COMP318

Ontologies and Semantic Web

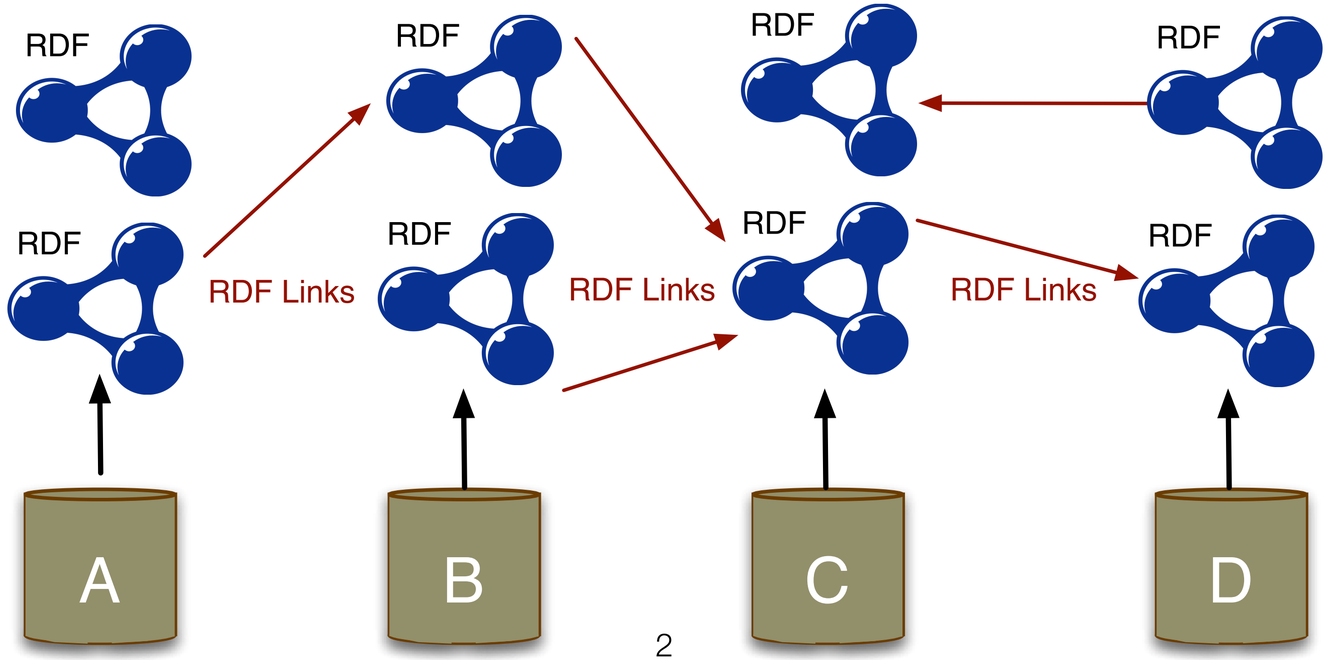
Linked Open Data- Part 1

Dr Valentina Tamma

V.Tamma@liverpool.ac.uk

Linked data: the idea

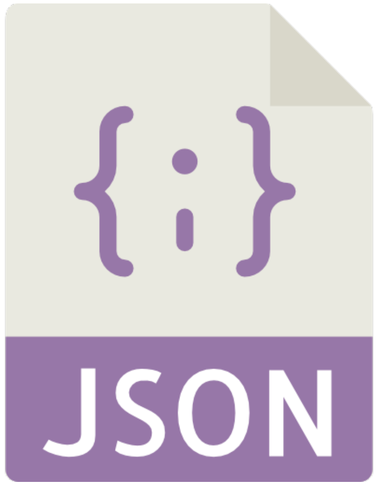
- Use the Web to create a global dataspace
  - RDF as a means to publish structured data on the Web
  - state explicitly links between data items from different data sources
  - Link data rather than documents!



Web of data

- Isn't this what we are doing now?
  - CSV, XML, key-value data, relational data

- But it's about the ***semantics*** rather than the surface or exchange syntax



In Other Words

- *"All kinds of conceptual things, they have names now that start with HTTP."*
- *"I get important information back. I will get back some data in a standard format which is kind of useful data that somebody might like to know about that thing, about that event."*
- *"I get back that information it's not just got somebody's height and weight and when they were born, it's got relationships. And when it has relationships, whenever it expresses a relationship then the other thing that it's related to is given one of those names that starts with HTTP."*

It is about:
a data (RDF)
and naming (URI)
model on Web

Tim Berners-Lee: Linked Open Data Presentation, TED 2009

What is linked data

- Linked Data refers to a set of best practices for publishing and interlinking structured data on the Web:
  - Based on 4 principles, following the web architecture:
    1. Use URIs as to identify things;
    2. Use HTTP URIs so that these things can be referred to and looked up ("dereferenced") by people and user agents.
    3. Provide useful information about the thing when its URI is dereferenced, using standard formats such as RDF/XML.
    4. Include links to related URIs in the exposed data to improve discovery of other related information on the Web.
       1. (a bit obvious, maybe) Link the data!

Tim Berners-Lee: http://www.w3.org/DesignIssues/LinkedData

1. Use URIs as to identify things

- Use URIs for naming everything you could think of:
  - URI references should be used to identify not just Web documents and digital content, but also:
  - real world objects and abstract concepts:
    - tangible things
      - e.g. people, places and cars,
    - more abstract things
      - e.g. the relationship type of knowing somebody, the set of all green cars in the world, or the colour green itself

- This principle extends the scope of the Web:
  - from online resources to encompass any object or concept in the world.

2. Use HTTP URIs allowing to refer and look up things on the Web

- HTTP protocol as the Web's universal access mechanism for people and user agents to dereference resources.
  - In the classic Web, HTTP URIs are used to combine globally unique identification with a simple, well-understood retrieval mechanism.
  - Thus, Linked Data advocates the use of HTTP URIs to identify objects and abstract concepts:
    - Dereferenceable URIs are recommended but not mandatory in linked data. When you do an HTTP GET (or type in a URL in your browser's address bar) something is returned, an RDF description in the best case!
    - Different servers are responsible for answering requests attempting to dereference HTTP URIs in many different namespaces.

3. Assign useful information to URIs using standards

- Provide useful information about the thing when its URI is dereferenced,
  - using standard formats such as RDF/XML or OWL
  - agreement on standardised content formats!

- Advocates the use of a single data model for publishing structured data on the Web
  - Resource Description Framework (RDF)
  - simple graph-based data model.

4. Include links to other URIs to improve discovery

- Following from principle 1, hyperlinks connect not only Web documents, but any type of thing.
  - e.g. a hyperlink may be set between a person and a place, or between a place and a company.

- Hyperlinks that connect things in Linked Data have types which describe the relationship between the things.
  - Unlike hyperlinks on the Web.
  - e.g. a hyperlink of the type friend of may be set between two people, or a hyperlink of the type based near may be set between a person and a place.
  - When an RDF link connects URIs in different namespaces, it ultimately connects resources in different data sets
    - **These typed hyperlinks are RDF properties interpreted as hyperlinks**

4a. Link the data!

- Across the Web, many different servers are responsible for answering requests attempting to dereference HTTP URIs in many different namespaces:
  - In Linked Data, this means returning RDF descriptions of the resources identified by these URIs.
    - If an RDF link connects URIs in different namespaces, it ultimately connects resources in different data sets.

- Linked Data uses hyperlinks to connect disparate data into a single global data space.
  - just as hyperlinks in the classic Web connect documents into a single global information space,

- These links, in turn, enable applications to navigate the data space.
  - e.g. a Linked Data application that has looked up a URI and retrieved RDF data describing a person may follow links from that data to data on different Web servers, describing the place where the person lives or the company for which the person works.

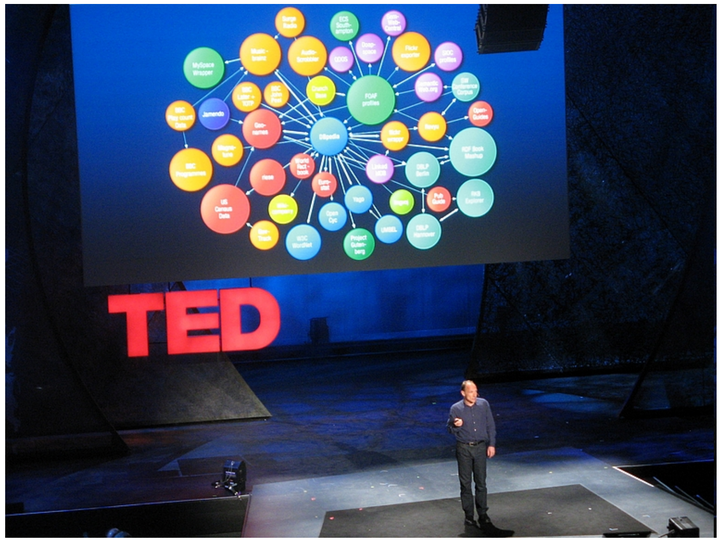
But TBL was talking about open data i.e. when the sw meets open data

- Over the last few years, the Linked Data paradigm has found a huge application when combined with the publication of data with liberal licenses.

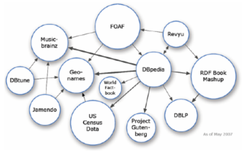
- The Open Data Movement, aims to release huge data sets often from local government authorities.

- It uses Linked Data technologies and best practices to publish a plethora of different interlinked data sets.
  - a bunch of data published with an open license is intended to be freely available to everyone to use and republished without restriction from copyright, patents or other restrictions.

- Where Open Data meets Linked Data, we have Linked Open Data.

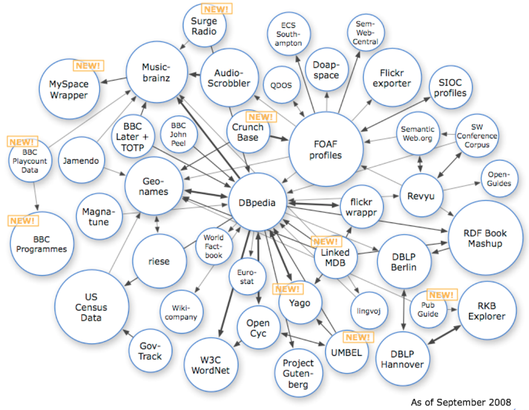


Linked Data Evolution: is there data out there?

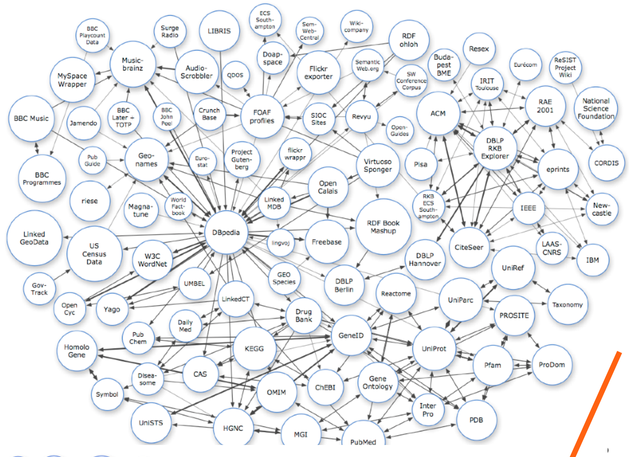


2007

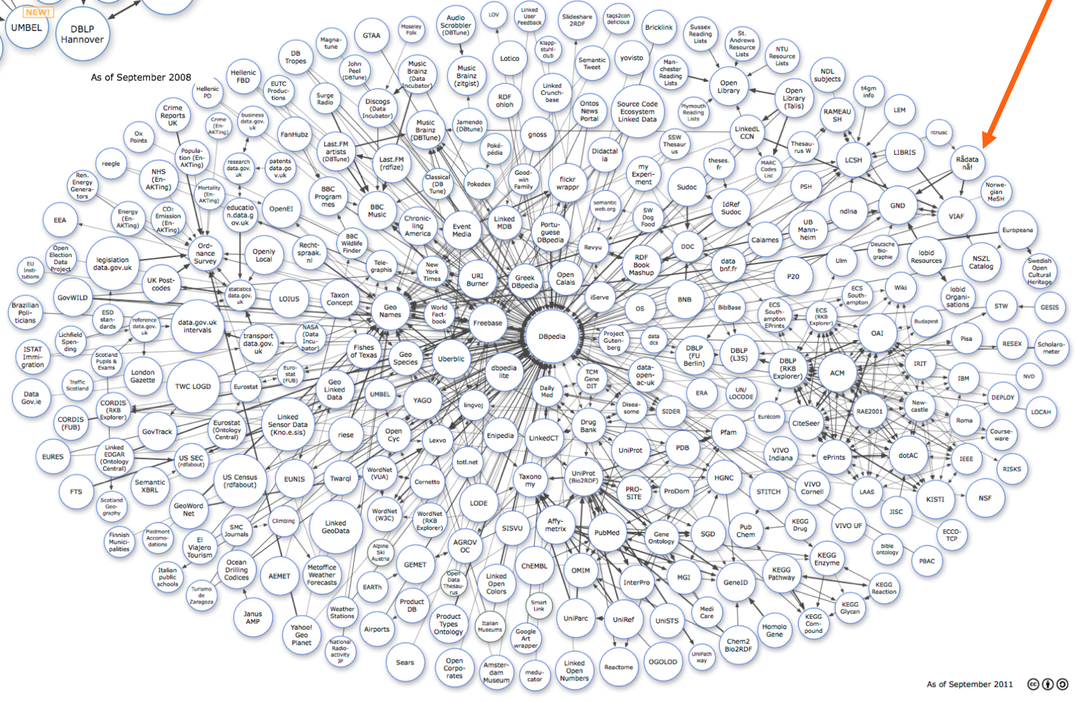
2008



2009



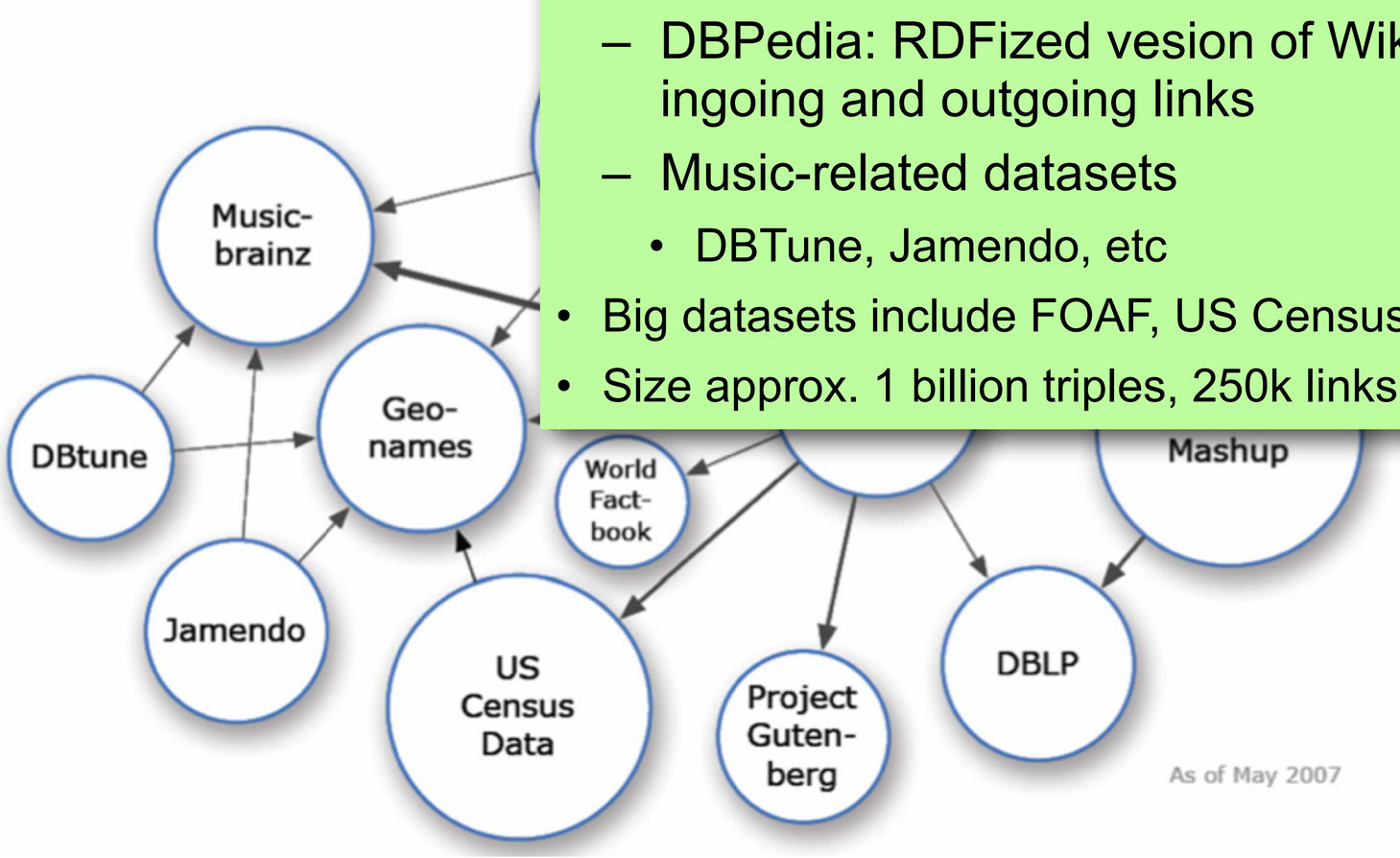
2011



May 2007

FACTS:

- Focal point based
  - DBPedia: RDFized vesion of Wikipiedia; many ingoing and outgoing links
  - Music-related datasets
    - DBTune, Jamendo, etc
- Big datasets include FOAF, US Census data
- Size approx. 1 billion triples, 250k links

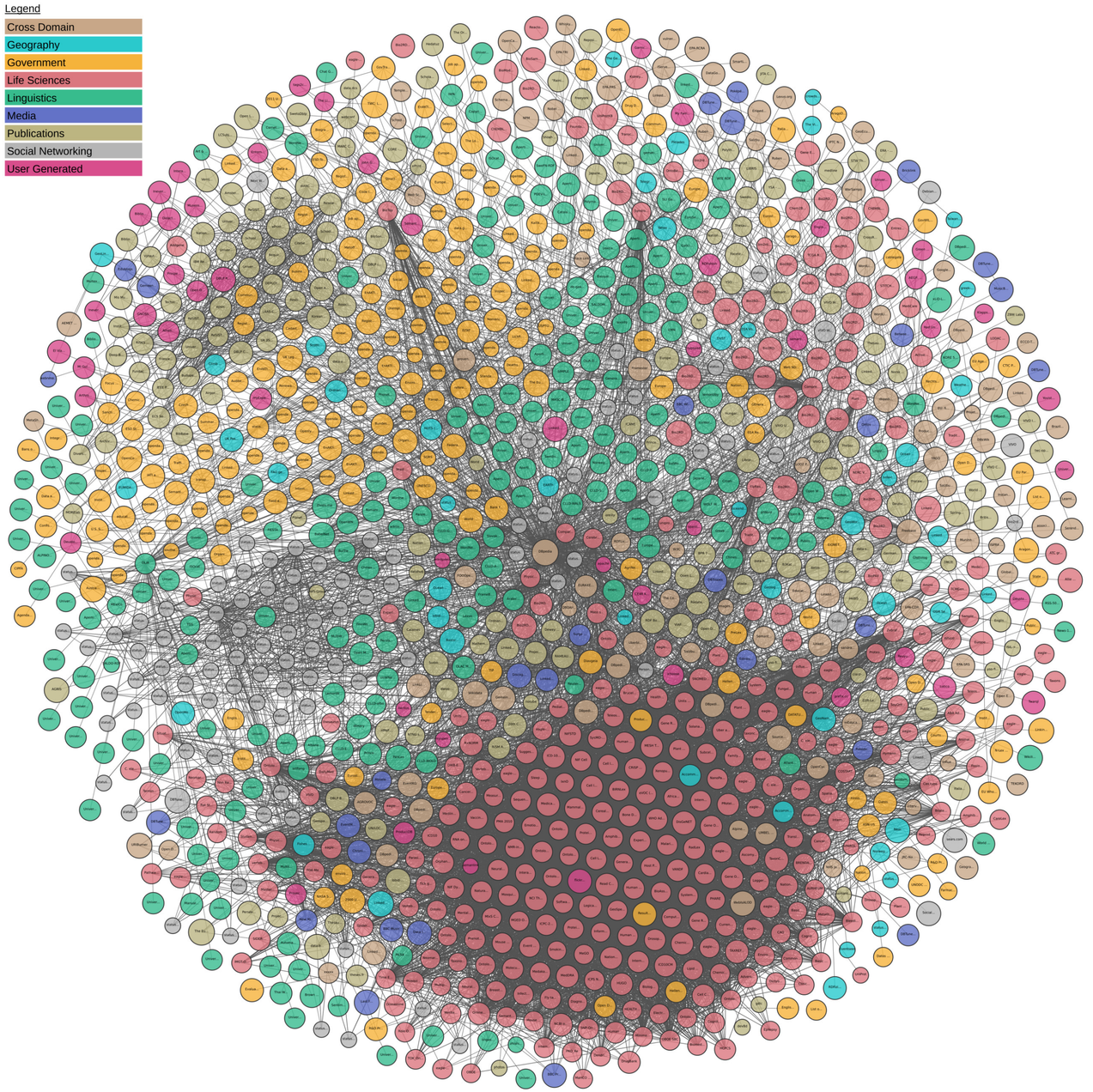


"Linking Open Data cloud diagram 2019. by A. Abele. J. P. McCrae. P. Buitelaar. A. Jentzsch and R. Cyganiak

March 2019

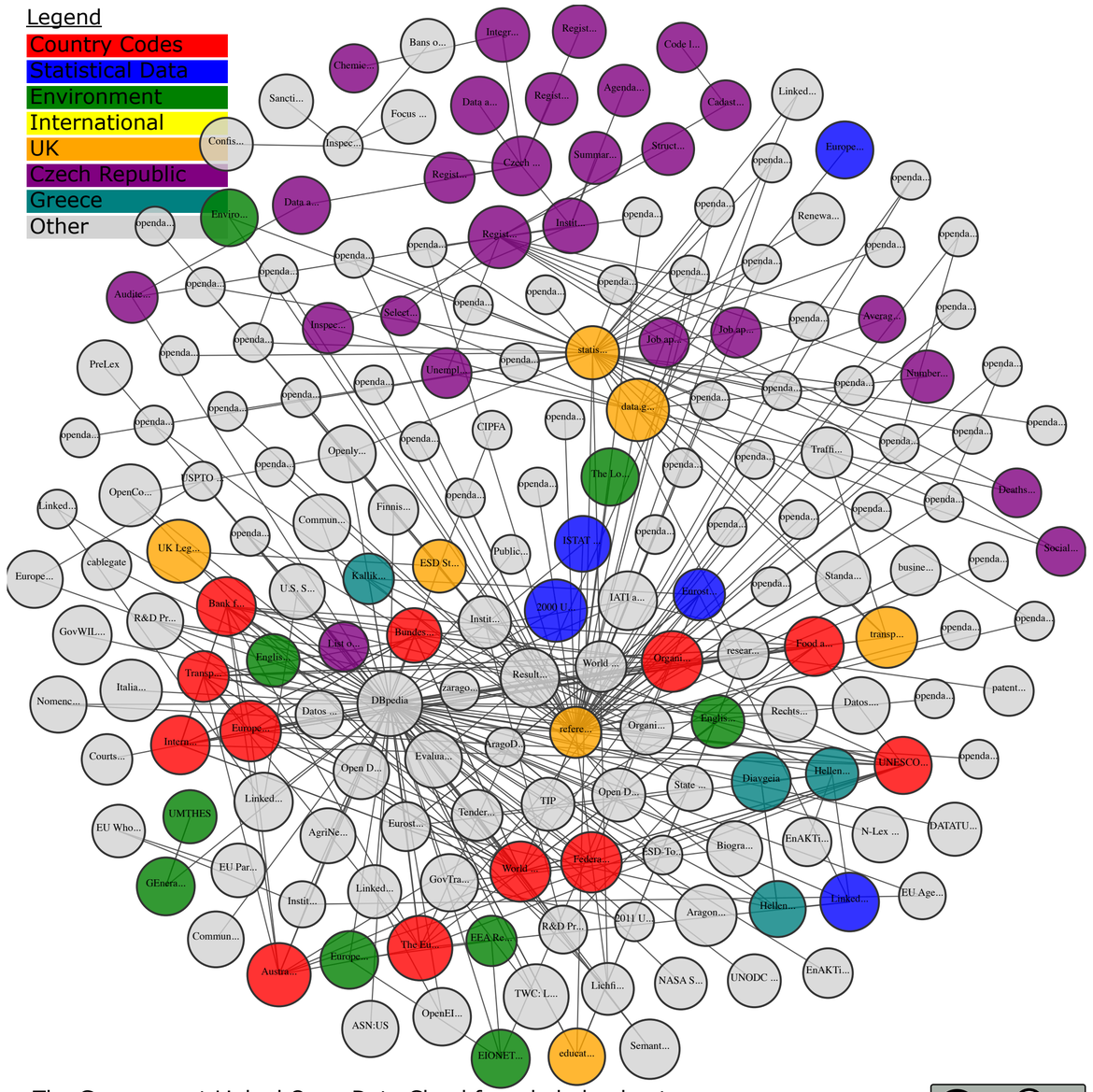
FACTS:

- 1,239 datasets with 16,147 links!



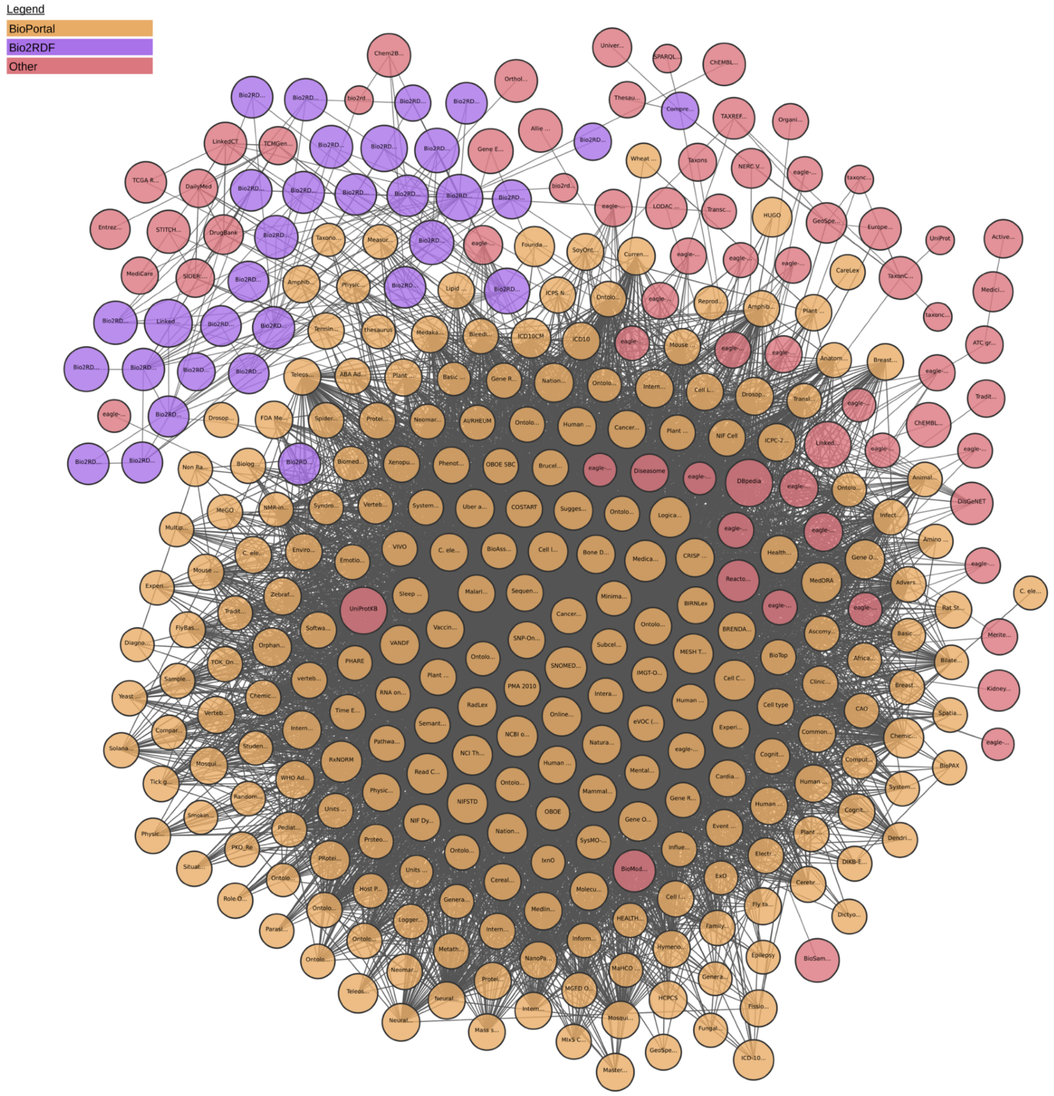
The Linked Open Data Cloud from lod-cloud.net

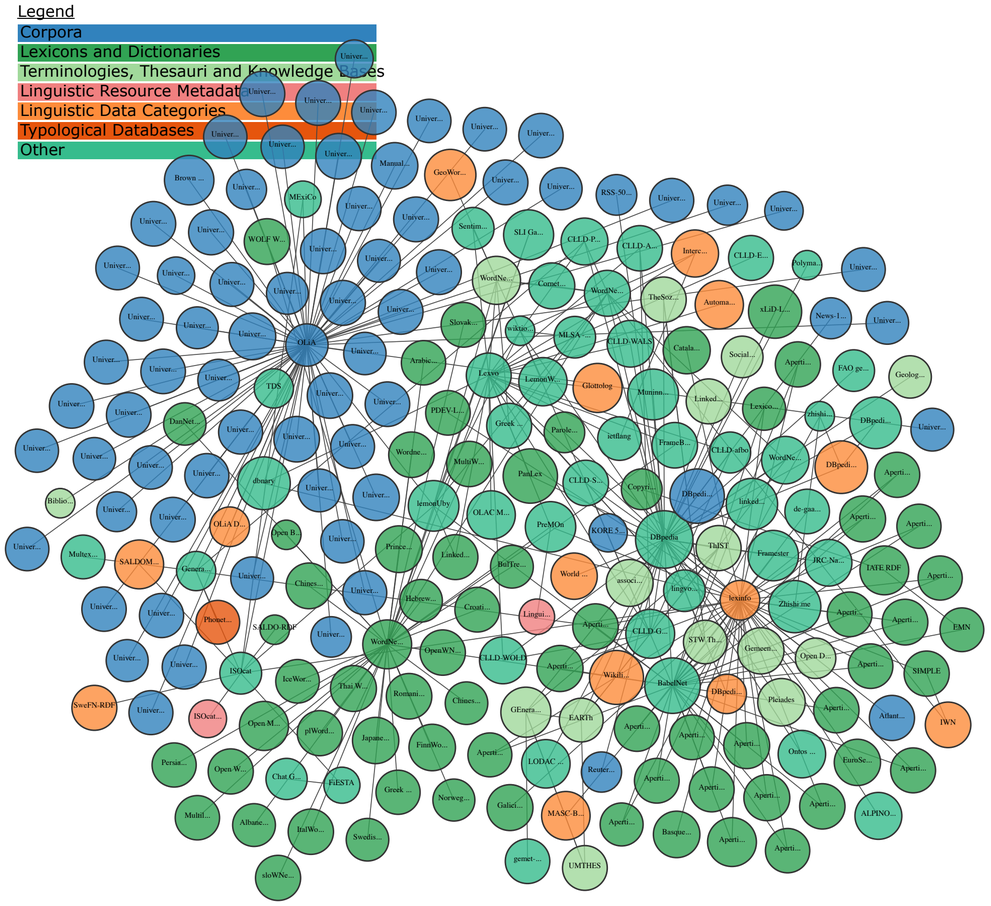
Governament data



The Government Linked Open Data Cloud from lod-cloud.net

Largest domain LODs





The Linguistic Linked Open Data Cloud from lod-cloud.net

The Life Sciences Linked Open Data Cloud from lod-cloud.net

COMP318

Ontologies and Semantic Web

End Linked Open Data- Part 1

Dr Valentina Tamma

V.Tamma@liverpool.ac.uk