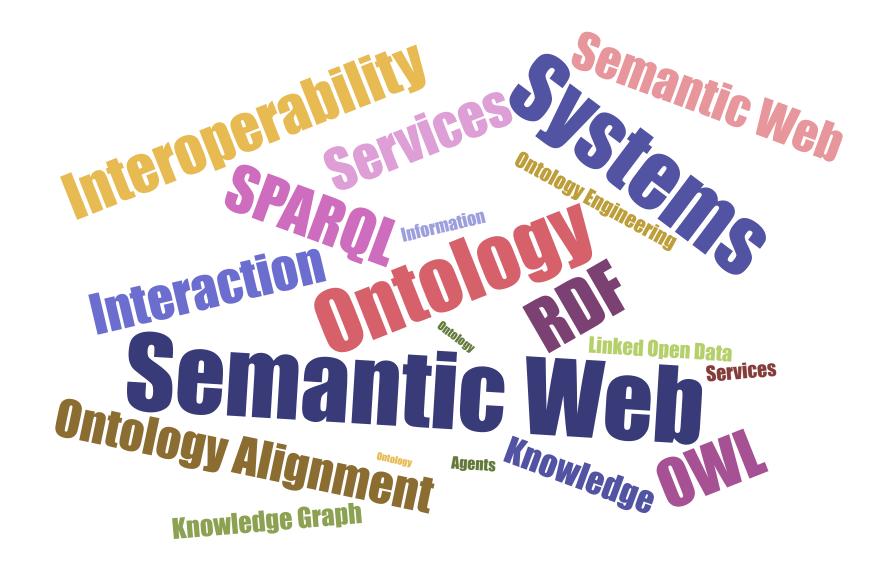
COMP318 Ontologies and Semantic Web



Linked Open Data- Part 2

Dr Valentina Tamma

V.Tamma@liverpool.ac.uk

Where is the data

Datasets

- **UMBEL** a lightweight reference structure of 20,000 subject concept classes and their relationships derived from OpenCyc, which can act as binding classes to external data; also has links to 1.5 million named entities from DBpedia and YAGO.
- FOAF a dataset describing persons, their properties and relationships
- VIAF (Virtual International Authority File) an aggregation of authority files (author names) from national libraries from around the world.
- Wikidata Wikidata is a free and open knowledge base, of more than 82,000,000 data items.
 It can be read and edited by both humans and machines.
- DBpedia a dataset containing extracted data from Wikipedia; it contains about 3.4 million concepts described by 1 billion triples, including abstracts in 11 different languages
- **GeoNames** provides RDF descriptions of more than 7,500,000 geographical features worldwide.

LOD applications building blocks

Using: Mashups

Mashups combine multiple datasets to create a new service, visualisation or information

Using: Search

Linked data search engines allow search across the web of data. Conventional search may present information derived from linked data.

Using: Productivity Linked data facilitates data integration for business intelligence or research.

Storing and publishing

Linked data can be published in simple flat files on a web server, in databases with a translation layer, or in specialised 'triple stores' built to store and share linked data.

Querying: SPARQL

SPARQL Protocol and RDF Query Language provides a way to run structured queries over linked data datasets.

Representing: Vocabularies

Vocabularies provide lists (and definitions) of common terms that can be used to describe the things and relationships in a dataset.

Representing: Ontologies

Ontologies are vocabularies that record the logical relationships between their

Identifying: URLs

Using HTTP Uniform Resource Locators (URLs) means that (a) data can be looked up across the Internet; (b) decisions about 'namespaces' for data are managed through the Domain Name System (DNS).

Interchanging: RDF

Resource Descriptor Framework (RDF) is a model for representing data as 'triples'. RDF can be serialised into a range of different file formats, including RDF-XML and text-based Turtle or N3 syntax.

Transporting: HTTP (The World Wide Web)

Data is hosted on servers that can talk Hypertext Transfer Protocol (HTTP) to each other and to browsers in order to exchange data across the Internet.

F. Bauer, M. Kaltenböck Linked Open Data: The Essentials A quick Start Guide for Decision

Makers

Elements of the Linked Open Data Puzzle (revision 2) - 2nd May 2011. CC BY-SA-NC

Draft sketch by Tim Davies (@timdavies / tim@practicalparticipation.co.uk) for IKM Working Paper on Linked Open Data for Development. Comments welcome. Search 'linked open data puzzle' on http://www.opendataimpacts.net for latest version.

Open data is made available under licenses (or is placed in t public domain) so that others can use and build upon it, free legal restrictions. Open standards for data files and interchan Licensing:

open data

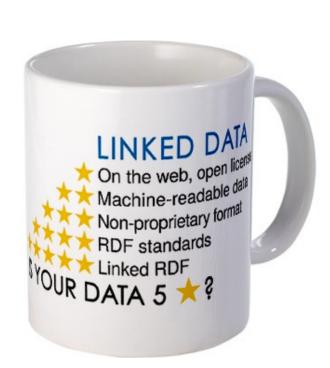
the of ange

How do we publish Linked Data

- Exposing Relational Databases or other similar formats into Linked Data
 - D2R
 - Triplify
 - R2O
 - NOR20
 - Virtuoso
 - Ultrawrap
 - ...



- Sesame
- Jena
- Owlim
- ..
- Incorporating it in the form of RDFa in CMSs like Drupal



How do we consume Linked Data

- Linked Data browsers: To explore things and datasets and to navigate between them.
 - Tabulator Browser (MIT, USA), Marbles (FU Berlin, DE), OpenLink RDF Browser (OpenLink, UK), Zitgist RDF Browser (Zitgist, USA), Disco Hyperdata Browser (FU Berlin, DE), Fenfire (DERI, Ireland)
- Linked Data mashups: Sites that mash up (thus combine Linked data)
 - Revyu.com (KMI, UK), DBtune Slashfacet (Queen Mary, UK), DBPedia Mobile (FU Berlin, DE), Semantic Web Pipes (DERI, Ireland)
- Search engines: To search for Linked Data.
 - Falcons (IWS, China), Sindice (DERI, Ireland), MicroSearch (Yahoo, Spain), Watson (Open University, UK), SWSE (DERI, Ireland), Swoogle (UMBC, USA)





BETA This is a new service – your feedback will help us to improve it

Search results

|--|

Filter by

254 results found

Best match ▼

Publisher

|--|

Organogram of Staff Roles & Salaries

Topic

Published by:

Serious Fraud Office

18 October 2016 Last updated:

Format

RDF

Organogram (organisation chart) showing all staff roles. Names and salaries are also listed for the Senior Civil Servants. Organogram data is released by all central government departments and...

Open Government Licence (OGL) only

Organogram of Staff Roles & Salaries

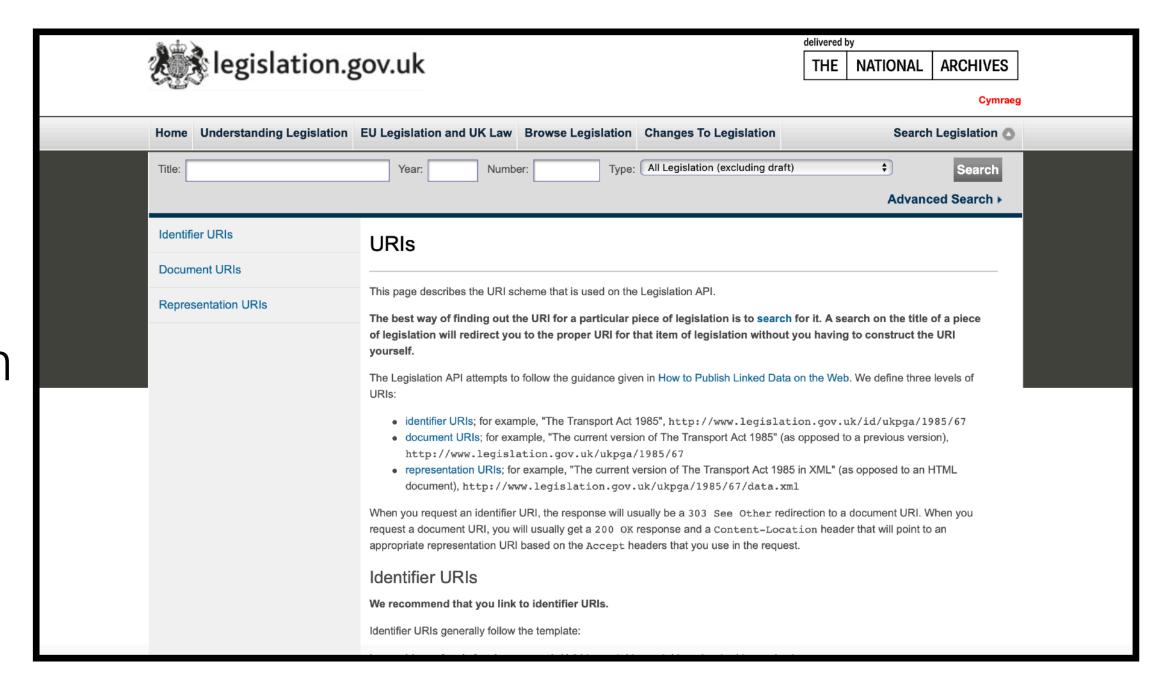
UK Supreme Court Published by:

18 October 2016 Last updated:

Organogram (organisation chart) showing all staff roles. Names and

legislation.gov.uk

- Official government archive of the UK, managed by National archives.
 - Best practice example of Open Government Data
 - Portal with access to all published UK legislation
 - Data from 1267 onwards, including recent changes in UK legislation
- Developed following W3C standards
 - Persistent URIs for legislative data
 - identifier URIs, document URIs and representation URIs.



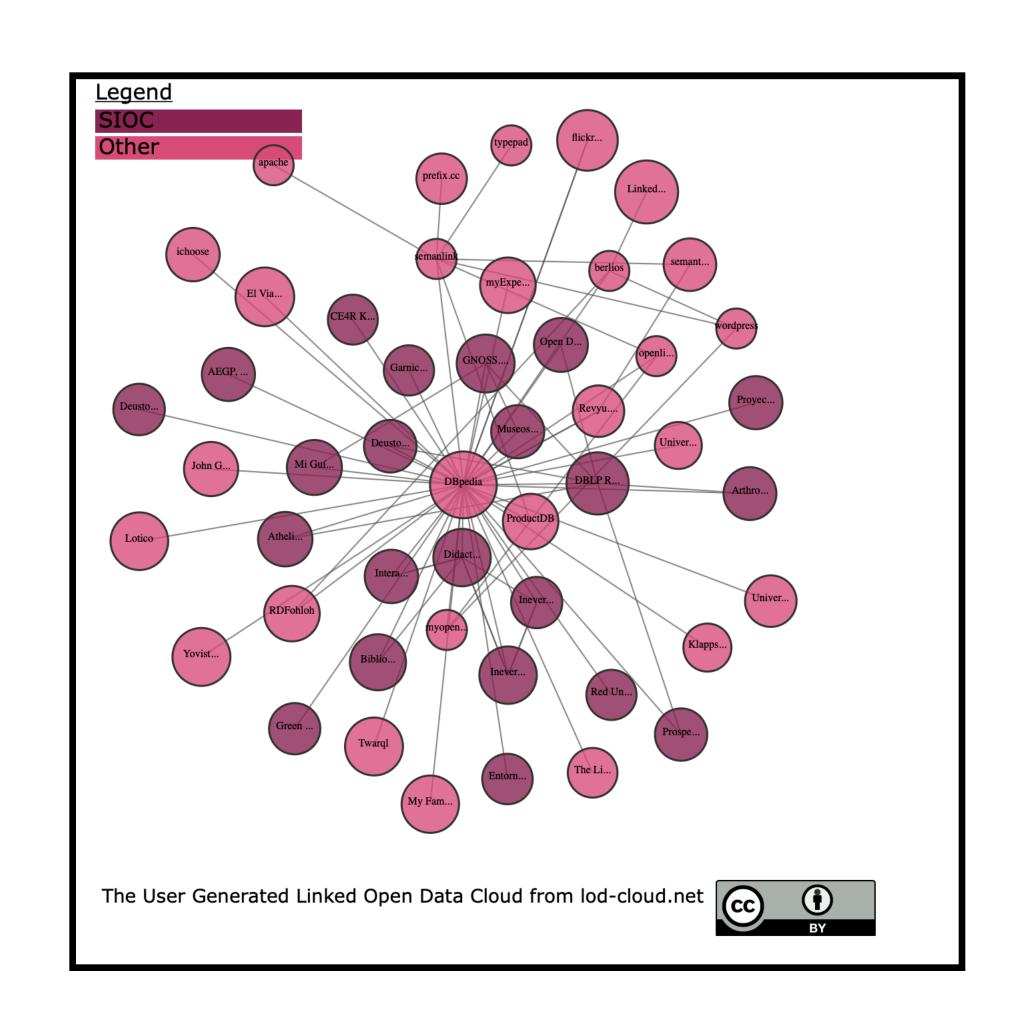
Is it all good

- Well, there are downsides...
 - The quantity of published Linked Data increases day by day, but...
 - Some of the data available might be either irregularly updated, or already available in other formats and APIs often becomes an issue
 - This is not happening with all the datasets, but it needs to be taken under consideration.
 - Additionally, more data needs to be available to share, extent and re-use.

- Data should be urgently published as Linked Data on the Web with appropriate licenses and provenance information.
 - Risk of creating RDF silos.
- More applications and tools needed to exploit Linked Data.
- Existing open issues make the development of Linked Data based applications a challenge:
 - difficulties to integrate data in different formats and from multiple sources,
 - the discovery of data
 - or the usability of user interfaces.

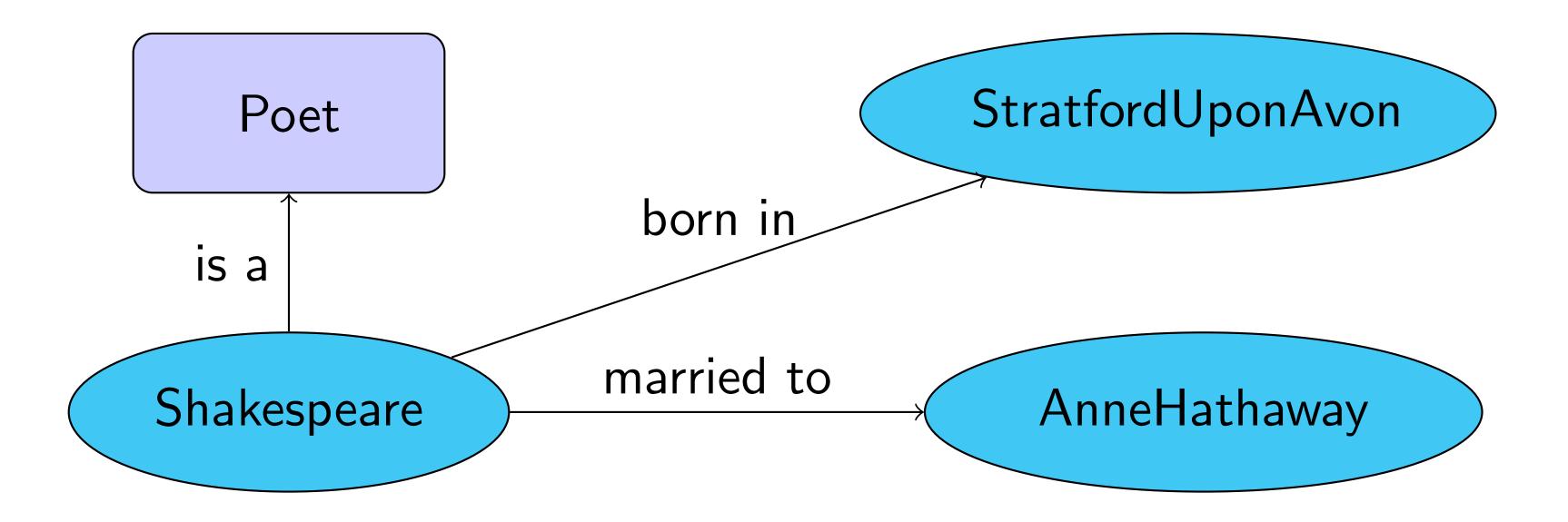
How to create Linked Data

- 4-step creation process
 - Add semantics to the data
 - Model general world knowledge
 - Acquire new knowledge from inference
 - Query consistent information from different sources

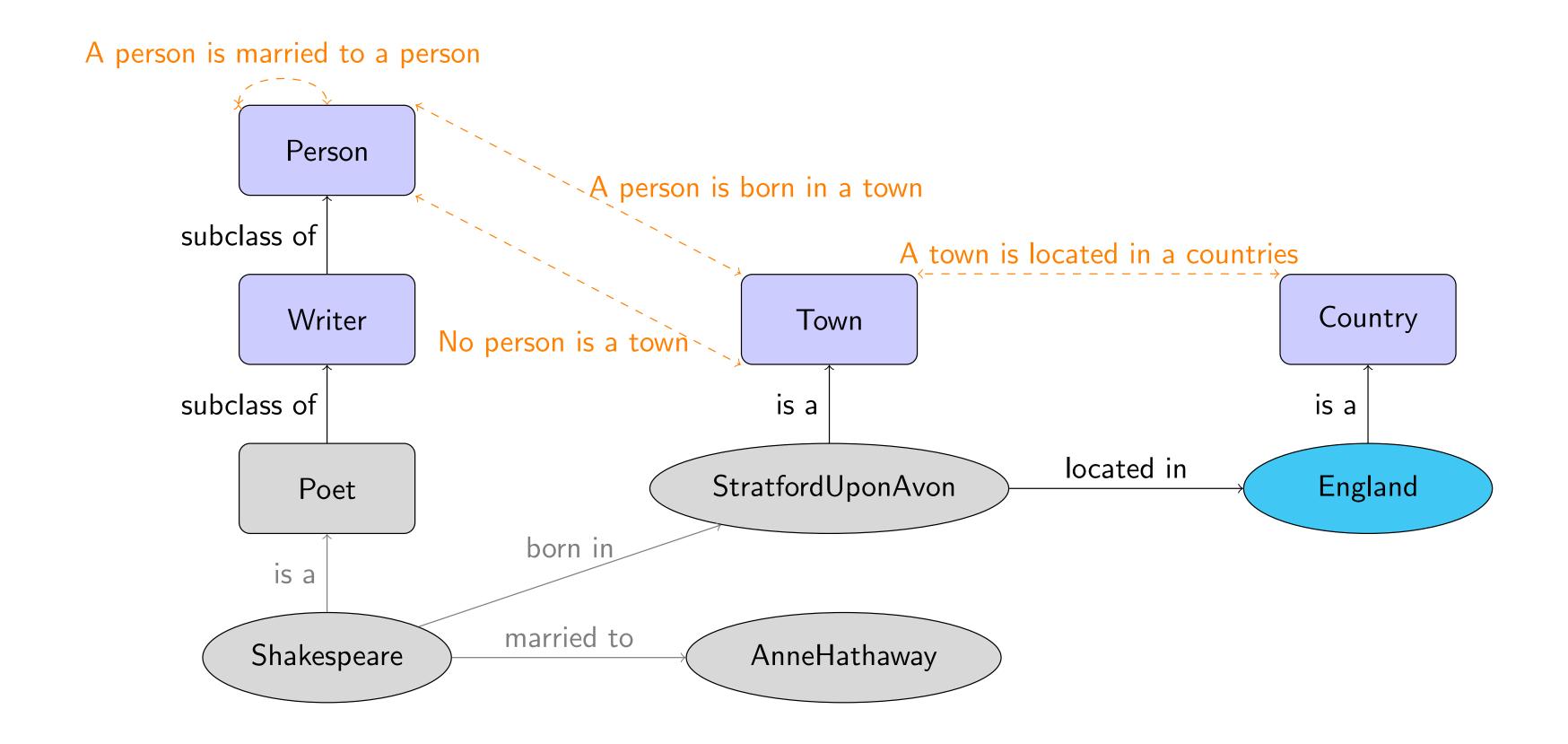


Step 1: Add semantics to the data

William Shakespeare was an English poet and playwright. Shakespeare was born and brought up in Stratford-upon-Avon. At the age of 18, he married Anne Hathaway, with whom he had three children: Susanna, and twins Hamnet and Judith. ...



Step 2: Model general world knowledge



Step 3: Acquire new knowledge from inference

- Is Shakespeare a writer?
 - facts we know:
 - Shakespeare is a poet
 - Every poet is a writer
 - facts we can infer or deduce →
 - Shakespeare is a writer
- Is Anne Hathaway a person?
 - Shakespeare is married to Anne Hathaway
 - A person is married to a person
 - facts we can infer or deduce →
 - Anne Hathaway is a person
- Is Anne Hathaway a town?
 - Anne Hathaway is a person
 - No person is a town
 - facts we can infer or deduce →
 - Anne Hathaway is NOT a town

```
:Shakespeare rdf:type :Poet
      :Poet rdfs:subClassOf :Writer
Infer :Shakespear rdf:type :Writer
      :Shakespear :marriedTo :Anne_Hathaway
      :Person :marriedTo :Person
Infer :Anne_Hathaway rdf:type :Person
      :Anne_Hathaway rdf:type :Person
       owl:DisjointWith(:Person,:Town)
      Anne_Hathaway rdf:type (not:Town)
```

Step 4. Query information from different sources

- List all persons . . .
 - from Wikipedia.
 - from IMDB.
 - from Wikidata.
 - from British Library

• . . .

- List all people who are authors on a paper together with a co-author of Paul Erdoss.
 - from FOAF-profiles on the web
- List all stores who sell fair trade items from Ecuador.
- List all doctors who treat dermatitis.

Recap

Motivation for linked data

- Link data principles
- Link data adoption
- Designing of linked data

COMP318 Ontologies and Semantic Web





Dr Valentina Tamma

V.Tamma@liverpool.ac.uk