

---

**Christian Becker**

is a researcher at Freie Universität Berlin, where he works in the area of Semantic Web user interfaces and ubiquitous Semantic Web applications. He has worked in several development and consulting positions and is currently a partner at MediaEvent Services, a consultancy focused on digital media ([www.beckr.org](http://www.beckr.org)).

**Peter Furness**

runs his own business specializing in the fields of decision analytics, modelling and data mining. A mathematician by background, Peter has been closely involved with the development and application of new analytical techniques in marketing and customer value management ([www.peterfurness.co.uk](http://www.peterfurness.co.uk)).

**Keywords:** Linked Data, Linking Open Data, RDF, Semantic Web, DBpedia, DBpedia Mobile

## The Semantic Web

## Linked Data enables data links on a global scale

## DBpedia and DBpedia Mobile

Christian Becker  
Freie Universität Berlin  
Web-based Systems Group  
Garystr. 21  
D-14195 Berlin, Germany  
Tel: +49 30 838 52612

# New Technology Briefing

## Linking spatial data from the Web

Christian Becker and Peter Furness

Received: 27th January 2010

### Abstract

The Semantic Web and, in particular, Linked Data, enable cross-database, cross-organizational and cross-domain links on a global scale. This paper introduces Linked Data projects that extract and apply location-related Linked Data, such as DBpedia. The case study of DBpedia Mobile, which provides a location-enabled Linked Data browser, is described. We conclude with examples of other applications, which can take advantage of location-related Linked Data from the Web.

*Journal of Direct, Data and Digital Marketing Practice* (2010) **11**, 317–323.  
doi:10.1057/dddmp.2010.10

### Introduction

The ‘Semantic Web’ was conceived by Web inventor Tim Berners-Lee and others in the 1990s as a development of the World Wide Web whereby the meaning (semantics) of information and services on the Web is defined, making it possible for the Web to understand and satisfy the requests of users.<sup>1</sup>

An emerging component of the Semantic Web is Linked Data, which enables cross-database, cross-organizational and cross-domain links on a global scale. Its development is fostered by the W3C (the World Wide Web Consortium) through the Linking Open Data (LOD) project.<sup>2–4</sup> This is a community effort to extend the Web with a data commons by publishing various open-licence data sets as Linked Data.<sup>3</sup> The LOD cloud has grown rapidly since 2007, and now encompasses more than 142 million data links.

In this paper we introduce Linked Data and describe one of the associated projects, DBpedia, being run jointly by the Freie Universität Berlin, the Universität Leipzig and OpenLink Software. DBpedia is involved with extracting structured data — such as spatial data — from Wikipedia to make it available as part of LOD.<sup>5</sup> It encompasses more than 409 million pieces of information and 4.9 million links to other LOD data sets. We give an example of the application of DBpedia to

## Spatial applications

local searching, through DBpedia Mobile, which presents data on a mobile phone browser.

We conclude with several examples of other applications of Linked Data involving complex spatial queries. Spatially referenced sources now include UK administrative and voting regions, Wikipedia, OpenStreetMap, GeoNames, US Census, EuroStat and the CIA World Factbook.

## Terminology: 'Linked Data' vs. 'Semantic Web'

### Linked Data

The term 'Semantic Web' encompasses a number of general design principles and standards, and has sparked many different areas of research. By contrast 'Linked Data' can be viewed as 'the semantic Web put into practice', and embodies a small set of core principles with the aim of getting data published, accessible and linked over the Web.

## Typed links connect data from different sources

Linked Data is about using the Web to create typed links<sup>6</sup> between data from different sources. These could range from databases maintained by two organizations in different geographical locations, to the situation of different siloed systems within one organization that, historically, have not been integrated at the data level. Technically, Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets and can, in turn, be linked to from other external data sets.

## RDF at the heart of Linked Data

At the heart of the Linked Data technical implementation is the Resource Description Framework (RDF). This provides a generic, graph-based model with which to structure and link data that describe things in the world.<sup>2</sup> The RDF model encodes data in the form of *subject*, *predicate*, and *object* triples. The subject and object of a triple are both Uniform Resource Identifiers (URIs)<sup>7</sup> that each identify a resource. The predicate specifies how the subject and object are related, and is also represented by a URI.

The following example of an RDF triple is taken from:<sup>2</sup>

*Subject:* <http://data.linkedmdb.org/resource/film/77>

*Predicate:* <http://www.w3.org/2002/07/owl#sameAs>

*Object:* [http://dbpedia.org/resource/Pulp\\_Fiction\\_%28film%29](http://dbpedia.org/resource/Pulp_Fiction_%28film%29)

This connects the description of the film Pulp Fiction in the Linked Movie Data Base (subject) with the description of the film provided by DBpedia (object). The predicate states that the subject and object both refer to the same real-world object, namely the film Pulp Fiction.

## Accessing Linked Data

To access Linked Data, browsers follow links using HTTP in a similar manner as document-based Web browsers do. Many data sets offer SPARQL<sup>8</sup> endpoints, which provide relational query capabilities similar to SQL interfaces offered by relational databases. Most data sets can be downloaded in their entirety.

## Linking Open Data

The LOD project is a community effort to extend the Web with a data commons by publishing various open-license data sets as RDF on the Web and by setting data links between data items from different data sources.<sup>3</sup> Since its inception in 2007, the LOD data sets available

on the Web have grown rapidly to encompass, as at December 2009, more than 13.1 billion RDF triples and 142 million data links. These span a large number of data domains, including life sciences, geographic data, demographic data, publications, online activities, music and media, and include many cross-domain data sets.

### **Linked Data as a technological platform**

Linked Data is an important new technological platform, which developers use to build applications that mask the underlying complexities of data retrieval from end-users. Some of these applications are described in later sections of this paper.

### **Linked Data vs. proprietary data integration solutions**

Despite the complexities involved, Linked Data offers many benefits that are complementary to the proprietary Web-based data integration solutions (such as Google Maps) that are already well established. These include:

- a large and rapidly growing universe of accessible data sources, published under open licences;
- links across databases, domains and organizations in a standard and readily usable form;
- access to raw data that can be retrieved in its most up-to-date form directly from the relevant publisher;
- an expanding number of applications that enable straightforward access to Linked Data for the end-user.

### **DBpedia publishes structured data from Wikipedia as Linked Data**

#### **DBpedia as a geospatial entry point to the Web of Data**

DBpedia is one of the community efforts within the LOD project.<sup>5</sup> Its objective is to extract structured information from Wikipedia and make this available on the Web under open licence. The DBpedia data set is served as Linked Data, allowing it to be interlinked with other open data sets on the Web. The principal contributors to DBpedia are the Freie Universität of Berlin, Universität Leipzig and OpenLink Software.

### **Overview of the DBpedia data set**

The DBpedia data set consists of RDF triples extracted from the ‘infoboxes’ commonly seen on the right-hand side of Wikipedia articles. It already contains identifiers for more than 2.9 million ‘things’, including over 280,000 persons and 339,000 geographic locations. The latter are interlinked with other spatially referenced data sources such as OpenStreetMap,<sup>9</sup> GeoNames, the US Census, EuroStat, the CIA World Factbook and Semantic CrunchBase. This makes DBpedia a valuable data source for location-based applications. It also contains short abstracts about places, which display well on mobile devices, in up to 30 languages. The current scope of the LOD Cloud is shown in Figure 1.

### **A location-aware Linked Data browser**

#### **Case study — DBpedia Mobile**

DBpedia Mobile<sup>10,11</sup> is a location-aware Linked Data browser that has been designed to be run on an iPhone or other suitable mobile device. Using the current GPS position of the mobile device, the application enables spatial integration of data about nearby locations from

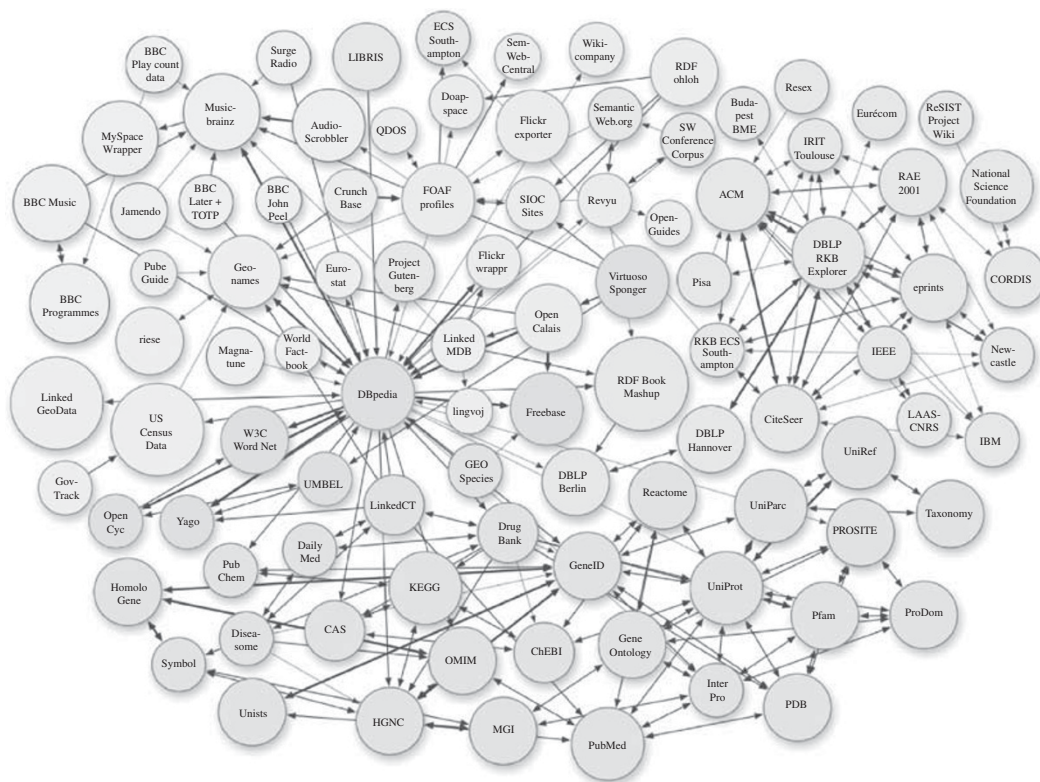


Figure 1: The LOD cloud

DBpedia, reviews from Revu (a generic reviewing and ratings site based on Linked Data<sup>12</sup>), and associated photos from Linked Data photo sources such as DBpedia and a Linked Data wrapper around the Flickr photo-sharing API.<sup>13</sup>

DBpedia has been designed with tourism applications in mind. Figure 2 shows DBpedia Mobile with a display from DBpedia and Revu about the Brandenburg Gate in Berlin.

In addition to accessing Web data, DBpedia Mobile enables users to publish their current location, photos and reviews to the Web as Linked Data, so that they are available to other applications and users. Thus, rather than just being tagged with spatial coordinates, this published content is linked with a nearby DBpedia resource and thus contributes to the overall richness of the Web of data.

### Publishing Linked Data from a mobile device

### Applications of Linked Data for spatial analysis and targeting

#### Using Linked Data for cross-domain analysis

DBpedia Mobile is just one of many actual and potential applications of Linked Data in the geographical domain. Linked Data is an especially interesting development for cross-domain analysis, and the publicly available data may soon reach a stage at which it can be used to answer queries such as ‘find all the stations on a particular rail route within a particular area’; ‘find all the hotels within areas with a high disposable income’; ‘find all competitor outlets and reviews within a certain range of my outlets’.<sup>14</sup>



Figure 2: Central Berlin on DBpedia Mobile

### Existing geographically referenced data sets

There is already a growing number of publicly accessible, geographically referenced data sets within the LOD project, which point to each other:

- The UK Ordnance Survey has started to publish topological information about administrative and voting regions in the UK as Linked Data<sup>15</sup>
- Wikipedia — through DBpedia, 339,000 locations are referenced
- OpenStreetMap — with over 350 million places through LinkedGeoData
- GeoNames — administrative hierarchies
- US Census
- EuroStat — general and economic data sets
- World Factbook

These and other spatially referenced Linked Data in the pipeline will open up a wide range of geographic and geodemographic applications.

### Conclusion

The Linked Data available over the Web is growing rapidly. Linked Data is based on Semantic Web standards approved by the W3C that cover cross-database, cross-organizational and cross-domain interoperability. While data sets and links are not perfect, they are moving beyond the research stage, with adoption by major public and commercial players, including the UK Government, the BBC, Thomson

### Linked Data is standards-based and moving past a research stage

## Technological platform that enables data links and access to raw data

## Partial availability of Linked Data

Reuters, New York Times, the Cyc Foundation and the life sciences community.

Although there are some complexities involved in data retrieval with Linked Data, developers are building powerful applications that mask these complexities for end-users. In the realm of spatial data, as with other areas, Linked Data allows links between data elements that can span databases, domains and organizations — and also gives access to raw data that can be retrieved up-to-date directly from the relevant publisher.

Despite much progress there are large amounts of data that have yet to be made available as Linked Data. This includes, for example, open government data and user-generated data such as reviews and locations. As applications of Linked Data such as DBpedia Mobile become more widely adopted, we can expect continued rapid growth in the future.

## Acknowledgments

This paper is partly based on a presentation given by the first author at the seminar ‘Target: 350 Million Consumers! How geodemographics is being applied across an ever expanding Europe’ organized by the Census and Geodemographics Group of the Market Research Society and held on Wednesday 1 April 2009 at The Society of Chemical Industry, London SW1 (Becker, C. (2009) ‘Linking spatial data from the Web’, The Market Research Society, CGG Conference April 2009, <http://www.mrs.org.uk/networking/cgg/downloads/euro/ChristineBaeckerLinking%20Spatial%20Data%20from%20the%20Web.pdf>). Additional information on this and related topics can be found at the Geodemographics Knowledge Base (The Census and Geodemographics Group of the Market Research Society (2009), ‘The geodemographics knowledge base’, <http://geodemographics.org.uk>).

## References

1. Berners-Lee, T. (2009) ‘Tim Berners-Lee on the next Web’, video available at TED (Technology, Entertainment, Design), [http://www.ted.com/index.php/talks/tim\\_berners\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/index.php/talks/tim_berners_lee_on_the_next_web.html).
2. Bizer, C., Heath, T. and Berners-Lee, T. (2009) ‘Linked Data – The Story So Far’, *International Journal on Semantic Web & Information Systems*, Vol. 5, No. 3, pp. S.1–22. <http://tomheath.com/papers/bizer-heath-berniers-lee-ijswis-linked-data.pdf>, <http://linkeddata.org/docs/ijswis-special-issue>.
3. Linking Open Data – W3C SWEO Community Project, <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>.
4. Linked Data – Connect Distributed Data Across the Web. Website for the Linked Data Community, <http://linkeddata.org/home>.
5. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R. and Hellmann, S. (2009) ‘DBpedia – A crystallization point for the Web of Data’, *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 7, No. 3, pp. S.154–165.
6. Wikipedia. (2009), [http://en.wikipedia.org/wiki/Typed\\_link](http://en.wikipedia.org/wiki/Typed_link).
7. Berners-Lee, T., Fielding, R. and Masinter, L. (1998) RFC 2396, Uniform Resource Identifiers (URI): Generic Syntax, <http://www.ietf.org/rfc/rfc2396.txt>.
8. World Wide Web Consortium. (2008) ‘SPARQL Query Language for RDF’, <http://www.w3.org/TR/rdf-sparql-query>.

9. OpenStreetMap – The Free Wiki World Map, <http://www.openstreetmap.org>. For the Linked Data version see <http://linkedgeo.org>.
10. Becker, C. and Bizer, C. (2009) 'Exploring the Geospatial Semantic Web with DBpedia Mobile', *Journal of Web Semantics*, Vol. 7, No. 4, pp. 278–286.
11. DBpedia Mobile. (2008) Website for the DBpedia Mobile project, <http://becker.org/DBpediaMobile>.
12. Heath, T. and Motta, E. (2008) 'Revyu: Linking reviews and ratings into the Web of Data', *Journal of Web Semantics*, Vol. 6, No. 4, pp. 266–273.
13. Becker, C., flickr™ wrapper (2007) <http://www4.wiwiw.fu-berlin.de/flickrwrapper/>, based on Flickr API at <http://www.flickr.com/services/api>.
14. Becker, C. (2009) 'Linking spatial data from the Web', The Market Research Society, CGG Conference April 2009, <http://www.mrs.org.uk/networking/cgg/downloads/euro/ChristineBaeckerLinking%20Spatial%20Data%20from%20the%20Web.pdf>.
15. Ordnance Survey: Linked Data for Ordnance Survey Data, <http://data.ordnancesurvey.co.uk/>.