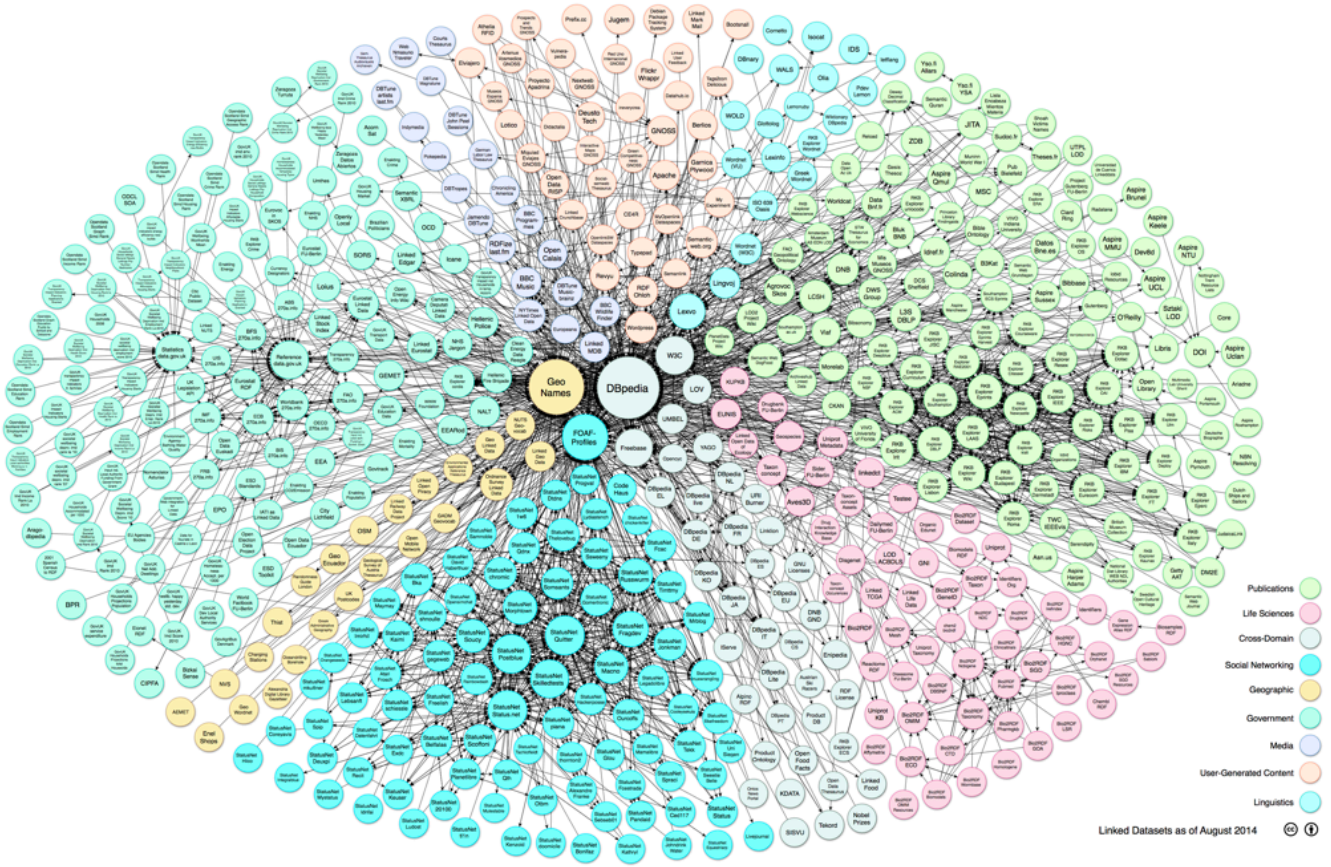
No Respect for the Ontology: The case of DBpedia and Freebase

Gianluca Demartini
University of Sheffield
gianlucademartini.net
The 4th UK Ontology Network Meeting

Linked Open Data



- "Linked Open Data" is Big Data
- Not just Volume but also **Data Quality**

Linked Open Data

- DBpedia has 4.6 million entities
  - DBpedia ontology has 1’109 properties defined
  - DBpedia data uses 1’368 properties
- Freebase has 47 million entities
  - Freebase ontology has 65’019 properties defined
  - Freebase data uses 18’841 properties

Problem

- Domain and Range defined in ontologies are not followed by RDF triples
- To identify problems, we compute **Domain Violation Rate** and **Range Violation Rate**

Example Problem

- DBpedia **dpo:gender** has no domain

- Used for both subjects of type
  - dpo:GivenName
  - dpo:School (e.g., it accepts only girls)

Solution

- Identify schema adherence problems
- Resolve them by
  - updating the underlying schema and/or
  - by modifying the data
- Identify properties that are used in different contexts
- Looking at **rdfs:type** attached to instances using the property
- 87-96% Precision of suggested schema modifications

Alberto Tonon, Michele Catasta, Gianluca Demartini, and Philippe Cudré-Mauroux**. Fixing Domain and Range of Properties in Linked Data by Context Disambiguation**. In: The 8th Workshop on Linked Data on the Web (LDOW2015) at WWW2015. Firenze, Italy, May 2015.

Conclusions

- The Web of Data is Big Data
- Data quality problems
- We focus on **schema adherence**
  - Identify problems
  - Fix schema or data
  - Look at how different entity types are used with the RDF property

We are hiring!

- PhD Studentship: **entities** in **big web archives**
  - in Sheffield
  - with The National Archive

http://bit.ly/entity-centric-exploration-PhD

gianlucademartini.net