# How to Publish Linked Data on the Web - Proposal for a Half-day Tutorial at ISWC2008

Tom Heath<sup>1</sup>, Michael Hausenblas<sup>2</sup>, Chris Bizer<sup>3</sup>, Richard Cyganiak<sup>4</sup>

<sup>1</sup>Talis Information Limited, UK <sup>2</sup>Joanneum Research, Austria <sup>3</sup>Freie Universität Berlin, Germany <sup>4</sup>DERI Galway, Ireland

tom.heath@talis.com, michael.hausenblas@joanneum.at, chris@bizer.de, richard@cyganiak.de

Abstract. The Web is increasingly understood as a global information space consisting not just of linked documents, but also of Linked Data. The Linked Data principles provide a basis for realizing this Web of Data, or Semantic Web. Since early 2007 numerous data sets have been published on the Web according to these principles, in domains as broad as music, books, geographical information, films, people, events, reviews and photos. In combination these data sets consist of over 2 billion RDF triples, interlinked by more than 3 million triples that cross data sets. As this Web of Linked Data continues to grow, and an increasing number of applications are developed that exploit these data sets, there is a growing need for data publishers, researchers, developers and Web practitioners to understand Linked Data community, this tutorial will address those needs, and provide participants with a solid foundation from which to begin publishing Linked Data from the Web.

## 1 Overview

The aim of this tutorial is to provide participants with a detailed conceptual understanding of how to publish Linked Data on the Web, and detailed exposure to the practical and technical steps that make up the publishing process. In addition to a conceptual introduction to Linked Data, the tutorial will cover best practices in topics such as minting URIs for published data sets, vocabulary selection, choosing what RDF data to expose, and interlinking distributed data sets. Specific patterns will be presented for publishing different forms of data set, such as data from static files, relational databases and existing Web APIs. Participants will also be shown how to debug published data.

The second focus of the tutorial will be applications that consume Linked Data from the Web. We will give an overview about existing Linked Data browsers, Web of Data search engines as well as Linked Data Mashups and cover the existing software frameworks that can be used to build applications on top of the Data Web. The tutorial will combine presentations by the tutors with demonstrations, interactive sessions, and group discussion. Other than a broad technical understanding of the Web development and Web publishing process, and a basic conceptual understanding of the Semantic Web, there are no pre-requisites for participation in the tutorial, which we expect to appeal to the full spectrum of ISWC2008 attendees, including researchers, developers, data managers/publishers and those seeking to commercially exploit the Semantic Web by publishing Linked Data.

### 2 Rationale

The publication since early 2007 of numerous interlinked RDF data sets on the Web has brought the Semantic Web concept from vision to reality. This Web of Linked Data already provides a basis for development and deployment of Semantic Web applications, which in turn are helping to stimulate publication of further Linked Data sets [1]. At this time, there is a strong need for guidance and education on which members of the Semantic Web community can build in publishing additional data sets and implement Web of Data client applications. By engaging in the Linked Data process through this tutorial, members of the core ISWC2008 audience can also better understand the upcoming research challenges raised by the existence of a Semantic Web.

Participants will benefit from the practical experience the tutors have gained in the W3C Linking Open Data community project [2]. The tutorial will also cover the latest development trends in the field that were discussed at the "Linked Data on the Web (LDOW2008)" workshop at WWW2008 [1] and the "Linked Data Planet" conference [3] in New York.

### **3** Content, Approach and Schedule

The tutorial will be based primarily on material from the "How to Publish Linked Data on the Web" online tutorial [4], the key resource in the field of Linked Data publishing, supplemented by practical examples to illustrate key issues and give participants hands-on experience of creating Linked Data.

## 4 Program Outline

#### 1. Introduction

- The Web of Data
- Linked Data Principles

#### 2. Publishing Linked Data on the Web

- Practical Task: enhancing a FOAF file with Linked Data
- Minting URIs and Content Negotiation
- Vocabulary Selection
- Manual and Automatic Interlinking

#### 3. Consuming Linked Data from the Web

- Linked Data Browsers
- Linked Data Search Engines
- Linked Data Mashups
- Libraries for implementing your own applicatons

#### 4. Conclusions and Outlook

- Linked Data prospects in the year ahead
- Upcoming directions and challenges for Linked Data research

# **5** Technical Requirements

- Projector
- Wireless internet connection usable by all participants
- Software requirements for participants:
  - Participants will be required to install the Firefox 2 Web browser and the Tabulator browser extension before/during the tutorial. These will be available to participants in advance of the tutorial via the tutorial Web site, and on CD/USB disk on the day of the tutorial.

### **6** Presenters

#### Tom Heath (primary contact)

Tom is a Researcher in the Platform Division of Talis, a company at the forefront of Semantic Web data storage, management and publishing, and a PhD candidate at KMi, The Open University. He is a leading member of the Linking Open Data community project, and co-author of "How to Publish Linked Data on the Web", the primary online tutorial for publishers of Linked Data. Tom has extensive experience of publishing Linked Data through services such as Revyu.com, the Linked Dataenabled reviewing and rating site and winner of the 2007 Semantic Web Challenge, and has spoken extensively on the subject of Linked Data and the Semantic Web.

Home page: http://kmi.open.ac.uk/people/tom/

#### **Michael Hausenblas**

Michael Hausenblas is a senior Semantic Web engineer at JOANNEUM RESEARCH, Austria, an applied research company. He is working in the area of multimedia semantics using Semantic Web technologies in a couple of national and international projects.

Within W3C he has been active in the Multimedia Semantics Incubator Group (2006/2007), the Semantic Web Deployment Working Group (since 2006), in the recently launched RDB2RDF Incubator Group and will participate in the W3C Video on the Web activity (http://www.w3.org/2008/01/video-activity.html).

Michael and his team have contributed to the Linking Open Data community project (http://riese.joanneum.at/). Michael regularly publishes on linked data issues and is involved in the development of linked data tools (http://143.224.254.32/irs/). In the realm of his PhD and his work he regularly gives lectures internally and for the Austrian-based Semantic Web Company (http://www.semantic-web.at).

Home page: http://sw-app.org/about.html

#### **Chris Bizer**

Prof. Dr. Bizer is the head of the Web-based Systems Group at Freie Universität Berlin (http://www4.wiwiss.fu-berlin.de/bizer/websys/). The group explores technical and economic questions concerning the development of global, decentralized information environments. Its current focus lies on the publication and interlinking of structured data on the Web using Semantic Web technologies.

The results of Prof. Bizer's research include the Named Graphs data model, which was adopted into the W3C SPARQL standard; the Fresnel display vocabulary implemented by several Semantic Web browsers, and the D2RQ mapping language which is widely used for mapping relational databases to the Semantic Web. He takes a leading role in the W3C Linking Open Data community effort and the DBpedia project, which both aim at interlinking large numbers of data sources on the Web. He has extensive of teaching experience, details of which are available on his Web site.

Home page: http://www.bizer.de

### **Richard Cyganiak**

Richard is a research assistant at the Digital Enterprise Research Institute, Galway, Ireland. He is an active member of the W3C Linking Open Data community project,

and has extensive experience as a publisher of Linked Data and as a developer of tools for publishing and consuming Linked Data. His software credits include D2R Server, the DBpedia project's Pubby server, and the Disco RDF browser. His latest project is the large-scale RDF crawler that feeds the Sindice search engine.

Richard is co-author of several tutorials and educational articles on the subject of publishing RDF on the Web, including "Cool URIs for the Semantic Web", published by the W3C's Semantic Web Education and Outreach group.

Home page: http://dowhatimean.net/

### References

[1] Bizer, C., Heath, T., Idehen, H., Berners-Lee, T. Linked Data on the Web (LDOW2008): Workshop Summary. In Proceedings , 17th International World Wide Web Conference, Beijing, China (2008)

[2] Linking Open Data project, http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

[3] Linked Data Planet conference, http://www.linkeddataplanet.com/

[4] Bizer, C., Cyganiak, R., Heath, T. How to Publish Linked Data on the Web, http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/ (2007)