

## Publishable Summary

---

**Introduction.** LOD2 is a large-scale integrating project co-funded by the European Commission within the FP7 Information and Communication Technologies Work Programme (Grant Agreement No. 257943). This 4-year project now going in its 3<sup>rd</sup> year comprises 15 leading Linked Open Data technology researchers, companies, and service providers from across 11 European countries and South Korea. LOD2 is coordinated by the AKSW research group at the University of Leipzig.

**General Goals.** The semantic web activity has gained momentum with the widespread publishing of structured data as RDF. The Linked Data paradigm has therefore evolved from a practical research idea into a very promising candidate for addressing one of the biggest challenges in the area of intelligent information management: the exploitation of the Web as a platform for data and information integration in addition to document search. In that context LOD2 targets a number of research challenges: improve coherence and quality of data published on the Web, close the performance gap between relational and RDF data management, establish trust on the Linked Data Web and generally lower the entrance barrier for data publishers and users. The LOD2 project tackles these challenges by developing:

- enterprise-ready tools and methodologies for exposing and managing very large amounts of structured information on the Data Web,
- a testbed and bootstrap network of high-quality multi-domain, multi-lingual ontologies from sources such as Wikipedia and OpenStreetMap.
- algorithms based on machine learning for automatically interlinking and fusing data from the Web.
- adaptive tools for searching, browsing, and authoring of Linked Data.

The LOD2 project integrates and syndicates linked data with large-scale, existing applications and showcases the benefits in the three application scenarios publishing, corporate data intranets and Open Government Data.

**Result of the second project year.** In the second year, the LOD2 consortium was enlarged by 5 additional partners from Eastern Europe and South Korea. The main result of the second project year is the LOD2 Stack – an integrated suite of software tools supporting the life-cycle of Linked Data on the Web. The LOD2 Stack was created as a compilation of Debian software packages which can be used with a unified Web user interface. This allows deploying the stack on various infrastructures ranging from virtual machines over full server instances to whole cloud infrastructures. The LOD2 Stack comprises tools covering the life-cycle of linked data on the Web.

- The LOD2 Stack was substantially improved in terms of tool coverage and inter-tool integration. It now comprises 37 components, which are connected via fine-grained dependencies; use the W3C provenance vocabulary to exchange information and the WebID protocol for authentication.
- Through the integration of previously experimental column store technology and other ground-breaking database innovations for RDF stores (i.e. dynamic query optimization, adaptive caching of joins, optimized graph processing, cluster/cloud scalability) the LOD2 consortium got much closer to its goal of closing the performance gap between RDF and relational data management.
- Various approaches and tools were developed for improving the extraction of RDF from unstructured, semi-structured and structured sources. With the NLP Interchange Format, for example, various Natural Language Processing tools can be loosely coupled and easier integrated to reach higher precision and recall. With SparqlMap relational databases can be connected to the Web of Linked Data.
- Data Linking on the Web was significantly simplified by employing active learning and genetic programming approaches for generating link specifications semi-automatically. With Sieve, LOD2

release a complete overhaul of the data quality assessment tool. LOD2 exploited its new Eastern European and Asian partners competences to improve multi-lingual interoperability and support for non-Latin alphabets.

- A number of visualization, exploration and browsing widgets were integrated into the LOD2 Stack in order to facilitate interactions with end users. Figure 1 shows widgets for the visualization of statistical and spatial data.
- With Publicdata.eu the consortium has released the first version of a pan-European data catalogue, which aggregates metadata from more than 30 national, local and regional data catalogues throughout Europe. Also, national data catalogues, such a Serbian one, were established by the project.

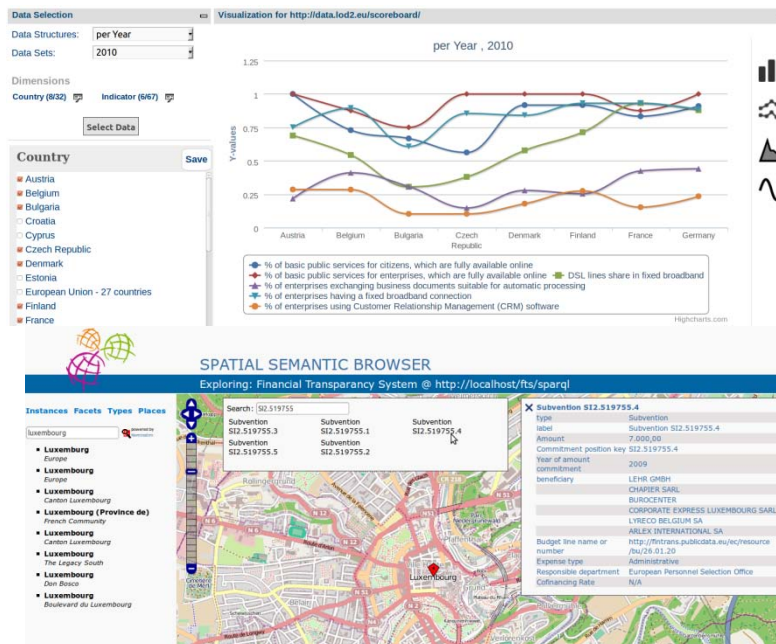


Figure 1: Two different data visualization widgets provided by the LOD2 Stack for statistical data (top) and spatial data (bottom).

In addition to the tools integrated into the LOD2 Stack tool suit also comprises a number of online services and complementary tools and datasets. These include:

- the **CKAN metadata repository**, which interacts bi-directionally with the LOD2 Stack,
- the **DBpedia and LinkedGeoData RDF extraction frameworks**, which have been complemented by LOD2 with Live-SPARQL endpoints interactively reflecting the current state of the two central LOD knowledge bases,
- the **DBpedia benchmark**, which measures the performance of triple stores with real data and real queries,
- the **Sindice entity search framework**, which won of the Yahoo Semantic Search Challenge 2011 and is used internally by several LOD2 Stack components.



Figure 2: Publicdata.eu

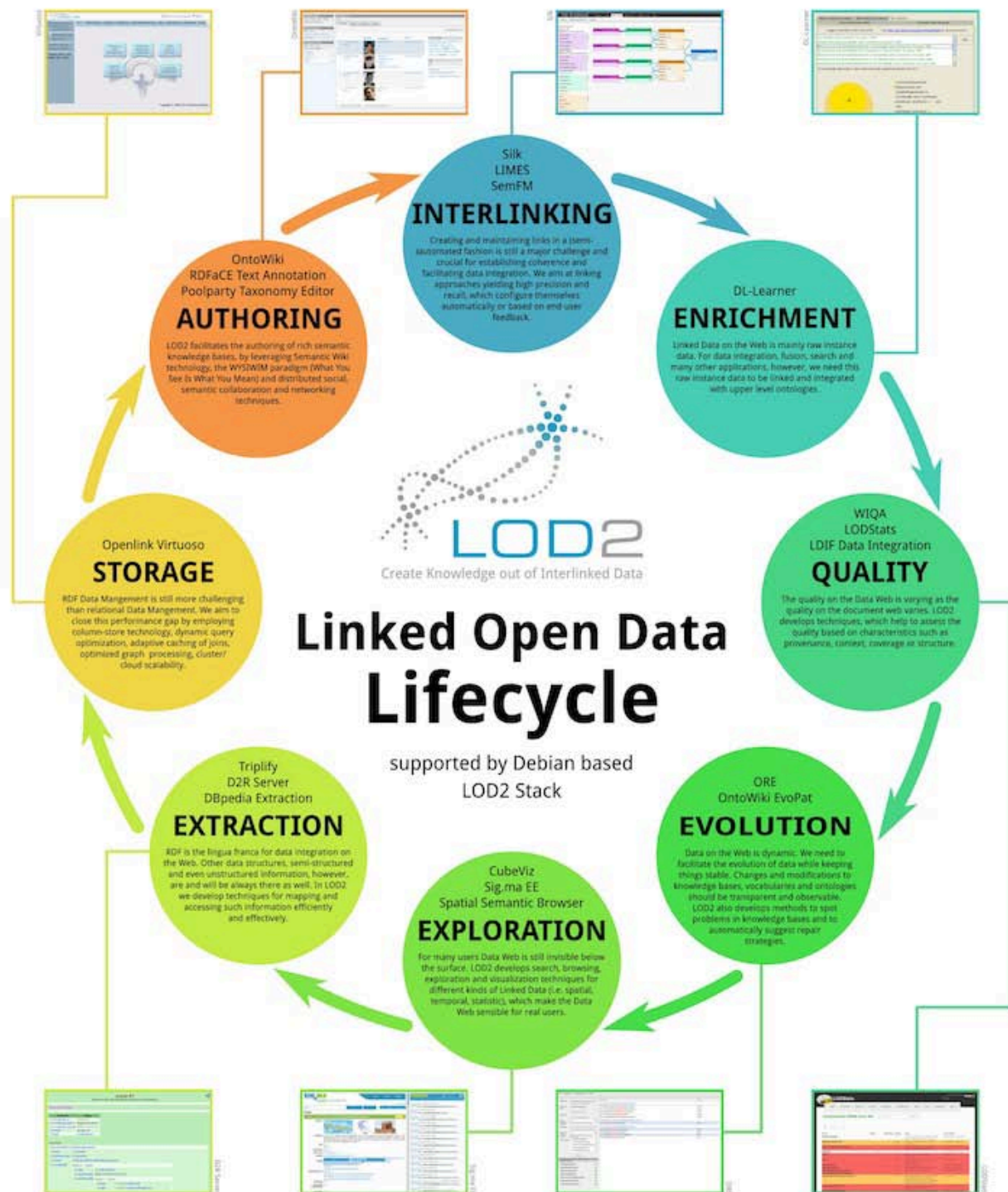
The project investigated the three LOD2 use-cases by developing initial application prototypes of the LOD2 Stack for the three application domains publishing, enterprise and governmental data.

During the second project year, the LOD2 members published more than 40 peer-reviewed scientific publications, won three Best-Paper awards, organized a number of events (e.g. I-Semantics, European Data Forum) and successfully supported a number of organizations in the publishing of Linked Data (e.g. Food and Agriculture Organization of UN and European Environment Agency). The consortium contributed to many W3C standards, organized a summer school and 10 webinars explaining the use and benefits of the LOD2 Stack tools.

**Outlook.** In the next year, the project will intensify its efforts to demonstrate the benefit of Linked Data in the three use cases. The refinement and integration of LOD2 Stack components will continue and new research approaches for exploiting synergies between the different phases of the Linked Data life-cycle. LOD2 will co-organize the second edition of the European Data Forum, which will take place on April 9-10 2012 in Dublin, Ireland.

**The LOD2 consortium** comprises with its academic partners Universität Leipzig, Freie Universität Berlin, University of Economics Prague, Institute Mihailo Pupin, Korean Advanced Institute of Science and Technology and National University of Ireland in Galway research expertise in Semantic Web technologies, ontological engineering, machine learning, Web search, information retrieval, databases and knowledge stores. With LOD2 partner CWI's reputation in the database realm, LOD2 aims to contribute to cross-fertilization between database and semantic web research. The LOD2 consortium has engineering power provided by the companies TenForce and OpenLink Software as well as dissemination and exploitation expertise provided by its innovative SMEs (Exalead, Openlink, TenForce, I2G and Semantic Web Company), a large corporation (Wolters Kluwer) and the eGovernment and Open Knowledge communities represented by Open Knowledge Foundation.

Further information can be found on the project homepage at: <http://lod2.eu> or from the coordinator Dr Sören Auer at Universität Leipzig (Phone: +49 (341) 97-32367, Fax: +49 (341) 97-32329, Email: [auer@uni-leipzig.de](mailto:auer@uni-leipzig.de)).



LOD2 is a large-scale integrating project co-funded by the European Commission within the FP7 Information and Communication Technologies Work Programme (Grant Agreement No. 257943). Commencing in September 2010, this 4-year project comprises leading Linked Open Data technology researchers, companies, and service providers (15 partners) from across 11 European countries (and one associated partner from Korea) and is coordinated by the AKSW research group at the University of Leipzig.



Figure 3: Poster illustrating the Linked Data life-cycle stages and the tools developed to support these.