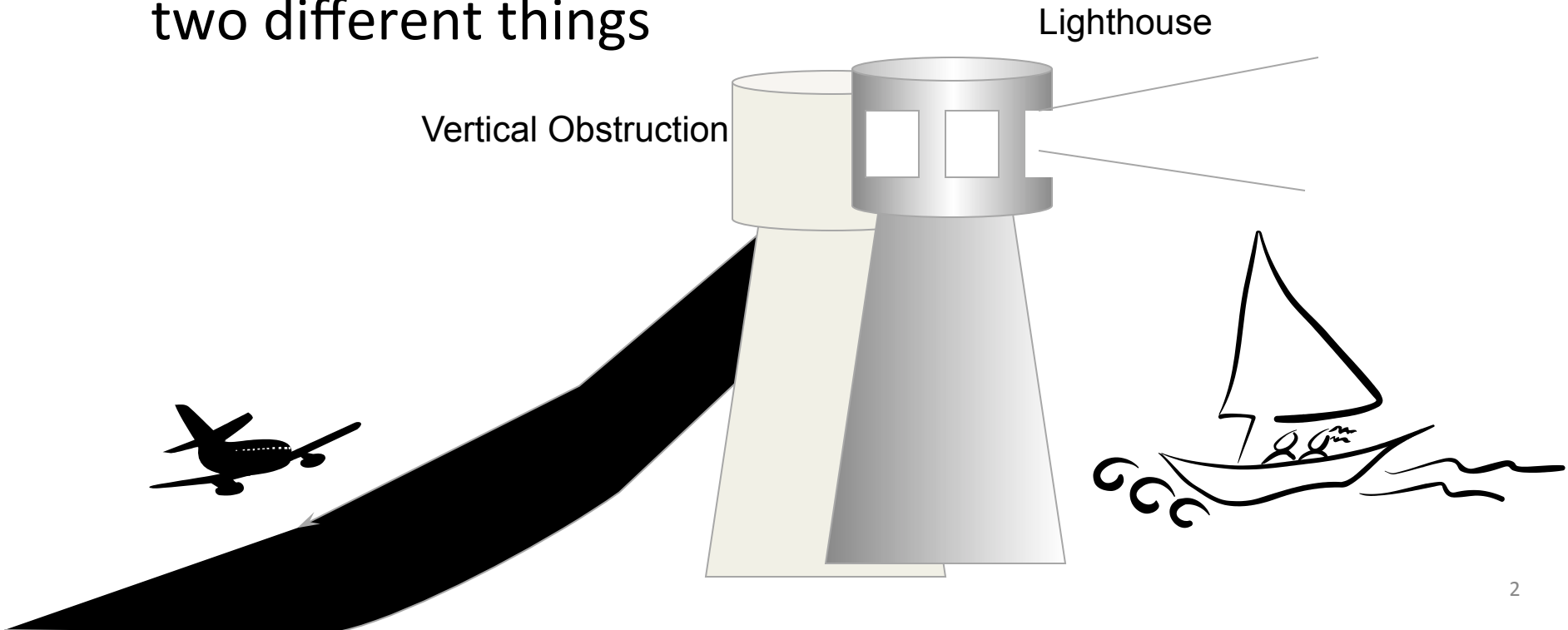


Problems of heterogeneity



- Semantics: two names for the same thing
- Semiotics: one name for two different things
- Schizophrenia (cognitive dissonance): two names for two different things



Semantic quandaries

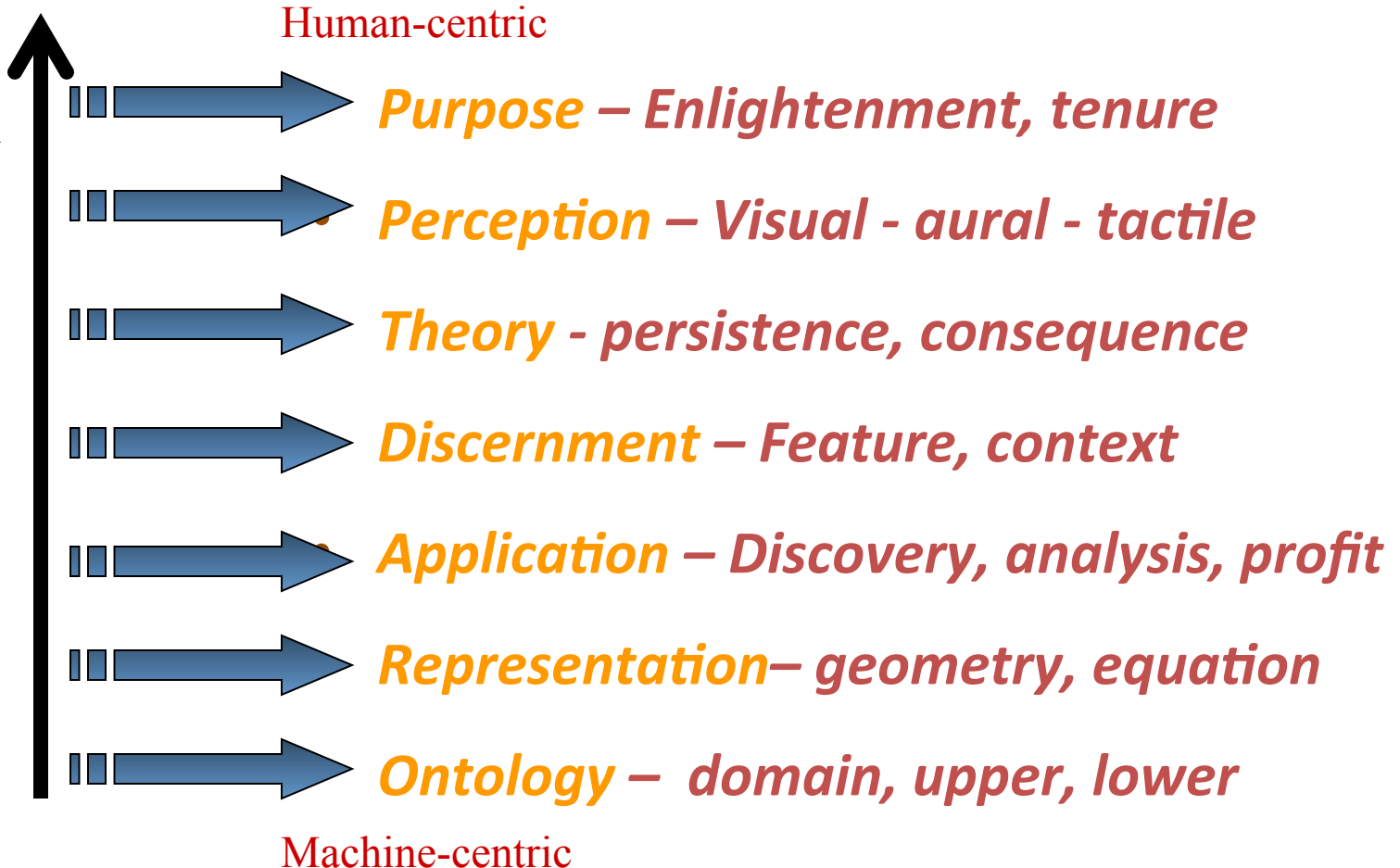


- Two co-located shapes -> “semantic heterogeneity”
- Agree “obstruction” equals “lighthouse” -> boat crashes into a water tower
- Agree lighthouse “is a” vertical obstruction -> ship ignores light buoy, hits shoal
- Agree vertical obstruction “is a” lighthouse -> plane hits watertower
- We have “cognitive heterogeneity” -> two theories for the same reality

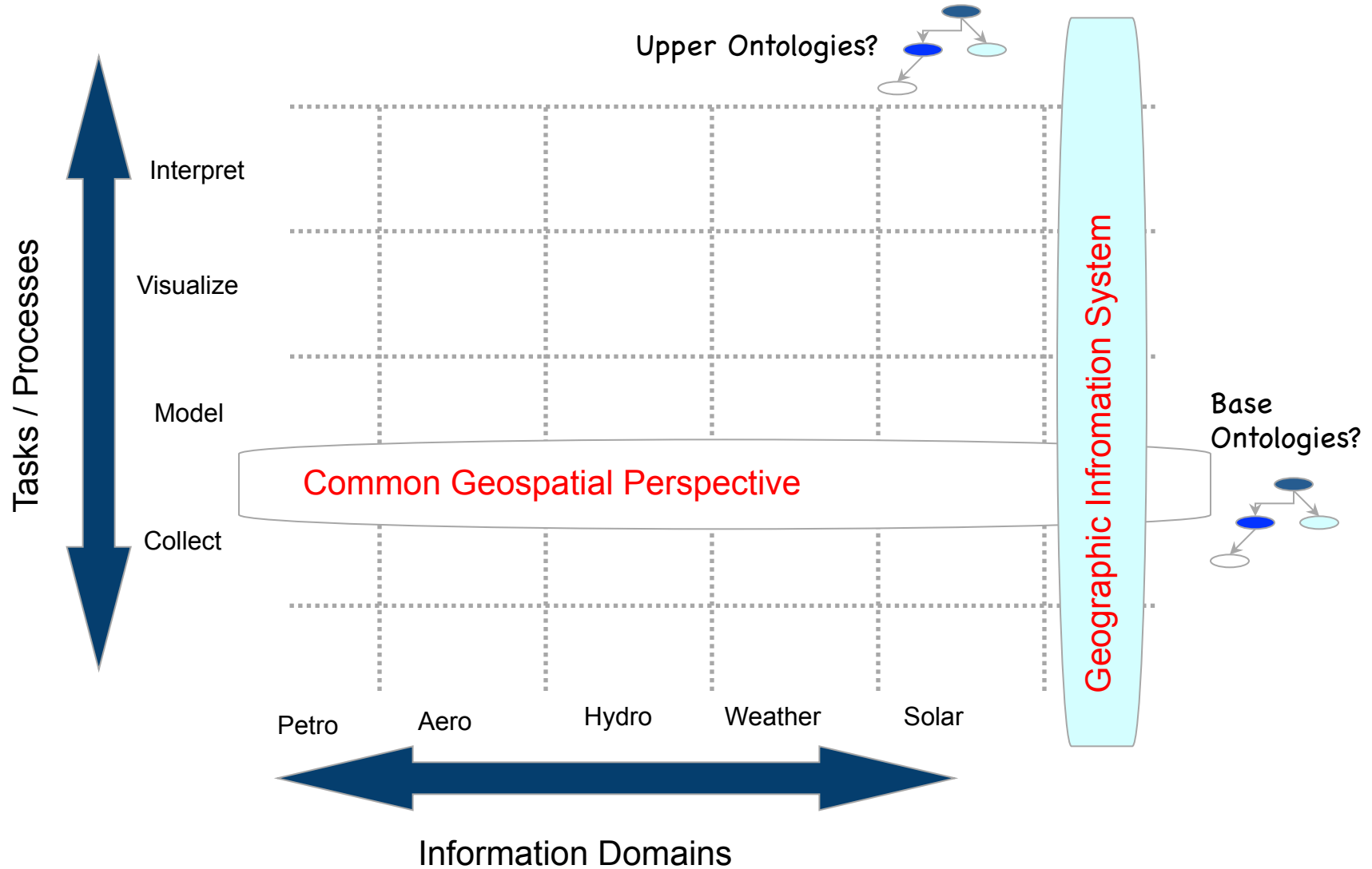
Top of the Interoperability Stack



Increasing /
higher level
interoperability



Which Geospatial Metadata Role?



We have the Data we Asked for



- The **Web of data** becomes a reality by connecting so far isolated islands or data silos such as
 - Enterprise silos (Disparate DBMS Engines & Development Frameworks)
 - Social Media silos (Social Networking Services, Discussion Forums, Blogs, Wikis etc.)
 - Scientific silos (in biology, physics, earth sciences, etc.)
- **Linked Open Data** (LOD) is a way of publishing data on the (Semantic) Web that:
 - Encourages **reuse**
 - Promotes its (real & potential) **inter-connectedness**
 - Enables network effects to **add value to data**



Properties of Linked Data

- ★ make your stuff available on the web (whatever format)
 - ★★ make it available as structured data (e.g. excel instead of image scan of a table)
 - ★★★ non-proprietary format (e.g. csv instead of excel)
 - ★★★★ use URLs to identify things, so that people can point at your stuff
 - ★★★★★ link your data to other people's data to provide context
- **Anyone of Linked Data**
 - Entities are connected by links
 - Creating a global data graph that spans data sources and enables the discovery of new data sources
 - **Data is self-describing**
 - If an application encounters data represented using an unfamiliar vocabulary the application can resolve the URIs that identify vocabulary terms in order to find their RDFS or OWL definition
 - **The Web of Data is open**
 - Applications can discover new data sources at run-time by following links



The mug is from Tim Berners-Lee's design note on Linked Data
<http://www.w3.org/DesignIssues/LinkedData.html>

Linked Open data



- The current Web is a Web of Documents intended for human interpretation. Other applications can not access and process these data unless...

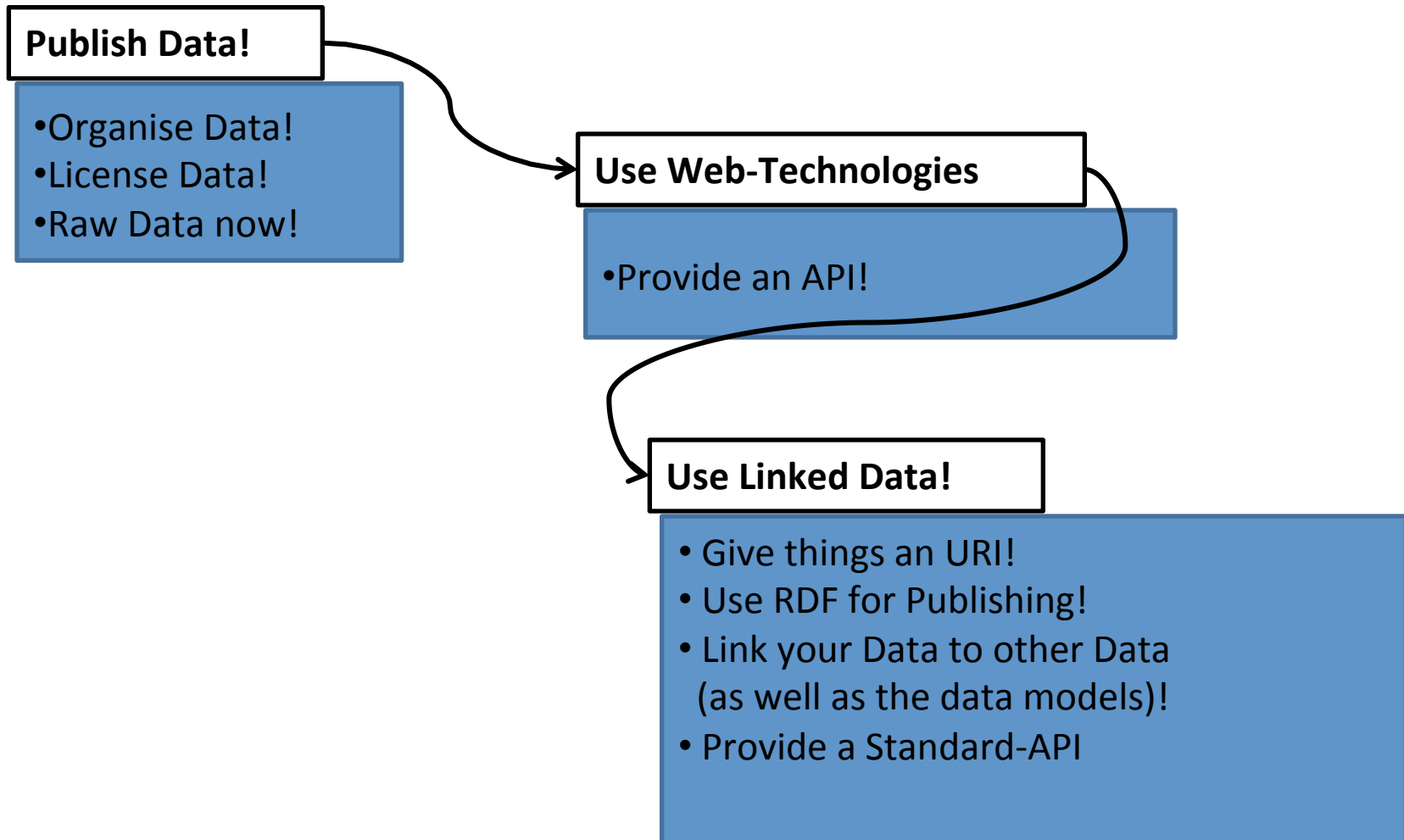
Linked Data Vision

- open the data silos and get rid of repository-centric mindset
- publish data of public interest on the Web in a way that other applications can access and interpret the data using common Web technologies



- A set of four 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

From Open Data to Linked Open Data

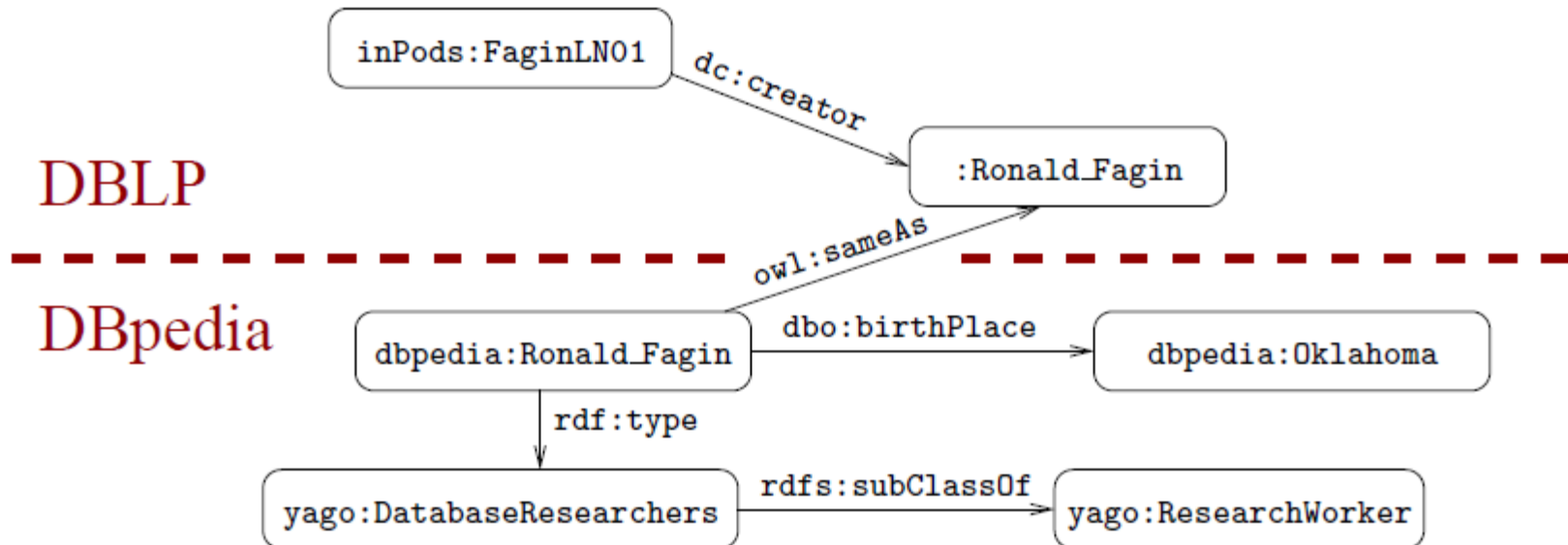


Example

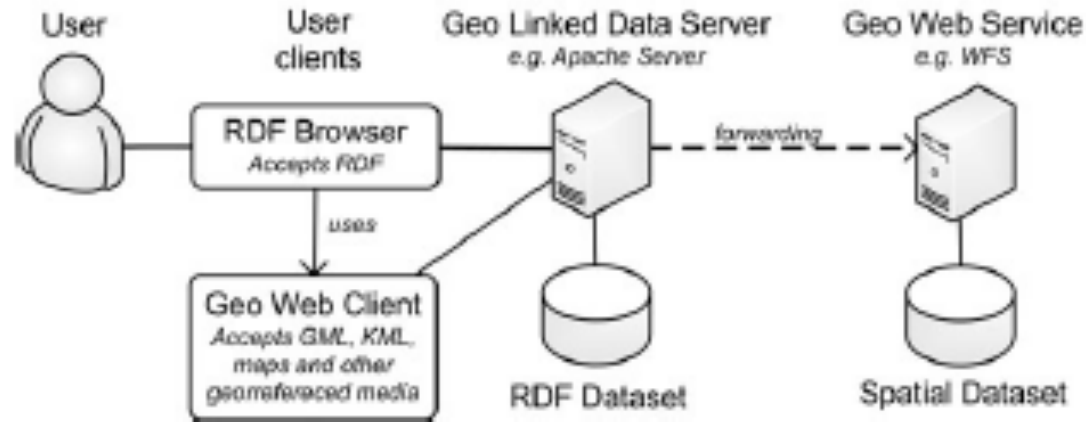
```

: <http://dblp.13s.de/d2r/resource/authors/>
dbpedia: <http://dbpedia.org/resource/>
rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
rdfs: <http://www.w3.org/2000/01/rdf-schema#>
owl: <http://www.w3.org/2002/07/owl#>
yago: <http://dbpedia.org/class/yago>
dbo: <http://dbpedia.org/ontology/>

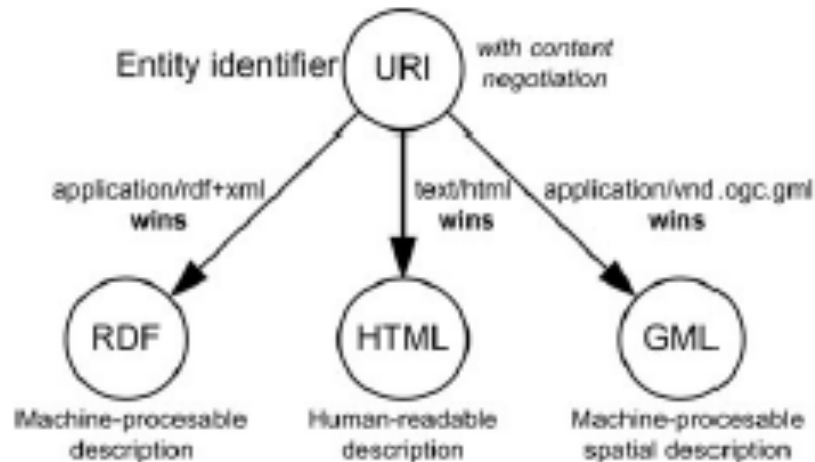
```



How to link data?



(a) Simple setup with a Geoweb data access service which serves geospatial resources in GML.



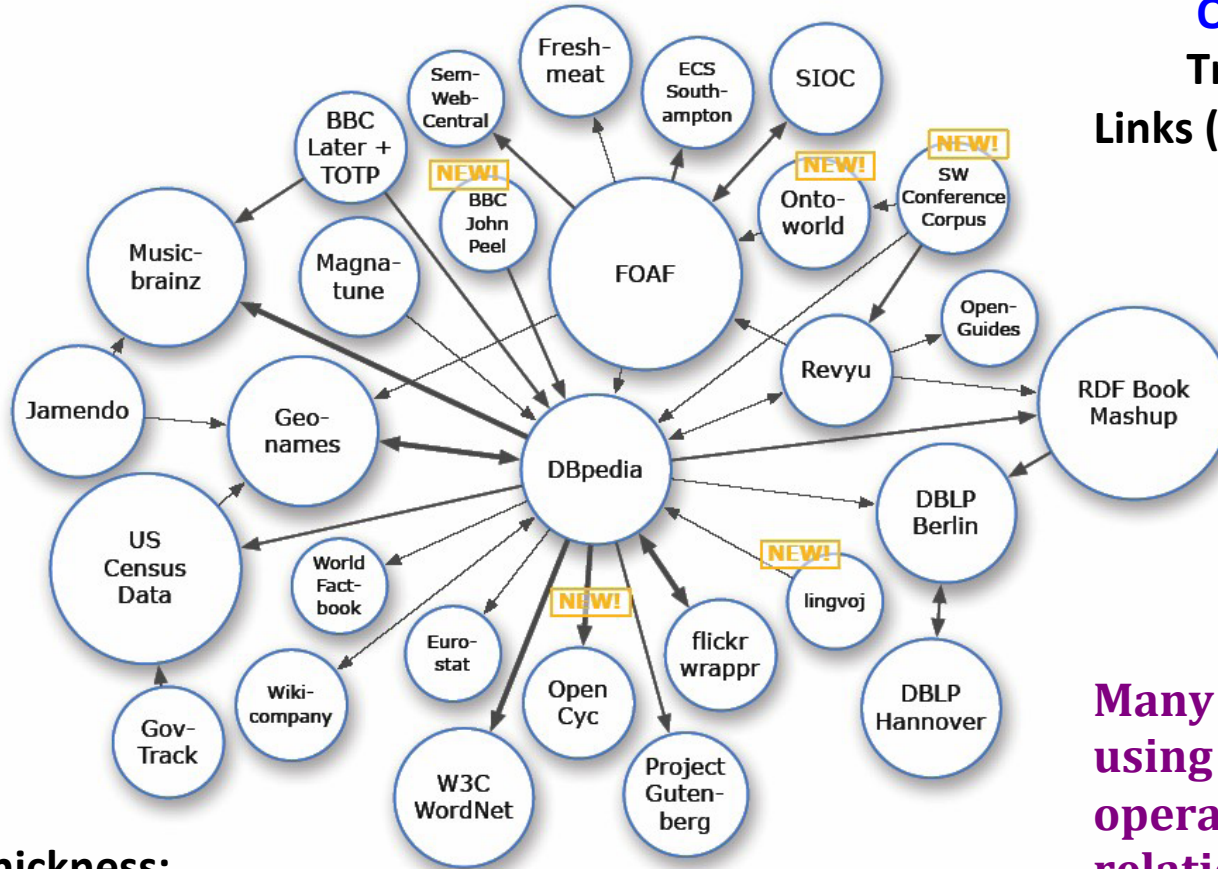
(b) The content negotiation model of Linked Data is extended with support to Geoweb MIME-types.

```

lgd-node:26890002      rdfs:comment          "Generated by Triplify V0.5" .
lgd-node:26890002      cc:license             cc:by-sa/2.0 .
lgd-node:26890002      lgd-vocabulary:attribution "This data is derived" .
lgd-node:26890002#id   rdf:type               lgd-vocabulary:node .
lgd-node:26890002#id   geo-wgs84:long        "13.7416"^^xsd:decimal .
lgd-node:26890002#id   geo-wgs84:lat         "51.0519"^^xsd:decimal .
lgd-node:26890002#id   lgd-vocabulary:created_by lgd:Potlatch+0.10e .
lgd-node:26890002#id   lgd-vocabulary:religion lgd:christian .
lgd-node:26890002#id   lgd-vocabulary:name   "Frauenkirche" .
lgd-node:26890002#id   lgd-vocabulary:tourism lgd:viewpoint .
lgd-node:26890002#id   lgd-vocabulary:amenity lgd:place_of_worship .
lgd-node:26890002#id   lgd-vocabulary:wikipedia%2525de
                        "http://de.wikipedia.org/wiki/Frauenkirche_(Dresden)" .
lgd-node:26890002#id   lgd-vocabulary:wikipedia%2525en
                        "http://en.wikipedia.org/wiki/Frauenkirche_Dresden" .
lgd-node:26890002#id   lgd-vocabulary:denomination lgd:lutheran .
lgd-node:26890002#id   lgd-vocabulary:url
                        "http://www.frauenkirche-dresden.de/" .
lgd-node:26890002#id   lgd-vocabulary:locatedNear lgd-way:23040893> .
lgd-node:26890002#id   lgd-vocabulary:locatedNear lgd-way:23040894> .

```

LOD Cloud as of November 2007



Overall Statistics

Triples #: 2 billion

Links (external) #: 3 million

Ratio: 0.001

Most of the links created based on *common identifiers* or *automatic wrapper scripts*

Many are published using tools that operate over relational data

Arrow Thickness:

thin = 100s

medium = 10k-100k

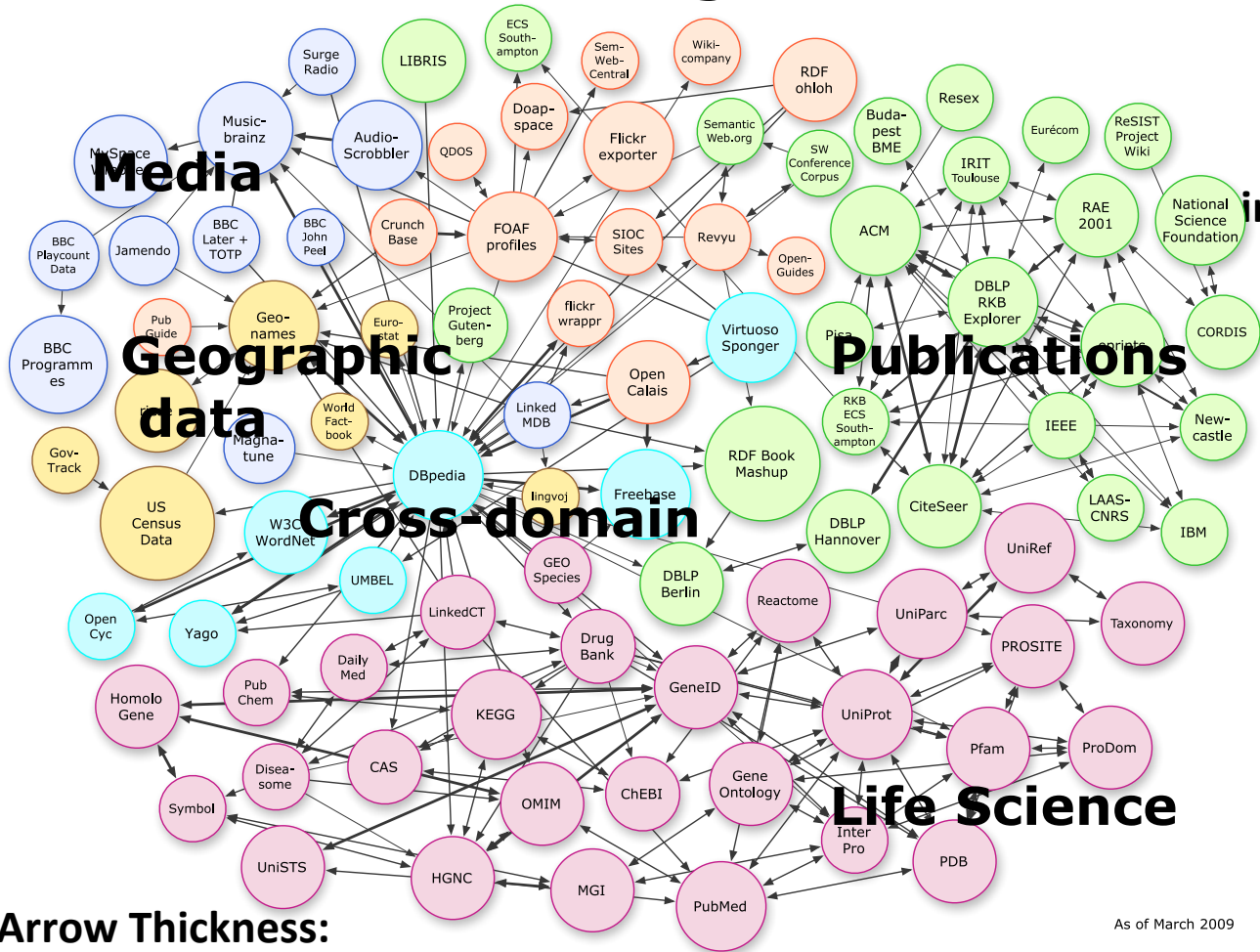
thick = 100k+



LOD Cloud as of August 2009



User-generated content



Overall Statistics

Triples #: 7.7 billion

Links(external) #: 142 million

Ratio: 0.018

Non-Bio:

Triples #: 5.3 billion

Links #: 9 million

Ratio: 0.002

Arrow Thickness:
 thin = 100s
 medium = 10k-100k
 thick = 100k+

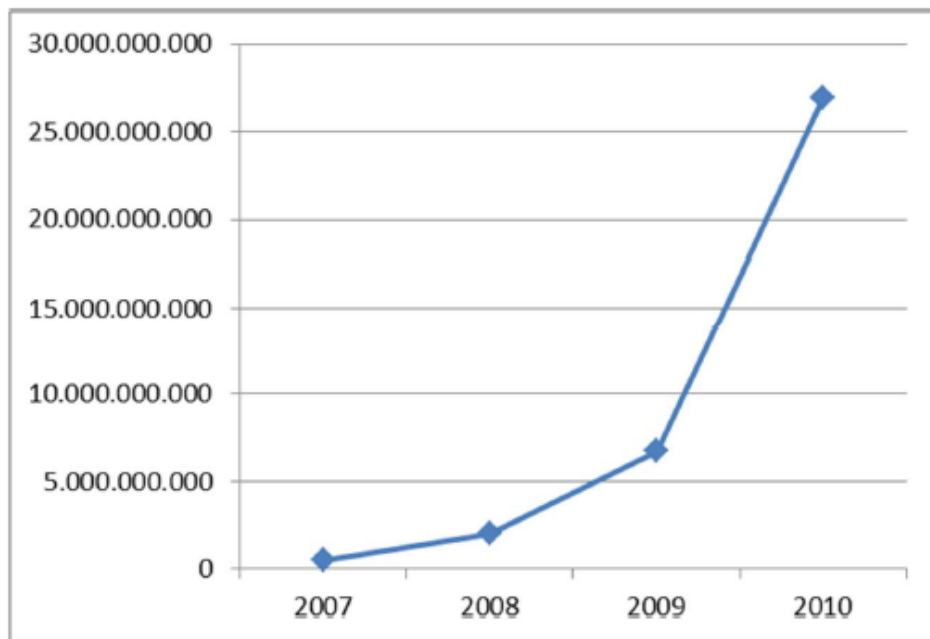
As of March 2009



The Growth in Number and Size



Year	Datasets	Triples	Growth
2007	12	500.000.000	
2008	45	2.000.000.000	300%
2009	95	6.726.000.000	236%
2010	203	26.930.509.703	300%

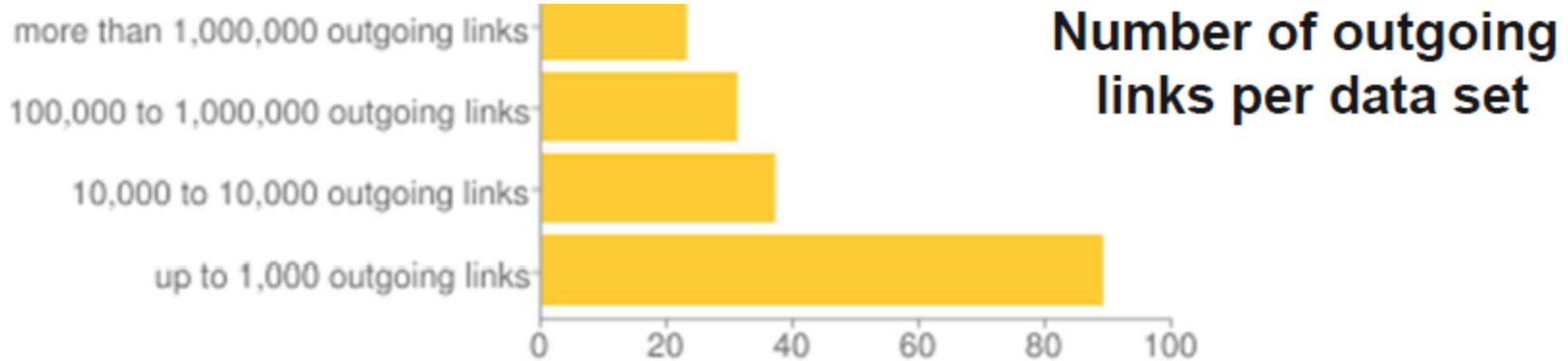


The Growth by Domain 2009-2010

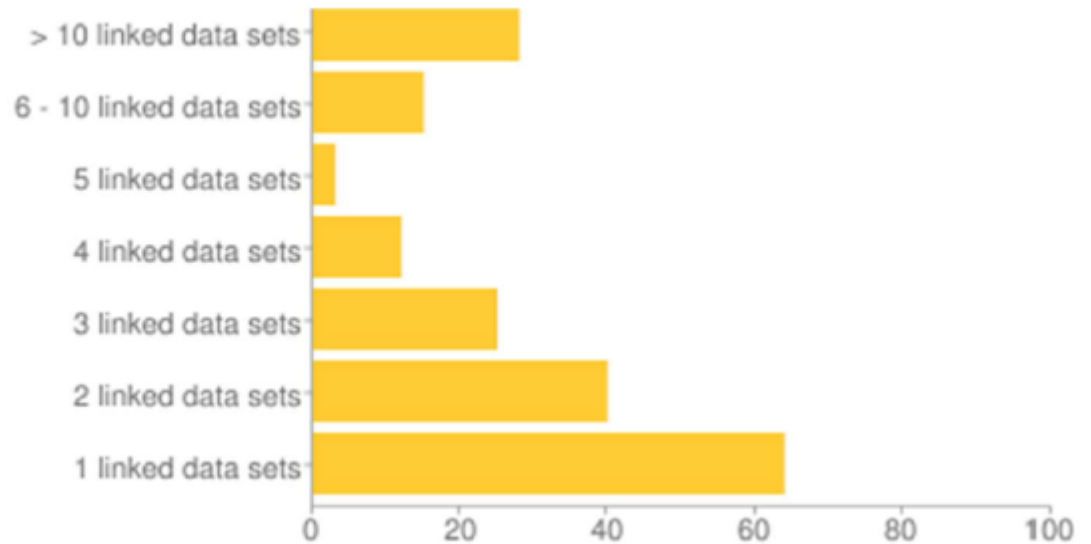


Domain	Triples (June 2009)	Triples (Nov 2010)	Growth
Geographic	3.097.000.000	5.904.980.833	91%
Libraries	212.000.000	2.237.435.732	955%
Media	698.000.000	2.453.898.811	252%
Life sciences	2.429.000.000	2.664.119.184	10%
Cross-domain	214.000.000	1.999.085.950	834%
User-generated	76.000.000	57.463.756	-24%
Government	0	11.613.525.437	-
Total	6.726.000.000	26.930.509.703	300%

Links between Datasets



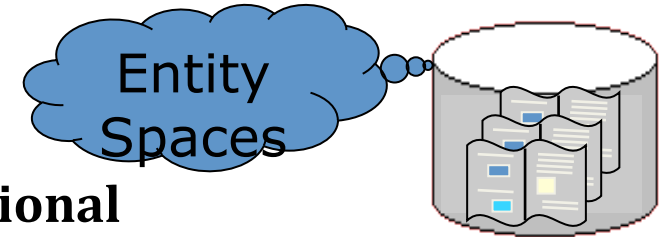
Number of linked target data sets



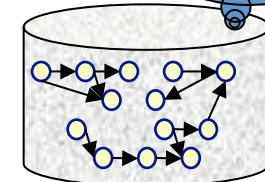
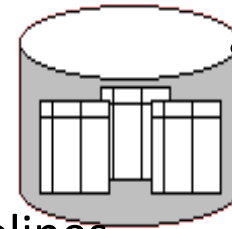
Towards a Convergence in Web 3.0



Unstructured/loosely structured documents



Relational Databases



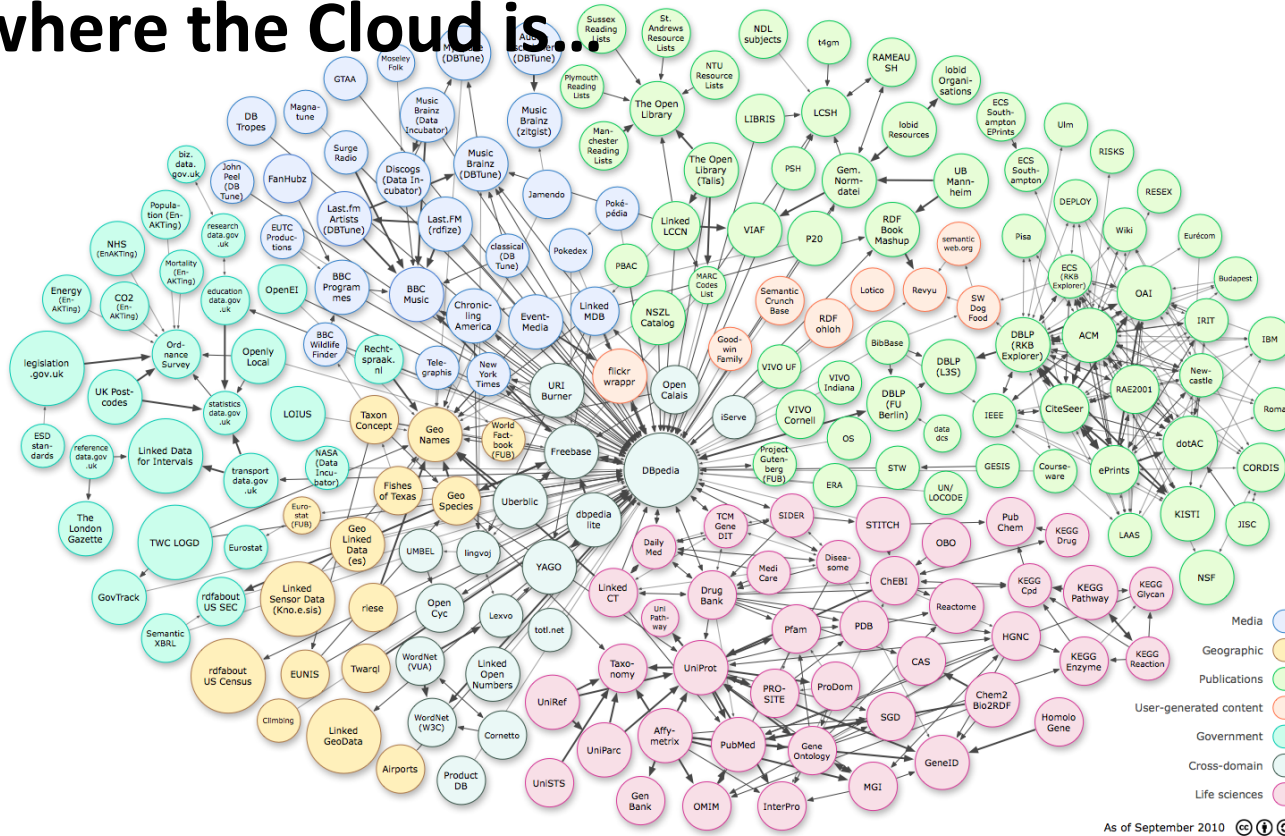
Semantic Web

- **Global information spaces** where a variety of data co-exist along with a minimal structuring and naming conventions
 - dataset participants + relationships (correspondences or mappings)
- **Participants**
 - data published according to LOD guidelines
 - generic representation of data as fine-grained triples
 - at best only partial knowledge of data semantics by ontologies/schemes
- **Relationships**
 - links constructed by humans (same as etc.) or by machines (“is view”, “schema mapping”, “created independently”, ...)

Linked Data Cloud



Here's where the Cloud is...



Where
Are We?

“Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch.

<http://lod-cloud.net/>”

So what is involved?



- Identifying Data (values) using URIs
- Using a unified upper ontology to describe the data (metadata)
- Incorporate in that: **INSPIRE**
 - INSPIRE already has a conceptual model
 - INSPIRE integrates standards
 - ISO
 - OGC
 - W3C
- INPSIRE uses other Ontologies
 - ISO 19101, 19115, 19119