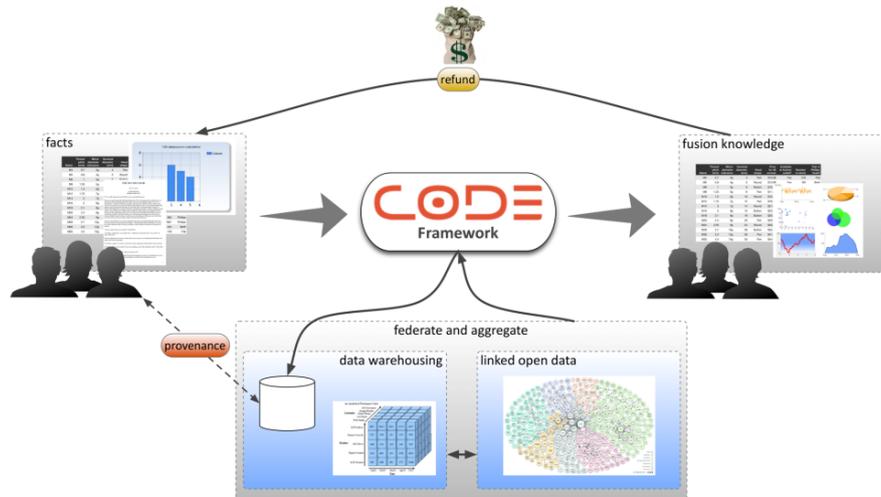


Publishable summary



CODE's vision is to establish the foundation for a web-based, commercially oriented ecosystem for Linked Open Data (LOD). Such an ecosystem will trigger new business models around linked data and new opportunities for leveraging the wealth of today's data.

Our use case focuses on research papers – the treasure chest of western knowledge society - as a source for mining facts and their integration into LOD repositories and light-weight ontologies. Hence, we will leverage the wealth of knowledge contained in research publications on a semantic, machine-readable level by creating the Linked Science Data cloud.



The project focuses on research and development in the following areas:

- Crowd-sourcing enabled semantic enrichment & integration techniques for integrating facts contained in unstructured information into the LOD cloud
- Federated, provenance-enabled querying methods for fact discovery in LOD repositories
- Web-based visual analysis interfaces to support human based analysis, integration and organisation of facts
- Socio-economic factors – roles, revenue-models and value chains – realisable in the envisioned ecosystem

Achieved results

The project developed three prototypes by analysing current data marketplaces and potential revenue models. The prototypes cover different kind of ecosystems around (Linked Open) Data for research, namely:

1. **Mendeley Research Paper Mining API¹**: The prototype extracts facts in form of tables and figures from research papers, links them to Open Data and enables their utilization through the Mendeley Desktop Client and the Mendeley API. The Mendeley API serves as platform for third party application provider to utilize extracted research data within their products. Currently over 1 Million Open Access articles have been enriched and are available through the Mendeley Desktop and API.

¹ <http://www.mendeley.com/r/desktop-about-summary-tab>

2. **MindMeister Semantic Mind Maps²**: This prototype emphasizes the consumption and provision of Linked Open Data through mind maps on the MindMeister platform. Mind maps allow for a user-friendly way of structuring knowledge and can be translated directly to the SKOS vocabulary. Two services foster the use of Open Data technology: First, a service called Wunderkind has been introduced that allows automatically extending mind maps with semantic web concepts. Second, on top of the CODE services for extracting structures from PDF documents, MindMeister offers functionalities to automatically create presentations from a research paper. MindMeister follows a freemium model, where Linked Open Data empowered services increase the number of paying users. Currently over 100,000 mind maps are available as LOD SKOS Thesaurus.
3. **42-data – A Data Flea market for Research**: As a third scenario we introduced a data “flea market” for research data called “42-data”³. The data flea market has been inspired by the fact that the monetary value for research data is rather low, but that there are potentially a high number of interested persons. Since data without interpretation remains without value, we implemented special discussion and bookmarking functions around data-centric resources with the aim of building up a community for discussing and sharing both, data and insights generated through data. So it is the first socio-data oriented ecosystem that aims to build a community around Linked Open Research Data. Although we haven’t succeeded in building a community around data yet, we still see the need and opportunity for opening up available data to social processes. However, a number of technical challenges like discoverability, availability and data quality have still to be solved in order to make this reliably possible. As outlined by Clay Shirkey, systems have to become technologically boring to become socially interesting. From our point of view, Linked Open Data has not achieved that status yet.

The above scenarios build on top of CODE services, which are also valuable contributions to potential future ecosystems. In particular we developed

- Scalable structure analysis and semantic annotation services for PDF documents to extract tabular data and annotate entities from computer science and the biomedical domain. During the project we annotated 1 Million open access research publications.
- The CODE Annotator Tool⁴ is a web based tool the showcases the developed extraction and enrichment technologies. It allows to interactively annotate data-sets imported from an research’s Mendeley library. The types of annotations can be defined via MindMeister mind maps. Finally, machine learning components can be applied on the annotation to produce a model, which then can be shared and re-used by others.
- Feedback-enabled, adaptive disambiguation services called DoSeR⁵ for integrating textual data with semantic concepts of the Linked Data Cloud. The service does not only cover traditional entity disambiguation but also supports type inference of table columns using Linked Open Data. The service has been made available under a permissive Open Source license.
- Balloon⁶ as Linked Data Aggregation and Consumption services which provide query federation mechanisms and common mining functions for Linked Open Data. When analysing existing data value chains we identified a possible future high impact scenario in

² <http://www.mindmeister.com/help/map/experimental>

³ <http://42-data.org/home>

⁴ <http://code-annotator.know-center.tugraz.at>

⁵ <https://github.com/Quhfus/DoSeR>

⁶ <http://schlegel.github.io/balloon/>

the re-use of Open Data in data analytical processes. Balloon fosters a first step into that direction. Balloon is also available under an Open Source license.

- Bacon⁷ as services for semi-automatically storing and merging the extracted and aggregated facts in the recently established RDF Data Cube Vocabulary – a schema for representing data warehouses like data in the Web of Data. Bacon is also available under a Open Source license.
- The CODE Query Wizard⁸ for easy Linked Open Data querying and discovery. The CODE Query Wizard provides an easy to use interface for non IT experts. The interface currently provides access and visualisation capabilities for over 10000 data cubes which have more than 2.5 Million observations.
- The CODE Visualisation Wizard⁹ to visualise and analyse RDF Data Cubes aggregated and created via CODE services (like the Query Wizard).

Services and service descriptions are made available via the CODE Homepage¹⁰. Although prototypes and services show potential, we found that there are still a number of key issues for ecosystems using Linked Open Data to be solved. Discoverability of data, quality of open data, technical reliability of the underlying services and, most importantly, ease-of-consumption for Linked Open Data needs to be improved such that Linked Open Data receives a broader uptake and a high impact.

Project Partners:

Know-Center (Graz, Austria; know-center.tugraz.at) The Know-Center is Austria's research centre for knowledge management and knowledge technologies. Since its establishment in 2001 it has been an innovation point at the interface between science and industry, successfully conducting over 300 application-oriented research projects in cooperation with academic institutions and companies. Know-Center develops innovative knowledge services utilizing semantic technologies and standards. For example, its KnowMiner software framework offers a rich set of knowledge discovery and visual analytics technologies targeting the analysis of large, heterogeneous data repositories. The Know-Center is being funded by Austria's Competence Center Program COMET.



University of Passau (Germany; uni-passau.de) Founded in 1978, the University of Passau combines a most up-to-date infrastructure with state-of-the-art technology, offering its over 10,000 students an ideal place to study. The two chairs involved in CODE, the chair of distributed information systems and the chair of media informatics, investigate distributed databases, multimedia system and automatic media analysis technologies to improve media access and management.



Mendeley (London, UK; mendeley.com) Mendeley is one of the world's largest research collaboration platforms, used by over 1.6 million researchers worldwide. Mendeley provides real-time statistics, trends by research area, and recommendations for related research based on its crowd-sourced database of over 225 million research documents. Launched in January 2009, the company has offices in London and New York, and its investors include former founders and executives of Skype, Last.fm and Warner Music Group.



⁷ <https://github.com/bayerls/bacon>

⁸ <http://codev.know-center.tugraz.at/search>

⁹ <http://codev.know-center.tugraz.at/vis>

¹⁰ <http://code-research.eu/code-results-at-a-glance>

MeisterLabs (Munich, Germany; meisterlabs.com) MeisterLabs develops and provides web-based productivity tools focusing on simplicity, usability and easy collaboration. Its flagship product MindMeister is the market-leading web-based mind mapping and brainstorming solution. Over 300 million ideas have been generated by individuals and businesses using its unique, award-winning interface. As a cloud-based solution, MindMeister functions in any web browser and comes complete with native mobile applications for iPhone, iPad and Android.



Contact

Project Coordinator of CODE

Univ.-Prof. Dr. Stefanie Lindstaedt
Know-Center; Graz University of Technology
slind@know-center.at
+43 316 873-30800

Scientific Coordinator of CODE

Prof. Dr. Michael Granitzer
University of Passau
michael.granitzer@uni-passau.de
+49 851 509 3305

Website: <http://code-research.eu/>