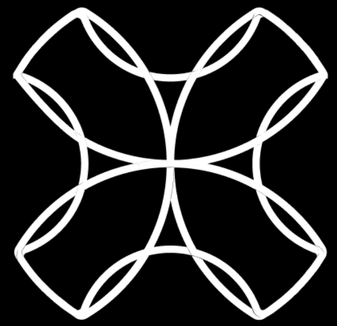


IKNEX

Knowledge Representation and Reasoning

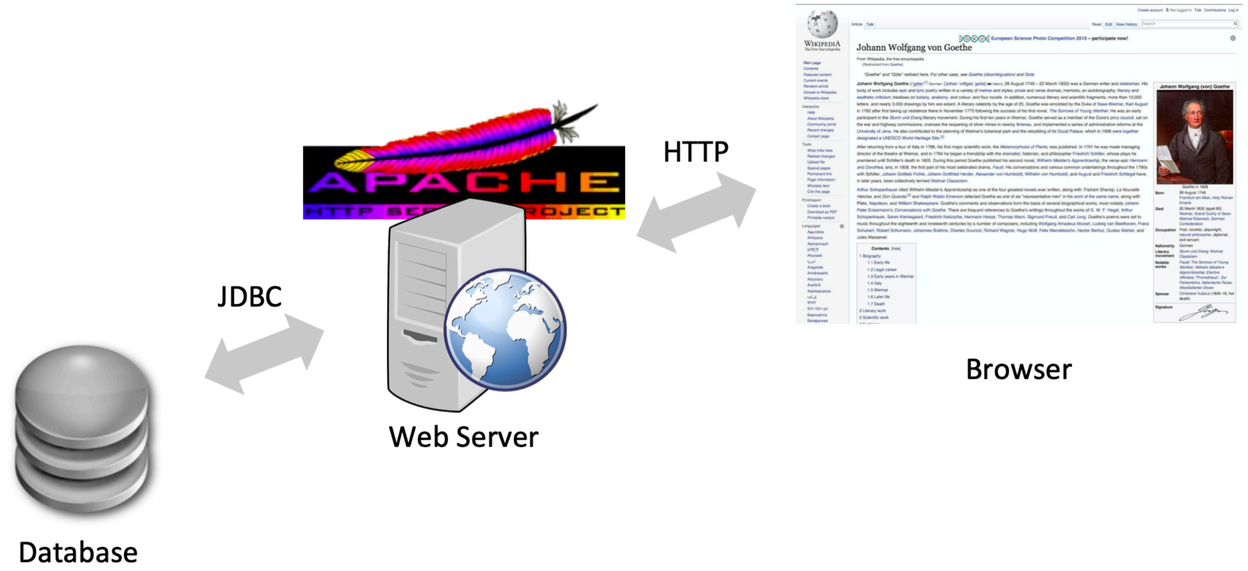
Fall 2023

Dr. Amna Basharat | Amna binte Kamran

Linked Data Engineering & Applications

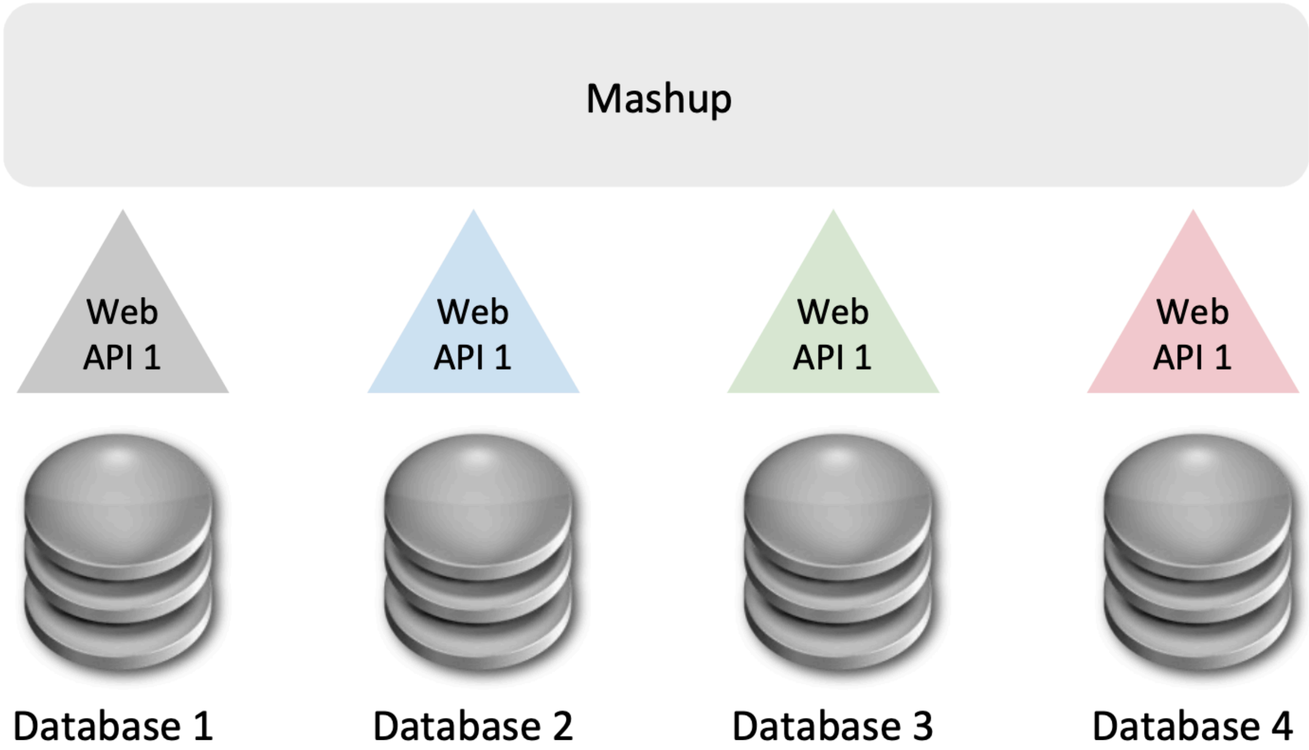
How To Get Data From the Web?

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



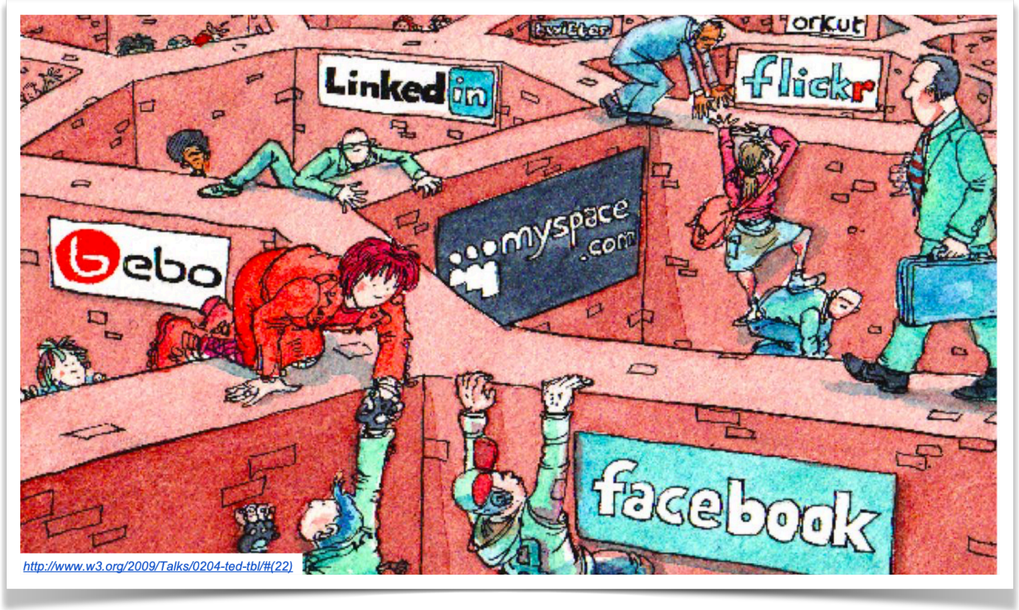
How To Get Data From the Web?

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



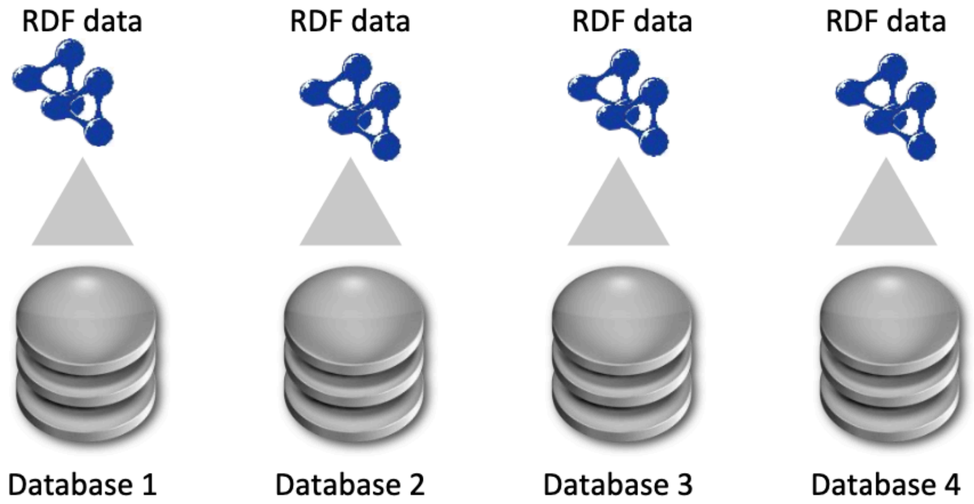
In the Web 2.0 and Even Now…

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



How To Get Rid of Closed Data Islands?

- Apply Semantic Web technologies
  - to publish (structured) data on the web
  - to draw connections from one data source to data from other data sources

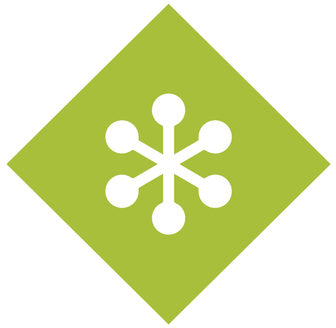


Linked Data and the ‘Web of Data‘

- Term refers to an idea originally from Tim Berners-Lee (Tim Berners-Lee, Linked Data, 2006, http://www.w3.org/DesignIssues/LinkedData.html)

- Set of best practices for publication and linking of structured data on the web

- Basic assumption: The value of data on the web increases when they are connected to other data sources



Linked Data Principles

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

Linked Data Principles (1/4)

Cd://Muhammad_Ali_Jinnah

- 1. Use URIs as names for things.
  - URIs do not only identify documents but also arbitrary objects
  - of the real world as well as abstract concepts

.

http://www.imdb.com/title/tt0017136/

https://viaf.org/viaf/24602065/

http://musicbrainz.org/artist/5700dcd4-c139-4f31-aa3e-6382b9af9032

Linked Data Principles (2/4)

- 2. Use HTTP URIs, so that people can look up those names.
  - HTTP URIs (URLs) as globally unique names enable dereferencing of associated information in the Web
    - *via http Content Negotiation*
    - *303 URIs*
      http Response Code 303 ‚See Other' (redirect)
    - *Hash URIs*
      http://example.com/Harald#me

Linked Data Principles (3/4)

- 3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
  - RDF as universal data model for publishing structured data on the Web
  - Make all URIs in the RDF graph dereferencible
  - Avoid RDF constructs that cause problems in Linked Data context
    - RDF Reification
    - RDF Collections und Containers
    - unnamed Blank Nodes

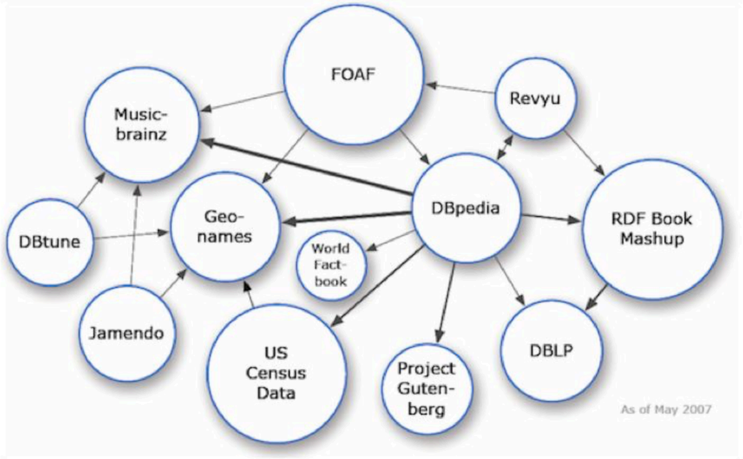
Linked Data Principles (4/4)

- Include links to other URIs, so that they can discover more things.
  - Link RDF references among data between different data sources, to find information related by content
  - **Relationship Links**
    Links to external LOD Entitites related with the original entity
  - **Identity Links**
    Links to external LOD Entities referring to the same object or concept
  - **Vocabulary Links**
    Links to definitions of the original entity

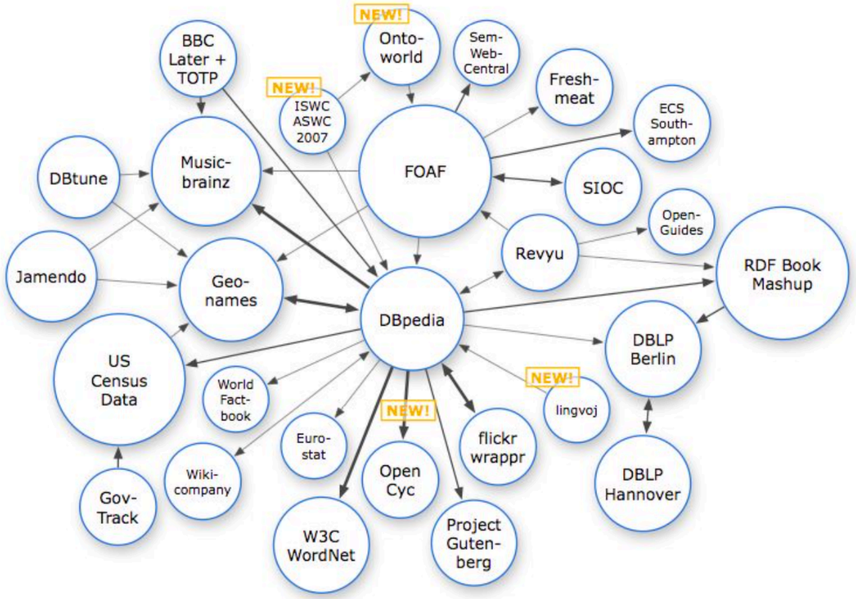
The Application of the Linked Data Principles Leads to a “Web of Data”

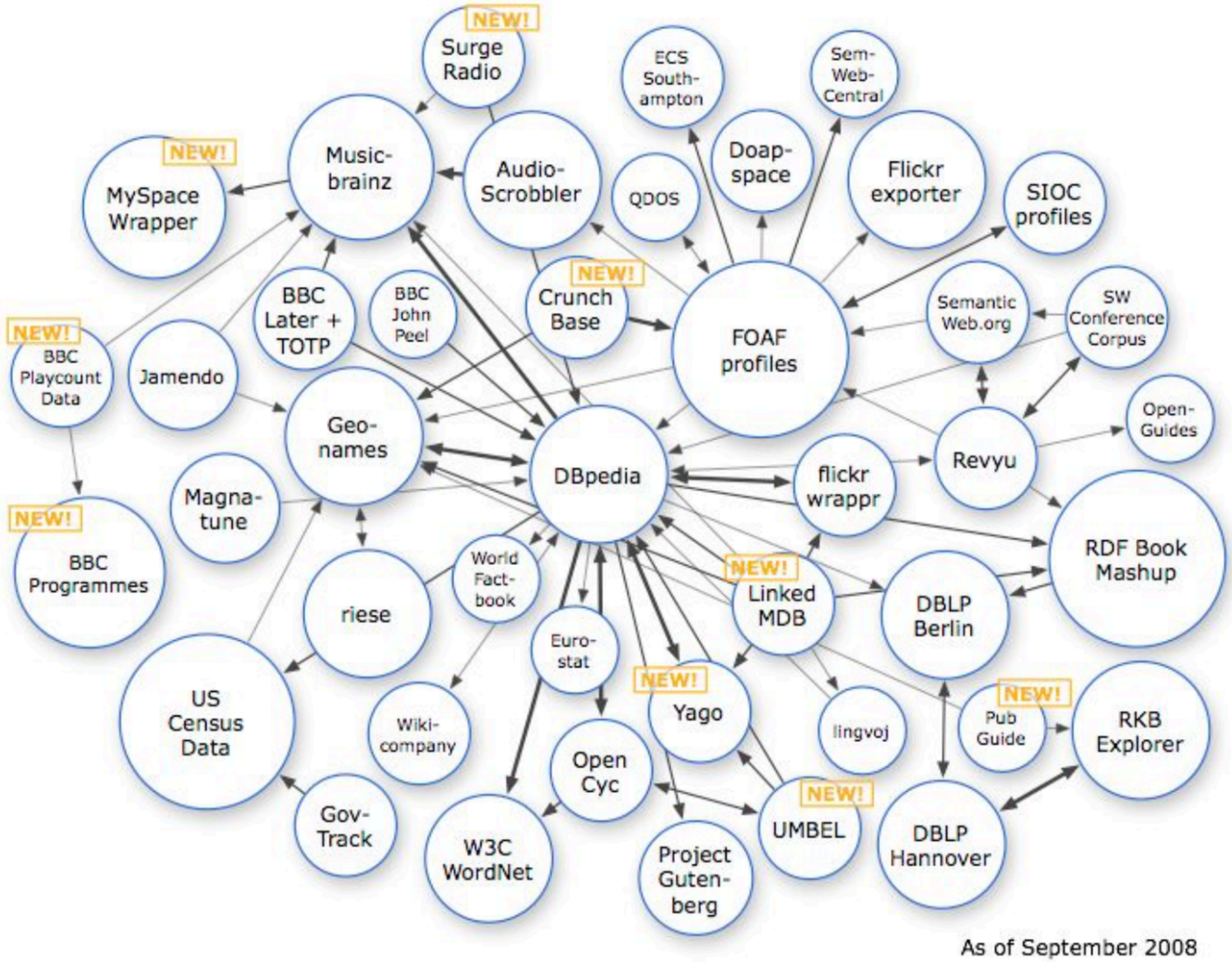
The Development of the Web of Data

The Development of the Web of Data

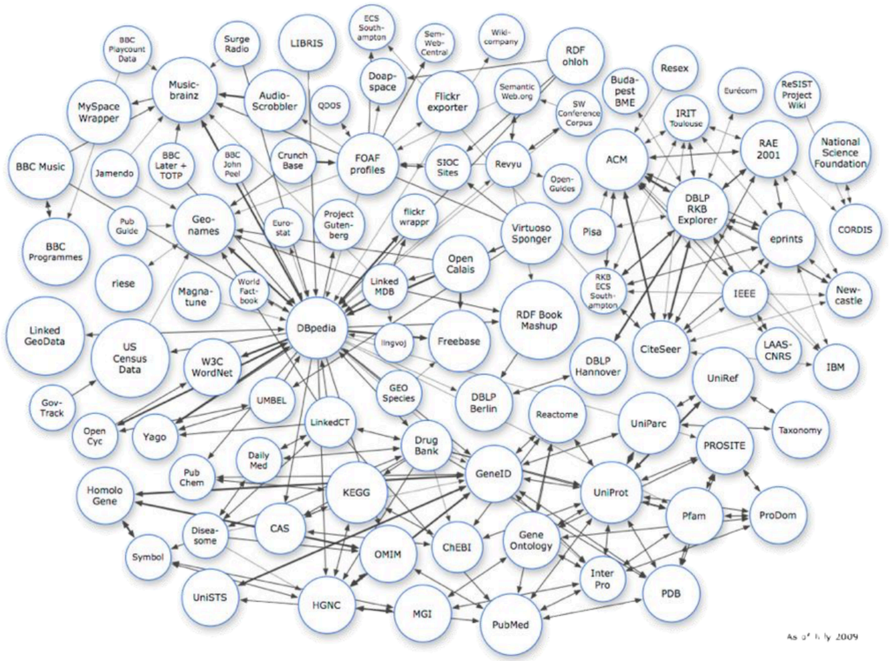


Nov 2007

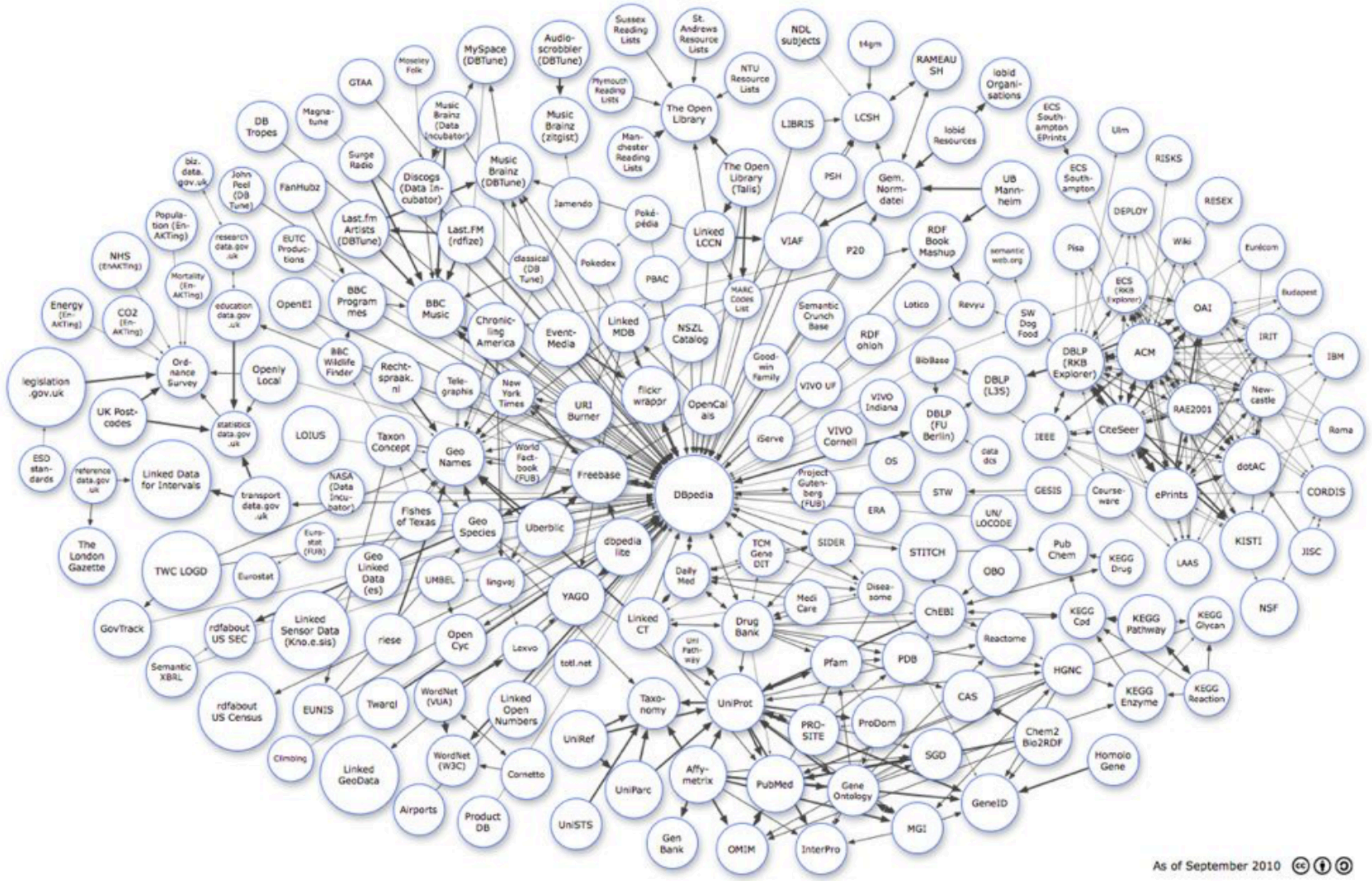

NEW!
BBC Later + TOTP
Onto-world
Sem-Web-Central
NEW!
ISWC ASWC 2007
Fresh-meat
ECS South-ampton
DBtune
Music-brainz
FOAF
SIOC
Open-Guides
Revyu
RDF Book Mashup
Jamendo
Geo-names
DBpedia
DBLP Berlin
US Census Data
World Fact-book
NEW!
lingvoj
NEW!
Euro-stat
flickr wrappr
Wiki-company
Open Cyc
DBLP Hannover
Gov-Track
W3C WordNet
Project Guten-berg


NEW!
Surge Radio
ECS South-ampton
Sem-Web-Central
Music-brainz
Audio-Scrobbler
Doap-space
Flickr exporter
NEW!
MySpace Wrapper
QDOS
SIOC profiles
BBC Later + TOTP
BBC John Peel
NEW!
Crunch Base
FOAF profiles
Semantic Web.org
SW Conference Corpus
NEW!
BBC Playcount Data
Jamendo
Geo-names
DBpedia
Revyu
Open-Guides
Magna-tune
flickr wrappr
RDF Book Mashup
NEW!
BBC Programmes
World Fact-book
NEW!
Linked MDB
DBLP Berlin
riese
Euro-stat
US Census Data
NEW!
Yago
NEW!
Pub Guide
RKB Explorer
Wiki-company
lingvoj
Open Cyc
NEW!
UMBEL
DBLP Hannover
Gov-Track
W3C WordNet
Project Guten-berg
As of September 2008

July 2009

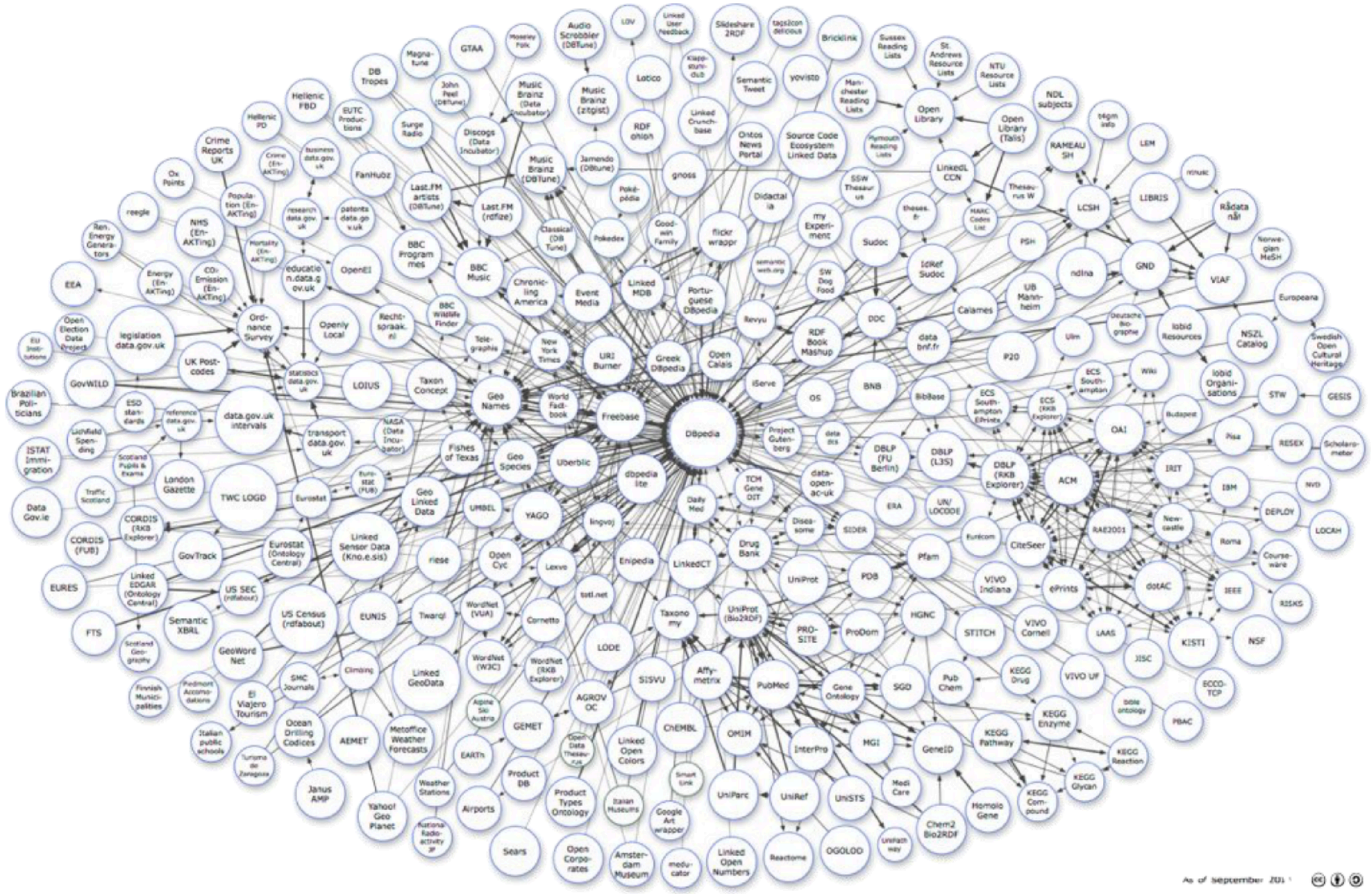

BBC Playcount Data
Surge Radio
LIBRIS
ECS South-ampton
Sem-Web-Central
Wiki-company
RDF ohloh
Resex
Music-brainz
Audio-Scrobbler
Doap-space
Flickr exporter
Semantic Web.org
Buda-pest BME
Eurécom
ReSIST Project Wiki
MySpace Wrapper
QOOS
SW Conference Corpus
IRIT Toulouse
BBC Music
BBC Later + TOTP
BBC John Peel
Crunch Base
FOAF profiles
SIOC Sites
Revyu
ACM
RAE 2001
National Science Foundation
Jamendo
Open-Guides
DBLP RKB Explorer
Pub Guide
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Virtuoso Sponger
Pisa
CORDIS
BBC Programmes
Open Calais
eprints
riese
RKB ECS South-ampton
IEEE
New-castle
Magna-tune
World Fact-book
Linked MDB
RDF Book Mashup
Linked GeoData
DBpedia
lingvoj
Freebase
CiteSeer
LAAS-CNRS
US Census Data
W3C WordNet
DBLP Hannover
IBM
UniRef
Gov-Track
GEO Species
UMBEL
DBLP Berlin
Reactome
UniParc
Taxonomy
Open Cyc
Yago
LinkedCT
Drug Bank
PROSITE
Daily Med
GeneID
Pub Chem
KEGG
UniProt
Homolo Gene
Pfam
ProDom
Disea-some
CAS
ChEBI
Gene Ontology
Symbol
OMIM
Inter Pro
UniSTS
HGNC
MGI
PDB
PubMed

Sept 2010



DBpedia
Freebase
YAGO
Geo Names
UniProt
Drug Bank
PubMed
GeneID
MusicBrainz (DBTune)
BBC Music
BBC Programmes
legislation .gov.uk
UK Post-codes
Linked Data for Intervals
TWC LOGD
GovTrack
rdfabout US Census
Linked GeoData
ACM
DBLP (L3S)
CiteSeer
RAE2001
ePrints
dotAC
KISTI
CORDIS
NSF
IBM
VIAF
LCSH
The Open Library
Linked MDB
flickr wrappr
OpenCyc
WordNet (VUA)
KEGG Pathway
As of September 2010

Sept 2011



Linked Open Data

Linked Open Data - 5 Star Criteria

- 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)

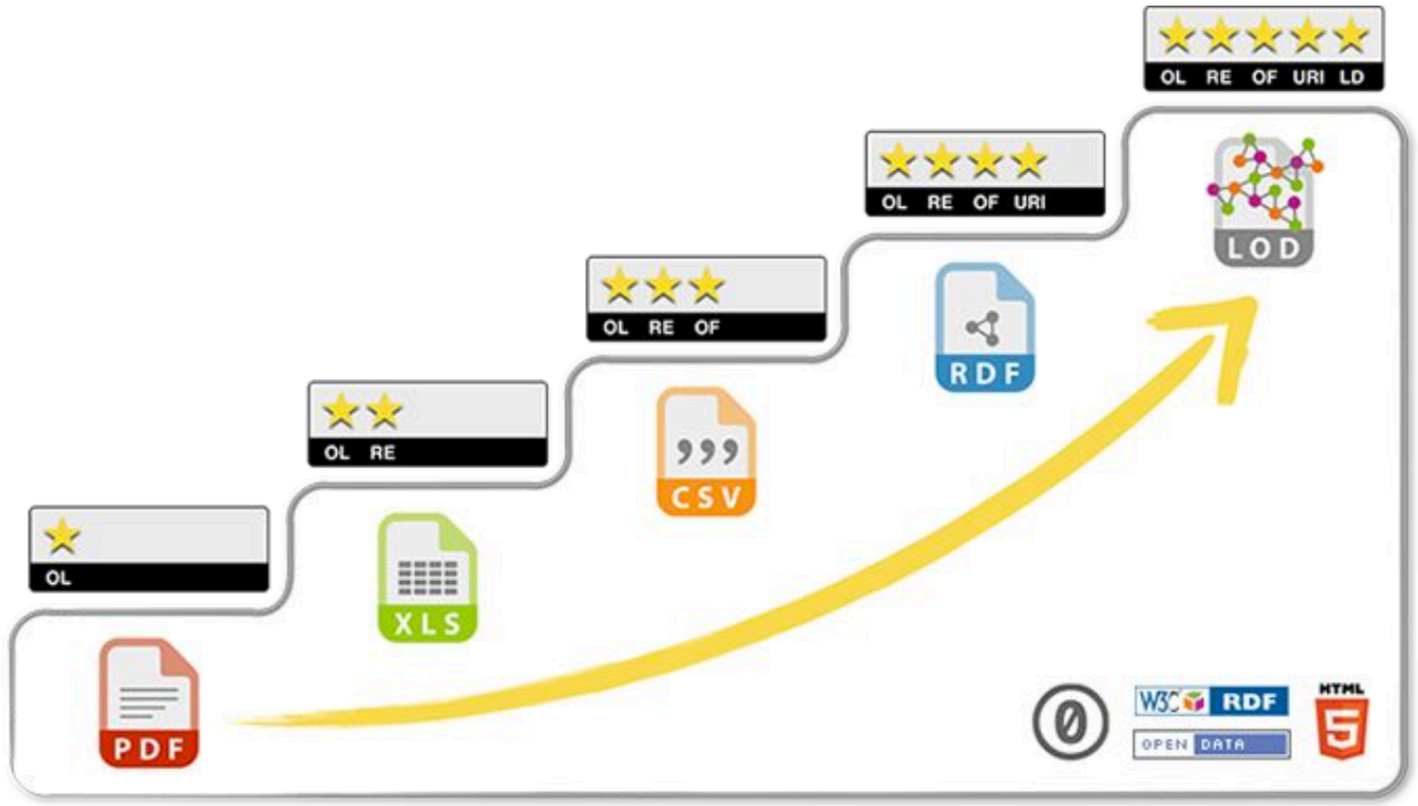
★★★ as (2) plus **non-proprietary format** (e.g. CSV instead of excel)

★★★★ 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

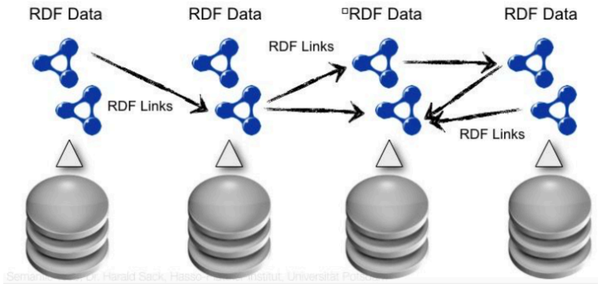
Linked Open Data

- http://5stardata.info/en



Advantages of Linked Open Data vs. APIs

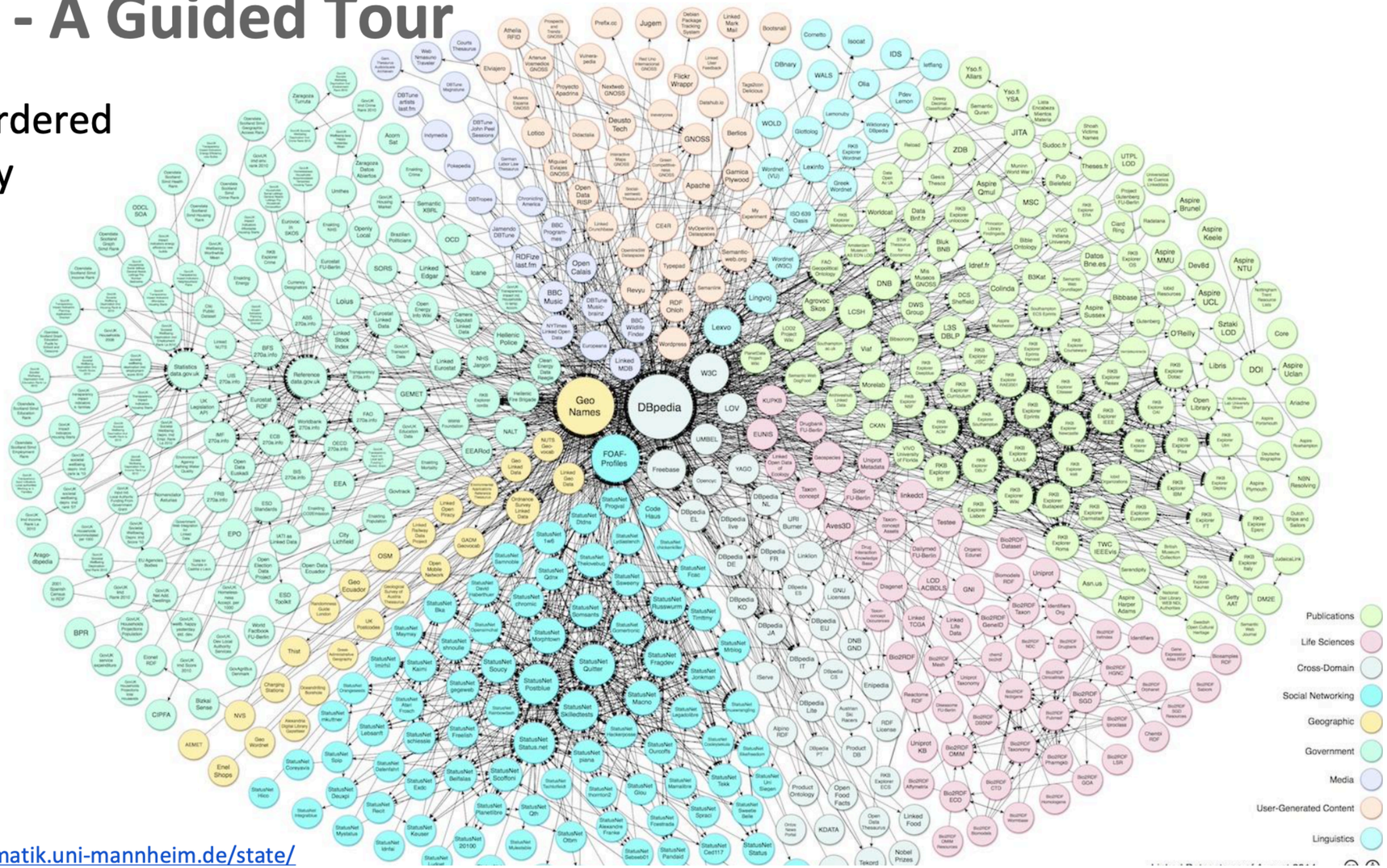
- Simple and generic API for various heterogeneous data sources enables simple reuse and data sharing among applications
- RDF Data model guarantees (simple) extensibility
- Transport via http, standard Port 80, prevents firewall adaption
- Ontologies enable meaningful connections between data sources
- Reasoning over Linked Data enables to generate new knowledge, i.e. inference from implicit to explicit knowledge



LOD - Tour

Linked Data - A Guided Tour

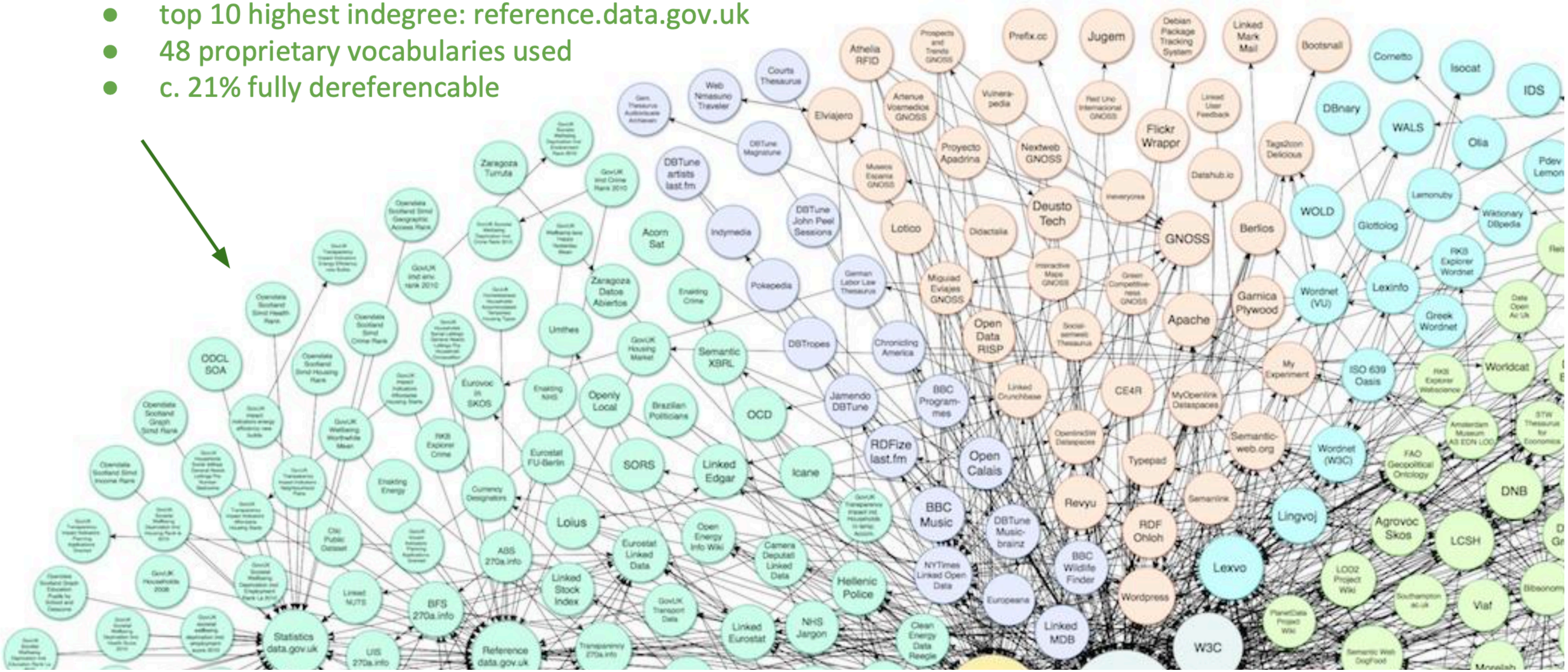
- Datasets ordered by category



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

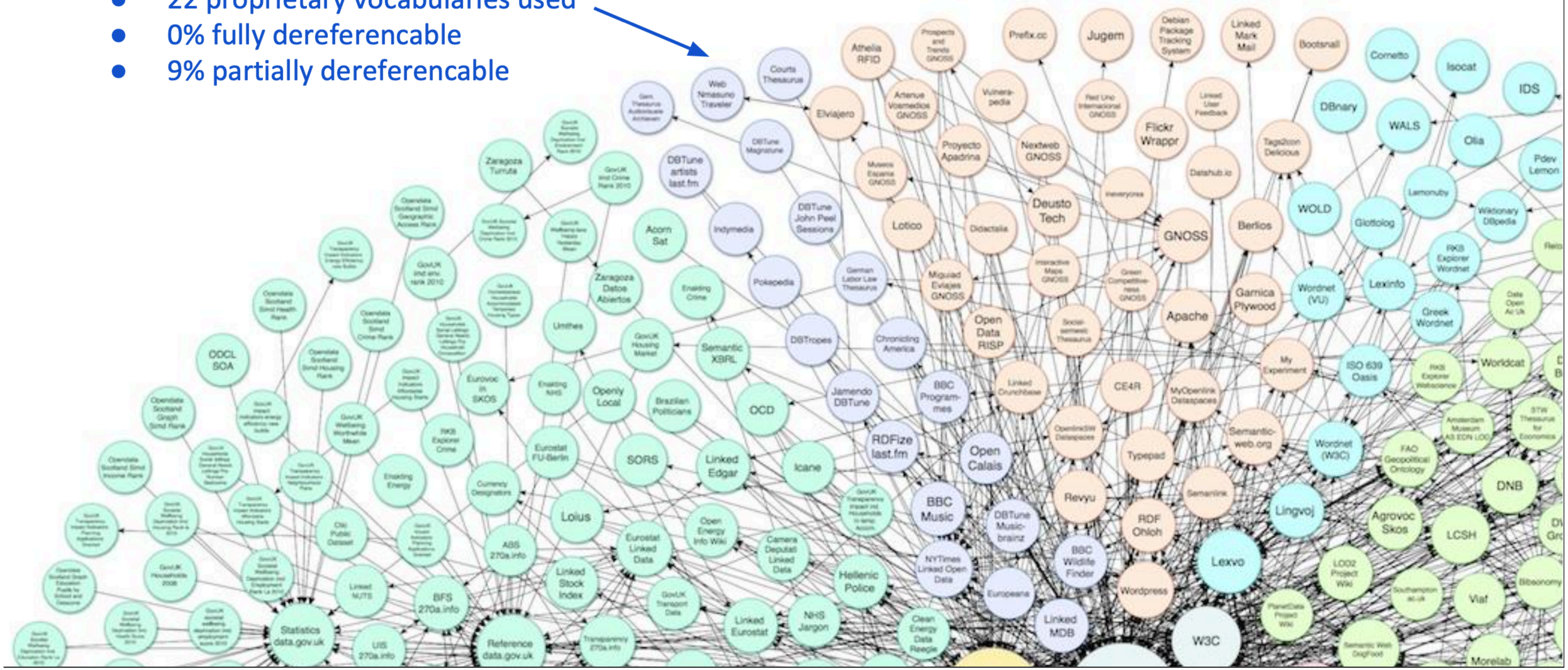
Government

- 183 datasets
- top 10 highest indegree: reference.data.gov.uk
- 48 proprietary vocabularies used
- c. 21% fully dereferencable



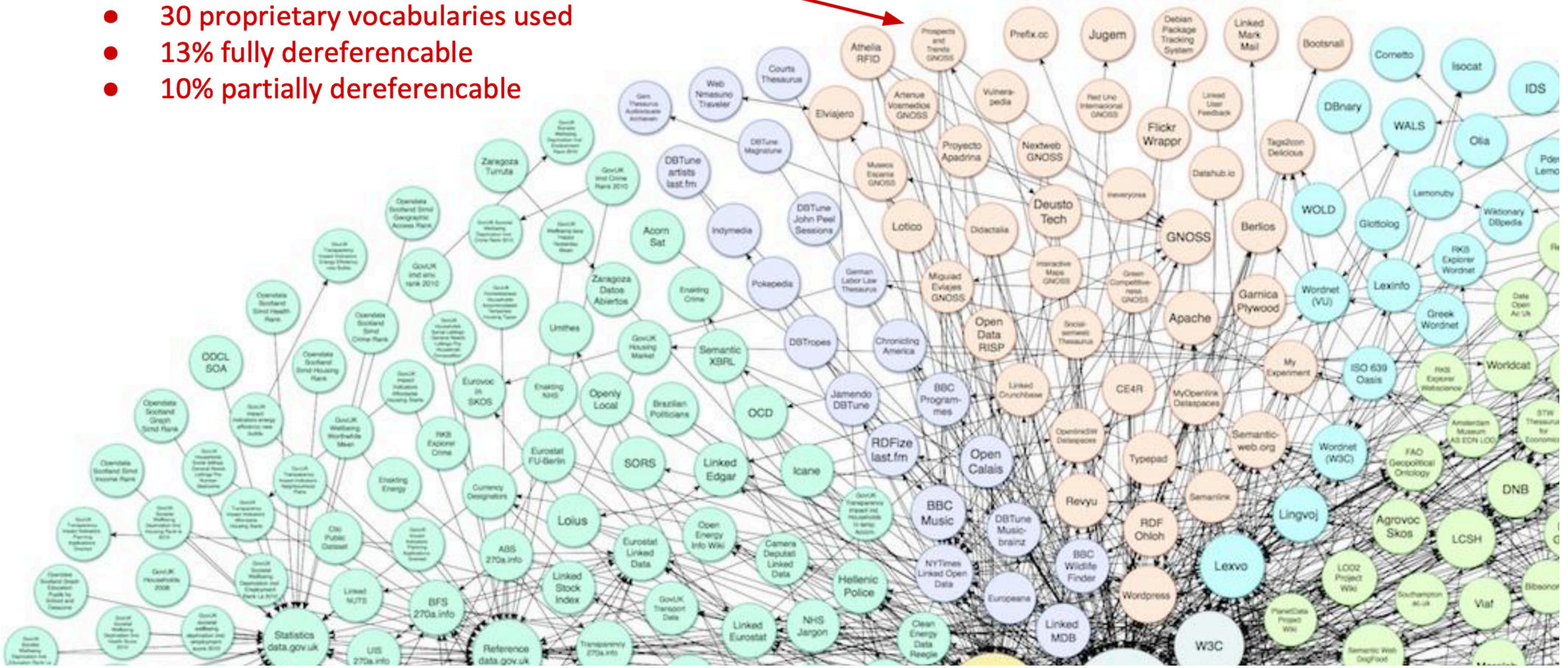
Media

- 22 datasets
- 22 proprietary vocabularies used
- 0% fully dereferencable
- 9% partially dereferencable



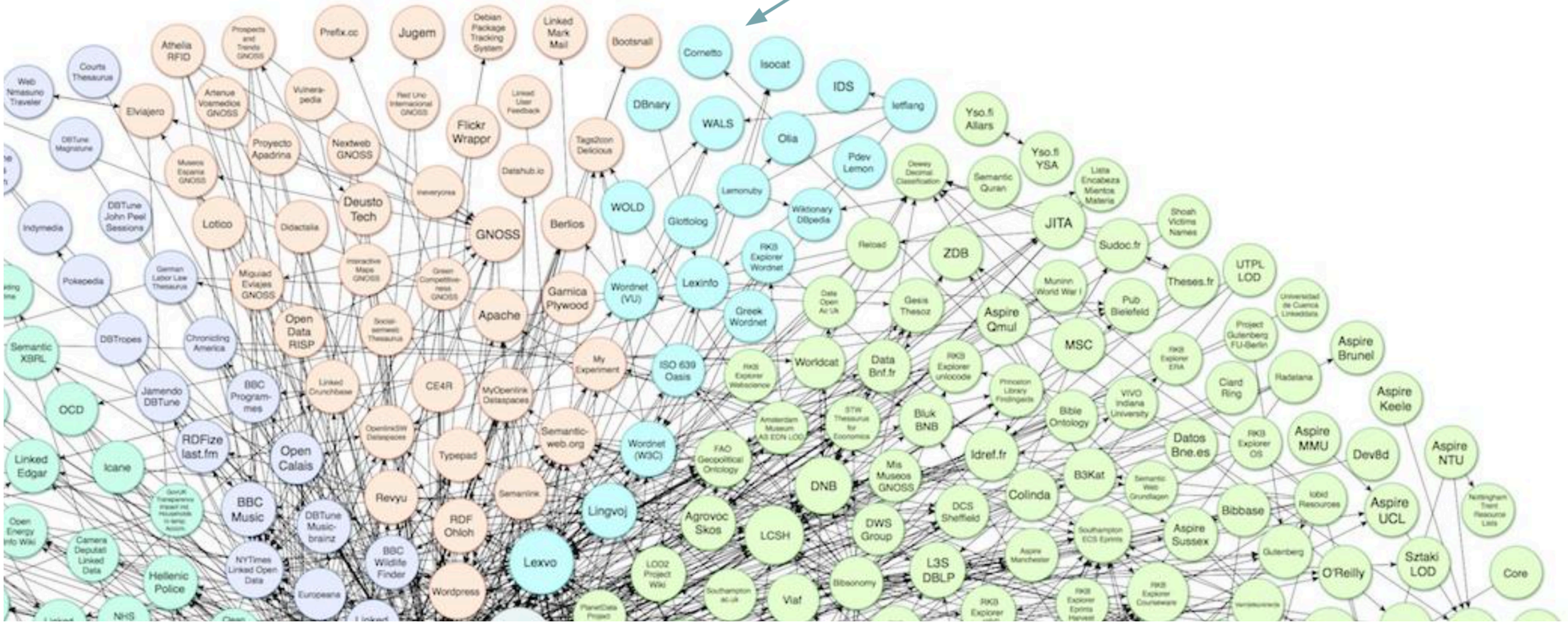
User Generated Content

- 48 datasets
- top 10 highest outdegree: semanticweb.org
- 30 proprietary vocabularies used
- 13% fully dereferencable
- 10% partially dereferencable



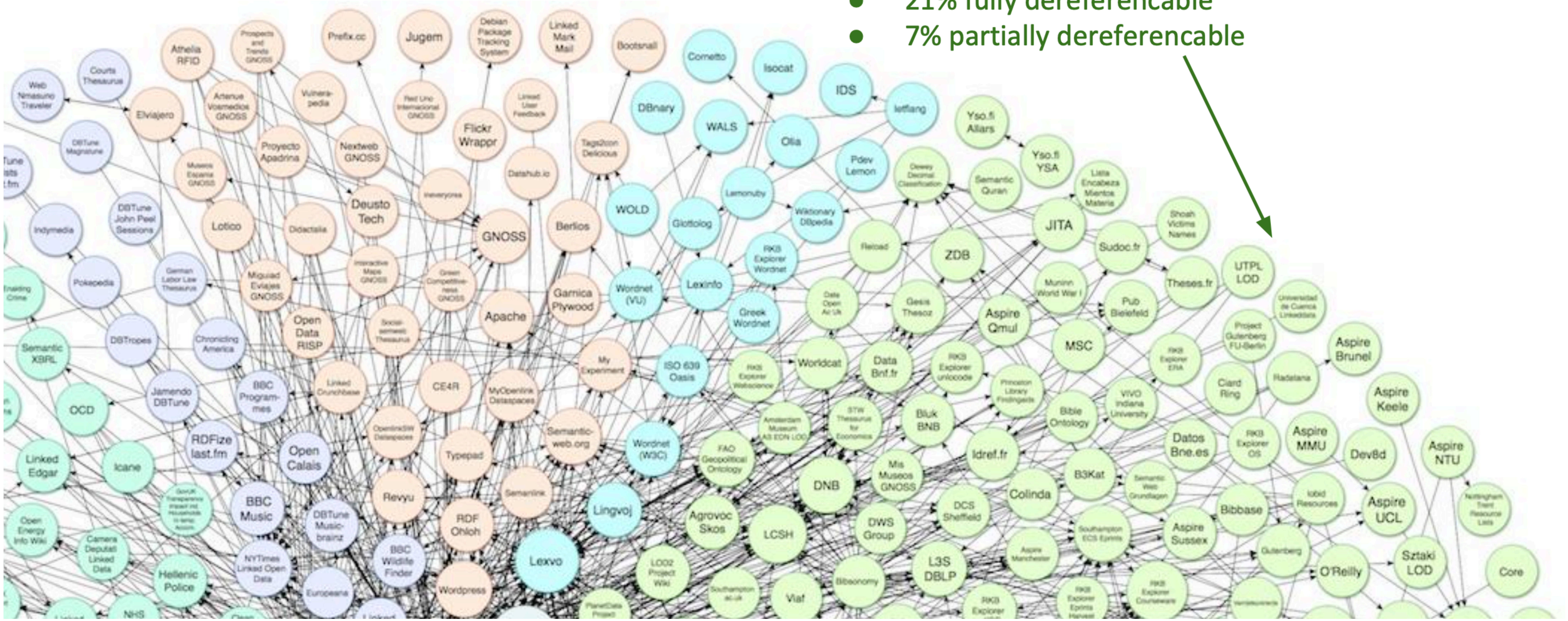
Linguistics

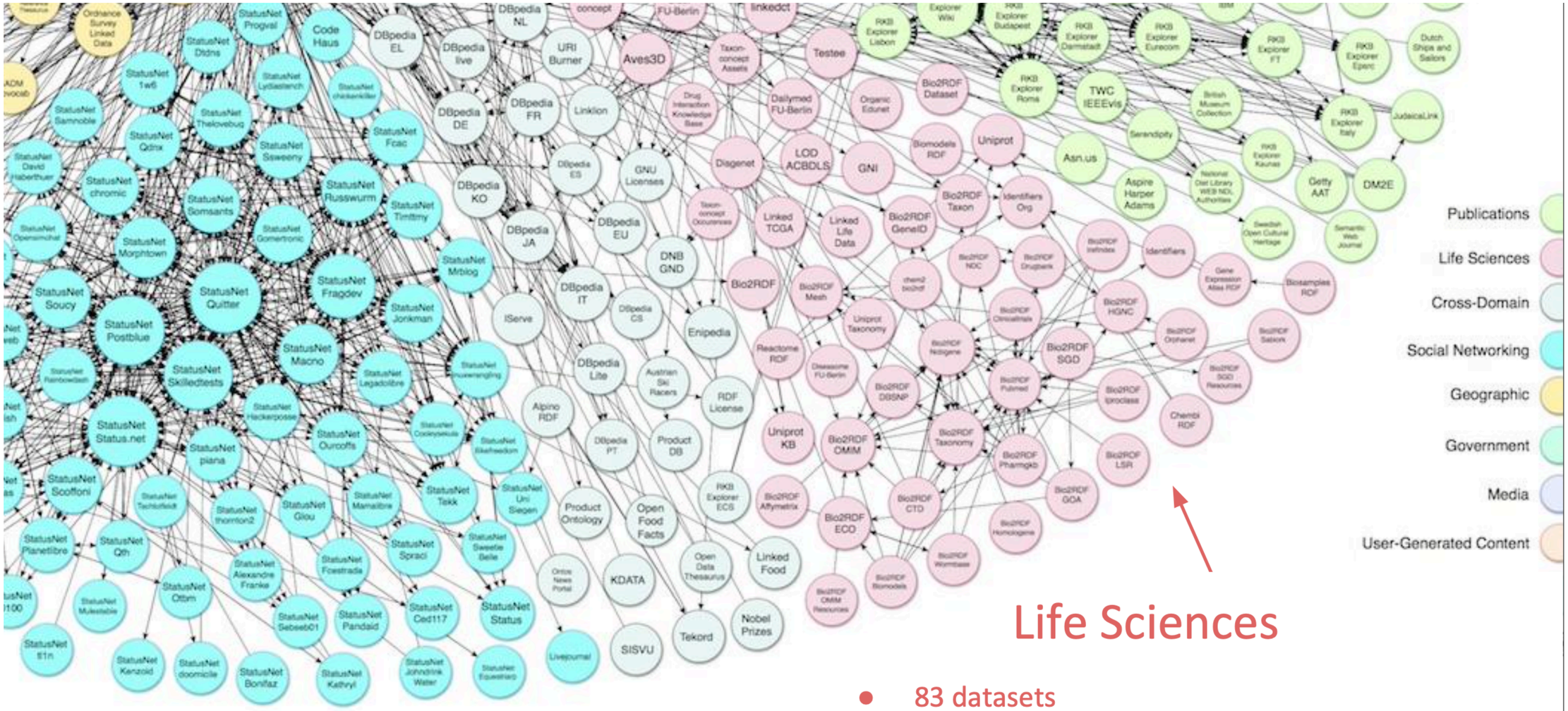
- no statistics available so far



Bibliographic Data

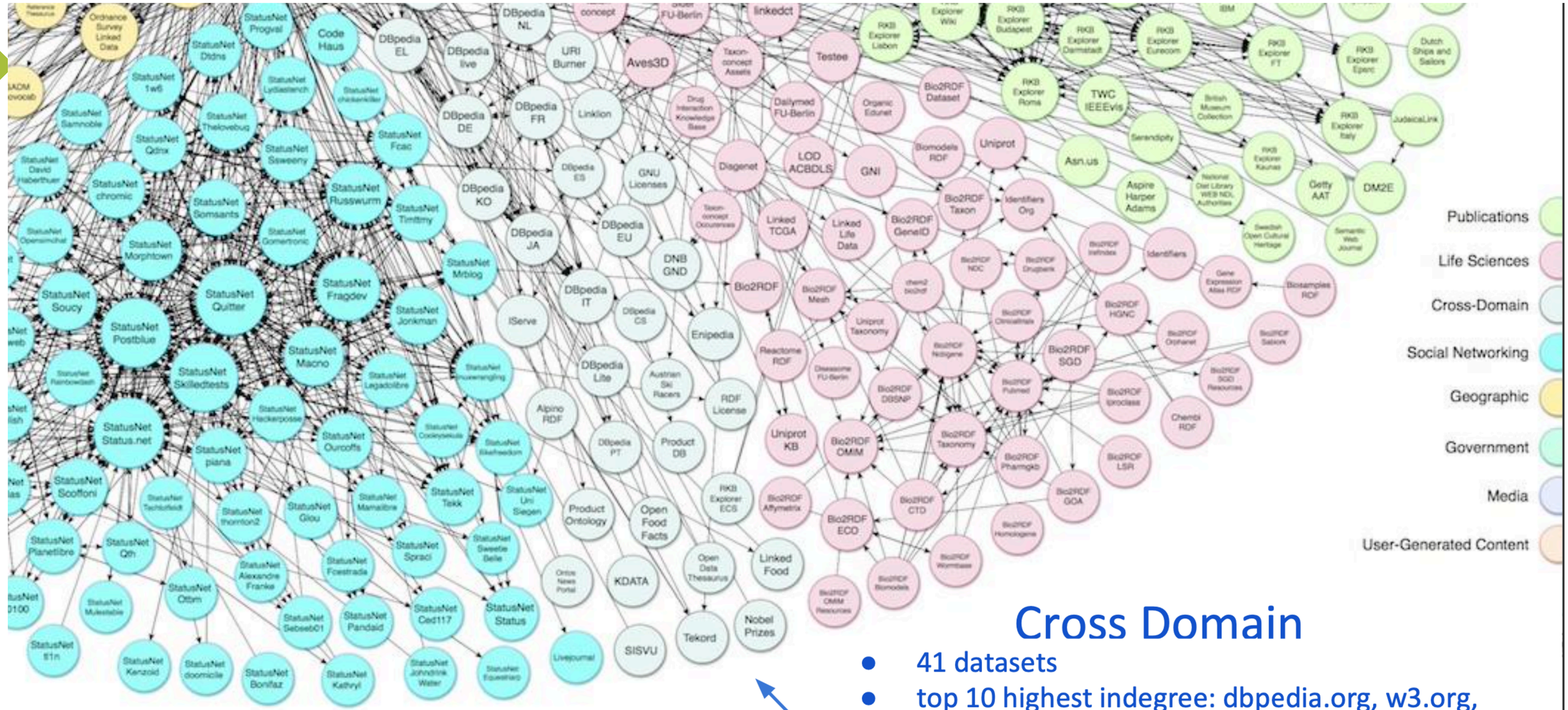
- 96 datasets
- top 10 highest indegree: data.semanticweb.org
- top 10 highest outdegree: bibsonomy.org
- 58 proprietary vocabularies used
- 21% fully dereferencable
- 7% partially dereferencable





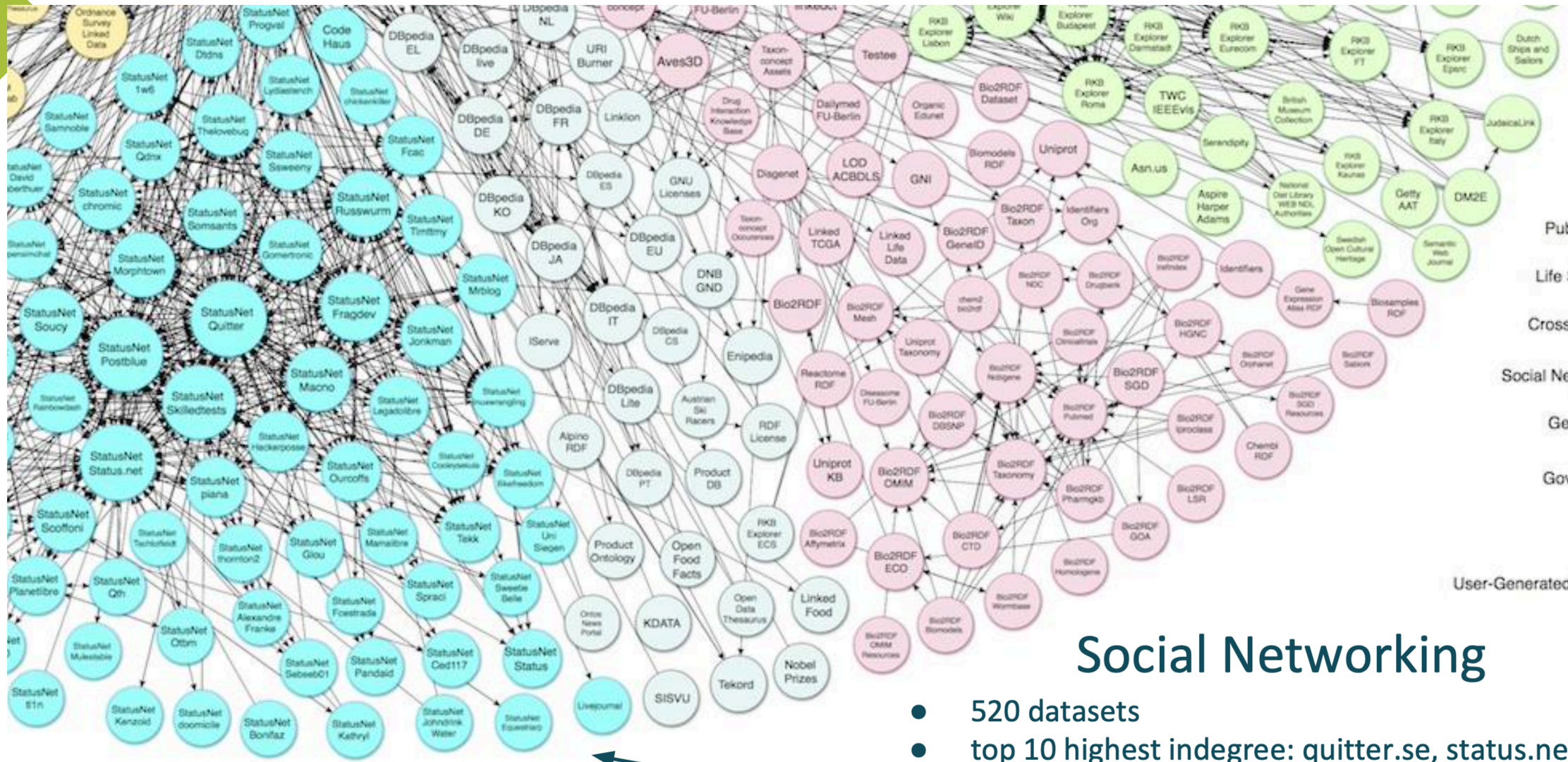
Life Sciences

- 83 datasets
- 35 proprietary vocabularies used
- 28% fully dereferencable
- 6% partially dereferencable



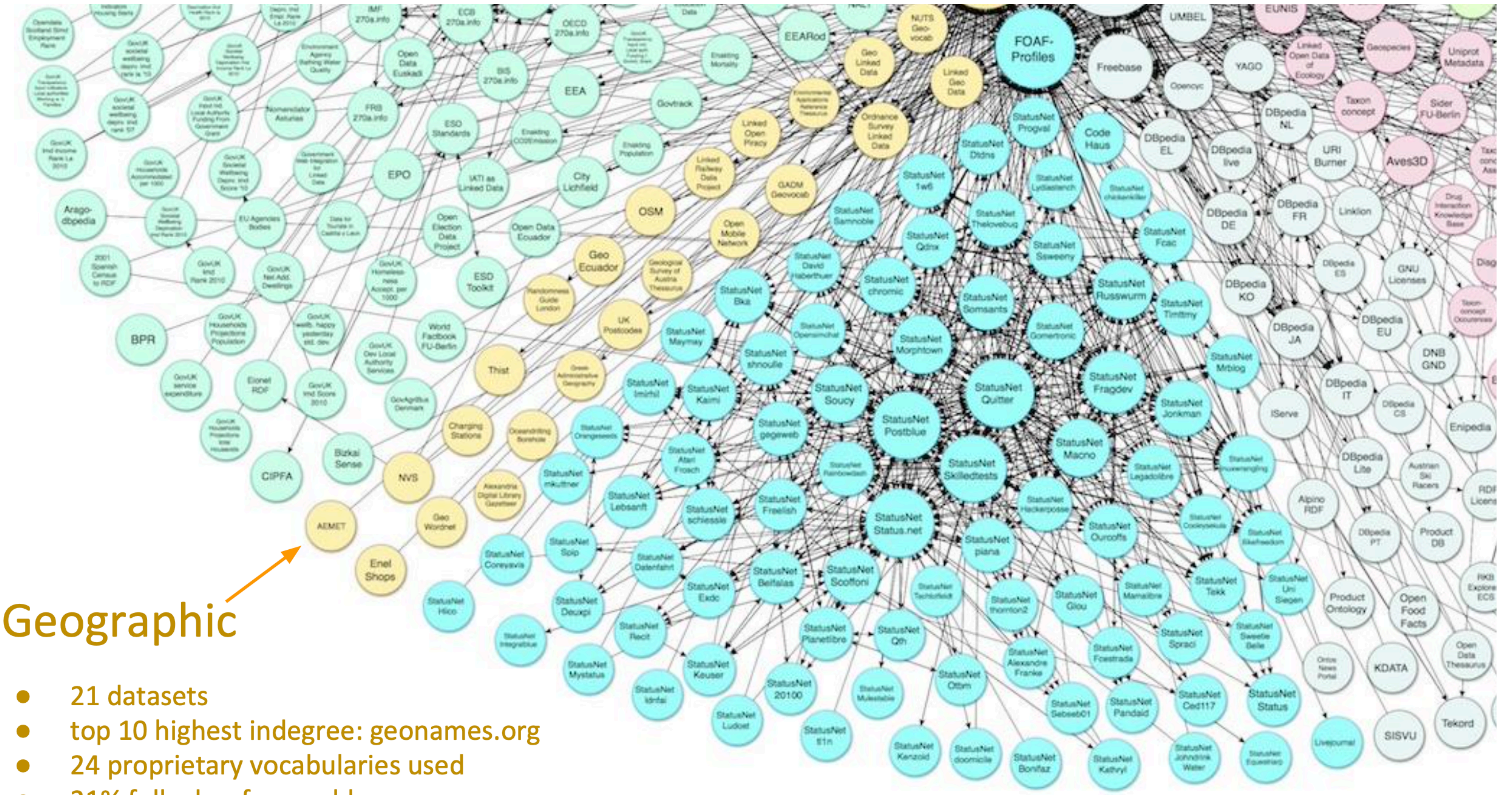
Cross Domain

- 41 datasets
- top 10 highest indegree: dbpedia.org, w3.org, lexvo.org
- 55 proprietary vocabularies used
- 27% fully dereferencable
- 11% partially dereferencable



Social Networking

- 520 datasets
- top 10 highest indegree: quitter.se, status.net, …
- top 10 highest outdegree: deri.org, harth.org,…
- 128 proprietary vocabularies used
- 16% fully dereferencable
- 6% partially dereferencable



Geographic

- 21 datasets
- top 10 highest indegree: geonames.org
- 24 proprietary vocabularies used
- 21% fully dereferencable
- 4% partially dereferencable

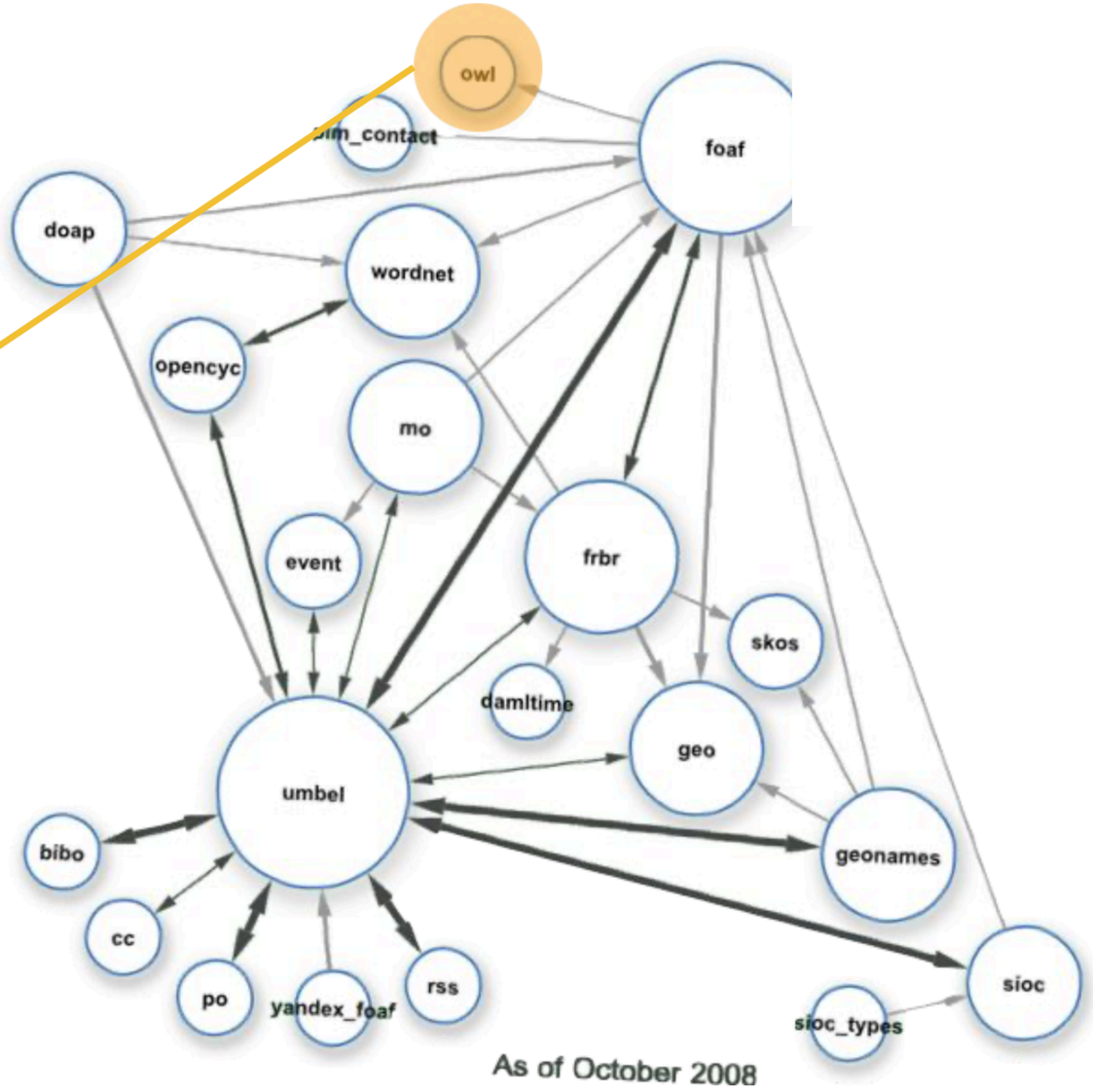
Linked Data Vocabularies

Linked Data Ontologies

Linked Data Ontologies

- Ontologies hold the Linked Data Cloud together

- **OWL**
  `owl:sameAs` connects identical individuals
  `owl:equivalentClass` connects equivalent classes

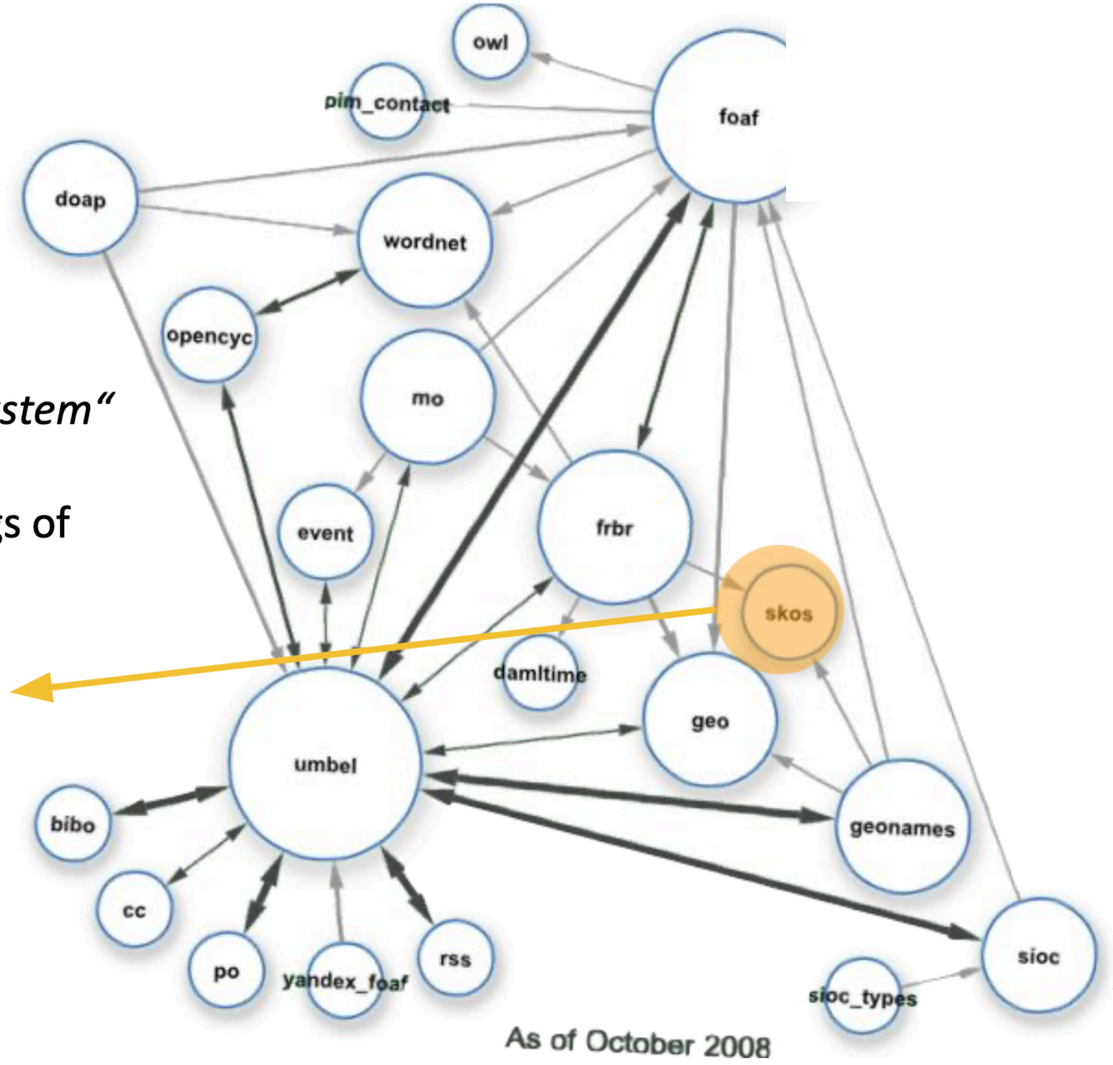


Linked Data Ontologies

Linked Data Ontologies

- Ontologies hold the Linked Data Cloud together

- **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`

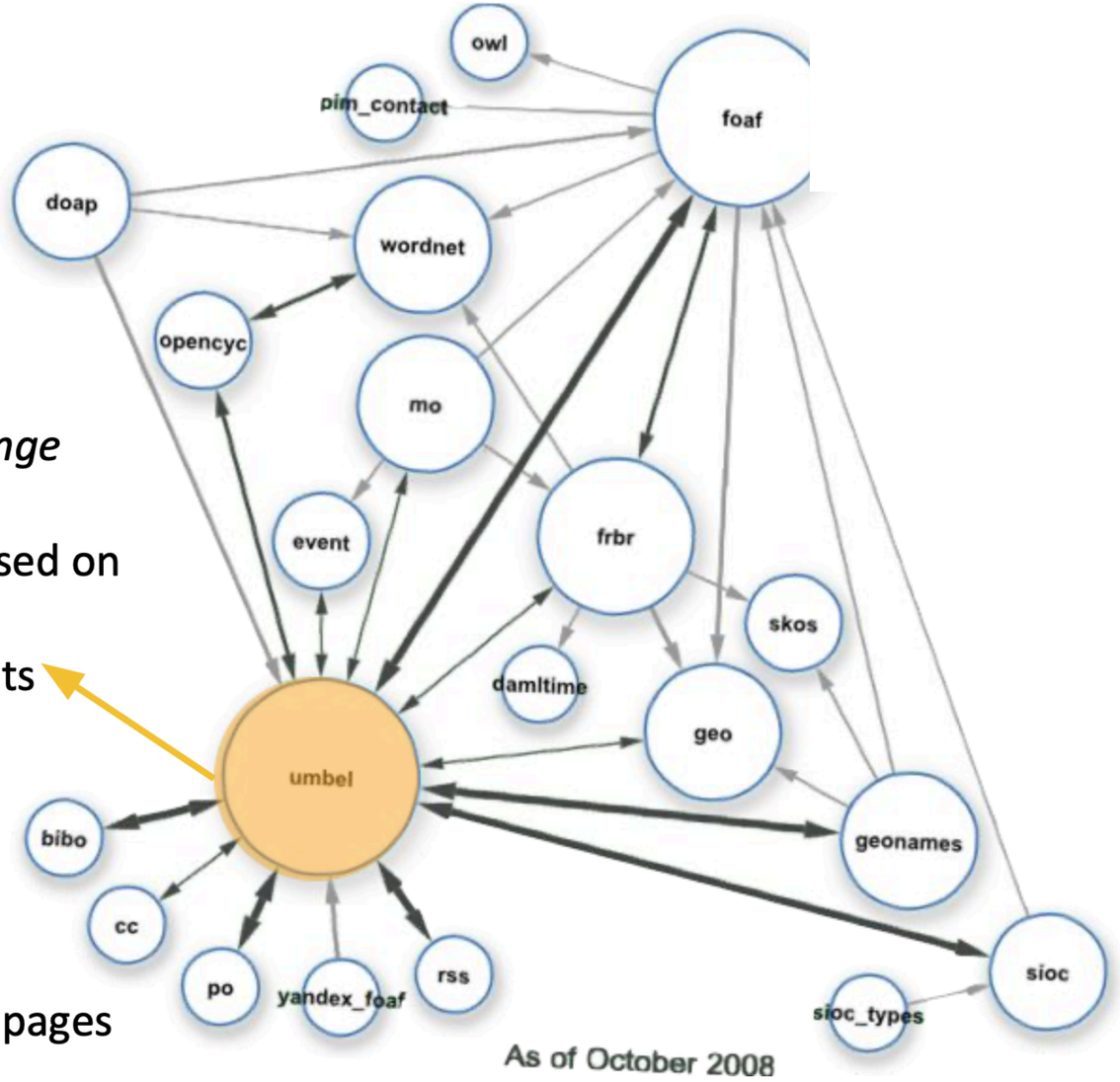


Linked Data Ontologies

Linked Data Ontologies

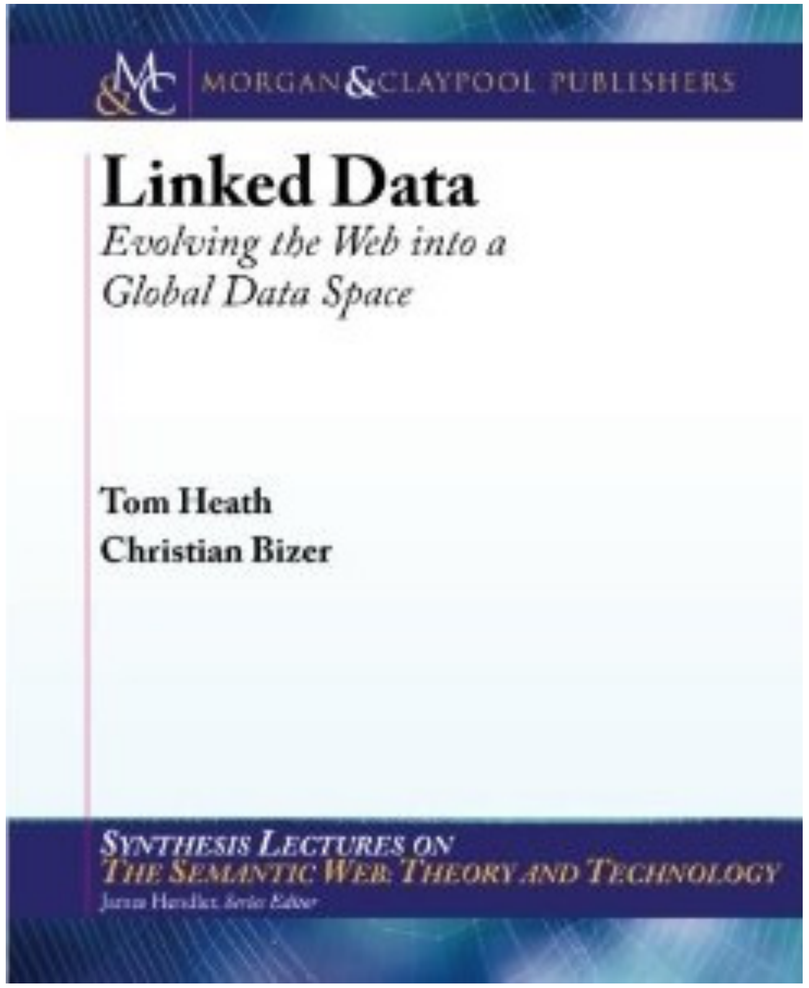
- Ontologies hold the Linked Data Cloud together

- **umbel**
  - *„Upper Mapping and Binding Exchange Layer“*
  - Subset of OpenCycas 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



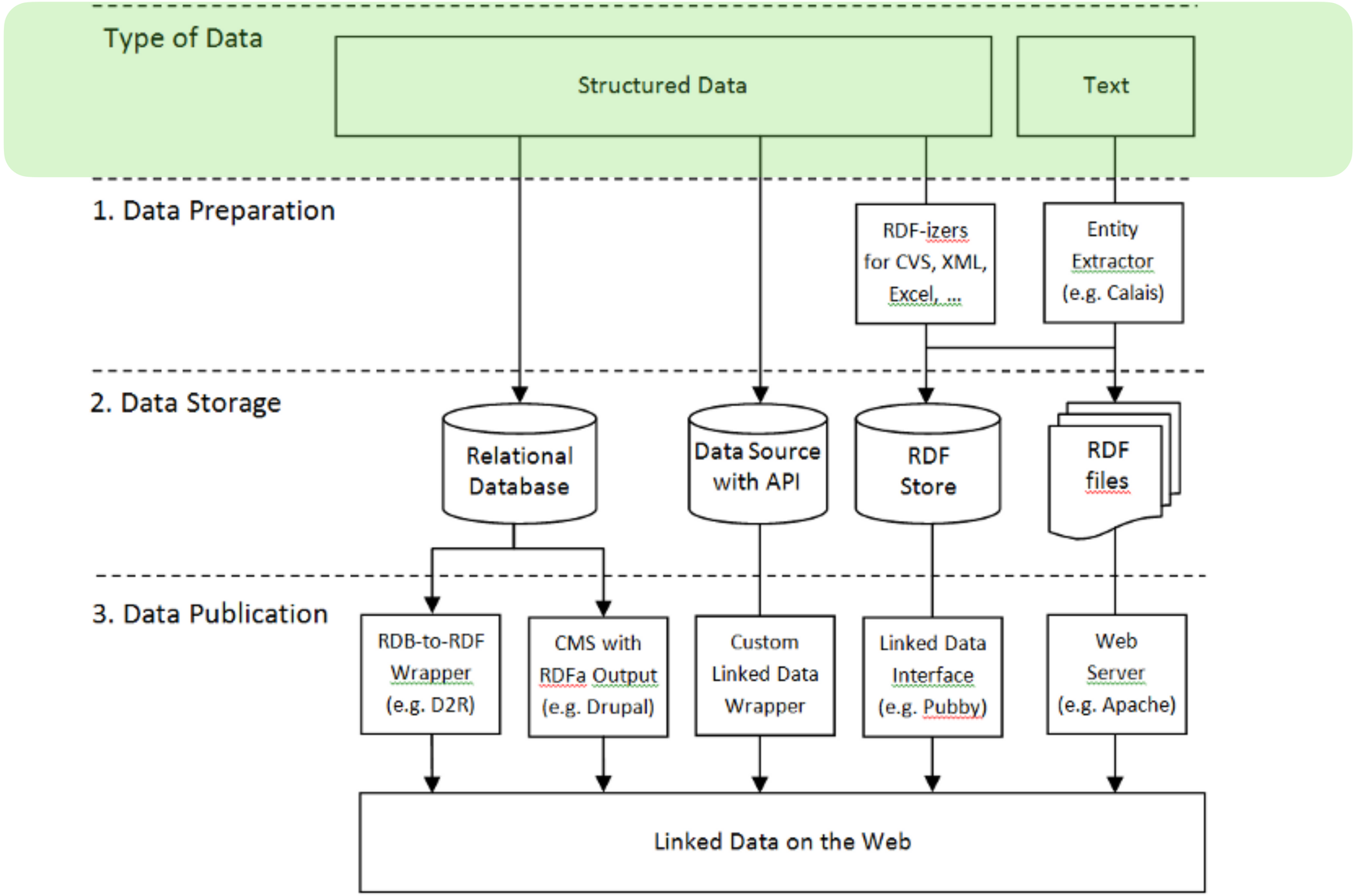
How To Publish Linked Data on the Web

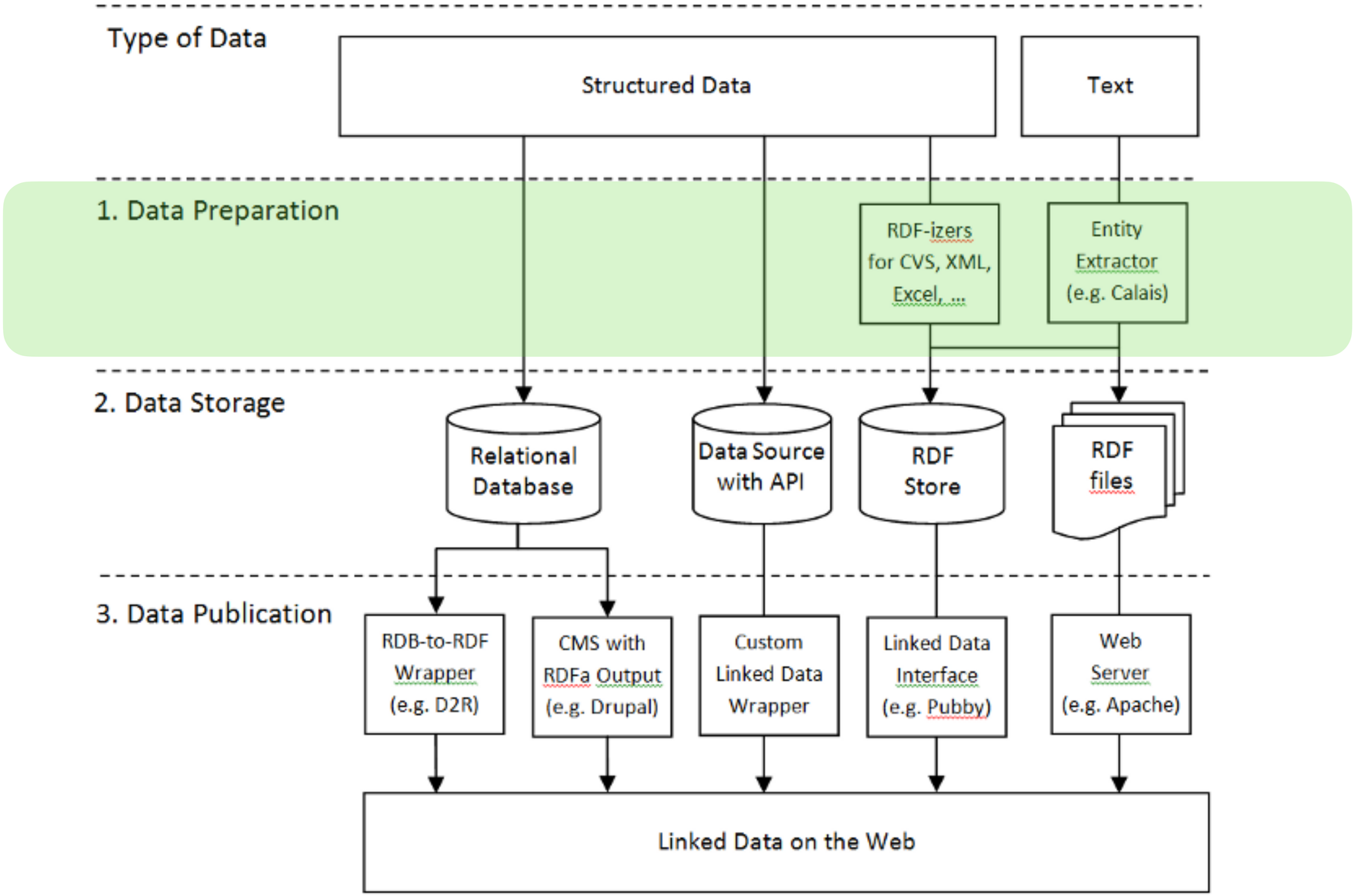
Must Read!

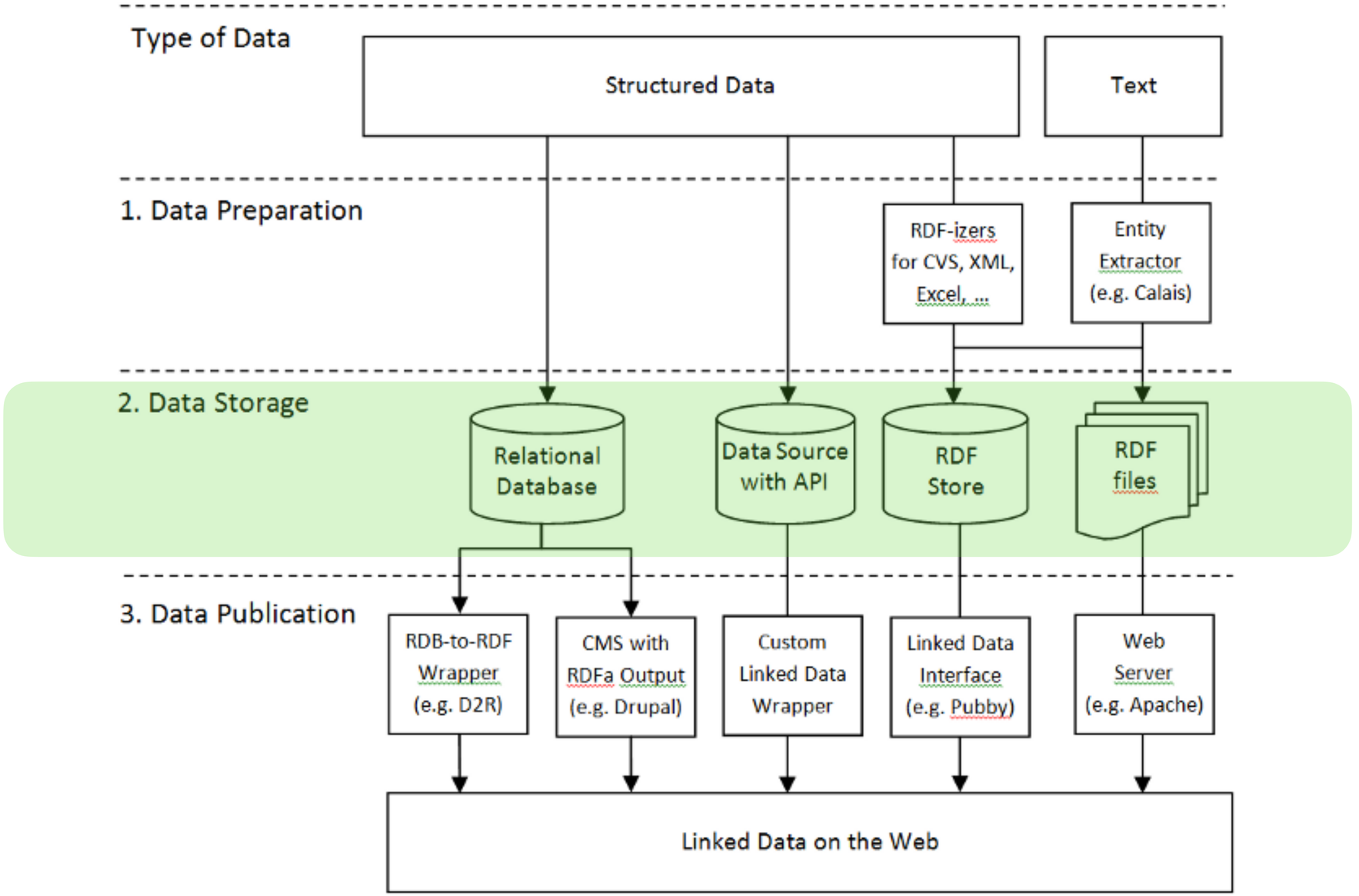


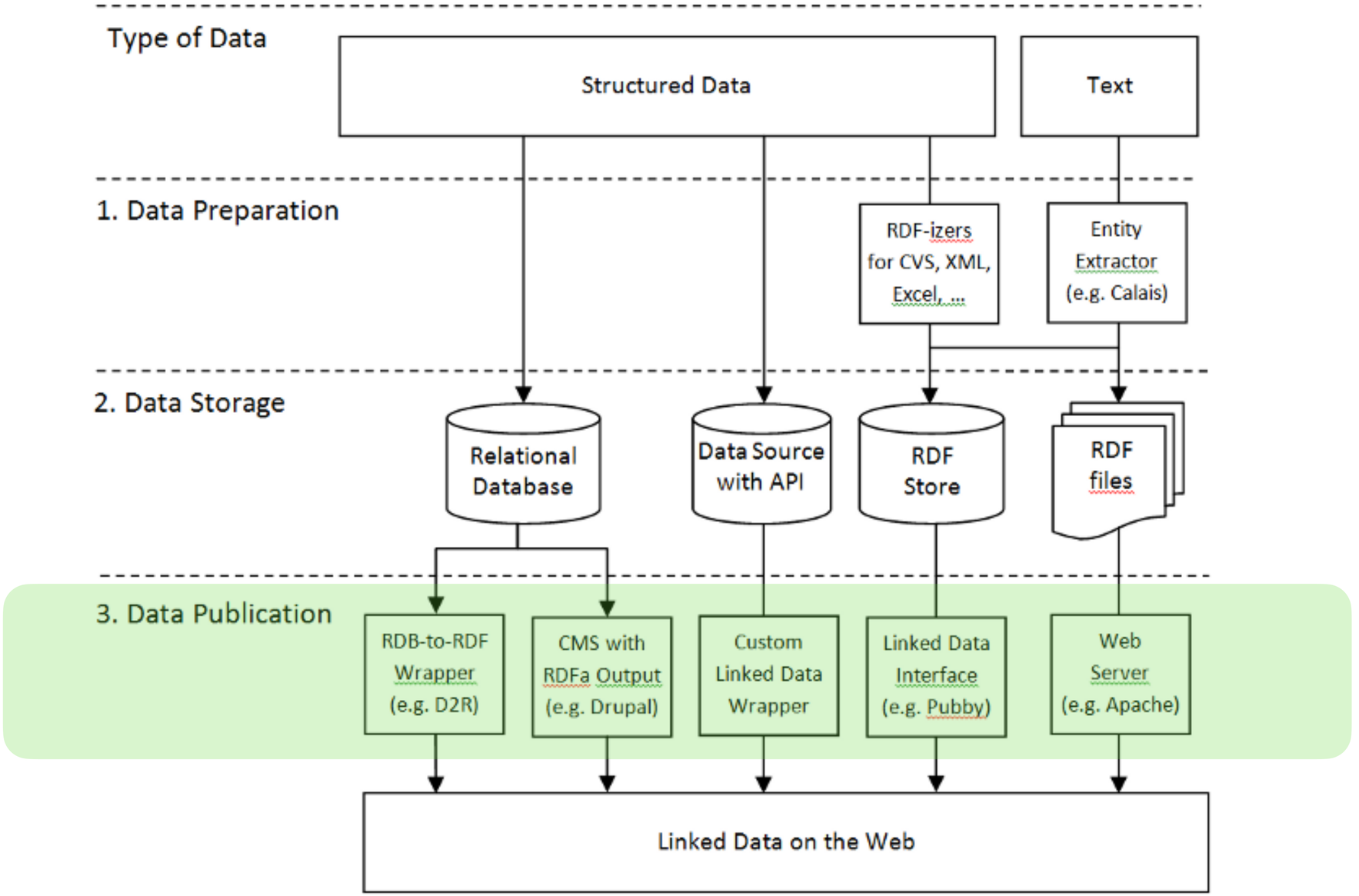
http://linkeddatabook.com/editions/1.0/

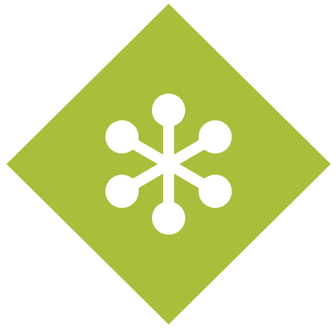
Linked Data Publishing Patterns




Type of Data
Structured Data
Text
1. Data Preparation
RDF-izers for CVS, XML, Excel, ...
Entity Extractor (e.g. Calais)
2. Data Storage
Relational Database
Data Source with API
RDF Store
RDF files
3. Data Publication
RDB-to-RDF Wrapper (e.g. D2R)
CMS with RDFa Output (e.g. Drupal)
Custom Linked Data Wrapper
Linked Data Interface (e.g. Pubby)
Web Server (e.g. Apache)
Linked Data on the Web


Type of Data
Structured Data
Text
1. Data Preparation
RDF-izers for CVS, XML, Excel, ...
Entity Extractor (e.g. Calais)
2. Data Storage
Relational Database
Data Source with API
RDF Store
RDF files
3. Data Publication
RDB-to-RDF Wrapper (e.g. D2R)
CMS with RDFa Output (e.g. Drupal)
Custom Linked Data Wrapper
Linked Data Interface (e.g. Pubby)
Web Server (e.g. Apache)
Linked Data on the Web


Type of Data
Structured Data
Text
1. Data Preparation
RDF-izers for CVS, XML, Excel, ...
Entity Extractor (e.g. Calais)
2. Data Storage
Relational Database
Data Source with API
RDF Store
RDF files
3. Data Publication
RDB-to-RDF Wrapper (e.g. D2R)
CMS with RDFa Output (e.g. Drupal)
Custom Linked Data Wrapper
Linked Data Interface (e.g. Pubby)
Web Server (e.g. Apache)
Linked Data on the Web

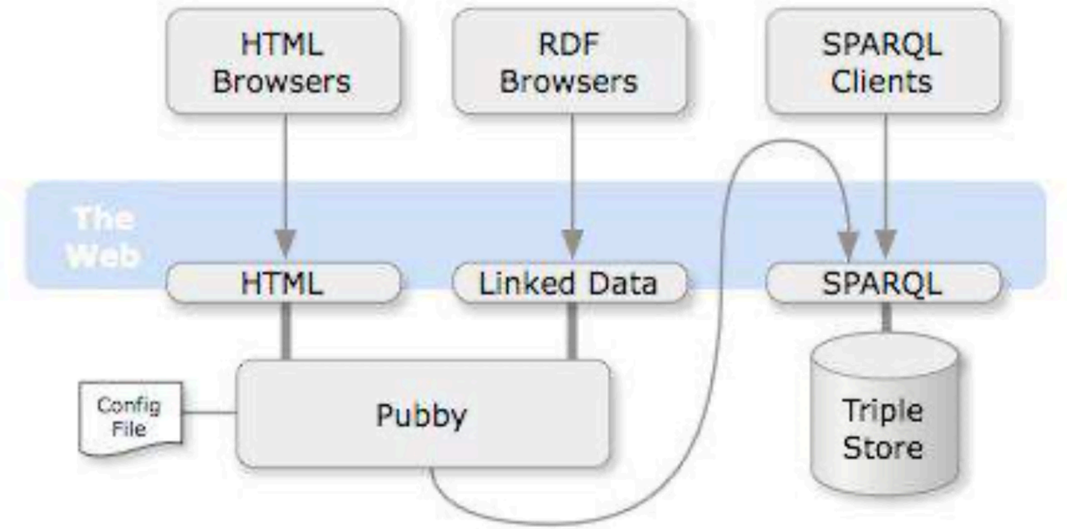


Native Publication

- **Native publication**
  - Triple Stores (SPARQL Endpoints)
    - *OpenLink Virtuosu, Sesame, Fuseki,….*
- SPARQL endpoints are **RESTful Web Services**
- Results available as
  - 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*
      - e.g. Accept: application/sparql-results+json (or via parameters of the SPARQL query)

Native Publication

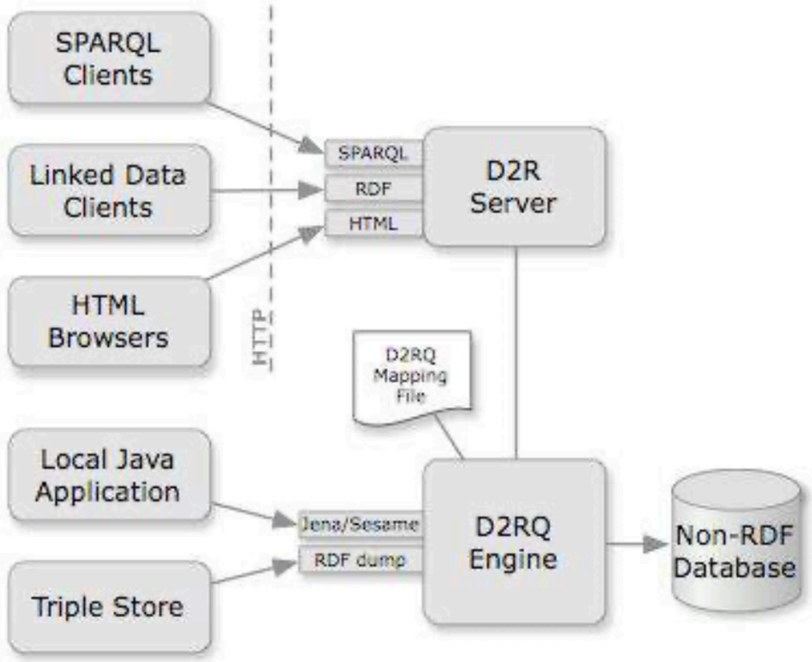
- **Native publication**
  - Triple Stores
    - OpenLink Virtuoso, Sesame, Fuseki, …
  - Linked Data Endpoints
    - Pubby, ...



http://wifo5-03.informatik.uni-mannheim.de/pubby

Native Publication

- **Native publication**
  - Triple Stores
    - OpenLink Virtuoso, Sesame, Fuseki, …
  - Linked Data Endpoints
    - Pubby, ...
  - D2R Servers
    - D2RQ, RDF2RDF, ...



http://d2rq.org

Linked Data Publishing - Best Practices

B. Hyland; D. Wood. The Joy of Data - Cookbook for Publishing Linked Government Data on the Web

Boris Villazón-Terrazas; et al.. Methodological Guidelines for Publishing Government Linked Data

How To Publish

Hands on Example

GAME OF THRONES