

# How to Consume Linked Data on the Web

## Tutorial Description

Olaf Hartig  
Humboldt-Universität zu Berlin  
Berlin, Germany  
hartig@informatik.hu-berlin.de

Jamie Taylor  
Metaweb, USA  
jamie@metaweb.com

Juan Sequeda  
University of Texas at Austin  
Austin, TX, USA  
jsequeda@cs.utexas.edu

Patrick Sinclair  
BBC, UK  
metade@gmail.com

### ABSTRACT

In the past two years, the amount of data published in RDF and following the Linked Data principles has increased dramatically. Everyday people are publishing datasets as Linked Data. However, applications that consume Linked Data are not mainstream yet. To overcome this issue, we present a beginners tutorial on consuming Linked Data. We will discuss existing techniques how users can currently consume Linked Data and use it in their current applications.

### Categories and Subject Descriptors

H.3.5 [Information Storage and Retrieval]:

### General Terms

Management

## 1. INTRODUCTION

Through conferences workshops [2, 3], tutorials [1, 5] and general evangelism, the Linked Data community has successfully encouraged the Web community to publish billions of triples in the Linked Open Data cloud<sup>1</sup>. This cloud of data has been the subject of numerous academic analysis [2, 3] and served as a platform for various data deployment experiments [4, 6].

While access to this data presents exciting opportunities for the next generation of semantic applications, there are currently very few applications that make use of this resource. While a few applications, such as the BBC's music guide<sup>2</sup>, have used the data to significant benefit, the deployment methodology has been to harvest the data of interest from the cloud to create a private, disconnected repository for each specific application. Because Linked Data is published by a large collection of autonomous parties the skills, tools and patterns necessary to consume Linked Data in situ are very different from situations when resource identifiers are known a priori, access to the repository is reliable, and

<sup>1</sup><http://richard.cyganiak.de/2007/10/lod/>

<sup>2</sup><http://www.bbc.co.uk/music/>

local, whole-repository queries are possible. This tutorial will walk attendees through the skills necessary for using Linked Data in situ and present useful design patterns that overcome some of the challenges present when using Linked Data in real world applications.

## 2. DESCRIPTION

The objective of this introductory tutorial is to provide participants with detailed exposure to the practical and technical steps to consume Linked Data. Our aim is that the participants can learn during the three hours what Linked Data is, how it can be consumed, what tools they can use, and what they can do with it.

In addition to the conceptual introduction of Linked Data, this tutorial will cover three types of consumption: by humans, machines, and applications. Participants will be able to learn about existing tools to graphically navigate Linked Data, understand the technical details on how the machine consumes Linked Data, and then expose them to tools which allow applications to consume Linked Data. This tutorial will also discuss issues beginners often struggle with and that need to be taken in account when consuming Linked Data. Furthermore, we will show different applications that are consuming Linked Data and how they are doing it.

The tutorial will be structured as follows:

1. Introduction
  - What is Linked Data
  - Linked Data principles
  - Motivation to consume Linked Data
2. Consuming Linked Data by humans
  - Linked Data browsers
  - Faceted browsers
3. Consuming Linked Data by machines and applications
  - Accessing and retrieving Linked Data
  - Querying Linked Data with SPARQL
  - Automated Linked Traversal
4. Getting started
  - Finding URIs

- Finding additional data
  - Finding SPARQL endpoints
5. Consuming and publishing Linked Data cycle
  6. Demos of Linked Data consuming applications
    - Freebase
    - BBC music / Music Bore
    - Researchers Map
    - Turn2Live
  7. Conclusions and questions
    - Current challenges and open issues
    - Future prospects

The tutorial will combine presentations by the tutors with demonstrations and group discussion. We expect to appeal to the full spectrum of WWW 2010 attendees, including researchers, developers, data managers/publishers and those seeking to commercially exploit the Semantic Web by consuming Linked Data. The pre-requisites for participation in this tutorial include a broad technical understanding of Web development and Web publishing, and basic conceptual understanding of the Semantic Web, RDF and SPARQL.

### 3. PRESENTERS

#### Olaf Hartig (primary contact)

Olaf is a Ph.D. student with the Database and Information Systems Group at Humboldt-Universität zu Berlin. His research focuses on query processing in the Web of Linked Data. His aim is to develop concepts that allow to query this Web of Linked Data as if it is a giant global database. As project maintainer and lead developer he is implementing these concepts in the free software projects *Semantic Web Client Library*<sup>3</sup> and the *SQUIN* query interface<sup>4</sup>. He presented the approach to consume Linked Data as realized in the Semantic Web Client Library during the tutorial on publishing Linked Data at the 7th International Semantic Web Conference. In addition to the development of tools that enable applications to consume Linked Data, Olaf also has extensive experience in realizing such Linked Data based applications itself. Recently, his *Database Researchers Map*<sup>5</sup>, a Linked Data based Web application, won the Semantic Scripting Challenge 2009<sup>6</sup>. During his work as a teaching assistant at Humboldt-Universität Olaf gives tutorials and seminars and he supervises students who work on their diploma, master, or semester thesis.

#### Juan Sequeda

Juan Sequeda is a Ph.D student at the University of Texas at Austin and has been an invited researcher at the Universidad Politecnica de Madrid. His research focuses on enabling the integration of relational databases to the Web of Linked

Data. He is also an invited expert in the W3C RDB2RDF working group<sup>7</sup>. He also works on the *Morphster* project<sup>8</sup>, which supports morphologically based phylogenetic studies through Semantic Web technologies. In conjunction with Olaf Hartig, Juan works on the *SQUIN* project: querying the Web of Data. He is now consulting for a local music search startup, Turn2Live.com, which is soon starting to consume and publish Linked Data.

#### Jamie Taylor

Jamie is co-author of the O'Reilly book "Programming the Semantic Web" [7]. His work involves Data Gardening and Community Building on Freebase<sup>9</sup>, a large scale, and collaborative semantic database. His interest in graph data grew while managing enterprise software projects, which used Dynamic Object patterns in his role as CTO and VP Engineering at Determine Software (now a part of Selectica). He was the founder of one of San Francisco's first ISP's and has a Ph.D. from Harvard University.

#### Patrick Sinclair

Patrick Sinclair is a software engineer at the Audio and Music Interactive at the BBC, where he has been involved in the BBC Music and BBC Programmes Linked Data initiatives including the Programmes Ontology. He was formerly a research fellow investigating the use of Semantic Web technology in the Cultural Heritage domain at the University of Southampton.

### 4. REFERENCES

- [1] C. Bizer, R. Cyganiak, T. Heath, K. Idehen, and T. Berners-Lee. How to Publish Linked Data on the Web. Online at <http://linkeddata.org/docs/how-to-publish>, 2007.
- [2] C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee. Linked Data on the Web (LDOW2008): Workshop Summary. In Proceedings of the 17th International World Wide Web Conference, Beijing, China, 2008.
- [3] C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee. Linked Data on the Web (LDOW2009): Workshop Summary. In Proceedings of the 18th International World Wide Web Conference, Madrid, Spain, 2009.
- [4] M. Hausenblas. Exploiting Linked Data to Build Web Applications. *IEEE Internet Computing*, 13(4):68-73, 2009.
- [5] T. Heath, M. Hausenblas, C. Bizer, R. Cyganiak, and O. Hartig. How to Publish Linked Data on the Web. Tutorial in the 7th International Semantic Web Conference, Karlsruhe, Germany, 2008.
- [6] G. Kobilarov, T. Scott, Y. Raimond, S. Oliver, C. Sizemore, M. Smethurst, C. Bizer, and R. Lee. Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections. In *Proceedings of European Semantic Web Conference*, 2009.
- [7] T. Segaran, J. Taylor, and C. Evans. *Programming the Semantic Web*. O'Reilly, Cambridge, MA, 2009.

<sup>3</sup><http://www4.wiwiiss.fu-berlin.de/bizer/ng4j/semwebclient/>

<sup>4</sup><http://squid.org/>

<sup>5</sup><http://researchersmap.informatik.hu-berlin.de>

<sup>6</sup><http://www.semanticscripting.org/SFSW2009/challenge.htm>

<sup>7</sup><http://www.w3.org/2001/sw/rdb2rdf/>

<sup>8</sup><http://www.morphster.org/>

<sup>9</sup><http://www.freebase.com/>