

Utilising Linked Open Data in Applications

Michael Hausenblas
Linked Data Research Centre, DERI, NUI Galway
IDA Business Park, Lower Dangan, Galway, Ireland
michael.hausenblas@deri.org

ABSTRACT

Over the past years, the Linked Data principles have successfully been applied, resulting in the Linked Open Data (LOD) cloud with over 200 datasets containing 26 billion statements, connected by 400 million typed links. Publishing into the LOD cloud is certainly a challenge, however various tutorials and guidelines exist.

This tutorial addresses the consumption part, that is, it gives an introduction how to use data from the LOD cloud in Web applications. We start with simple access methods (follow-your-nose, using dumps) over querying LOD data using SPARQL, to client-side access in the browser (using JavaScript) as well as show the integration steps into a popular Content Management System (Drupal).

Categories and Subject Descriptors

H.3.5 [Information Storage and Retrieval]: Online Information Services—*Data sharing*; H.3.5 [Information Storage and Retrieval]: Online Information Services—*Web-based services*

General Terms

Experimentation, Standardization, Verification

Keywords

Linked Data, Web of Data, Applications, Consumption

1. LINKED DATA WEB

Linked Data [2] refers to a set of best practices for publishing and interlinking structured data on the Web, introduced in 2006 by Tim Berners-Lee. In a nutshell, Linked Data applies the general architecture of the World Wide Web to the task of sharing structured data on a global scale. Technically, Linked Data is about employing Uniform Resource Identifiers (URIs), the Resource Description Framework (RDF) and the Hypertext Transfer Protocol (HTTP)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WIMS'11, May 25–27, 2011 Sogndal, Norway

Copyright 2011 ©ACM 978-1-4503-0148-0/11/05 ... \$10.00.

to publish structured data on the Web and to connect related data that is distributed across multiple data sources. Similar to hyperlinks in the Web of Documents creating a single global information space, Linked Data provides for setting typed links between data items in different data sources, thereby connecting these sources into a single global data space—the *Web of Data* [3].

Increasing numbers of data providers and application developers have begun to adopt Linked Data [4]. In doing so they have created a global, interconnected data space. Just as the classic document Web is diverse in its topical coverage, so the Web of Data spans numerous topical domains, containing data about people, companies, films, music, locations, books and other publications, online communities, as well as an increasing amount of scientific and government data.

Figure 1 gives an overview of the Web of Data. Each node in the diagram represents a distinct data set published as Linked Data. The arcs indicate the existence of links between items in the two data sets. As of September 2010, the Web of Data is estimated to contain 26 billion RDF triples and over 400 million cross-data set links [5].

Relatively recent additions to the Web of Data, such as *Ordnance Survey*, *New York Times* and various government data sets from the UK and US, demonstrate how the Web of Data is evolving from data publication primarily by third party enthusiasts and researchers, to data publication *at source* by large media and public sector organisations.

A key challenge now is to build applications based on the Web of Data [7, 8]—this tutorial is one first step to get researchers and practitioners up to speed with the possibilities and their shortcomings.

2. TUTORIAL TOPICS

The tutorial focuses on the consumption part of Linked Open Data. In the following, some typical scenarios are discussed, which form the core topics of the tutorial.

2.1 Discovery

One main tenant of Linked Data are URIs. However, quite often one is only equipped with a term, that is, for example a word in a certain language. In such a situation a *look-up service* can be used that takes a term as an input and returns one or more URIs as result. We note that look-up service might also take URIs as input and deliver documents or resources that use these input URIs. Several types of look-up services can be distinguished:

- Simple, dataset-specific look-up services, such as DB-

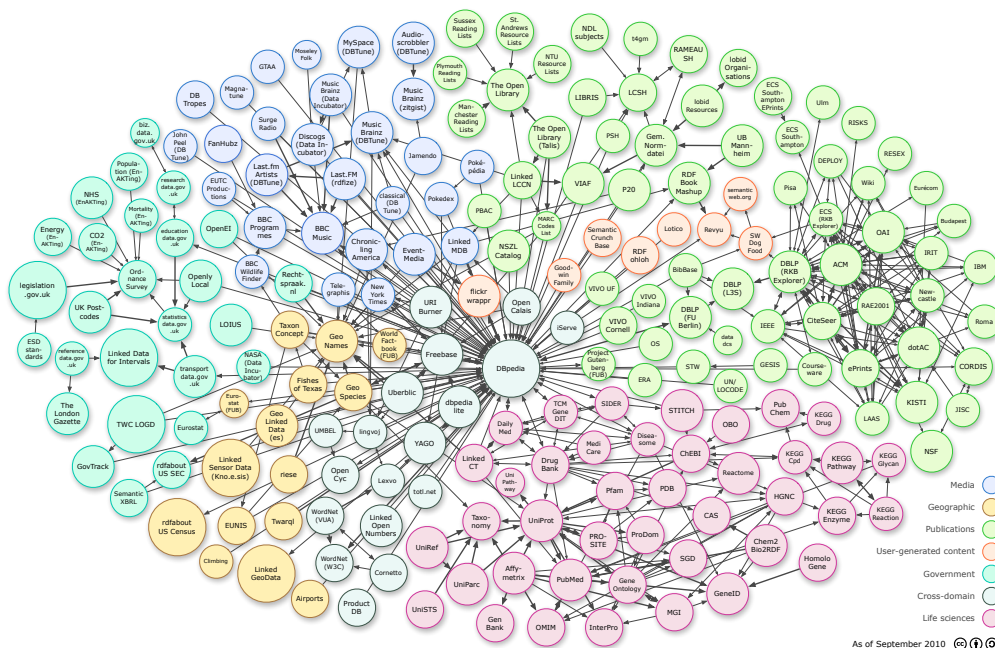


Figure 1: Topology of the Web of Data (September 2010) with more than 26 billion facts and over 400 million cross-data set links from over 200 datasets.

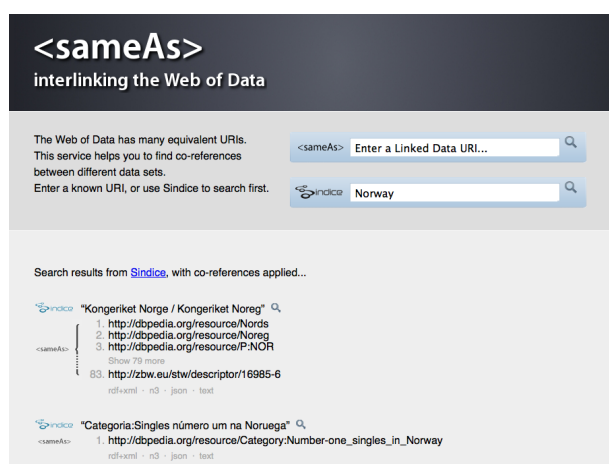


Figure 2: The sameAs.org look-up service.

pedia Lookup¹. These services typically take a term as an input and return one or more URIs of resources that potentially are of interest regarding this term;

- Generic look-up services (or semantic indexer), such as Sindice [9] or Watson [6] that allow to look-up both terms and URIs, typically covering large parts of the LOD cloud;
- Specialised, link type or entity look-up services, such as sameAs.org², see Fig. 2.

¹<http://lookup.dbpedia.org/>

²<http://sameas.org>

The discovery process on the dataset-level can be made more effective and efficient using metadata. A widely deployed way to express dataset metadata is based on the “Vocabulary of Interlinked Datasets” (VoID) [1]. VoID descriptions³ can contain dataset characteristics and interlinking information, as well as licensing, provenance and statistics about the dataset.

Community-driven services, such as the *Comprehensive Knowledge Archive Network* (CKAN)⁴ (Fig. 3) allow users to find datasets, based on topics, license and quality indicators. Additionally, these services typically also expose the dataset metadata in VoID, enabling an automation of the discovery task.

2.2 Follow-Your-Nose

Linked Data has an intrinsic discovery mechanism, often referred to as “Follow-Your-Nose” (FYN)⁵. The FYN method is a native, but not very efficient way to discover resources in RDF⁶. There are also other issues with it, such as with the trust aspect, but for certain use cases FYN is ideal.

Based on FYN, the Linked Data Web can be accessed using Linked Data browsers and aggregators, such as Sig.ma (Fig. 4). However, instead of following links between HTML pages, Linked Data browsers enable users to navigate between different data sources by following RDF links. This allows the user to start with one data source and then move through a potentially endless Web of data sources connected by RDF links.

³<http://www.w3.org/TR/void/>

⁴<http://ckan.net/>

⁵<http://esw.w3.org/topic/FollowLinksForMoreInformation>

⁶<http://webodata.wordpress.com/2009/03/01/effective-and-efficient-discovery/>

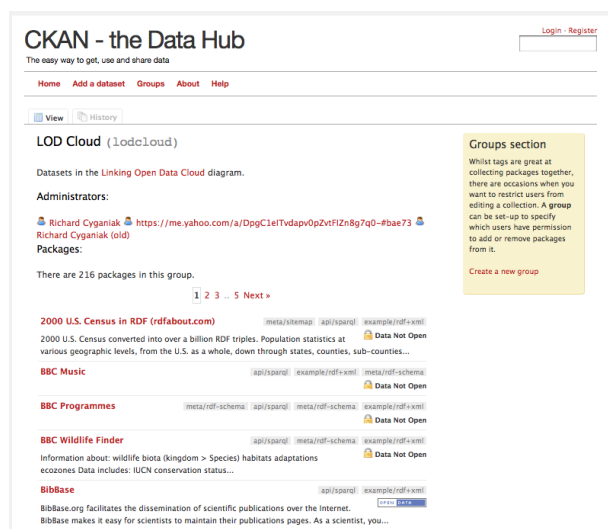


Figure 3: CKAN, a community-driven dataset discovery service.

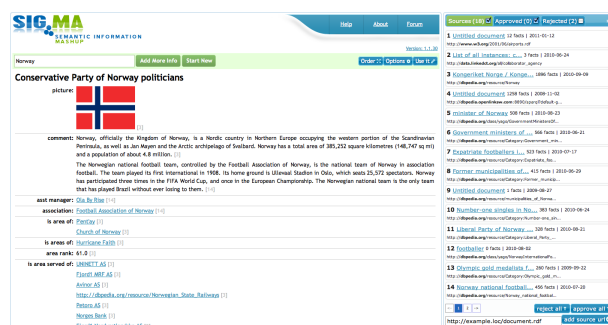


Figure 4: Sig.ma, the Web of Data aggregator.

Additionally, Linked Data Search engines crawl the Web of Data, based on FYN, and provide sophisticated query capabilities on top of the complete data space, similar to those provided by conventional relational databases.

2.3 Access

Concerning the data access, one needs to think about the physical access on the one hand and about the *optimal* access method on the other hand. While for the former issue, certain converter and extractor tools, such as OpenLink's URIBurner⁷ or the Java library any23⁸ can be utilised, for the latter issue there is no general applicable method.

VoID-based metadata may help in understanding which access methods (dump, SPARQL endpoint, etc.) a dataset offers, however, this typically can not take over the task to decide what is best for the agent accessing the dataset.

2.4 Applications

Taken discovery and access together, one can advance to build a concrete application based on Linked Data [7]. Besides the overall issues what kind of application (for example an Ajax-based application or a desktop application) one plans to develop, there exist orthogonally certain cross-cutting concerns. This includes the utilisation of a local RDF store or synchronisation policies concerning dynamic datasets [10].

3. TUTORIAL OUTLINE

The tutorial is of practical nature, with a number of hands-on sessions included. The topics as outlined above will be covered in the following order:

1. Linked Data 101
 - (a) HTTP URIs
 - (b) RDF data model
 - (c) Vocabularies
2. The Linked Open Data cloud
 - (a) Domains and coverage
 - (b) Community-driven discovery (CKAN)
 - (c) Licenses and provenance
3. Discovery and Access of Linked Data
 - (a) Using Look-up services
 - (b) Dumps and FYN
4. Querying Linked Data
 - (a) SPARQL Introduction
 - (b) SPARQL Endpoints
 - (c) Recipes
5. Developing an Ajax application with Linked Data
6. Using Linked Data from Drupal 7

All materials from the tutorial, including slides and source code as well as references to advanced topics and tools will be made available via <http://latc-project.eu/>.

⁷<http://linkeddata.uriburner.com/>

⁸<http://code.google.com/p/any23/>

4. PRESENTER

Dr. Michael Hausenblas is a Research Fellow at DERI, NUI Galway where he is coordinating the Linked Data Research Centre. He focuses on Linked Data discovery and Dataset Dynamics as well as the write-enabled Linked Data Web, on the one hand and application domains—including Linked Government Data and streamed Linked Data—on the other hand. Michael has been contributing to the W3C SWEO Linking Open Data project since mid 2007 and is currently the co-ordinator of the FP7 European Commission Support Action LOD Around The Clock (LATC).

He has published over 45 papers at conferences, workshops and in journals, served on over 40 Programme Committees as well as held tutorial at various international conferences, including the World Wide Web conference and the International Semantic Web Conference. Since 2006, Michael has contributed to a number of W3C standardisation activities, including RDFa, co-chairing the RDB2RDF Working Group, and only recently publishing the VoID guide as a Semantic Web Interest Group Note.

5. ACKNOWLEDGEMENTS

The tutorial has partially been made possible due to funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement No. 256975, LOD Around-The-Clock (LATC) Support Action. The author would like to thank Richard Cyganiak and Lin Clark for their valuable input and support throughout.

6. REFERENCES

- [1] K. Alexander, R. Cyganiak, M. Hausenblas, and J. Zhao. Describing Linked Datasets - On the Design and Usage of void, the 'Vocabulary of Interlinked Datasets'. In *WWW 2009 Workshop: Linked Data on the Web (LDOW2009)*, Madrid, Spain, 2009.
- [2] T. Berners-Lee. Linked Data—Design Issues, 2006. <http://www.w3.org/DesignIssues/LinkedData.html>.
- [3] C. Bizer. The Emerging Web of Linked Data. *IEEE Intelligent Systems*, 24(5):87–92, 2009.
- [4] C. Bizer, T. Heath, and T. Berners-Lee. Linked Data—The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 2009.
- [5] C. Bizer, A. Jentzsch, and R. Cyganiak. State of the LOD Cloud, 2010. <http://www4.wiwi.fu-berlin.de/lodcloud/state/>.
- [6] M. d'Aquin, E. Motta, M. Sabou, S. Angeletou, L. Gridinoc, V. Lopez, and D. Guidi. Toward a New Generation of Semantic Web Applications. *IEEE Intelligent Systems*, 23(3):20–28, 2008.
- [7] M. Hausenblas. Exploiting Linked Data to Build Web Applications. *IEEE Internet Computing*, 13(4):68–73, 2009.
- [8] T. Heath. How Will We Interact with the Web of Data? *IEEE Internet Computing*, 12(5):88–91, 2008.
- [9] E. Oren, R. Delbru, M. Catasta, R. Cyganiak, H. Stenzhorn, and G. Tummarello. Sindice.com: a document-oriented lookup index for open linked data. *International Journal of Metadata, Semantics and Ontologies*, 3(1):37–52, 2008.
- [10] J. Umbrich, B. Villazon-Terrazas, and M. Hausenblas. Dataset Dynamics Compendium: Where we are so far!

In *Consuming Linked Data Workshop (COLD 2010) Workshop at International Semantic Web Conference (ISWC) 2010*, Shanghai, China, 2010.