

Publishable Summary

of the Work performed in the first project year of the project „LOD2 – Creating Knowledge out of Interlinked Data“

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). Commencing in September 2010, this 4-year project comprises leading Linked Open Data technology researchers, companies, and service providers from across 7 European countries and is coordinated by the AKSW research group at the University of Leipzig.

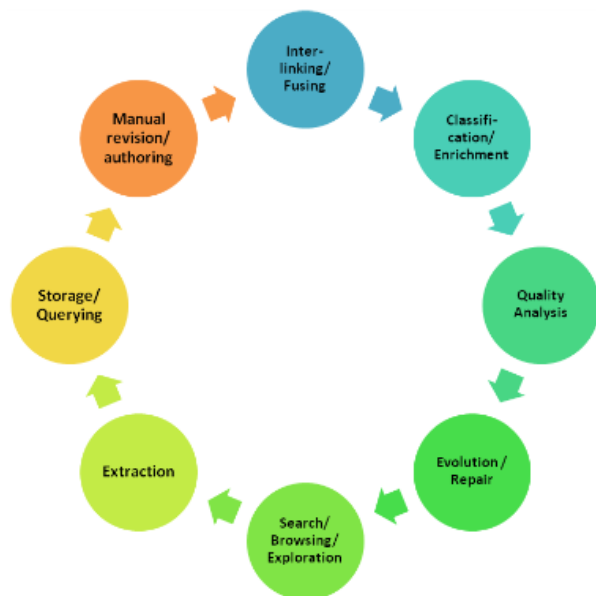


Figure 1: The Life-cycle of Linked Data on the Web.

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. To translate this initial success into a world-scale reality, the following research challenges are addressed by LOD2: 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. With partners among those who initiated and strongly supported the Linked Open Data initiative, the LOD2 project aims at tackling 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 media and publishing, corporate data intranets and eGovernment.

Result of the first project year. The main result of the first 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 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 in particular the following tools, which were substantially advanced and partially newly developed in the first project year:

- **Virtuoso** is a knowledge store and virtualization platform that transparently integrates Data, Services, and Business Processes across the enterprise. The open-source data integration server and the highly efficient and scalable RDF triple store implementation in Virtuoso was extended by LOD2 to (1) allow the dynamic partitioning and distribution of RDF data across a cluster of machines and (2) increase the compression and

scalability by integrating column-oriented RDF storage similar to MonetDB (as provided by LOD2 partner CWI).

- **DL-Learner** is a supervised Machine Learning framework for OWL and Description Logics. In LOD2, DL-Learner was extended in such a way, that it can be used by knowledge engineers for ontology enrichment: Given a SPARQL endpoint, it gives suggestions (with corresponding confidence scores) on how to extend the existing schema, e.g. suggest domain and ranges of properties, super classes, definitions and many more. Those methods will be enhanced and integrated in the **ORE** toolkit and user interface. ORE allows users to easily detect and repair inconsistencies and modelling errors in knowledge bases (e.g. loaded in Virtuoso) as well as improve their schemata in general.
- The **Silk Linking Framework** supports data publishers in setting explicit RDF links between data items within different data sources. Using the declarative Silk - Link Specification Language, developers can specify which types of RDF links should be discovered between data sources as well as which conditions data items must fulfil in order to be interlinked. SILK was substantially extended and improved in the course of the first LOD2 project year with a comprehensive user interface, blocking strategies for increasing the scalability and a distributed computation model based on Apache Hadoop for boosting the performance.
- **OntoWiki** is a user interface tool providing support for agile, distributed knowledge engineering scenarios. It enables intuitive authoring of Linked Data, with an inline editing mode for editing RDF content, similar to WYSIWIG for text documents. OntoWiki was integrated into the LOD2 Stack and extended with a browsing widgets for spatial-faceted browsing.

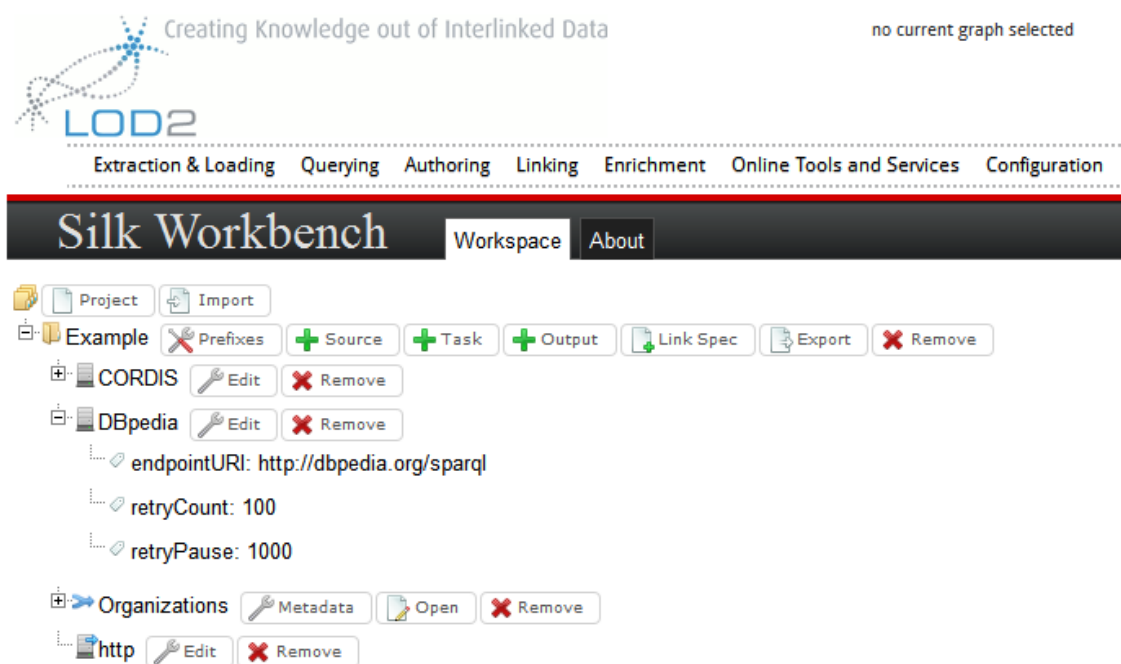


Figure 2: The user interface of the LOD2 Stack showing the SILK workbench for creating link specifications.

In addition to these main tools being integrated into the LOD2 Stack the 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.

Last but not least the project has started to investigate the three LDO2 use-cases by identifying requirements (e.g. with the Open Government Data Stakeholder Survey) and performing first experiments and developing initial prototypes (e.g. with publicdata.eu).

Already during the first project year, the LOD2 members published more than 20 peer-reviewed scientific publications, organized a number of events (e.g. I-Semantics, Open Data Camp) and successfully supported a number of organizations in the publishing of Linked Data (e.g. Deutsche Bibliografie, INFOSO Digital Agenda Scoreboard).

Outlook. After having a stable framework for the integration and combination of different tools and technologies available with the LOD2 Stack we aim to further enrich the Stack with additional tools and algorithms. We also plan to support with next development iterations of the stack complete domain specific linked data management and integration workflows. Since 4 new eastern European-partners are joining the LOD2 consortium through a consortium enlargement in the next year we will complement the LOD2 work programme with an Linked Data technology internationalization effort and an additional use case targeting public procurement.



Figure 3: The LOD2 consortium during the kick-off meeting in September 2010 in Leipzig.

The LOD2 consortium comprises with its academic partners Universität Leipzig, Freie Universität Berlin 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 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).