

Linked Data: the royal road to Open Government

Dov Winer

MAKASH – Advancing ICT Applications

dov.winer @ gmail.com

Summary of a proposed presentation at the WG2 Meeting of the action

COST TU1204 – People friendly cities in a data rich world

Madrid, December 16th 2013

The Linked Data model will be presented to stimulate a discussion within the TU1204 community concerning its critical role in making open data usable and as a trigger for the development of new services able to increase social cohesion and citizen's participation.

Lee and Kwak (2011)¹ describe four stages of open government: (1) increasing data transparency (2) improving open participation (3) enhancing open collaboration (4) realising ubiquitous engagement. Making government information available increases transparency by breaking down information barriers and allowing citizens to participate in a meaningful way.² This openness allows for closer collaboration with the community and increasingly involved citizens. Open government can form a new relationship between public administrations and citizens through increased transparency, accessibility of information and responsiveness to citizen's needs.³

Peter Hinssen points out that there is no use in providing information if it cannot be understood or used. Publishing proprietary or closed data and calling it 'Open Data' is a no go. You need open formats that can be linked with one another, and with other openly available data that comes from other sources than yours. It is precisely in this 'Linked Data' that real insights can be found.⁴

The main concepts involved in the Linked Data approach will be reviewed and then some summary indications of its applications in local government and other actions concerning Linked Data in government indicated. The identification of examples of local government applications will continue and is expected to be expanded by the time of WG2 meeting.

Linked Data Concepts

Linked Data is a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF⁵. Since 2007 it became a

¹ G. Lee, Y.H. Kwak (2011). An Open Government Implementation Model: Moving to Increased Public Engagement. IBM Center for the Business of Government. Retrieved on November 14, 2013 from:

http://observgo.quebec.ca/observgo/fichiers/73044_3.pdf

² Open Data Guidebook: final report of the Eurocities working group on open data. October 2013. Retrieved on November 14, 2013 from:

http://nws.eurocities.eu/MediaShell/media/WG%20open%20data%20report_open%20data%20guidebook%20_%20EUROCITIES%202013.pdf

³ Public Sector Modernisation: Open Government. OECD Policy Brief. Retrieved on November 14, 2013 from:

<http://www.oecd.org/gov/pem/34455306.pdf>

⁴ P. Hinssen (Editor). Open Data Powers Smart Cities. Data Science Series. EMC Greenplum.

http://datascienceseries.com/assets/blog/Greenplum-Open_Data_Power_Smart_Cities-web.pdf

⁵ Linked Data – Connect Distributed Data Across the Web <http://linkeddata.org/>

http://en.wikipedia.org/wiki/Linked_data

bandwagon – large databases are published every week - with more than 40 billion assertions now available.

Tim Berners-Lee summarized the main Linked Data guidelines as follows⁶:

- Use URIs as names for things.
- Use HTTP URIs so that people can look up those names.
- When someone looks up a URI, provides useful information, using the standards (RDF, SPARQL)
- Include links to other URIs, so that they can discover more things.

The article “[How to Publish Linked Data on the Web](#)” (Bizer, Cyganiak and Heath, 2007) provide detailed instruction. Two recent articles provide a summative evaluation of the process, identify pending issues and providing reference to the software tools that can be employed: Bizer, Heath, Berners-Lee (2009); Hausenblas and Karnstedt (2010). Here I will summarily describe the basic concepts involved:

URI

A Uniform Resource Identifier⁷ (URI) is a compact segment of characters that identifies an abstract or physical resource. The URI syntax defines a common grammar that is a superset of all valid URIs, allowing an implementation to parse the common components of a URI reference without knowing the scheme-specific requirements of every possible identifier. The term “resource” is used in a general sense for whatever might be identified by a URL – electronic resources as well as human beings, corporations and books; abstract concepts can be resources, as for example types of relationships (e.g., “parent” or “employee”) or numerical values. An identifier embodies the information required to distinguish what is being identified from all other things within its scope of identification.

While URIs may or may not be useful as locators in practice⁸, a URI scheme definition must be clear as to how it is expected to function. Schemes that are not intended to be used as locators should describe how the resource identified can be determined or accessed by software that obtains a URI of that scheme.

In many cases new URI schemes are defined as ways to translate between other namespaces or protocols and the general framework of URIs. For example the “ftp” URI scheme translates into the FTP protocol. For such schemes the description of the mapping must be complete, and in sufficient detail so that the mapping in both directions is clear: how to map from a URI into a set of protocol actions or name in the base namespace and how legal name values or protocol interactions might be represented in a valid URI.

As part of the definition of how a URI identifies a resource, a URI scheme definition should define the applicable set of operations that may be performed on a resource using the URI as its identifier.

HTTP URIs

The HTTP protocol is used by web servers and clients to request representations of Web documents and send back responses. HTTP includes capabilities to offer different formats and language versions of the same Web document and this process is known as “content negotiation”. This capability provides a recommended solution to the problem of access to identifiers (URI) of resources that are not documents such as persons, corporations, books, concepts. Such identifiers should be distinguished from representations of documents and photographs. This makes it possible assertions resident at distinct systems through the Semantic Web, that are made about the same object or concept, to be retrieved and processed, enabling the richness of services that may be offered.

The W3C Interest Group note of December 2008, “Cool URIs for the Semantic Web” recommends two

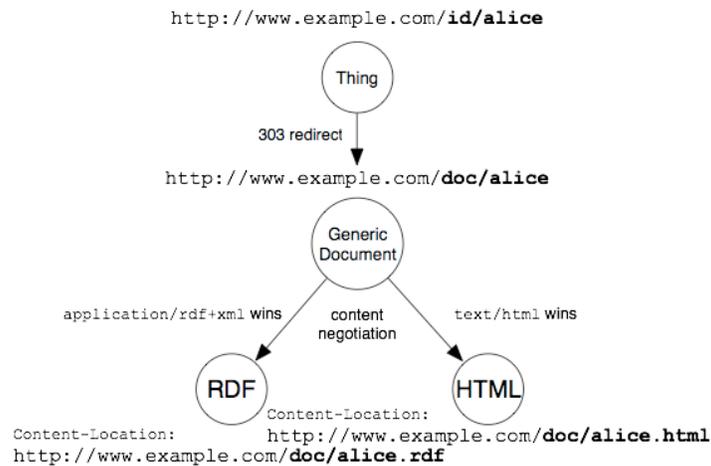
⁶ Tim Berners-Lee (2006). Linked Data. <http://www.w3.org/DesignIssues/LinkedData.html>

⁷ Uniform Resource Identifier (URI): Generic Syntax.RFC 3986, January 2005 <http://tools.ietf.org/html/rfc3986>

⁸ Guidelines and Registration Procedures for New URI Schemes. RFC 4395, February 2006 (page 6)
<http://tools.ietf.org/html/rfc4395>

approaches for making such references, Hash URIs and 303 Re-direction URIs. The following picture taken from that document nicely describes the procedure called Content Negotiation.

The URI for Alice the person is `http://www.example.com/id/alice`. The Web client that addresses this URI negotiates with the server; if the client is a human user using a regular browser the 303 redirect procedure leads him to a document that may be a description of Alice. If the client is a machine able to process RDF, it is redirected to an RDF representation of Alice that can, for example, be fetched and integrated in the service that originated the lookout – a database or a mashup for example.



Namespace

Namespace is an abstract container providing context for the items (names, or technical terms, or words) it holds and allowing disambiguation of homonym items having the same name (residing in different namespaces). A namespace is also called a context, as the valid meaning of a name can change depending on what namespace applies. Names in it can represent objects as well as concepts. For many programming languages, a namespace is a context for identifiers. In an operating system, an example of namespace is a directory. It contains items which must have unique names.⁹

RDF

RDF (Resource Description Framework) provides a way to express simple statements about resources, using named properties and values.¹⁰ It can be used to represent information about things that can be **identified on the Web**, even when they **cannot be directly retrieved** on the Web. RDF is intended for situations in which this information needs to be processed by applications, rather than being only displayed to people.

RDF is based on the idea of identifying things using Web identifiers (called Uniform Resource Identifiers, or URIs) and describing resources in terms of simple properties and property values. This enables RDF to represent simple statements about resources as a graph of nodes and arcs representing the resources, and their properties and values. This is represented in the form of subject-predicate-object expressions and are known as *triples* in RDF terminology.

For example, one way to represent the notion "The sky has the color blue" in RDF is as the triple: a subject denoting "the sky", a predicate denoting "has the color", and an object denoting "blue". RDF is an abstract

⁹ Namespace <http://en.wikipedia.org/wiki/Namespace>

¹⁰ RDF Primer – W3C Recommendation 10 February 2004 <http://www.w3.org/TR/rdf-primer/>

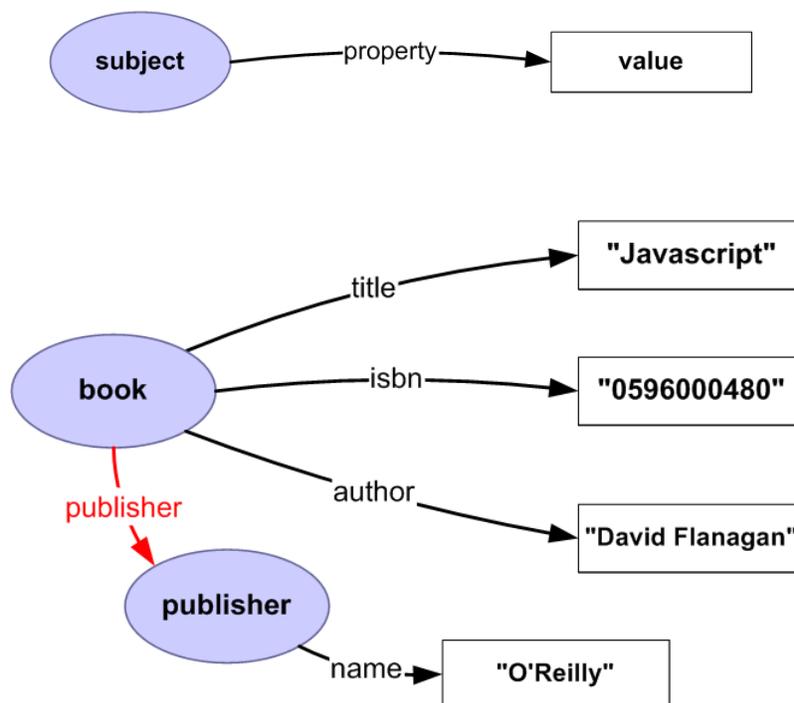
model with several serialization formats (i.e., file formats), and so the particular way in which a resource or triple is encoded varies from format to format.¹¹

A good tutorial on RDF was prepared by Ian Davis from Talis Research¹². The following example is taken from this tutorial.

In a relational database model the intersection of a row and a column in a table gives the value of a property (in the example below the Title) for a given thing (the book whose ISBN is 0596000480). The **book** has a **title** with a value of **"Javascript"**

isbn	title	author	publisherID	Pages
0596002637	Practical RDF	Shelley Powers	7642	350
0596000480	JavaScript	David Flanagan	3556	936
...
...
...

This is the essence of RDF: the (s,p,v) triple



An (s,p,v) triple can be viewed as a labeled edge in a graph. Nodes in graph are things, arcs are relationships between things. Such graphs are amenable to graph algebraic operations like those described in the specifications of the RDF query language, SPARQL¹³

¹¹ Resource Description Framework – Wikipedia entry - http://en.wikipedia.org/wiki/Resource_Description_Framework

¹² An Introduction to RDF by Ian Davis, Talis Information Ltd, <http://research.talis.com/2005/rdf-intro/>

¹³ SPARQL Query Language for RDF – W3C Recommendation 15 January 2008 - <http://www.w3.org/TR/rdf-sparql-query/>

Linked Data in Local Government

1. Putting Government Data Online by Tim Berners-Lee¹⁴

Government data is put online typically for 3 reasons:

1. Increasing citizen awareness of government functions to enable greater accountability;
2. Contributing valuable information about the world; and
3. Enabling the government, the country, and the world to function more efficiently.

Each of these purposes is best served by using Linked Data techniques. In general Linked Data is:

Open: Linked Data is accessible through an unlimited variety of applications and applications because it is expressed in open, non-proprietary formats.

Modular: Linked Data can be combined (mashed-up) with any other piece of Linked Data. For example, government data on health care expenditures for a given geographical area can be combined with other data about the characteristics of the population of that region in order to assess effectiveness of the government programs. No advance planning is required to integrate these data sources as long as they both use Linked Data standards.

Scalable: It's easy to add more Linked Data to what's already there, even when the terms and definitions that are used change over time.

The essential message is that whatever data format people want the data in, and whatever format they give it to you in, you use the RDF model as the interconnection bus. That's because RDF connects better than any other model.

- It uses URIs and so allows linking of things and concepts
- It allows separate systems designed independently to be later joined at the edges
- It allows interoperability to be added where cost-effective
- It allows any data to be expressed in a mixture of vocabularies.

That's enough about why it is useful. That is elaborated elsewhere, but it can be difficult for those familiar with other technologies to understand the difference. Sometimes it is better just to do it.

2. W3C Government Linked Data (GLD) Working Group¹⁵

The **mission** of the W3C Government Linked Data (GLD) Working Group¹⁶ is to provide standards and other information which help governments around the world publish their data as effective and usable Linked Data using Semantic Web technologies. The group, a part of the eGovernment Activity¹⁷ and closely connected with the Semantic Web Activity, will collect and make available information about government Linked Data activities around the world.

¹⁴ <http://www.w3.org/DesignIssues/GovData.html>

¹⁵ http://www.w3.org/2011/gld/wiki/Main_Page

¹⁶ <http://www.w3.org/2011/gld/>

¹⁷ <http://www.w3.org/2007/eGov/>

3. Open Data Communities: Open Access to Local Data

This site of the UK Department for Communities and Local Government's official Linked Open Data site¹⁸ provides a selection of statistics on a variety of themes including Local Government finance, housing and homelessness, wellbeing, deprivation, and the department's business plan as well as supporting geographical data. All of the data is available as fully browsable and queryable Linked Data, and the majority is free to re-use under the Open Government Licence.

4. Local Government Association: Local Transparency/Spending¹⁹

Linked data: making better use of data²⁰

[Local Government Service List](#) – on the Electronic Service Delivery (ESD) Standards website

A general 'ontology', or 'linked data schema', for payments has been devised by the Local e-Government Standards Body and data.gov.uk. This can be used to define local expenditure data. Publishing the data according to a given standard – as proposed in the template in appendix A – will help to build a linked data set and make use of linkages to other datasets, such as the Local Government Service List.

[Guide to the payments ontology](#) – on data.gov.uk

The Electronic Service Delivery (ESD) Toolkit is developing and hosting a simple tool to express spending data as open linked data. The tool will also convert data into a linked data format.

The ESD Toolkit already provides datasets that enable local authorities to compare performance to a consistent standard. Local authorities may find these useful to compare their service performance and improve services. The ESD Toolkit also links to other datasets and sources – for example socio-demographic and lifestyle data provided – such as 'output area classification' (OAC), Mosaic or Acorn.

The pilot ESD Toolkit Linked Data Tool is freely available to local authorities. You can access it, and a sample spreadsheet used for upload, on the ESD Toolkit website. The tool will allow you to upload expenditure data in CSV format according to the expenditure template ([see Appendix A](#)). The tool will link the data to other information held in the Toolkit by other organisations. Examples of this include authority type, total expenditure, local authority service code. It will then output the data in linked format.

[Payments open data](#) – on the ESD Toolkit

With this additional information linked to the expenditure data, further analysis can be carried out on expenditure type. For example you could compare expenditure in a service area with another local authority, or compare company spend with those in other authorities etc.

5. Zaragoza

Since 2003, the city of Zaragoza²¹ has been developing an ambitious project, the aim of which is to become a knowledge society. The project covers a wide variety of areas ranging from the implementation of telecommunication infrastructures to Electronic Administration and the encouragement of entrepreneurship and innovation. Some of the purposes of its programme for Knowledge Society Development include:

- Use the new technologies to provide more efficient public services.

¹⁸ <http://opendatacommunities.org/>

¹⁹ <http://localspending.readandcomment.com/>

²⁰ <http://localspending.readandcomment.com/4-linked-data-making-better-use-of-data-2/>

²¹ Open government strategy in the Digital City 2012-2015. Smart Citizenship, Ayuntamiento Zaragoza, June 2012. Retrieved on November 14, 2013 from: <http://www.zaragoza.es/contenidos/sectores/tecnologia/Estrategia-Ciencia-Tecnologia-en.pdf>

- Guarantee all citizens, entities, companies and institutions the exercise of their full rights of access to the communications networks.
- Make the most of technological change to develop a new, cohesive city planning and create economic opportunities.

The aim of Zaragoza City Council's Open Data policy is that the data be: unique, shared, accessible and re-usable by third parties.

What has Zaragoza done in terms of Open Data?

<http://datosabiertos.zaragoza.es> is a 2010 Zaragoza City Council initiative to encourage the **re-use of the information posted on its web site**. This project includes:

Data catalogue. It includes all of the data sets (posted on the web site) ordered by subject, tags and by representation formats. Re-user agents can search for sets of data according to their preferences. The idea is to offer any public item of information in as many formats as is possible, always seeking the most open ones in order to avoid restricting usage.

SPARQL Protocol and RDF Query Language (SPARQL). This affords developers/professionals enormous potential and flexibility when it comes to constructing quality applications. In addition to providing top quality data in the Linked Data format (Web 3.0 or Semantic Web Technology), it has query mechanisms for these data. As far as SPARQL is concerned, this is a standard query language that enables access to all of the data that are provided in Linked Data format.

Applications. The developer community can notify the City Council about the applications (apps) that they have developed. These are the ones that go to make up the catalogue of applications that use the city's public data.

SOLR classified data. The Apache Solr platform has been implemented as a search platform for indexing and posting Open Data on the web site.

Citizen Query. A forum that encourages dialogue and participation of any party interested in the related issues. This medium was opened on <http://consultadatos.zaragoza.es> at the beginning of July 2010. During the first month over 30 ideas were submitted, and it received 500 votes or more.