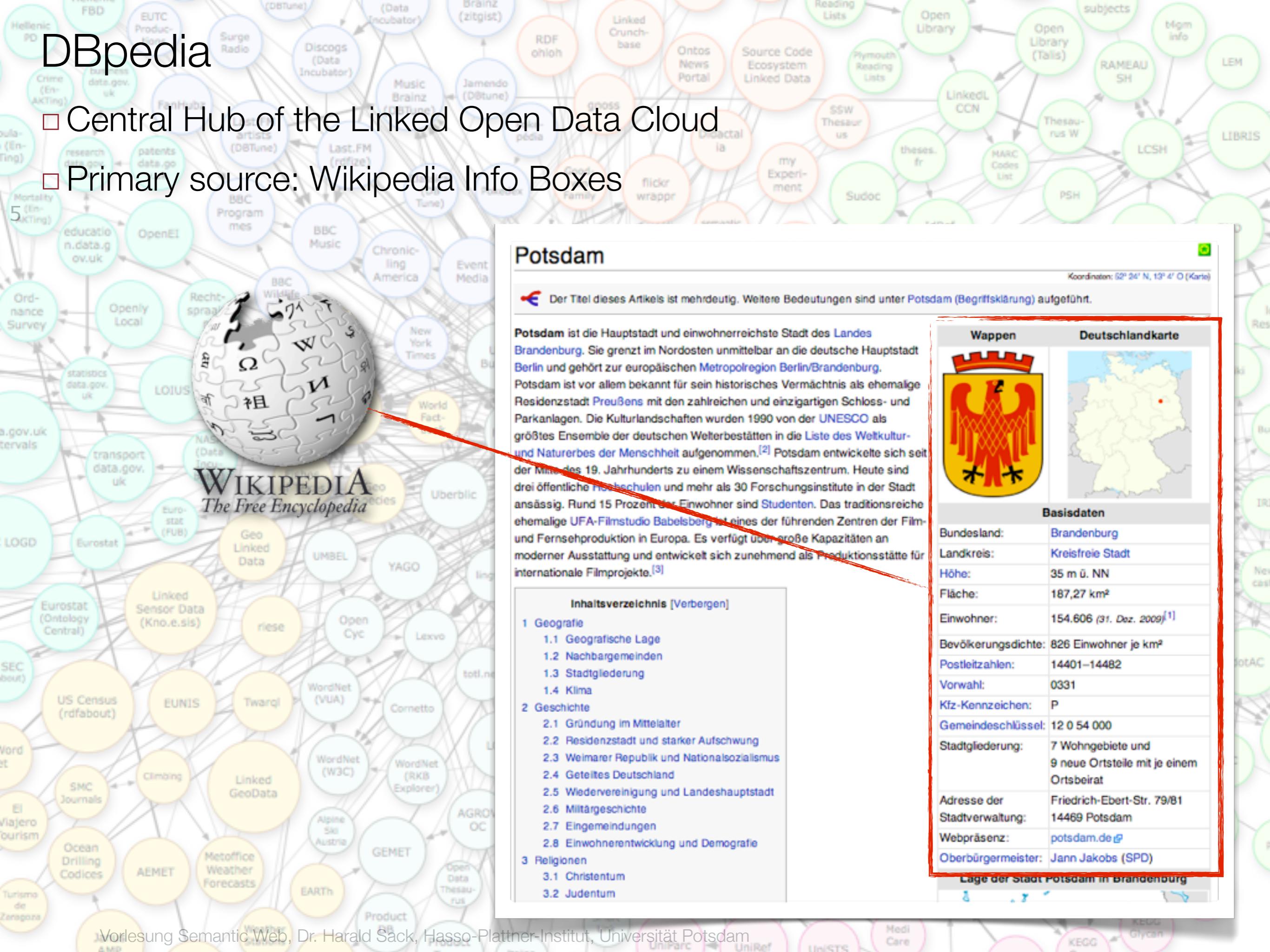
<http://lod-cloud.net/>

ordered by categories



DBpedia

- ❑ Central Hub of the Linked Open Data Cloud
 - ❑ Primary source: Wikipedia Info Boxes



Potsdam

 Der Titel dieses Artikels ist mehrdeutig. Weitere Bedeutungen sind unter Potsdam (Begriffsklärung) aufgeführt.

Potsdam ist die Hauptstadt und einwohnerreichste Stadt des Landes Brandenburg. Sie grenzt im Nordosten unmittelbar an die deutsche Hauptstadt Berlin und gehört zur europäischen Metropolregion Berlin/Brandenburg. Potsdam ist vor allem bekannt für sein historisches Vermächtnis als ehemalige Residenzstadt Preußens mit den zahlreichen und einzigartigen Schloss- und Parkanlagen. Die Kulturlandschaften wurden 1990 von der UNESCO als größtes Ensemble der deutschen Welterbestätten in die Liste des Weltkultur- und Naturerbes der Menschheit aufgenommen.^[2] Potsdam entwickelte sich seit der Mitte des 19. Jahrhunderts zu einem Wissenschaftszentrum. Heute sind drei öffentliche Hochschulen und mehr als 30 Forschungsinstitute in der Stadt ansässig. Rund 15 Prozent der Einwohner sind Studenten. Das traditionsreiche ehemalige UFA-Filmstudio Babelsberg ist eines der führenden Zentren der Film- und Fernsehproduktion in Europa. Es verfügt über große Kapazitäten an moderner Ausstattung und entwickelt sich zunehmend als Produktionsstätte für internationale Filmprojekte.^[3]

Wappen	Deutschlandkarte
	
Basisdaten	
Bundesland:	Brandenburg
Landkreis:	Kreisfreie Stadt
Höhe:	35 m ü. NN
Fläche:	187,27 km ²
Einwohner:	154.606 (31. Dez. 2009) [1]
Bevölkerungsdichte:	826 Einwohner je km ²
Postleitzahlen:	14401–14482
Vorwahl:	0331
Kfz-Kennzeichen:	P
Gemeindeschlüssel:	12 0 54 000
Stadtgliederung:	7 Wohngebiete und 9 neue Ortsteile mit je einem Ortsbeirat
Adresse der Stadtverwaltung:	Friedrich-Ebert-Str. 79/81 14469 Potsdam
Webpräsenz:	potsdam.de ↗
Oberbürgermeister:	Jann Jakobs (SPD)

DBpedia

- Primary source:

- Wikipedia info boxes contain structured data

Ordnance Survey

F K Erweitert Sonderzeichen Hilfe

[[Begriffsklärungshinweis]]

[[Infobox Gemeinde in Deutschland

|Art = Stadt

|Wappen = Coat of arms of Potsdam.svg

|Breitengrad = 52/23/45/N

|Längengrad = 13/03/41/E

|Lageplan = Brandenburg_P.svg

|Lageplanbeschreibung = Lage der Stadt Potsdam in Brandenburg

|Bundesland = Brandenburg

|Landkreis = Kreisfreie Stadt

|Höhe = 35

|Fläche = 187.27

|PLZ = 14401-14482

|Vorwahl = 0331

|Kfz = P

|Gemeindeschlüssel = 12054000

|Gliederung = 7 Wohngebiete und
9 neue Ortsteile mit je einem Ortsbeirat

|Adresse = Friedrich-Ebert-Str. 79/81
14469 Potsdam

|Website = [http://www.potsdam.de/ potsdam.de]

|Bürgermeister = [[Jann Jakobs]]

|Bürgermeistertitel= Oberbürgermeister

|Partei = SPD

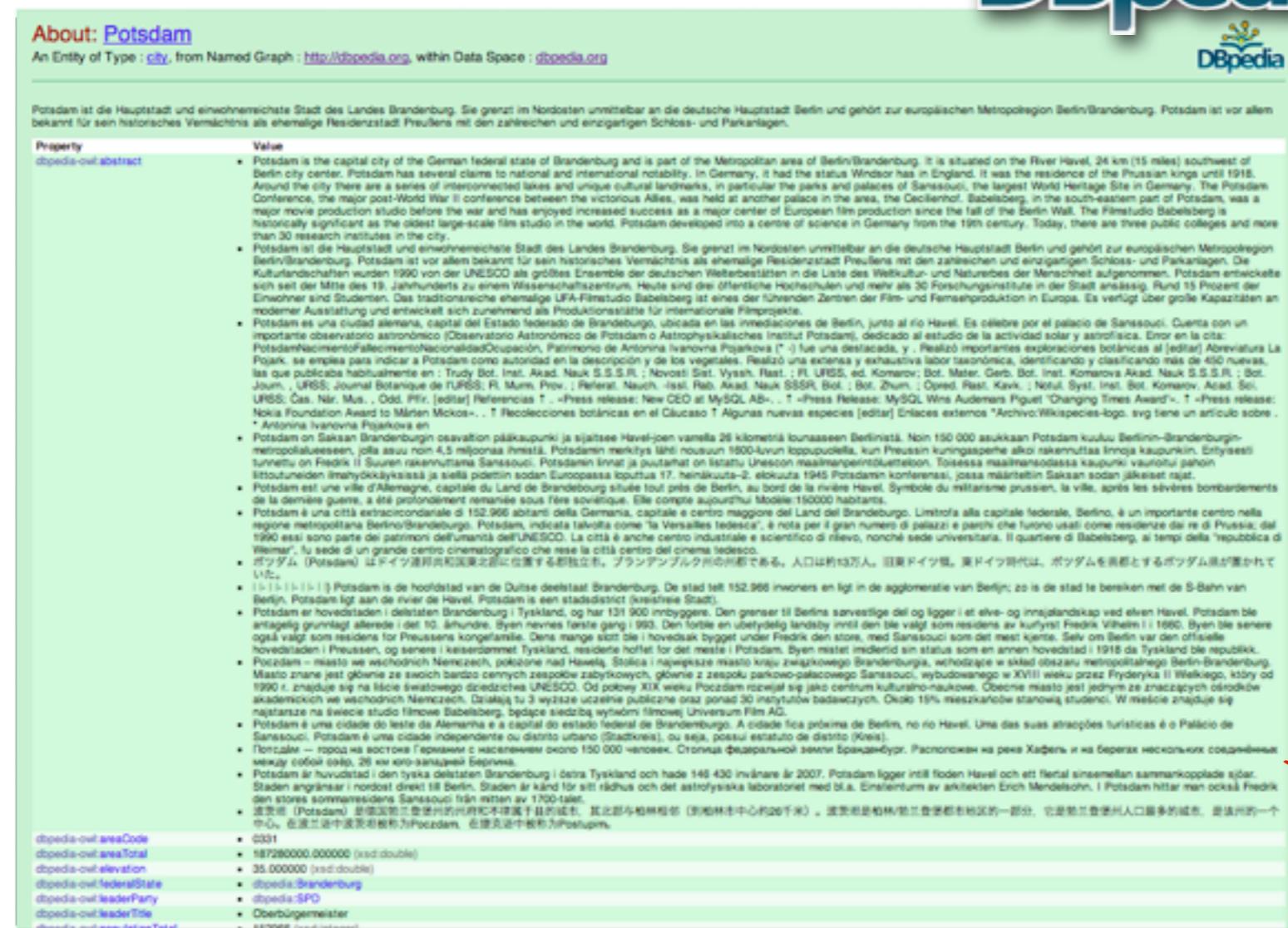
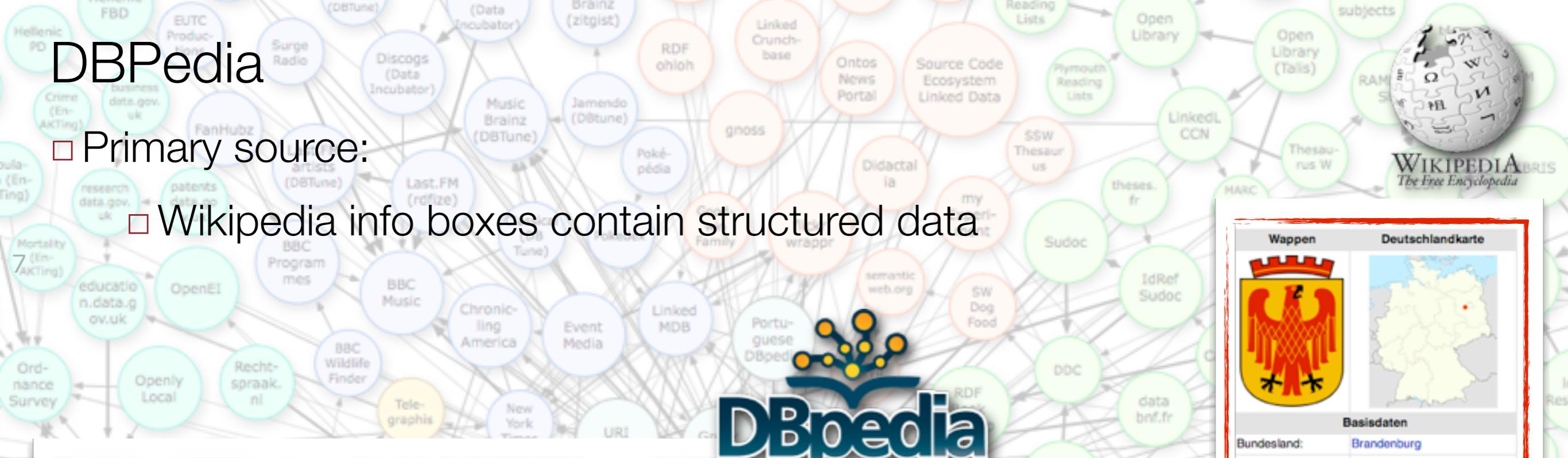
}}



Wappen	Deutschlandkarte
	
Basisdaten	
Bundesland:	Brandenburg
Landkreis:	Kreisfreie Stadt
Höhe:	35 m ü. NN
Fläche:	187,27 km²
Einwohner:	154.606 (31. Dez. 2009) ^[1]
Bevölkerungsdichte:	826 Einwohner je km²
Postleitzahlen:	14401–14482
Vorwahl:	0331
Kfz-Kennzeichen:	P
Gemeindeschlüssel:	12 0 54 000
Stadtgliederung:	7 Wohngebiete und 9 neue Ortsteile mit je einem Ortsbeirat
Adresse der Stadtverwaltung:	Friedrich-Ebert-Str. 79/81 14469 Potsdam
Webpräsenz:	potsdam.de
Oberbürgermeister:	Jann Jakobs (SPD)

DBpedia

- Primary source:
 - Wikipedia info boxes contain structured data



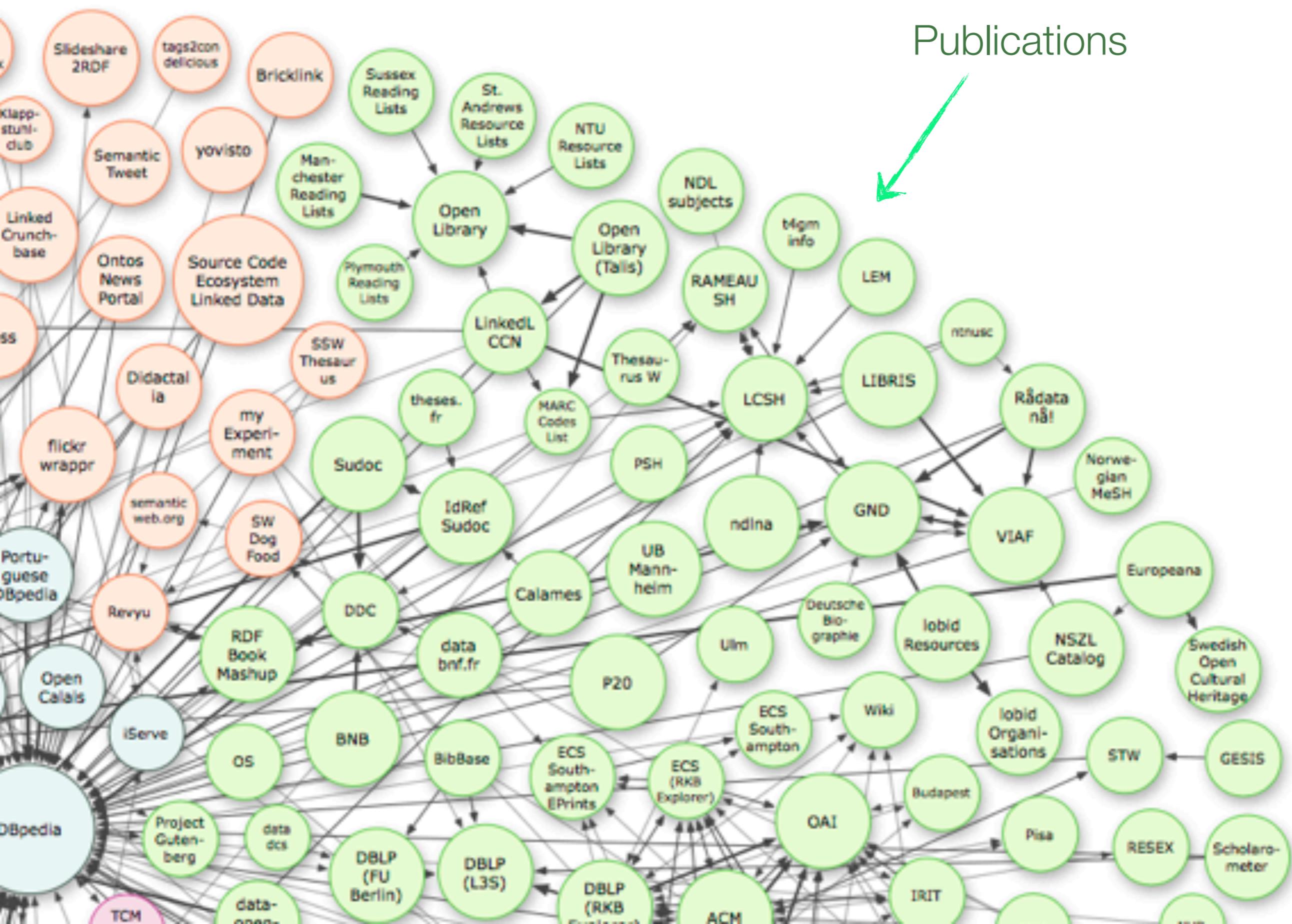
Linked Open Data

User Generated Content

Media



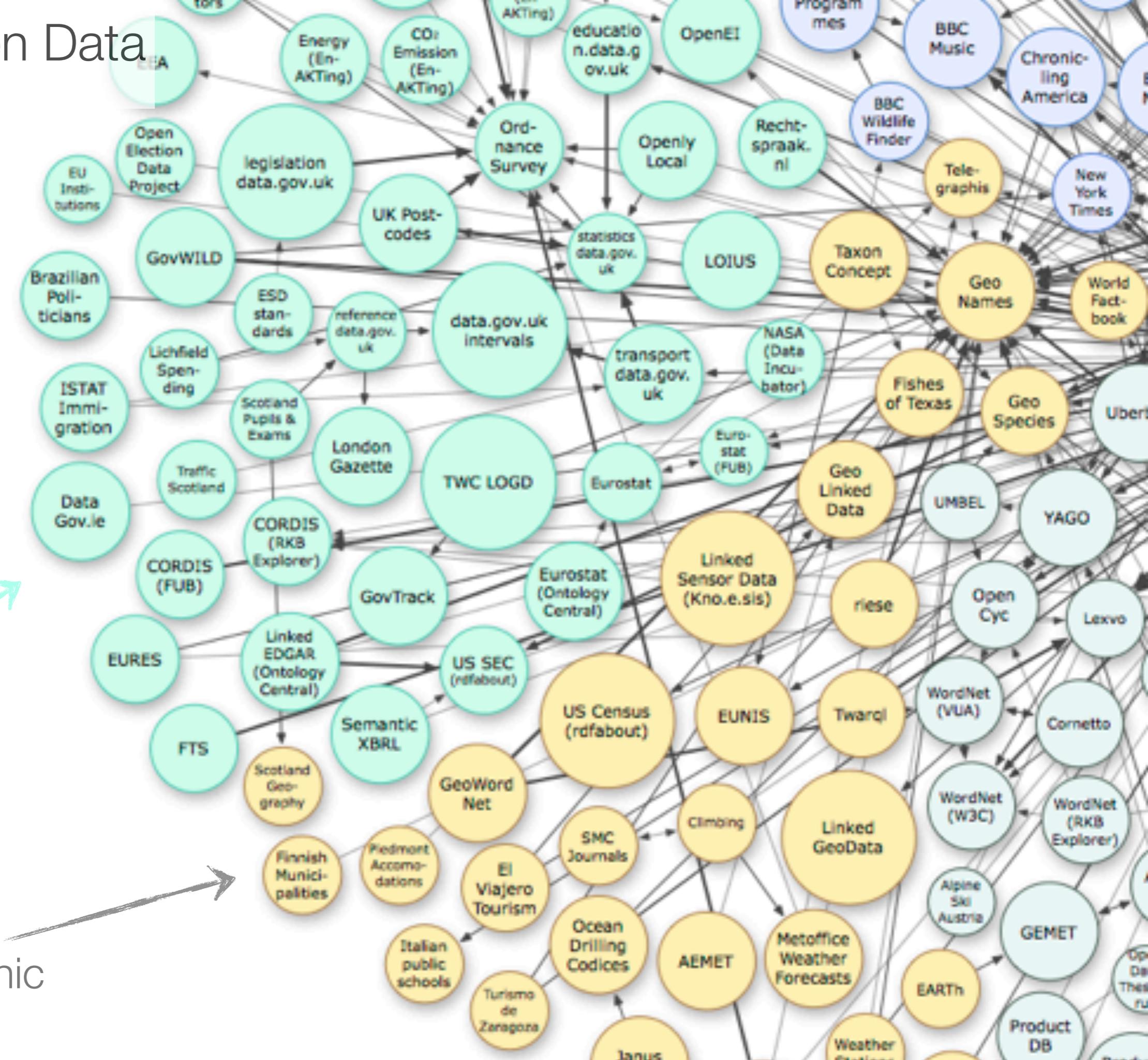
Linked Open Data



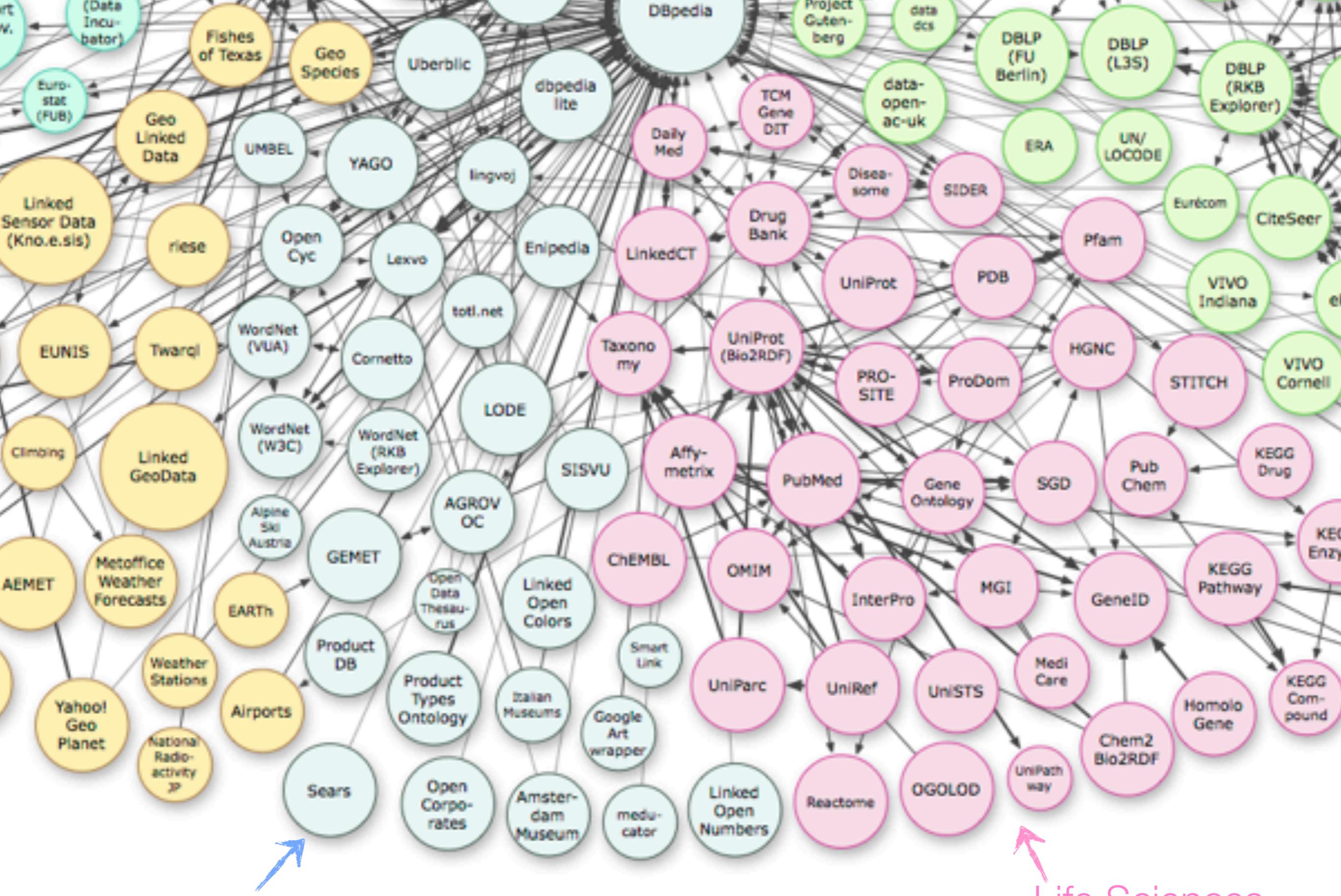
Linked Open Data

10

Government



Geographic



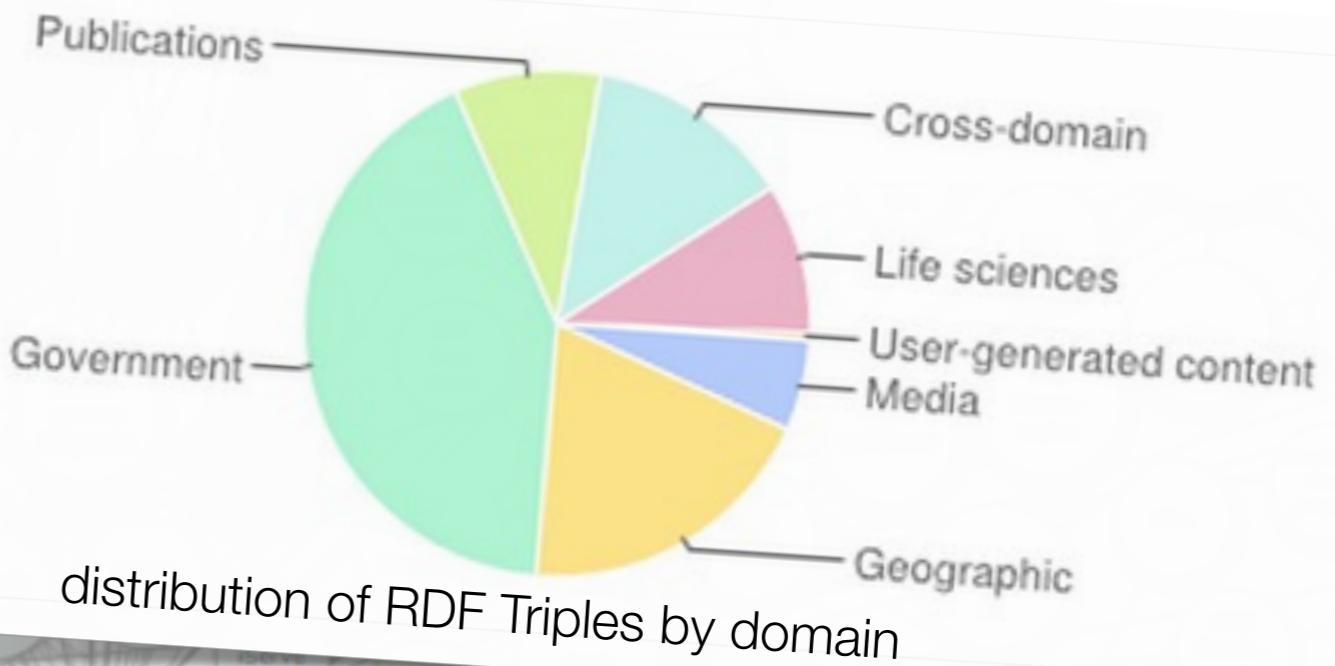
Cross-Domain

Life Sciences

Linked Open Data

Linking Open Data

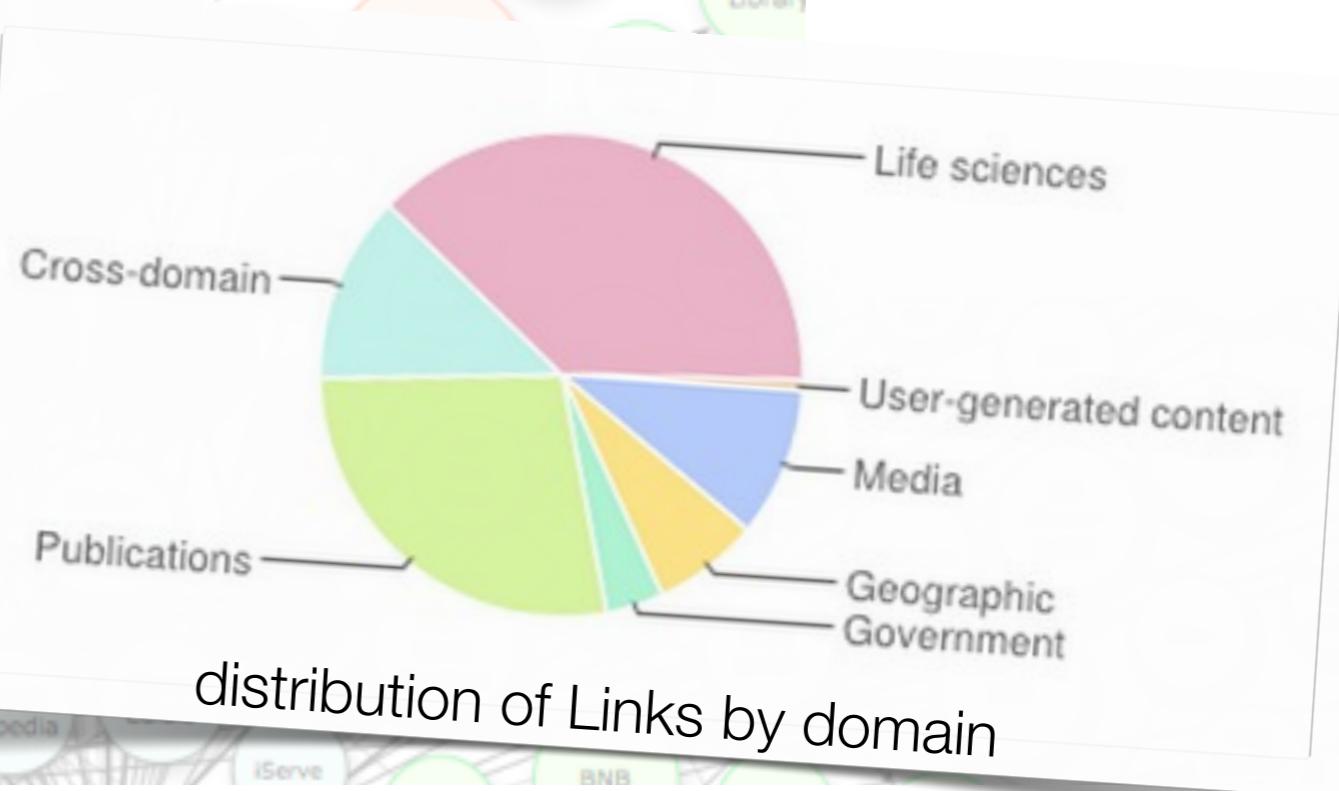
■ Some statistics (as of 09/2011)



Domain	Number of datasets	Triples	%	(Out-)Links	%
Media	25	1,841,852,061	5.82 %	50,440,705	10.01 %
Geographic	31	6,145,532,484	19.43 %	35,812,328	7.11 %
Government	49	13,315,009,400	42.09 %	19,343,519	3.84 %
Publications	87	2,950,720,693	9.33 %	139,925,218	27.76 %
Cross-domain	41	4,184,635,715	13.23 %	63,183,065	12.54 %
Life sciences	41	3,036,336,004	9.60 %	191,844,090	38.06 %
User-generated content	20	134,127,413	0.42 %	3,449,143	0.68 %
Total	295	31,634,213,770		503,998,829	

Linking Open Data

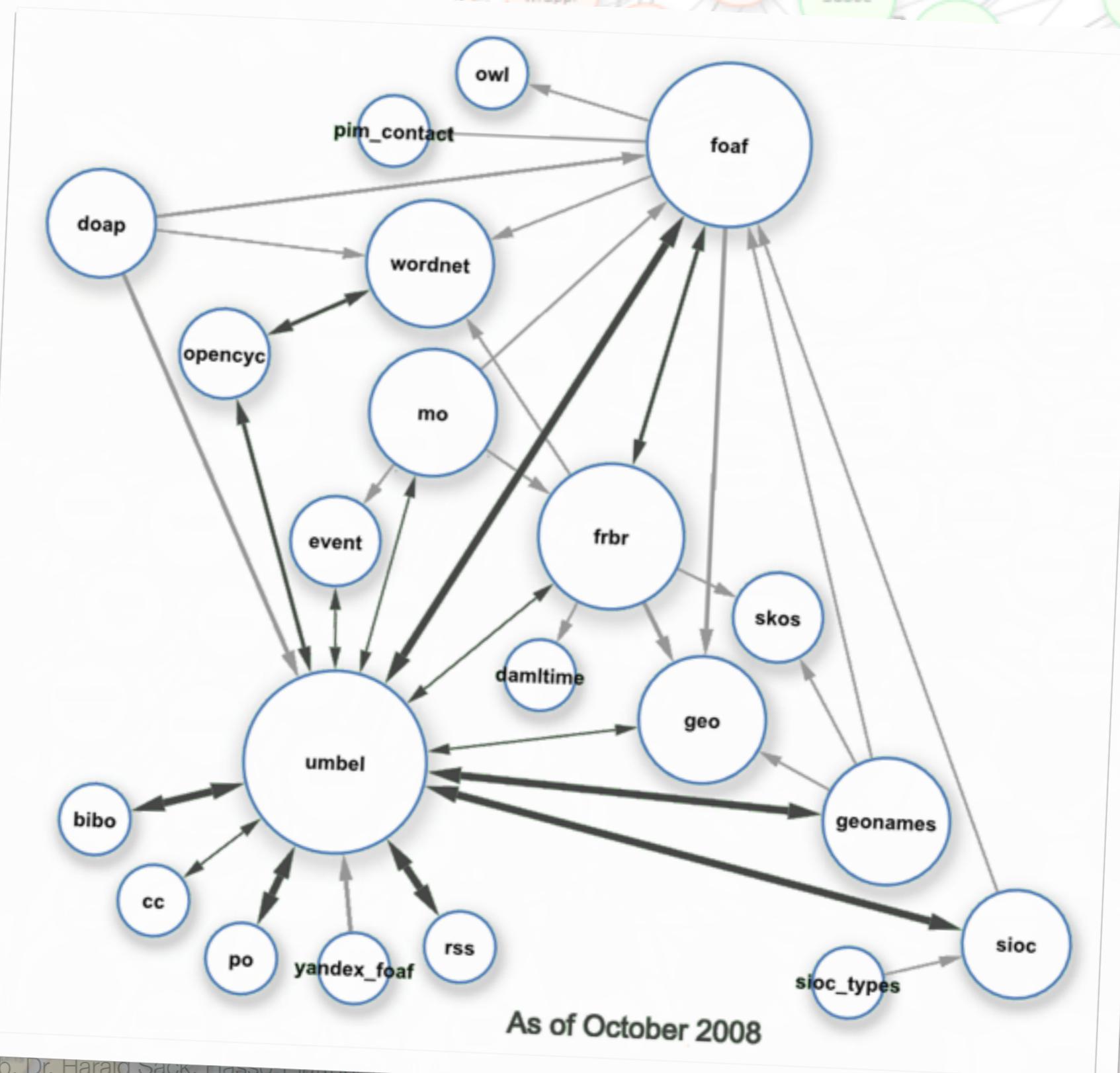
■ Some statistics (as of 09/2011)



Domain	Number of datasets	Triples	%	(Out-)Links	%
Media	25	1,841,852,061	5.82 %	50,440,705	10.01 %
Geographic	31	6,145,532,484	19.43 %	35,812,328	7.11 %
Government	49	13,315,009,400	42.09 %	19,343,519	3.84 %
Publications	87	2,950,720,693	9.33 %	139,925,218	27.76 %
Cross-domain	41	4,184,635,715	13.23 %	63,183,065	12.54 %
Life sciences	41	3,036,336,004	9.60 %	191,844,090	38.06 %
User-generated content	20	134,127,413	0.42 %	3,449,143	0.68 %
	295	31,634,213,770		503,998,829	

Linked Data Ontologies

- Ontologies hold the Linked Data Cloud together

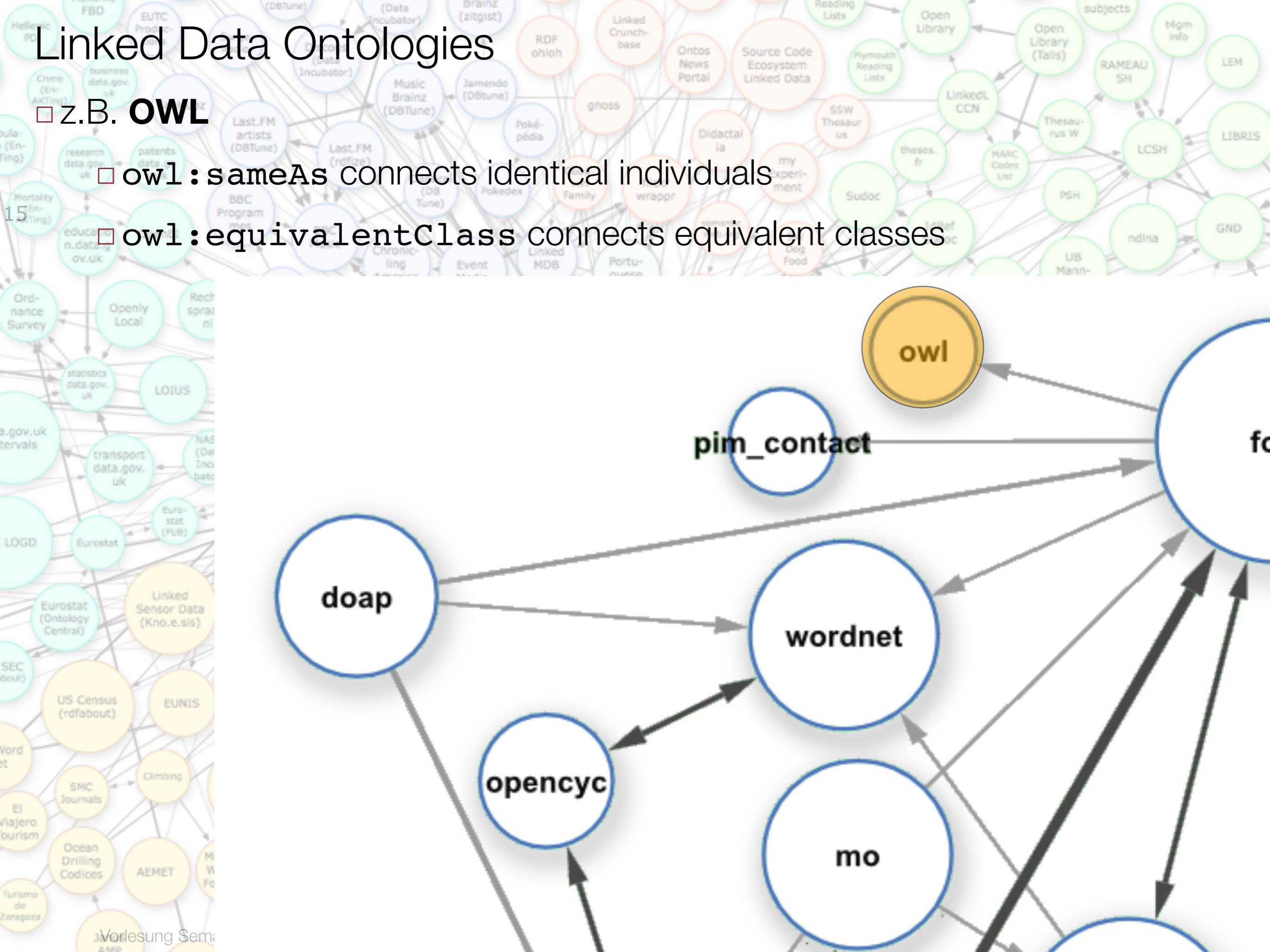


Linked Data Ontologies

z.B. OWL

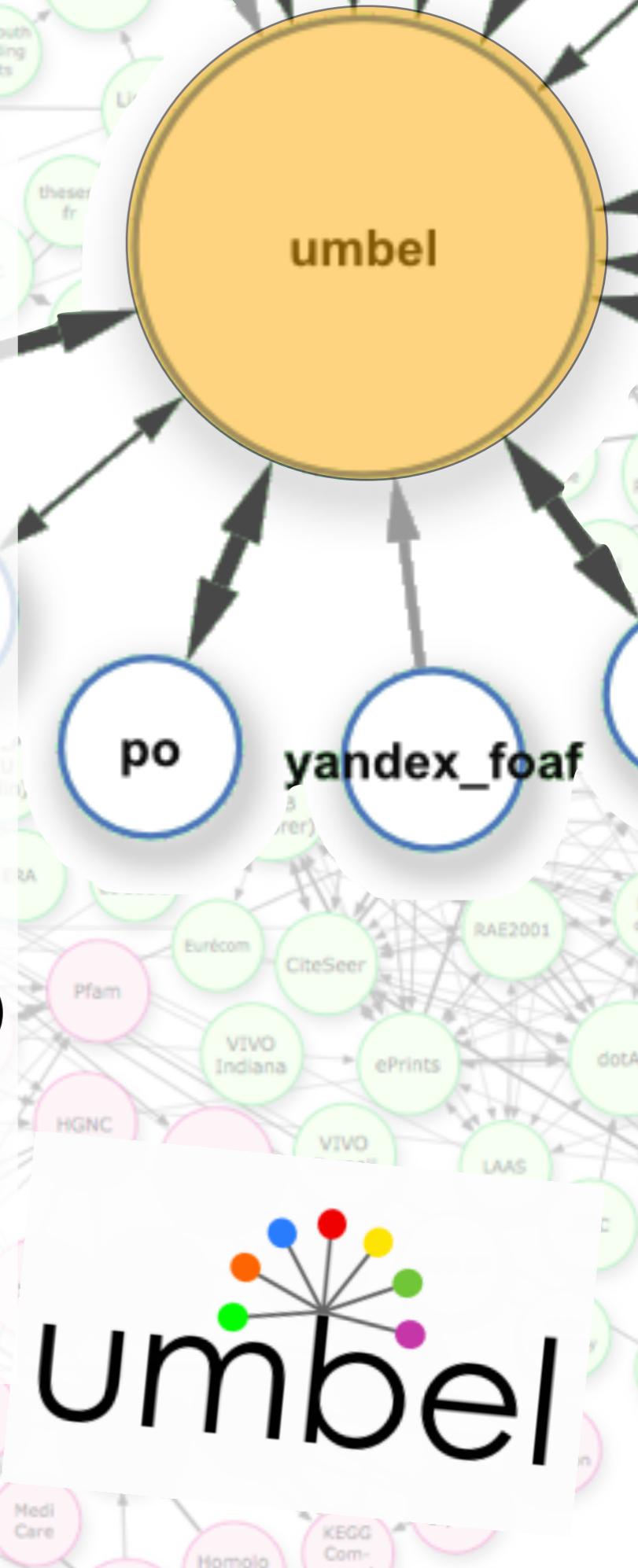
owl:sameAs connects identical individuals

owl:equivalentClass connects equivalent classes



Linked Data Ontologies

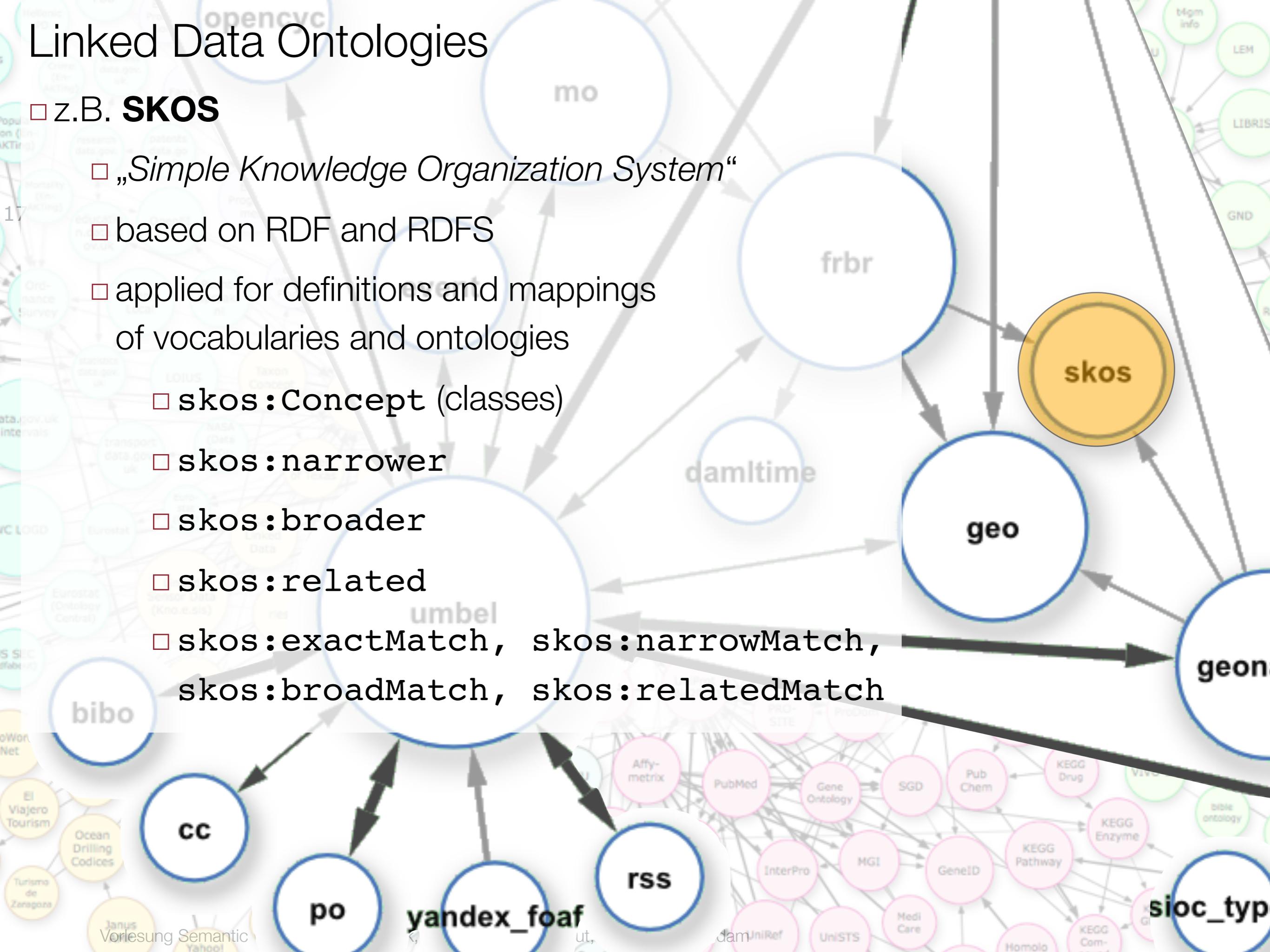
- z.B. **umbel** (version 1.0, Feb. 2011)
 - „Upper Mapping and Binding Exchange Layer“
 - Subset of OpenCyc as RDF Triples based on SKOS and OWL2
 - Upper Ontology with 28.000 concepts (`skos:Concept`)
 - 46.000 Mappings into DBpedia, geonames, e.a. (`owl:equivalentClass`, `rdfs:subClassOf`)
 - Links to more than 2 Mio Wikipedia pages



Linked Data Ontologies

□ z.B. **SKOS**

- „Simple Knowledge Organization System“
- based on RDF and RDFS
- applied for definitions and mappings of vocabularies and ontologies
 - **skos:Concept** (classes)
 - **skos:narrower**
 - **skos:broader**
 - **skos:related**
 - **skos:exactMatch**, **skos:narrowMatch**,
skos:broadMatch, **skos:relatedMatch**



Linked Data Sources in the Web

□ Native publication

- D2R-Server, OpenLink Virtuoso, Pubby, etc.

□ Implementation of Wrappers around existing applications / APIs

- SIOC Exporter for Wordpress, Drupal, phpBB, ...
- RDF Book Mashup (Amazon API, Google Base-API, ...)

□ Linking Open Data Project

- Semantic Web Education and Outreach W3C working group
- Catalogue of all known sources of linked data with an open source license
 - » DBPedia, Flickr, Open-Cyc, FOAF, SIOC, GeoNames, ...

Browser for Linked Data

■ Differences to arbitrary RDF-Browsers

- RDF Data to be visualized does not necessarily reside in local repository, but is distributed in the Web
 - requires dynamic reload of RDF resources

■ Tabulator (Tim Berners-Lee, MIT-)

(T. Berners-Lee et al.: Tabulator: Exploring and analyzing linked data on the semantic web, in Proc. 3rd Int. Semantic Web User Interaction Workshop, 2006, <http://swui.semanticweb.org/swui06/papers/Berners-Lee/Berners-Lee.pdf>)

■ OpenLink RDF Data Explorer

- enables visualization as graph, timeline, map, etc.
<http://ode.openlinksw.com/>

■ Zitgist Browser

<http://browser.zitgist.com/>

■ DISCO Browser

<http://sites.wiwiss.fu-berlin.de/suhl/bizer/ng4j/disco/>

Search Engines for Linked Data

- Crawler-based, follow links in datasets to create an index that can be queried
 - Swoogle
 - keyword-based full text search (Apache-Lucene), uses only limited semantic annotation
<http://swoogle.umbc.edu/>
 - Semantic Web Search Engine (SWSE)
 - additionally uses rdf:type properties as search filter
<http://swse.deri.org/>
 - Sindice
 - <http://www.sindice.com/>
 - Falcons
 - with data browser for result analysis
<http://iws.seu.edu.cn/services/falcons/>
 - Sig.ma - Semantic Information Mashup (based on Sindice)
 - <http://sig.ma/>

Linked Data Driven Web Applications

- Required Components:

- **Local RDF Store**

- caching of results
 - permanent storage

- **Logic** (Controller) and
User Interface (=Business Logic)

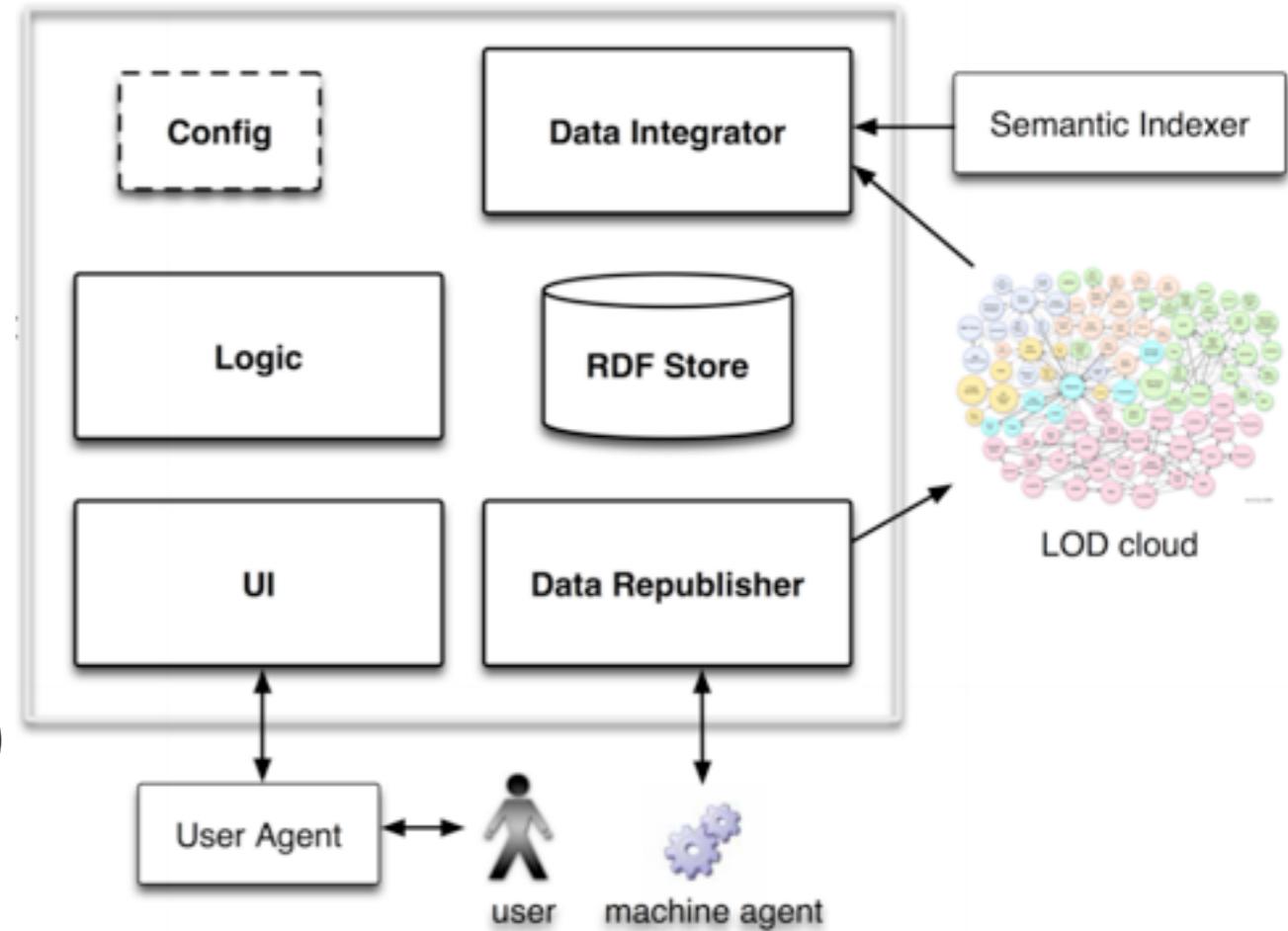
- (not LOD specific)

- **Data Integration component**

- get data directly from LOD-Cloud or
 - via Semantic Indexer (sindice, etc.)

- **Data Republishing component**

- write back application dependent data into the Web of Data



Linked Data Driven Web Applications

- Access to Linked Data via **SPARQL endpoints**
- ...but where do I find SPARQL endpoints?
 - W3C: Currently Alive SPARQL Endpoints
<http://esw.w3.org/SparqlEndpoints>
- SPARQL endpoints are a **RESTful Web Services**
 - HTTP GET Request with SPARQL query
 - Result available as
 - XML, JSON, plaintext (SPARQL Select/Ask)
 - RDF/XML, NTriples, Turtle, N3 (SPARQL Describe/Construct)
 - Data format can be determined via HTTP Accept Header z.B. `Accept: application/sparql-results+json`
 - (or via parameters of the SPARQL query)

Linked Data Driven Web Applications

- The easiest way is to make use of a suitable library:

- SPARQL Javascript Library

http://www.thefigtrees.net/lee/blog/2006/04/sparql_calendar_demo_a_sparql.html

- ARC for SPARQL (PHP)

<http://arc.semsol.org/>

- RAP - RDF API für PHP

<http://www4.wiwiss.fu-berlin.de/bizer/rdfapi/index.html>

- Jena/ARQ (Java)

<http://jena.sourceforge.net/>

- Sesame (Java)

<http://www.openrdf.org/>

- SPARQL Wrapper (Python)

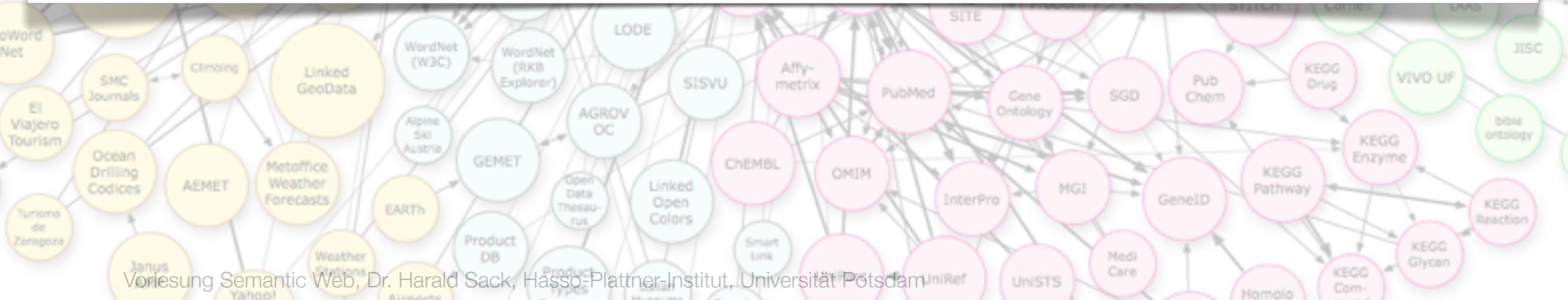
<http://sparql-wrapper.sourceforge.net/>

- ...

Linked Data Driven Web Applications

- Simple example with Jena ARQ:

```
import com.hp.hpl.jena.query.*;  
  
String service = "..."; // address of the SPARQL endpoint  
String query = "SELECT ..."; // your SPARQL query  
QueryExecution e = QueryExecutionFactory.sparqlService(service, query)  
  
ResultSet results = e.execSelect();  
while ( results.hasNext() ) {  
    QuerySolution s = results.nextSolution();  
    // ...  
}  
  
e.close();
```





06 - Named Entity Recognition

Open HPI - Course: Semantic Web Technologies - Lecture 6: Applications in the Web of Data