Presentation Topic 11: Open and linked data

Group K:
Giles Howard
Jacob Kittley-Davies
S. Mahmoud Miri
Velizar Velkov
Vlad Velici
Henry Wilkes
Marcus Ojerinde-Ardalla

Tutor: Tim Chown
What is Linked Open Data?
Introduction

- Linked and Open Data - Two related concepts that are increasingly important in our data centric society.
- Our aim of this presentation is to explain what Open data and Linked Data are, and examine their implementations.
- Also, we will examine arguments for and against open data, and put our own argument forward for it.
Definitions (Open Data)

- “Open data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.” [1]
- Open data is not a specific implementation or method of accessing data – instead, it is an idea of exactly how it should be used and redistributed.
Definitions (Linked Data)

- "The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data. "[2]
- Linked Data is a specific implementation of data, which aims to make it more useful by making it easier for a user to quickly access more relevant data.
Background

- Open Data - A general movement towards greater and easier data access, important for academic study and scientific research (Mertonian science)
- Linked Data - Term coined by Tim Berners Lee to describe a hypothetical next step for the WWW, currently various examples of implementation using the RDF and URI specifications laid out by the World Wide Web Consortium
- Both concepts are closely linked; linked data technology can be used to access newly open data with greater ease
Implications
Implications by sector

Business
Implications by sector

Business

Personal
Implications by sector

Business

Government

Personal
Implications by sector

Business

Government

Personal

Academic
Business

I don't want to lose my competitive advantage.
£16 Billion
+ UK Economy
Personal

How about my private data?
Government

Is the government really doing a good job?
WHERE DOES MY MONEY GO?
Showing you where your taxes get spent

Source: http://wheredoesmymoneygo.org/
Academic

Scholarly applications
Implementation
The technical side of linked data

- Presenting data in a *machine-parsable form*, but also maintain links between *data sets*.
- Generally represented by triples in the format:
  <source object> <relationship> <target object>
- Rules of these representations depend on implementation.
Data as a graph!
RDF

- Uses XML as major format - good for machine interaction.
- N-Triples/Turtle/N3 for human readability.
- Designed to represent data such as:

"Crafting a rich and personal blending learning environment: an institutional case study from a STEM perspective" has author Su White’
RDF/N-Triples

<http://eprints.soton.ac.uk/id/eprint/346929>
<http://www.loc.gov/loc.terms/relators/AUT>
<http://eprints.soton.ac.uk/id/person/ext-44780>.
<rdf:Description rdf:about="http://eprints.soton.ac.uk/id/eprint/346929">
  <nsx:AUT xmlns:nsx='http://www.loc.gov/loc.terms/relators/' rdf:resource="#epid;person/ext-44780" />
</rdf:Description>
SPARQL

- **SPARQL Protocol and RDF Query Language**

- Query language for RDF data
- Federated Query support
  - Send the same query to more SPARQL endpoints
  - Process and gather all the results
- Implementations: Jena, 4store, OpenAnzo, AllegroGraph, Seasme etc.
The Web seen by a human

John Smith

Date of Birth: 3 Jan. 1980
Occupation: Programmer at Some-Company
The Web seen by a machine

Some title

some text some text
some text some text some text
some text
Solution: Semantic Web

- Collaborative movement led by W3C
- Promotes common data formats on the web
- E.g.: having a standard way of describing a person on the web
- Machines (including search engines) will be able to understand the data on the web.
- Is a web of data.
Semantic web (projects/solutions)

- **Microformats** (HTML attributes *class*, *rel*, *rev*); e.g.: hCard, hProduct, hCalendar
- **RDFa** - **W3C Recommendation** (HTML attributes *vocab*, *property*, *rel*, *rev* etc.); adds complex metadata to (x)HTML and XML documents.
- **Microdata** (Schema.org) - **Collaboration** between Google, Yahoo! and Microsoft. Aims to provide a single way of describing data on the web, supported by all major search engines.
Conclusions

- Open data can help make businesses more efficient and profitable, and academic work more insightful. Linked data can help make an open data society possible.
- Use of standardised formats such as RDF can make implementing linked data on a large scale easy.
References

http://opendatahandbook.org/en/what-is-open-data
http://www.w3.org/TR/2004/REC-rdf-primer-20040210/
http://www.w3.org/DesignIssues/LinkedData.html
http://www.wais.ecs.soton.ac.uk/projects
http://eprints.soton.ac.uk/cgi/export/eprint/346929/RDFXML/eps-eprint-346929.rdf
http://eprints.soton.ac.uk/cgi/export/eprint/346929/RDFNT/eps-eprint-346929.nt
http://id.ecs.soton.ac.uk/docs/
http://www.w3.org/TR/rdf-sparql-query/
http://microformats.org/wiki/Main_Page
http://www.w3.org/TR/xhtml-rdfa-primer/
http://schema.org/
http://www.w3.org/2001/sw/
http://www.deloitte.com/view/en_GB/uk/market-insights/deloitte-analytics/de0d882044ad7310VgnVCM3000001c56f00aRCRD.htm
http://wheredoesmymoneygo.org/