

Linked Data: Present & Future

Autor : Philippe Cudré-Mauroux

Datum : 1. Februar 2017



In this text, I give a short overview of Linked Data technologies, describing their main characteristics as well as their adoption. I also risk making a few predictions on the future of Linked Data.

What is Linked Data?

Linked Data can be seen as a simplified (and pragmatic) implementation of the Semantic Web vision. Sir Tim Berners-Lee, the inventor of the Web, coined the term Linked Data in 2006 to prescribe a simple method of publishing data using web standards. The method can be summarized in three points:

1. All data items should have names that start with http
2. When looked up online, the http names should return some data in a standard format to describe the items
3. The description of the items should also contain relationships to other pieces of data.

In technical terms, this means that data items are identified by URIs, so that they can be dereferenced through HTTP, and can refer to other items using their HTTP URI-based

identifiers. The language used to express such data is often called the Resource Description Framework (RDF).

Linked Data & Me

I've been a close observer of the emergence of Linked Data. Publicly, I was involved in a number of forums and meetings dealing with Linked Data. I have co-organized the [ISWC](#), the main research venue for Linked Data, a number of times since 2007 (I was, for instance, PC Chair of ISWC 2012 in Boston and will be In-Use Chair this year in Vienna). Privately, I regularly leverage Linked Data in my own research, either to better grasp content (e.g. to understand text better) or to serialize output data (e.g. to publish datasets).

Linked Data Today

The adoption of Linked Data has been phenomenal. Linked Data is used in two main ways today: i) to create webs of data that can be accessed and queried by anyone, and ii) to add metadata to Web pages.

The most prominent web of data created through Linked Data is called the Linked Open Data (LOD) cloud (see Figure 1). It is conceptually similar to the World Wide Web, but contains interlinked data instead of interlinked documents. The LOD cloud includes thousands of different datasets from a wide range of domains: from governmental data to geographic, life-science or bibliographic data. Each of these datasets contains a myriad of data items and links, is fully open, and can be queried using a standard query language (SPARQL). Other important webs of data exist besides the LOD Cloud, such as Wikidata or Google's Knowledge Graph.

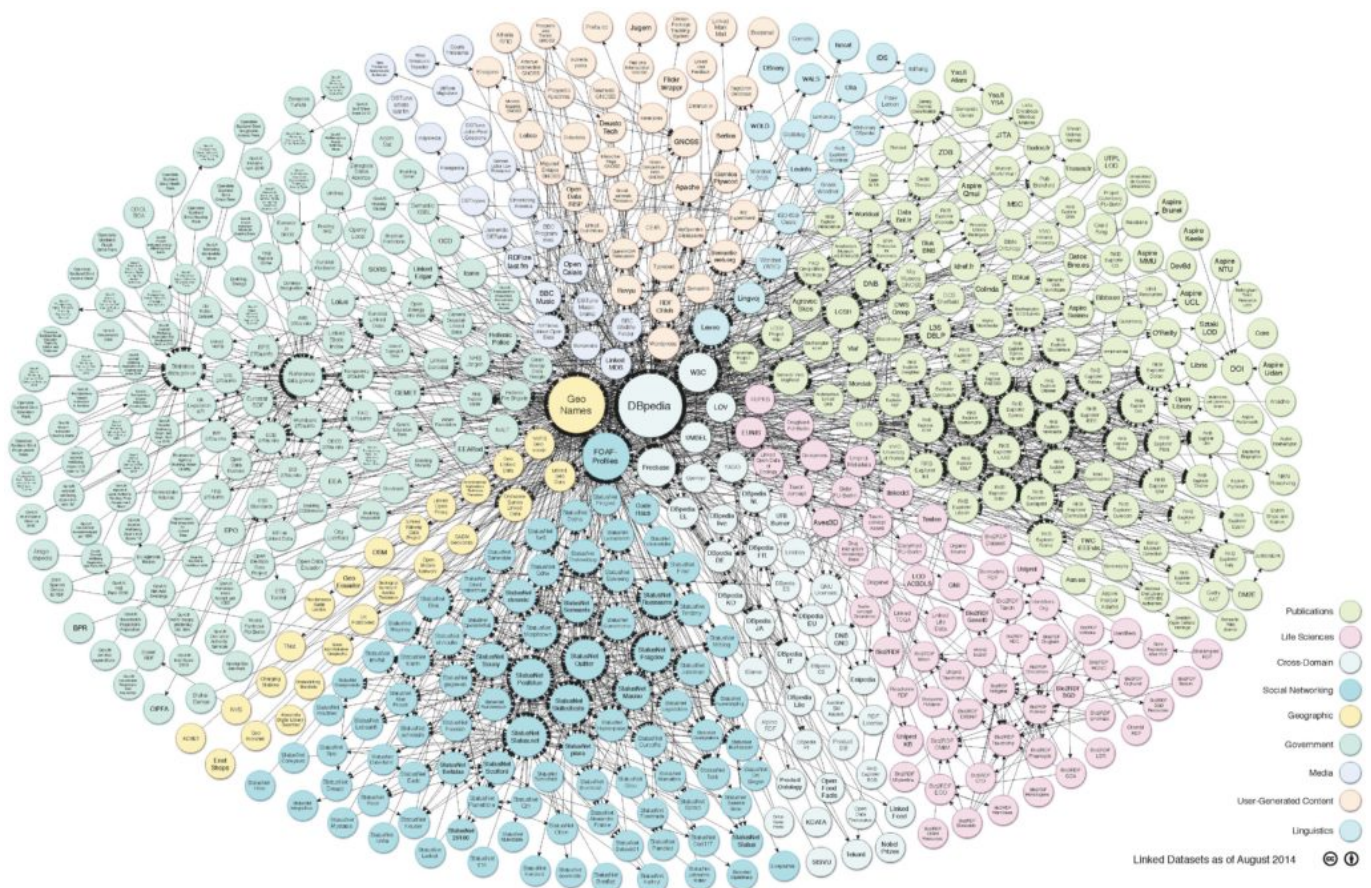


Figure 1: the Linked Open Data (LOD) cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>

In addition, Linked Data is also used to add metadata to Web pages. The main format used in that sense is called schema.org, which is supported by a number of prominent companies including Google, Microsoft, Yahoo and Yandex. This format allows all sorts of data to be added to a Web page, to describe for example people, products, events, or reviews that are contained in that Web page. Those data can then be used to summarize, describe, or manipulate the Web page (e.g. to create rich snippets on a search engine). Today, millions of websites use this format to describe their pages.¹

Linked Data Tomorrow

Linked Data is widely available today, in the LOD cloud and on Web pages. However, the development of applications using Linked Data has been hampered by a series of technical issues, from data quality to complex standards. In the following paragraphs, I give my own vision of the evolution of Linked Data.

1. Agile standards: RDF and its applications are governed by a monolithic and complex set of standards that are revamped every few years. In that context, agile and incremental efforts like schema.org will be increasingly popular and important as they correct, update or try out features on a continuous basis, akin to methodologies used for agile software.
2. Smart clients: using Linked Data productively is often more complex than it seems, as one typically has to spend considerable time selecting, aligning and cleaning up data (which is a common issue in Big Data and Data Science). Increasingly, Machine Learning methods will be able to automate such processes to create smart Linked Data clients capable of ingesting, aligning and cleaning up raw Linked Data using sophisticated supervised models.
3. Unification: Linked Data is available today from several distinct and heterogeneous platforms (the LOD cloud, Wikidata, HTML pages, etc.) In the future, bridges will be built to integrate those platforms and create more extensive webs of data. Fribourg's VoldemortKG is a first effort in that direction, as it interlinks schema.org data to the LOD cloud.