CODE
Commercially Empowered
Linked Open Data Ecosystems in Research

D 7.2 – Exploitation and Sustainability Plan

Summary: This deliverable reports the plan of actions for future exploitation of the achieved results and for guaranteeing the sustainability of the developed platforms. The document is based on contributions coming from all partners describing their planed exploitation activities, the target groups addressed by these activities, as well as the relevance and the expected impact to their organisations. Additionally, we introduce joint plans involving academic and commercial exploitation. Exploitation along the following routes is considered: exploiting services and prototypes developed in the project, exploiting new technologies used in the project, exploiting data and contributing to the data economy, and finally contributing source code to the open source community.

Project Acronym	CODE
Grant Agreement number	296150
Project Title	Commercially Empowered Linked Open Data Ecosystems in Research
Date	2014-04-30
Nature	R (Report)
Dissemination level	RE (Restricted to a group specified by the consortium, including the Commission Services)
WP Lead Partner	MindMeister
Revision	Final revision
Authors	Vedran Sabol, Nina Simon, Kris Jack, Michael Granitzer, Michael Hollauf

Consortium: KNOW Center   UNIVERSITÄT PASSAU   MENDELEY   meister LABS

SEVENTH FRAMEWORK PROGRAMME

Project Officer & Project Coordinators

Project Officer	Stefano Bertolo	European Commission DG CONNECT Unit G3 EUROFORUM 01/293 - 10 rue Robert Stumper, L-2557 Luxembourg stefano.bertolo@ec.europa.eu
Project Coordinator	Stefanie Lindstaedt	Inffeldgasse 21a, 8010 Graz, Austria +43 (316) 873-9250 (phone) +43 (316) 873- 9254 (fax) slind@know-center.at
Scientific Coordinator	Michael Granitzer	Innstrasse 33, D-94032 Passau +49(0)851-509-3305 michael.granitzer@uni-passau.de

Document Revision History

Revision	Date	Author	Organization	Description
1st Draft	2014-04-07	Vedran Sabol	Know-Center	Document structure, Summary, Introduction
2nd Draft	2014-04-16	V. Sabol, N. Simon	Know-Center	Ideas Know-Center exploitation
3rd Draft	2014-04-17	Kris Jack	Mendeley	Mendeley exploitation
4th Draft	2014-04-20	Michael Granitzer	Uni Passau	Uni Passau exploitation
5th Draft	2014-04-21	Vedran Sabol	Know-Center	Know-Center exploitation
6th Draft	2014-04-23	Nina Simon	Know-Center	Know-Center exploitation feedback
7th Draft	2014-04-25	Vedran Sabol	Know-Center	Joint exploitation
8th Draft	2014-04-28	Michael Hollauf	MeisterLabs	MeisterLabs exploitation
Final Version	2014-04-30			

Table of Contents

1 Introduction

The central objective of the CODE project is to establish the foundations for a web-based, commercially oriented ecosystem for Linked Open Data (LOD). A multitude of services, prototypes and platforms supporting this objective were developed within the project. The project's use cases focus on two main data sources: i) research papers as a source for mining facts and their integration into LOD repositories and ii) the wealth of semantic data already available in the Linked Open Data Cloud. Using these sources the CODE services produce and store new sematic data sets which are subsequently made available to users for further consumption.

In the following, plans for exploiting the prototypical services and the produced data are described. Each member of the consortium outlines the planned exploitation activities and the addressed target groups and provides a statement on the expected impact of CODE results on their organisation. Possible routes of exploitation include (but are not limited to):

- Exploiting services and prototypes developed in the project
- Exploiting new technologies the partners came in contact with during the project
- Exploiting data and contributing to the data economy
- Contributing to the open source community

After the project end, we identified the following main routes of exploitation

- Due to the uptake of the Research Paper Mining features, Mendeley plans to take the developed services in-house and to develop it further.
- Also, MeisterLabs will utilize CODE services in their product and offer them as part of their Freemium package
- University of Passau plans to continue 42-data especially in educational processes. University of Passau considers a spin-off together with the Balloon platform as underlying Linked Data Mining Services, where the focus is on integrating open data into data mining processes
- Both, Know-Center and University of Passau will exploit the individual technical services developed in future project and research. Since all frameworks are open source we have realized all possibilities to allow exploitation by third parties.

2   Exploitation Plans

In this chapter project partners describe their individual exploitation plans. Also, joint exploitations plans are briefly outlined.

2.1   University of Passau

The exploitation strategy at University of Passau is twofold: First, we publish results as Open Source in order to give others the opportunity to uptake and refine prototypes and to re-use it in other research projects. Second we plan on a start-up around 42-data and Balloon.

University of Passau made three prototypes open source:

1. **Balloon** for crawling and aggregating linked data[1]
2. **Bacon** as framework for the integration and merging of RDF Data Cubes[2]
3. **DoSeR** as a prototype for word sense and table disambiguation[3].

DoSeR will be re-used in future research projects and probably integrated into the Apache Marmotta Open Source project. We plan to continue our efforts in disambiguating tables from research publications and publish them into our Linked Open Data Endpoints. The current set of 3000 tables will be extended by tables extracted from further arXiv publications.

Balloon and Bacon will be also continued in future research projects. However, beyond this we identified Linked Data Analytics as a Service as potential start-up opportunity on top of the developed frameworks. In particular, services provided by Balloon, like for example identifying semantic similarity between two resources/strings, or endpoint discovery etc. can be made available via a service level agreement. We will particularly target the field of Big Data Analytics and offer services to extend Data Mining techniques and Big Data data sets with Linked Open Data. By injecting Open Data into Big Data analytics we see a large potential future business opportunity. We already engaged with a few companies in the area of Passau and conducted informal talks. One company is a large retailer requiring Open Data to boost their internal sales analytics. The second company is an analytical company that conducts energy analytics for energy heavy industries. Both have been interested in using our services to gather and integrate Open Data.

42-data will be also part of the start-up strategy for building a community around open data in research and a sustainable platform. However, this requires further development of the platform and corresponding seed funding. Both Chairs at University of Passau commit parts of their resources to the further development of the platform for at least one year. Most crucial will be the development of corresponding content, i.e. data and textual discussions around data and to build up a larger community. In order to do so we will utilize the platform in educational processes and engage in potential industry applications. In the next half year we aim to overcome the current usability hurdles in the platform to open it for a wider audience.

---

[1] http://schlegel.github.io/balloon/

[2] https://github.com/bayerls/bacon

[3] https://github.com/Quhfus/DoSeR

The exploitation route for 42-data may also involve other partners, in particular the Know-Center in Graz, where we aim to conduct joint dissemination to academia and to take up 42-data also in educational processes at Graz University of Technology.

2.2 Know-Center

Know-Center's exploitation plans primarily focus on two groups of prototypes we have developed within the project:

1) **Enrichment tools**
   a. **Enrichment Service** is a Web-based service for extracting information, such as tables, figures, references, document structure and named entities from scientific publications in PDF format.
   b. **Annotator Tool** enables experts to annotate documents, apply machine learning techniques and finally share created models with others.
2) **CODE Wizards** provide easy to use discovery, exploration and visualisation of Linked Open Data (LOD) targeting non-IT experts.
   a. **Query Wizard** supports Google-style searching in Linked Data and setting up of the resulting data set in the form of a table.
   b. **Visualisation Wizard** provides automatic visualisation of statistical data sets (Data Cubes) through intelligent mapping of data onto visual properties of different visualisations.

Our plans for exploiting CODE results are two-fold and will be targeting the academia and the industry. Both exploitation strategies will contribute to maintaining and further development of the prototypes.

Academic/Scientific Exploitation

We will use CODE tools and services for teaching at the Graz University of Technology. In lecture exercises as well as in Bachelor and Master's Theses we will employ our prototypes, letting students improve available and implement new features, and test novel methods for creating, interacting with and analysing Linked Data.

Finally, student engagement with the tools will trigger the creation of new Linked Data and insights. For example, the Enrichment Service produces new data which can be integrated into LOD using services from University of Passau. Query Wizard can be used to create new statistical data sets from the wealth of data already available in the LOD cloud. 42-data question and answering platform, developed by the University of Passau, will be used by students and lecturers as a tool for, data analysis, discussion and presentation of findings and conclusions.

Refinement and further development of our prototypes and the creation of Linked Data sets will be exploited for future research endeavours. In addition, the strategy to grow the user community through scientific events and university lectures is the basis to achieving a critical mass of users making commercial exploitation viable in the future (e.g. in a Freemium model).

Commercial Exploitation

Know-Center's strategy for deploying the developed prototypes in commercial settings is based on the Austrian COMET programme, which is managed by the Austrian Research Promotion Agency (FFG). COMET supports applied research projects to promote the transfer of research results from

the academia into the industry. The Know-Center is a COMET-funded research centre. Within this programme, applied research projects are partially funded by COMET and partially by the Know-Center's industry partners. We plan to intensify the demonstrations our CODE prototypes in B2B meetings with our industry partners with the goal of acquiring new COMET projects.

Extraction service will be exploited within cooperation projects with Mendeley. The goals are to improve the quality of extraction results and deploy the service at the respective companies. For projects with other industry partners we plan to adapt the extraction service to specific industry branch needs. Extraction algorithms developed for the enrichment service can be tuned to application domains other than the currently supported Computer Science and Bio-Med, such as Mobility/Automotive.

CODE Wizards are independent of the application domain and can be used in a variety of scenarios where ad-hoc search and analysis of Linked Data is required, whereby their applicability depends on the availability and quality of the data. For example, open governmental and statistical data (such as EU Open Data, Vienna Linked Open Data or Eurostat Linked Data) fulfil these two requirements. Potential target groups include:

- Media and data journalists reporting on politics and finance will be interested in utilising the wealth of data published in the LOD Cloud to support their statements. CODE Wizards provide an easy to use tool chain for accessing and selecting the data of interest, and visually communicating the findings in a manner understandable to the public.
- The management and the business development departments in companies are interested in relating their own, internal data with the data openly available in the LOD Cloud. CODE wizards will enable business managers to explore market opportunities and will provide decision support based both on the LOD and the internal data. A potential application arises when, for example, a company plans to export in a new country/region. In such a scenario open governmental data and the company's business plan need to be analysed, compared and matched.

Integration of developed technologies in our commercial product lines is also planned. For example the PDF extraction technology will be integrated into Know-Center's upcoming Sensium[4] Software as a Service (SaaS) platform. Sensium is a scalable data mining and knowledge discovery platform targeting primarily analysis of text data.

Open Source Contributions

Source code produced during the project is or will be released under an open source licence. Know-Center pursues a dual-licensing scheme where access to CODE results is available either through a commercial licence or through the GNU Affero General Public License version 3 (AGPL3).

As with April 2014 the following prototypes have been released as open source[5]:

- **Extraction Service**
- **Annotation Tool**

As recently approved by the Know-Center management, following prototypes will be released under open source in the near future:

- **Query Wizard**
- **Visualisation Wizard**

---

[4] www.sensium.io
[5] https://knowminer.at/svn/opensource/projects/code/trunk

2.3 Mendeley

The CODE project has allowed Mendeley to explore the use of Linked Data in a commercial setting with an active user community. Mendeley developed a platform, in the form of an API, which exposes semantically enhanced research data, and takes it to a large community of researchers through the vehicle that is Mendeley Desktop. Usage of these features by real users indicates that they will prove to be popular and, as such, Mendeley intends to continue supporting them and developing them further.

Originally, it was envisioned that the platform Mendeley would integrate all of the CODE tools together. As the project progressed, however, it became obvious that there was too much variety in the tools developed in order for them to fit naturally within Mendeley's community platform. As such, the main platform that integrates the tools together is now 42-data. Mendeley takes on the role of being a data provider, providing the semantically enriched research data to 42-data. This fits with Mendeley's platform-based model, providing data upon with other applications can be built.

The CODE partners produced the research data used by Mendeley. Not all of this data has been included in the API and therefore Mendeley Desktop due to quality issues. Overall, the data provided by CODE partners includes the table of contents, tables, figures and entities extracted from Mendeley's collection of Open Access research articles. The entities were not only extracted but also disambiguated against Open Linked Data sources.

Mendeley exploited the data by integrating it into its API and developing new features that expose it in through Mendeley Desktop. The data integrated into the API is available for all users to access and make use of for commercial and non-commercial applications. Through user testing with non-functional prototypes through to a fully functional prototype, Mendeley was able to test how useful the features were to real users. The usage report provided in deliverable D5.3 indicates that the features are being used enough to warrant keeping them in Mendeley Desktop and exploring future development of them into fully developed features. Although Mendeley was not able to make use of the disambiguated entities, as the quality of the disambiguation was so poor that it couldn't be released to real users on a large scale and thus enable the crowdsourcing phase, it will continue to explore such research. Due to the positive response that Mendeley has had to these features, it is expected that it will have an impact in increasing the attractiveness of Mendeley's Freemium offering.

In the future Mendeley plans to move the extraction services run by the Know Center in-house so that they can be fully developed to a production-level of development and run with the operational guarantees expected from Mendeley's services. Many of the costs of running the services are one off costs per document processed and are currently low enough that they justify continuing to run them. Mendeley also has similar services that it currently runs allowing costs to be shared over various projects.

Mendeley will continue to use the tools that have been developed and prototyped during the CODE projects. In particular, allowing users to see the table of contents, tables and figures that appear in research papers will continue. Mendeley hopes to expand the number of papers that are currently processed and the quality of the data that is extracted. By moving these technologies in-house, Mendeley can continue to provide such semantic data to researchers both through Mendeley Desktop's premium plan and through Mendeley's API.

2.4   MeisterLabs

The CODE project gave MeisterLabs a long-sought opportunity to conduct research into the scientifically very attractive areas of semantic technologies and Linked Open Data. As provider of the leading web-based topic structuring tool we receive many requests to provide semantic services for mind mappers, ideally using open web resources to do so.

During the course of the projects a multitude of features and technologies were prototypically implemented, of which three major functionalities made it into a beta stage and were deployed to the users. These are automatic topic suggestions via **WunderKind**, import of **PDF research papers** with table of content extraction, and **RDF export**. Of the other technologies tested, topic disambiguation and semantic entity enrichment was implemented until alpha stage, including a special user interface, but abandoned due to insufficient results quality and source data inconsistencies (DBPedia). Other smaller features have already been released to the public, such as the automated presentation generation, a by-product of the Know-Center Visualization Wizard integration.

WunderKind

As a commercial company we're most excited about the WunderKind feature, a CODE outcome that bears the potential to be the first step towards an automated "mind mapping assistant" and could substantially increase creativity and productivity of users on our platform. This feature has been released as part of a group of "experimental features" to the entire MindMeister user community of about 2.2 million users in late April 2014. We will continue to work on quality improvements with the CODE partners, especially focusing on improved automated disambiguation through usage of additional context (map parent, map title). After the experimental stage, which we expect to last about 6 months, we will launch the feature as part of our Freemium offering, accompanied by an extensive marketing push (press release, videos, etc). Free users will be able to perform a limited number of queries for testing and receive an upgrade trigger once these are exceeded. If successful we will integrate the feature also into the other MindMeister clients for iOS and Android. Target users are not only researchers but all MindMeister users who work with our tool to structure topics and domains, collect data or perform creative writing in mind maps.

PDF research papers

MindMeister prides itself on its many import and export facilities. An area where the tool has been lacking is the import of non-trivial text data, such as PDF documents. This gap is now closed through the CODE PDF extraction, which allows uploading research papers (and also many other similar PDF documents) and obtaining a mind map containing the table of contents. This is also available already as experimental feature and will be integrated into the main feature set after a short test phase (2-3 months). As the extraction service is currently hosted at the Know-Center in Graz, thereby taking control out of our hands, we have plans to in-source the mechanism before launching in in production. Service costs for this are low and should be covered by the benefits of the extra functionality. The extraction technologies developed during the code project bear further potential also for other applications within the mind mapping area, something we're eager to explore with the Know-Center after conclusion of the project. For this feature the current target audience is mainly researchers. However, if the extraction can be generalized to more document genres, its target audience will surely increase.

RDF export

Publishing of mind maps in RDF format is on of the earliest outcomes of the CODE project. It resulted in the enrichment of the Linked Open Data cloud with MindMeister's over 120,000 public mind maps (or structured thesauri) and will continue to do so with all new maps published. While the technology itself is only interesting to a small subset of users, potential exploitation routes are along data analysis of the provided data. We will therefore work towards a more structured and quality-assured library of public maps where domain-specific subsets could be offered commercially as structured thesauri. Furthermore, the RDF export implementation lends itself to developing an OPML export with little additional effort, something which has larger applications among the general public (OPML is a popular exchange format for hierarchical data).

2.5   Joint Exploitation Plans

Joint exploitation plans target the scientific community, driven by University of Passau and Know-Center, and commercial utilisation within a joint project between Mendeley and Know-Center.

University of Passau and Know-Center plan to organise events, such as workshops and challenges to generate awareness in the scientific community, trigger usage of CODE prototypes, and collect valuable feedback allowing us to improve the developed technologies. In particular, at i-KNOW 2014 conference we will organise the ALOA Challenge[6] on Getting Answers through LOD. Within the challenge, the CODE tool chain shall be used to generate insights from data. As the i-KNOW attracts both scientists and entrepreneurs, we also expect to gain visibility in industry circles which will our aid commercial exploitation efforts. Further, University of Passau and the Know-Center will support the European Youth Award 14 by using special campaigns in 42-data. The aim is to give young people in Europe the opportunity to conduct data centric discussions using Europe's Open Data portal like the EU Digital Agenda Scoreboard or Eurostats.

Know-Center and Mendeley are in the process of setting up a COMET project for commercially exploiting CODE results. In this cooperation the Extraction Service, which is currently running at the Know-Center, will be deployed within Mendeley. The service shall be developed further in order to achieve production-level stability required by Mendeley. Additionally, the Know-Center will extend the extraction functionality to support extraction of vectors graphics and citations.

---

[6] http://i-know.tugraz.at/i-science/call-for-contributions-aloa-challenge

- 9 -

3 References

[1]   CODE Description of Work, Version Date 20012-01-23