JISC cetis

centre for educational technology
& interoperability standards

The Semantic Web, Linked and Open Data

A Briefing Paper

By Lorna M. Campbell and Sheila MacNeill

The development of the Semantic Web has been a long running project championed by the inventor of the web, Sir Tim Berners-Lee. It is built around his concept of the 'Web of data' which means moving on from the existing document centric view of the Web to a data centric one. In this vision of the Web, data and the relationships between data are key. Coupled with these ongoing Semantic Web developments there has also been growing interest in the related areas of linked and open data.

Between 2009 and 2010 both the UK and US governments launched high profile projects to release a wide range of publicly funded government information as open data sets. There is considerable potential for the education sector to use and contribute to these data sets as they become available. There are also potential benefits for institutions in using the principles of open and linked data in a number of key areas such as institutional administration, teaching, learning and research. However, there is still a degree of confusion regarding the key concepts of the Semantic Web and linked and open data, as well as a range of views on approaches to implementation.

This briefing paper will provide a high level overview of key concepts relating to the Semantic Web, semantic technologies, linked and open data; along with references to relevant examples and standards. The briefing is intended to provide a starting point for those within the teaching and learning community who may have come across the concept of semantic technologies and the Semantic Web but who do not regard themselves as experts and wish to learn more. The examples and links are intended as starting points for further exploration.

http://jisc.cetis.ac.uk

The Semantic Web

DEFINITION

"The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C. It is based on the Resource Description Framework (RDF)."
W3C, http://www.w3.org/2001/sw/

OVERVIEW

The World Wide Web Consortium (W3C) define the main goals of the Semantic Web as follows:

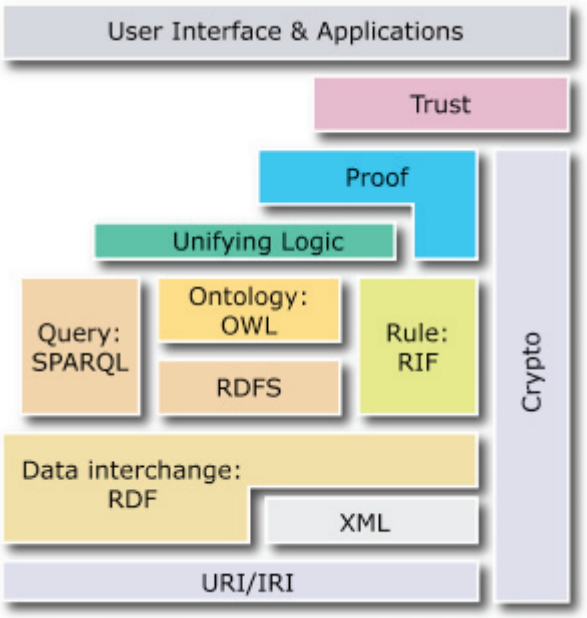
"The Semantic Web is a Web of Data. The vision of the Semantic Web is to extend the principles of the Web from documents to data. Data should be accessed using the general Web architecture using, e.g., URIs; data should be related to one another just as documents (or portions of documents) are already. This also means creation of a common framework that allows data to be shared and reused across application, enterprise, and community boundaries, to be processed automatically by tools as well as manually, including revealing possible new relationships among pieces of data.

In order to achieve theses goals, it is necessary to define and describe the relations among data on the Web. This is not unlike the usage of hyperlinks that connect the current page with another page: the hyperlinks defines a relationship between the current page and the target. On the Semantic Web, such relationships can be established between any two named resources or values and the relationship itself (i.e, the link) is also named.

By contrast, a link on the traditional Web is not named, which means that the significance or meaning of that link needs to be deduced by the human reader. The naming and defining those relations explicitly enables better and automatic interchange of data. RDF, which is one of the fundamental building blocks of the Semantic Web, gives a formal definition for that interchange."
W3C, http://www.w3.org/2001/sw/SW-FAQ

This ability to dynamically describe complex and evolving concepts, resources and relationships could have a profound impact on how we all access and use data.

The Semantic Web is frequently represented by the following "layer cake" diagram:



http://www.w3.org/2007/03/layerCake-small.png

RELEVANT STANDARDS:

Uniform Resource Identifier (URI): http://www.ietf.org/rfc/rfc2396.txt

Resource Description Framework: http://www.w3.org/TR/REC-rdf-syntax/

RDF Vocabulary Description Language 1.0: RDF Schema: http://www.w3.org/TR/rdf-schema/

SPARQL Query Language for RDF: http://www.w3.org/TR/rdf-sparql-query/

OWL 2 Web Ontology Language: http://www.w3.org/TR/owl2-overview/

Rule Interchange Format Core Dialect: http://www.w3.org/TR/rif-core/

Simple Knowledge Organisation System: http://www.w3.org/TR/skosreference/

EXAMPLES:

A Linked Open Data Resource List Management Tool for Undergraduate Students: http://www.w3.org/2001/sw/sweo/public/UseCases/Talis/

FURTHER READING:

W3C Semantic Web Activity: http://www.w3.org/2001/sw/

W3C Semantic Web FAQ: http://www.w3.org/2001/sw/SW-FAQ

Semantic technologies

DEFINITION

"Semantic technologies encode meaning separately from data and content files, and separately from application code. This enables machines as well as people to understand, share and reason with them at execution time."

Wikipedia, http://en.wikipedia.org/wiki/Semantic_technology

OVERVIEW

There are no hard and fast rules as to what does and does not constitute semantic technologies, rather semantic technologies may be placed along a continuum. Depending on how information and semantics are modelled and expressed some technologies are capable of considerably greater semantic processing and reasoning than others.

RELEVANT STANDARDS:

All the Semantic Web standards are relevant to semantic technolgies although few technologies will support them all. Other relevant standards include:

RDF Site Summary 1.0: http://web.resource.org/rss/1.0/spec

Topic Maps ISO / IEC 13250: http://www.isotopicmaps.org/

EXAMPLES:

PROWE: developed a recommended metadata profile related to a Friend Of A Friend (FOAF) vocabulary to allow remote, part-time tutors to collaborate using 'blikis' (a combination of blogs and wikis).
http://www.prowe.ac.uk/

COMPENDIUM is a semantic, visual hypertext tool for supporting collaborative domain modelling and real time meeting capture. It provides an intuitive visual map of issues, ideas and arguments. It can be thought of as real time concept mapping, but with additional native semantic hypertext functionality. Compendium is interoperable via XML with other technologies.
http://compendium.open.ac.uk

AWESOME is a semantic wiki run by the University of Leeds with partners in the Universities of Coventry and Bangor under the sponsorship of JISC. AWESOME aims at supporting the academic writing process through social online mediated environments. AWESOME uses both Web 2.0 technologies and Semantic Web technologies such as tagging, folksonomies, content annotation, blogs, chat, ontologies and semantically enabled search.
http://awesome.alexlebek.co.uk/wiki/

FURTHER READING:

JISC Semantic Technologies in Teaching and Learning (SemTech) Project: http://www.semtech.ecs.soton.ac.uk/

JISC SemTech Project Report: http://www.jisc.ac.uk/media/documents/projects/semtech-report.pdf

CETIS Semantic Technology Resources: http://jisc.cetis.ac.uk/topic/semantic_technologies

Linked Data, linked data, linkable data

DEFINITION

"To make the Semantic Web or Web of Data a reality, it is necessary to have a large volume of data available on the Web in a standard, reachable and manageable format. In addition the relationships among data also need to be made available. This collection of interrelated data on the Web can also be referred to as Linked Data. Linked Data lies at the heart of the Semantic Web: large scale integration of, and reasoning on, data on the Web."
W3C, http://www.w3.org/standards/semanticweb/data

OVERVIEW

There is currently considerable ambiguity as to the exact nature of Linked Data. The debate primarily centres around whether Linked Data must adhere to the four principles outlined in Tim Berners-Lee's "Linked Data Design Issues", and in particular whether use of RDF and SPARQL is mandatory. Some argue that RDF is integral to Linked Data, others suggest that while it may be desirable, use of RDF is optional rather than mandatory. Some reserve the capitalised term Linked Data for data that is based on RDF and SPARQL, preferring lower case "linked data", or "linkable data", for data that uses other technologies. There is currently no definitive practice.

The fact that the "Linked Data Design Issues" paper is a personal note by Tim Berners-Lee, and is not formally endorsed by W3C also contributes to the ambiguity. The note states:

- **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.**

"I'll refer to the steps above as rules, but they are expectations of behavior. Breaking them does not destroy anything, but misses an opportunity to make data interconnected. This in turn limits the ways it can later be reused in unexpected ways. It is the unexpected re-use of information which is the value added by the web."

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

In the recent JISC "Linked Data Horizon Scan" Paul Miller notes:

"Whilst the exact wording of these statements has changed slightly since first expressed in 2006, and there remains some doubt as to the strength of the requirement for specific standards, the acronyms mask a simple yet powerful set of behaviours;

- **Name objects and resources, unambiguously;**
- **Make use of the structure of the web;**
- **Make it easy to discover information about the named object or resource;**
- **If you know about related objects or resources, link to them too.**

There is much to gain in embracing the philosophy behind these rules, separately to adopting the standards and specifications required to realise their full potential."
Miller, http://linkeddata.jiscpress.org/

Since the vision of Linked Data pre-supposes a means for any piece of data to be able to link to any other piece, perhaps the distinction is simply between data that has been converted into Linked Data already, and data that has not been converted yet, but could be quite easily, perhaps on the fly.

RELEVANT STANDARDS:

Uniform Resource Identifier (URI): http://www.ietf.org/rfc/rfc2396.txt

Resource Description Framework: http://www.w3.org/TR/REC-rdf-syntax/

Hypertext Transfer protocol (HTTP): http://www.w3.org/Protocols/

SPARQL Query Language for RDF: http://www.w3.org/TR/rdf-sparql-query/

EXAMPLES:

DBpedia in essence, makes the content of Wikipedia available in RDF. DBpedia not only includes Wikipedia data, but also incorporates links to other datasets on the Web, e.g. Geonames.
By providing those extra links (in the form of RDF triples) it allows applications to exploit the extra (and possibly more precise) knowledge from other datasets. More importantly, by exploiting the links within and between the datasets, applications can find new information that would be very difficult to extract in any other way.
http://www.w3.org/standards/semanticweb/data#examples
http://dbpedia.org

Linked Data which connects linked data across the web, is a site that provides "a home for, or pointers to, resources from across the Linked Data community". Information, tutorials, data sets, tools, events are available.

http://linkeddata.org

FURTHER READING:

When is Linked Data not Linked Data? A summary of the debate: **http://blogs.cetis.ac.uk/lmc/2010/03/16/when-is-linked-data-not-linked-data-a-summary-of-the-debate/**

W3C Semantic Web Linked Data: **http://www.w3.org/standards/semanticweb/data**

Tim Berners-Lee's Linked Data Design Issues: **http://www.w3.org/DesignIssues/LinkedData.html**

Linked Data - the story so far: **http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf**

Linked Data Horizon Scan: **http://linkeddata.jiscpress.org**

Open Data

DEFINITION

There is no agreed standard definition for the term open data, however the Open Knowledge Definition is beginning to gain acceptance:

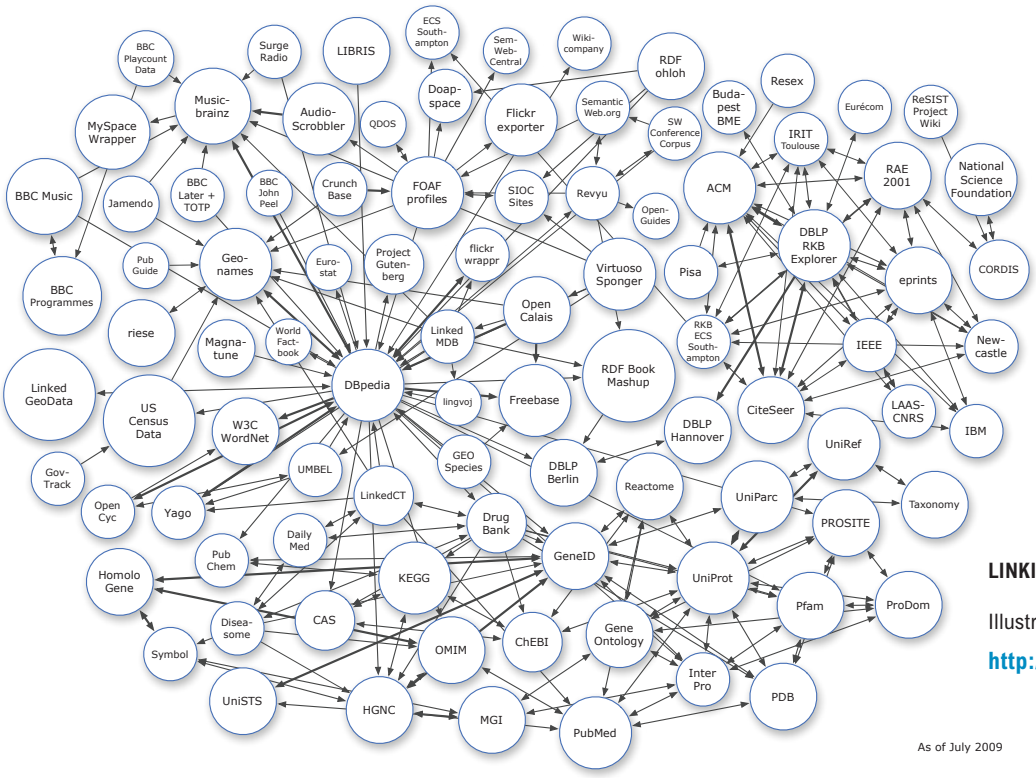
The Open Knowledge Definition (OKD) sets out principles to define "openness" in knowledge – "from sonnets to statistics, genes to geodata". This covers all forms of data as well as content such as music, films, books, images and other material. It can be summed up in the statement that "A piece of knowledge is open if you are free to use, reuse, and redistribute it".

OKD, **http://www.opendefinition.org**

OVERVIEW

"Open Data is a philosophy and practice requiring that certain data are freely available to everyone, without restrictions from copyright, patents or other mechanisms of control. It has a similar ethos to a number of other "Open" movements and communities such as open source and open access. However these are not logically linked and many combinations of practice are found. The practice and ideology itself is well established but the term "Open Data" itself is recent."

Wikipedia, **http://en.wikipedia.org/wiki/Open_data**



As of July 2009

LINKING OPEN DATA SET CLOUD

Illustration copyright Richard Cyganiak July 2009

http://richard.cyganiak.de/2007/10/lod/

THE SEMANTIC WEB, LINKED AND OPEN DATA

The relationship between Linked Data and Open Data is currently ambiguous. The JISC Linked Data Horizon Scan states:

"Linked Data may be Open, and Open Data may be Linked, but it is equally possible for Linked Data to carry licensing or other restrictions that prevent it being considered Open, or for Open Data to be made available in ways that do not respect all of Berners-Lee's rules for Linking."
Miller, http://linkeddata.jiscpress.org/

The concept of open data has risen in prominence in the UK following the government's 2009 / 2010 data.gov.uk initiative which aims to open access to public-sector government data, and slowly turn them into Linked Open Data where appropriate.

RELEVANT STANDARDS:

In addition to those standards relevant to Linked Data, arguably the standards that are most critical to Open Data are open licences. Licensing of data needs to be explicit to avoid any ambiguity in terms of use and re-use. There are many differences worldwide relating to the copyright of data. For example in the European Union database rights exist automatically, whereas in the USA data is not covered by existing copyright law. Guidance is available from a number of sources including the following:

Creative Commons: http://creativecommons.org/

GNU Free Documentation License: http://www.gnu.org/copyleft/fdl.html

Public Domain Dedication and License: http://www.opendatacommons.org/licenses/pddl/

Science Commons Database Protocol: http://sciencecommons.org/resources/faq/database-protocol

Freedom to Research: Keeping Scientific Data Open, Accessible, and Interoperable: http://sciencecommons.org/wp-content/uploads/freedom-to-research.pdf

EXAMPLES:

MusicBrainz is a project that aims to create an open content music database by capturing information about artists, their recorded works, and the relationships between them. MusicBrainz's core data is in the public domain, and additional content including moderation data is placed under the Open Audio License. The server software is covered by the GNU General Public License. The MusicBrainz client software library is licensed under the GNU Lesser General Public License.
http://musicbrainz.org/

Guardian Data Store is a growing database of open data sets used by the Guardian and available to download in a variety of formats. It supports active community engagement and sharing of how data sets are being used and reused.
http://www.guardian.co.uk/data-store

OpenStreetmap provides free worldwide geographical information such as maps. Currently the site uses creative commons licensing. Global community contribution is actively encouraged.
http://wiki.openstreetmap.org/wiki/Main_Page

OS Open Data provides a selection of open Ordnance Survey mapping datasets for Great Britain. In addition, an open API (OS OpenSpace) also allows the integration of OS maps with other websites. The OS Open Data licence is designed to be interoperable with Creative Commons to enable the recombination of data.
http://www.ordnancesurvey.co.uk/oswebsite/opendata/index.html

BBC Backstage is a collection of open BBC data feeds and APIs available for re-mixing and mash-ups. The site aims to "encourage participation and support creativity through open innovation." There are some licence restrictions (mainly around commercial use) which are clearly defined on the site.
http://ideas.welcomebackstage.com/data

FURTHER READING:

Open Data Commons: http://www.opendatacommons.org/

Open Knowledge Definition: http://www.opendefinition.org/okd/

Open Knowledge Foundation: http://www.okfn.org/

data.gov.uk: http://data.gov.uk/

JISC CETIS, the Centre for Educational Technology and Interoperability Standards, provides advice to the UK Higher and Post-16 Education sectors on educational technology and standards.

 PRODUCED JUNE 2010