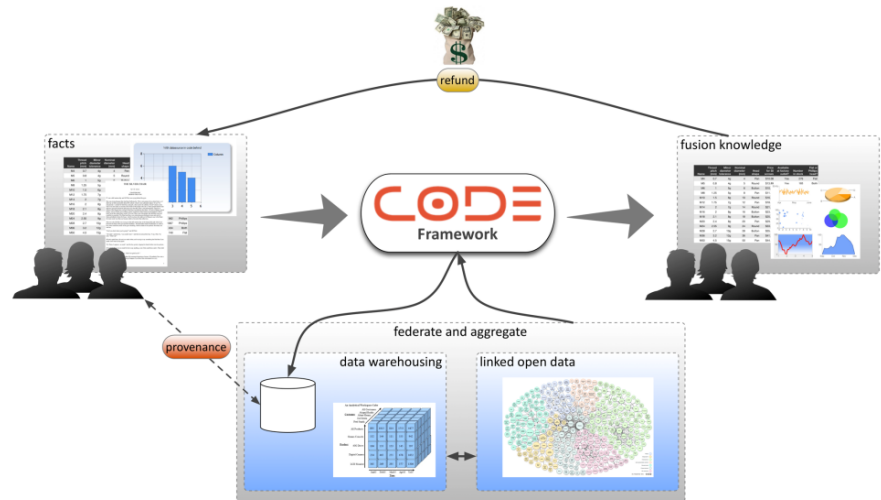


Publishable summary



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 – realizable in the envisioned ecosystem

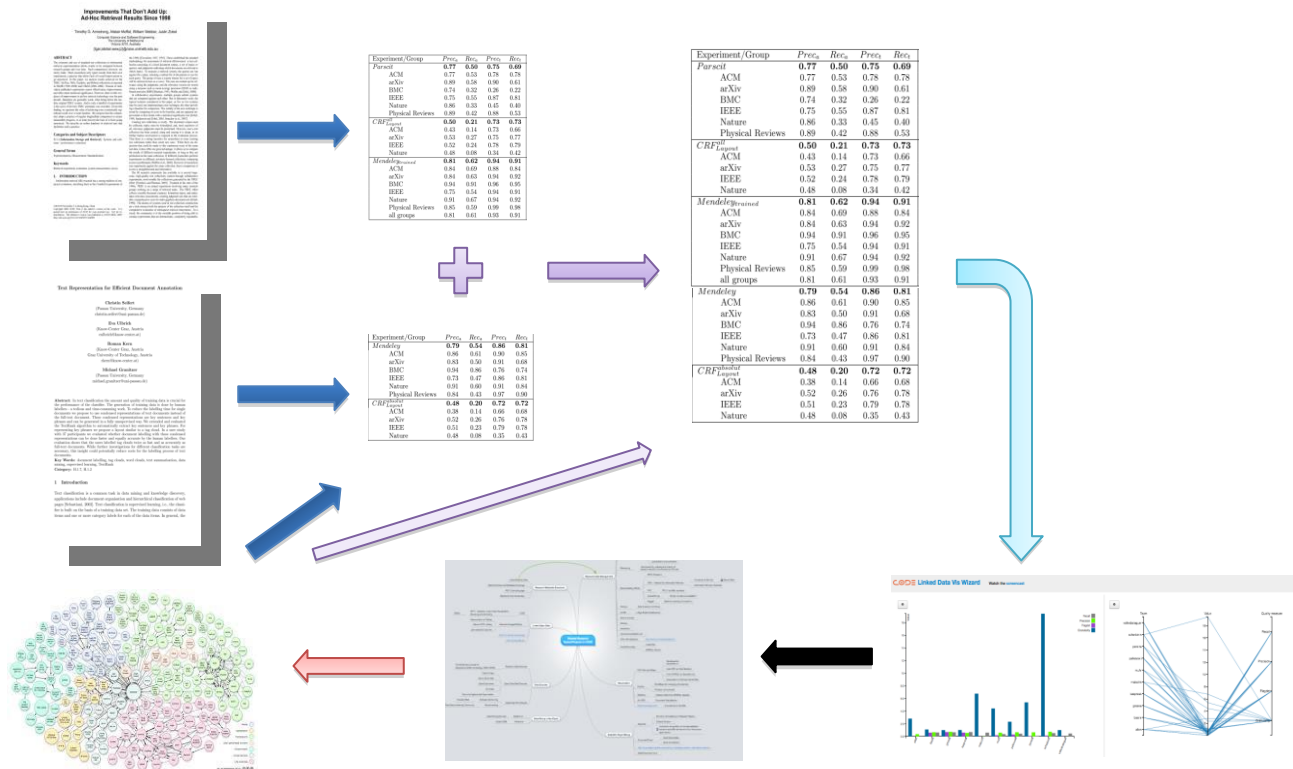
The project has started on 01.05.2012 with a planned duration of two years. In the first year, a set of web-based services was realized for the extracting, integrating and analysing of research data from unstructured (i.e. research papers) and structured data (i.e. linked open data). Particularly the following services have been made available and been partially integrated:

- Scalable structure analysis and semantic annotation services for PDF documents to extract tabular data and annotate entities from computer science and biomedicine
- Feedback-enabled, adaptive disambiguation services for integrating textual data with semantic concepts of the Linked Data Cloud
- Basic query federation mechanisms for automatically distributing SPARQL queries over more than 100 Linked Open Data Repositories
- Services for semi-automatically storing and merging the extracted and aggregated facts using the recently established RDF Data Cube Vocabulary – a schema for representing data warehouses like data in the Web of Data

- Easy to use Visual Analytics interfaces for querying and analysing linked data repositories, including data aggregated and created via CODE services

Services and service descriptions are made available via the CODE Homepage¹ and integrated into a special release of the “Mendeley Desktop”².

From the data-centric point of view (see image bellow) our services enable users to extract facts in tabular-form from scientific publications (PDFs), to semantically annotate the extracted facts and have them stored for further analysis. CODE semantic store persists these facts and mirrors selected parts of the LOD cloud offering federated search functionality. Users can search for facts with a simple to use user interface and merge the retrieved information into new factual data sets. Finally, these factual data sets can be visually analysed and organised, and stored into the LOD cloud providing an added value for the research community.



An important goal of the project is to investigate the opportunities for establishing a value-creation-chain among traditional (e.g. data provider and consumer) and non-traditional (e.g. data analyst) roles in data marketplaces. We estimated sustainability factors for data marketplaces in research through extensive surveys and expert interviews. Results show the potential for three marketplace concepts: first, a data marketplace on data warehousing and visual analytics concepts for public data in the field of data journalism; second, a non-commercial data marketplace for semantic research papers, and third, a marketplace for data-centric mind-maps and mind-map based presentations.

The exploitation and commercialisation will be based on two rationales: i) reduce the costs for enriching Linked Open Data Sets through crowd-sourcing in combination with the development of

¹ <http://code-research.eu/code-results-at-a-glance>
² www.mendeley.com

intelligent mining technologies, and ii) increase the value of the data by establishing an ecosystem around smart, visual services utilizing the data and through a commercial value creation chain. Such an ecosystem shall provide sustainable growths with a minimum of costs, as it is shown by the Web 2.0 success stories.

The second year of the project will begin by completing the integration of all components and services into an alpha version of the CODE platform, which will be made available to a selected circle of friendly users. Subsequently the platform will undergo two refinement phases and, with the increasing quality of the services, will be made available to a larger audience of researchers.

Project Partners:

Know-Center (Graz, Austria; know-center.tugraz.at) The Know-Center is Austria's research center for knowledge management and knowledge technologies. Since its establishment in 2001, it has been an innovation point at the interface between science and industry. The Know-Center develops innovative knowledge services utilizing semantic technologies and standards. It 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.



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.



Contact

Project Coordinator of CODE

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

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

Scientific Coordinator of CODE

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