

Facilitation the publication of Open Governmental Data with the LOD2 Stack

Sören Auer*, Michael Martin*, Phillip Frischmuth*, Bastiaan Deblieck+

* Universität Leipzig, Institut für Informatik, AKSW/BIS, Postfach 100920, 04009 Leipzig
+ Tenforce, Haachtsesteenweg 378, 1910 Kampenhout, Belgium

One of the biggest challenges in the area of intelligent information management is the exploitation of the Web as a platform for data and information integration as well as for search and querying. Just as we publish unstructured textual information on the Web as HTML pages and search such information by using keyword-based search engines, we will soon be able to easily publish structured information, reliably interlink this information with other data published on the Web and search the resulting data space by using expressive querying. The Linked Data paradigm has evolved as a powerful enabler for the transition of the current document-oriented Web into a Web of interlinked Data and, ultimately, into the Semantic Web. The term Linked Data here refers to a set of best practices for publishing and connecting structured data on the Web. These best practices have been adopted by an increasing number of data providers over the past three years, leading to the creation of a global data space that contains many billions of assertions - the Web of Linked Data.

In order to make the advances, we will attain in the LOD2 project with respect to the different research challenges, readily available to applications and end users, we will implement and integrate them as components of the LOD2 Stack. The LOD2 Stack will be a distribution of integrated software tools and components enabling corporations, organizations and individuals to employ Linked Data technologies with minimal initial investments. The LOD2 Stack will be made available among other formats as a cloud hosted software as service distribution on Amazon EC2, Elastichosts, Virtualbox and possibly other cloud and virtualization infrastructures. This delivery format will eliminate installation and configuration problems and will provide instant value, especially when combined with pre-loaded common data sets. The stack will comprise in particular the following components:

- large-scale RDF data management based on significant extensions and merging of the open-source MonetDB¹ and Openlink Virtuoso²,
- methods for fusing information on the Data Web possibly combined with local information based on FUB's WIQA framework,
- tools and methods for **interlinking schema and data** in LOD knowledge bases using supervised and unsupervised algorithms, i.e. with and without interaction with

¹<http://monetdb.cwi.nl/>

²<http://monetdb.cwi.nl/>

knowledge base maintainers,

- **machine learning algorithms** for enrichment and repair of knowledge bases, which will make use of DL-Learner³ as a baseline,
- **various user interfaces** for interacting with the above components as well as with the Data Web as a whole. These include integrated and substantially extended versions of Poolparty⁴, OntoWiki⁵ and Sig.ma.

In addition to the LOD2 Stack components locally deployable at corporations, organizations or as special-purpose Web services, we will enhance and develop a number of *services and data sets, which work as crystallization points for the Web of Data*. These services and data sets include improvements to the Sindice Semantic Web index, the search interfaces Sig.ma, the faceted browser Sparallax, the RDF publication tool Triplify, as well as a number of data sets serving as interlinking hubs on the Web of Data, such as the Wikipedia-derived DBpedia and the RDF representation of OpenStreetMap - LinkedGeoData.

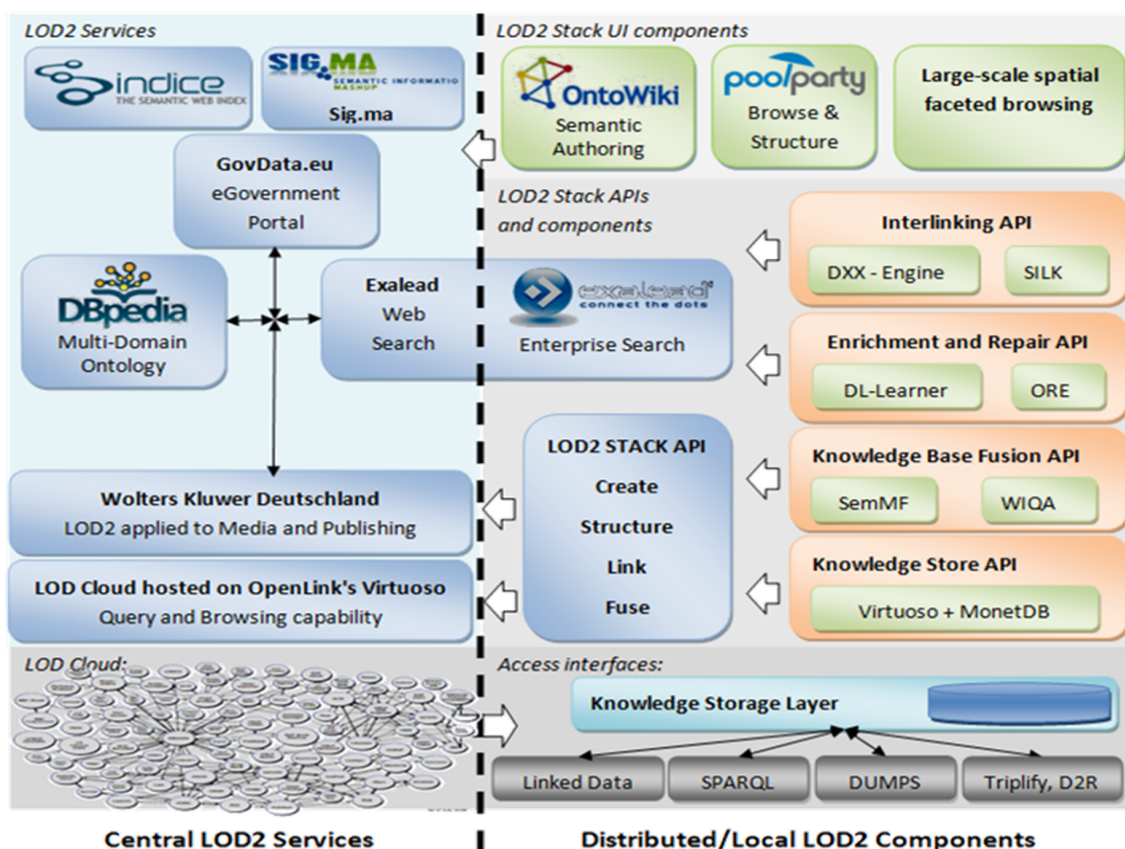


Figure 1: High-level architecture of the LOD2 Stack.

One of the target user groups of the LOD2 Stack are governments and public administrations aiming to publish governmental data as Linked Data. The LOD2 Stack will facilitate the whole life-cycle of Linked Data Management from extraction, authoring over linking, enrichment and repair to browsing and exploration. A first release of the LOD2 Stack is expected for September 2011. However, in addition to the LOD2 Stack, which can be deployed on the premises of an organization, the LOD2 project is also developing

³<http://dl-learner.org>

⁴<http://poolparty.punkt.at/>

⁵<http://ontowiki.net>

PublicData.eu - a decentralized network of public sector dataset registries based on the CKAN.net infrastructure.

The data sets are catalogued on PublicData.eu according to various dimensions, e.g. geographical coverage, temporal coverage, type of data, origin etc. In addition to cataloguing the data sets, we will provide functionality to submit tools and mashups for exploration and browsing of the data sets. All data sets will be available in formats adhering to open-standards, in particular RDF and Linked Data. For convenience we will deploy automatic conversion tools to other data formats, such as CSV, Excel, KML. We plan the addition of personalization features to the PublicData.eu website. Citizens will be enabled to register and login (also via OpenId and FOAF+SSL). Once logged in users can rate and comment on existing data sets; they will be enabled to provide a wish list of missing data sets, add additional data sets with respective descriptions; they can upload revised data set versions and develop visualization tools and mashups on top of the data sets. We will also work on a notification and subscription service, which will enable users to be notified once new data sets concerning their personal preferences (e.g. region, language, type of data etc.) are added to PublicData.eu.



Figure 2: Current prototype of the PublicData.eu portal.

By adapting the LOD2 Stack and developing the PublicData.eu portal in an LOD2 use case, we aim to increase the ability of the public to find, browse, and easily use information that is generated and held by various governmental branches and institutions in Europe. PublicData.eu will expand creative use of the data beyond the walls of government by encouraging innovative ideas such as semantic mashups and widgets. PublicData.eu will make government more transparent and will help to create an unprecedented level of openness in government. This openness will strengthen European democracy and promote efficiency and effectiveness in Government.