Summary: The project has been successfully completed on 30th April 2014. We developed 3 prototype ecosystems and a set of services underlying those ecosystems. Our services have been integrated with Mendeley and MindMeister, whereby especially the Mendeley prototype achieved a wider uptake of the CODE features. Based on year 1 review feedback, the project roadmap was updated and marketplace scenarios have been revisited, which resulted in delaying service integration and deployment to a wider public by 1 month. Additional development work necessary for the 42-data platform further reduced the time for community building and evaluation, hindering the platform’s uptake. However, due to the developed services and ecosystems we have established a foundation for future uptake especially through our two industry partners.

<table>
<thead>
<tr>
<th>Project Acronym</th>
<th>CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grant Agreement number</td>
<td>296150</td>
</tr>
<tr>
<td>Project Title</td>
<td>Commercially Empowered Linked Open Data Ecosystems in Research</td>
</tr>
<tr>
<td>Date</td>
<td>2014-04-30</td>
</tr>
<tr>
<td>Nature</td>
<td>R (Report)</td>
</tr>
<tr>
<td>Dissemination level</td>
<td>PU (Public)</td>
</tr>
<tr>
<td>WP Lead Partner</td>
<td>WP8 Project Management, Know-Center</td>
</tr>
<tr>
<td>Revision</td>
<td>Final Version</td>
</tr>
<tr>
<td>Authors</td>
<td>Michael Granitzer, Kris Jack, Michael Hollauf, Florian Stegmaier, Roman Kern, Patrick Höfler, Maya Hristakeva, Vedran Sabol</td>
</tr>
</tbody>
</table>

Consortium:

Statement of originality: This document contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

This document reflects only the author’s views and the European Community is not liable for any use that might be made of the information contained herein. © CODE Consortium, 2012
Project Officer and Project Coordinators

<table>
<thead>
<tr>
<th>Position</th>
<th>Name</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Officer</td>
<td>Stefano Bertolo</td>
<td>European Commission</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DG CONNECT Unit G3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EUROFORUM 01/293 - 10 rue Robert Stumper, L-2557 Luxembourg</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:stefano.bertolo@ec.europa.eu">stefano.bertolo@ec.europa.eu</a></td>
</tr>
<tr>
<td>Project Coordinator</td>
<td>Stefanie Lindstaedt</td>
<td>Know-Center GmbH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inffeldgasse 13, 8010 Graz, Austria</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+43 316 873-30800 (phone)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+43 316 873-1030800 (fax)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:slind@know-center.at">slind@know-center.at</a></td>
</tr>
<tr>
<td>Scientific Coordinator</td>
<td>Michael Granitzer</td>
<td>University of Passau</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Innstrasse 33a, D-94032 Passau</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+49(0)851-509-3305</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:michael.granitzer@uni-passau.de">michael.granitzer@uni-passau.de</a></td>
</tr>
</tbody>
</table>

Document Revision History

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Author</th>
<th>Organization</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Draft</td>
<td>2014-03-30</td>
<td>Vedran Sabol</td>
<td>Know-Center</td>
<td>Structure, publications, events, deliverable lists</td>
</tr>
<tr>
<td>3rd Draft</td>
<td>2014-04-25</td>
<td>Michael Granitzer</td>
<td>University of Passau</td>
<td>Project Overview and Overall Status Report</td>
</tr>
<tr>
<td>4th Draft</td>
<td>2014-04-28</td>
<td>Patrick Höfler</td>
<td>Know-Center</td>
<td>WP4</td>
</tr>
<tr>
<td>5th Draft</td>
<td>2014-04-29</td>
<td>Michael Hollauf</td>
<td>MeisterLabs</td>
<td>WP7</td>
</tr>
<tr>
<td>6th Draft</td>
<td>2014-04-30</td>
<td>Roman Kern</td>
<td>Know-Center</td>
<td>WP2</td>
</tr>
<tr>
<td>7th Draft</td>
<td>2014-05-02</td>
<td>Vedran Sabol</td>
<td>Know-Center</td>
<td>Management, next steps, showcase materials, summary</td>
</tr>
<tr>
<td>Draft</td>
<td>Date</td>
<td>Author</td>
<td>Organization</td>
<td>Comments</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>-------------------</td>
<td>--------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>8th</td>
<td>2014-05-05</td>
<td>Florian Stegmaier</td>
<td>University of Passau</td>
<td>WP3</td>
</tr>
<tr>
<td>9th</td>
<td>2014-05-06</td>
<td>Michael Granitzer</td>
<td></td>
<td>WP6, added success factors to WP5</td>
</tr>
<tr>
<td>10th</td>
<td>2014-05-16</td>
<td>all</td>
<td>all</td>
<td>Resource consumption data</td>
</tr>
<tr>
<td>Final</td>
<td>2014-05-16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table of Contents

1 Executive Summary - 3 -
   1.1 PROJECT OVERVIEW - 3 -
   1.2 OVERALL STATUS - 6 -
   1.3 MAIN EVENTS - 7 -

2 Project Report - 8 -
   2.1 WORK PROGRESS AND ACHIEVEMENTS - 10 -
      2.1.1 WORK PACKAGE 1 – REQUIREMENTS DEFINITION - 10 -
      2.1.2 WORK PACKAGE 2 – SEMANTIC ENRICHMENT & INTEGRATION - 10 -
      2.1.3 WORK PACKAGE 3 – DATA QUERYING, AGGREGATION AND PROVENANCE - 12 -
      2.1.4 WORK PACKAGE 4 – WEB-BASED VISUAL ANALYTICS INTERFACES - 14 -
      2.1.5 WORK PACKAGE 5 – PLATFORM INTEGRATION & DEPLOYMENT - 16 -
      2.1.6 WORK PACKAGE 6 – SOCIO-ECONOMIC MECHANISMS IN DATA MARKETPLACES - 19 -
      2.1.7 WORK PACKAGE 7 – DISSEMINATION & EXPLOITATION - 21 -
   2.2 PROJECT MANAGEMENT - 25 -
      2.2.1 CONSORTIUM - 26 -
      2.2.2 PROJECT PLAN - 26 -
   2.3 MAIN EVENTS - 27 -
      2.3.1 MEETINGS - 28 -
   2.4 DELIVERABLES AND MILESTONES - 30 -
      2.4.1 DELIVERABLES - 30 -
      2.4.2 MILESTONES - 32 -
   2.5 NEXT STEPS - 33 -

3 References - 34 -

4 Appendix - Showcase Materials - 35 -
1 Executive Summary

1.1 Project Overview

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 different prototypes covering different kind of ecosystems around (Linked Open) Data for research, which are based on services developed in the project. In detail we have developed three ecosystem prototypes, namely

1. **Mendeley Research Paper Mining API**: 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

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

2. **MindMeister Semantic Mind Maps**\(^2\): 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.
- Feedback-enabled, adaptive disambiguation services called DoSeR\(^4\) 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.
- Balloon\(^5\) 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 the re-use of

---

\(^2\) [http://www.mindmeister.com/help/map/experimental](http://www.mindmeister.com/help/map/experimental)

\(^3\) [http://42-data.org/home](http://42-data.org/home)

\(^4\) [https://github.com/Quhfus/DoSeR](https://github.com/Quhfus/DoSeR)

\(^5\) [http://schlegel.github.io/balloon/](http://schlegel.github.io/balloon/)
Open Data in data analytical processes. Balloon fosters a first step into that direction. Balloon is also available under an Open Source license.

- Bacon\(^6\) 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 Wizard\(^7\) 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 Wizard\(^8\) 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 Homepage\(^9\). 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

---

\(^{6}\) https://github.com/bayerls/bacon

\(^{7}\) http://codev.know-center.tugraz.at/search

\(^{8}\) http://codev.know-center.tugraz.at/vis

\(^{9}\) http://code-research.eu/code-results-at-a-glance
and New York, and its investors include former founders and executives of Skype, Last.fm and Warner Music Group.

**MeisterLabs** (Munich, Germany; [meisterlabs.com](http://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**

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

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

Website: [http://code-research.eu/](http://www.code-research.eu/)

1.2 Overall Status

The project has been successfully finished on 30th of April 2014. Based on the feedback from the 1 year project review meeting, the project roadmap has been adapted and marketplace scenarios have been revisited. While the revisiting of marketplace scenarios do not impact the refinement of existing services, it delayed the integration of the services into the different scenarios and hence the capability to face a wider public. By the end of the project, we developed 3 prototype ecosystems and a set of underlying services as basis for those ecosystems. Especially the Mendeley prototype seems to achieve a wider uptake of the CODE features.

However, the end of Phase II of the project, the Beta deployment, ended by 1 month delayed. After the Beta deployment we identified some critical errors and functions missing especially for the 42-data platform, which hindered a wider uptake. The most critical bugs have been corrected, but the 4 months planned for community building, dissemination and evaluation have been reduced to 1 month. The reduced evaluation time had also impact on our success factors, where some of those could not be achieved.

Although we did not meet all success factors, the project can be considered as successful from the consortium points of view. Our prototypes have been integrated with Mendeley and MindMeister, as requested in the first year’s review. Especially the Mendeley prototype reaches wide uptake and constitutes a first step towards factual representations of research results. Further we analysed and identified success factors for data marketplaces and found a missing link on socialising open (research) data. The developed concept and prototype serves as starting point for a hopefully broad community uptake. During the process of developing and evaluating the prototypes we also identified challenges that have to be solved in future Linked Open Data ecosystem in order to
become successful. Central to such ecosystems is data quality, service quality and easier ways for consuming and aggregating Linked Open Data. Core challenges to be solved for sustainable Linked Open Data ecosystems. Although the CODE project could not address all of them, our ecosystems and services may serve as a first starting point. For example, for consuming and aggregating Linked Open Data we extended the scope of the Balloon service to inject open data in data analytics processes. This already triggered interesting cooperation with analytical companies beyond the lifetime of the CODE project.

1.3 Main Events
In the last six months the following main events where CODE has been presented can be reported:

1. Science 2.0 Policy Workshop organised by DG Research & Innovation C2 and ERIAB, 4th November 2013, Brussels
2. ASIS&T 2013 Annual Meeting, 5th November 2013 in Montreal, Canada: project presentation
3. First Standardization and collaboration meeting with Erik Mannes from University Gent, December 17th 2013, University of Passau, Germany
5. European Data Forum 2014, March 19-20, Athens, Greece: project results presentation and exhibit
7. Science 2.0 PhD Spring School, 24th-27th 2014, March, Hamburg, Germany
8. 5th International Workshop on Data Engineering Meets the Semantic Web, co-located with ICDE 2014, April 2014, Chicago, USA: paper presentation

For the full list of events is available in Section 2.3.
2 Project Report

The project has been finished on 30th April 2014 with a final meeting of the consortium. In the project we have analysed existing data-centric and research-centric ecosystem in WP 6 and realised three ecosystems on top of the underlying CODE services (WP 2-5). Two ecosystems have been built around the portals of the two SME partners, namely Mendeley and MeisterLabs, and the third ecosystem introduces a new data flea market called 42-data. We give a brief description of those ecosystems in the following.

- For the Mendeley ecosystem\textsuperscript{10} we utilized CODE Text Mining services for extracting facts in the form of tables, figures and entities from research papers and provide them (i) to users of the Mendeley Desktop and (ii) to third parties via the Mendeley API. We also developed a prototype called CAT (Code Annotator Toolkit) to enable power users in annotating research papers and creating their own annotation models. However, we found that it is beyond the skills of the current user base to utilize this prototype. The ecosystem follows a freemium model with the API as enable for third party applications. The services have been made publicly available as preview in February and as release in March 2014. Till end of April, over 8000 users already engaged with the features and over 500,000 API calls have been made.

- For the MindMeister ecosystem\textsuperscript{11} we emphasized on the consumption and creation of Linked Open Data in form of Semantic Mind Maps. Mind maps can be created through user input, utilizing CODE Linked Data services or from the extracted structures of a research paper. The CODE services are part of the new experimental MindMeister features. Mind maps provide a special presentation mode, where researchers can use mind maps for giving scientific presentations and the Wunderkind services allows to semi-automatically extend mind maps with Linked Data resources. In order to increase the re-use of Linked Open Data, mind maps can be exported as RDF SKOS Vocabulary. The ecosystem follows again a freemium model, where sustainability is envisioned through making MindMeister services more attractive by using Linked Open Data. Most of the experimental features have been made available to the public end of April, where Wunderkind will be released in May.

- As data marketplace we have developed the 42-data data flea market for research data. It capitalizes on the fact that today’s Open Data portals lack the ability to jointly generate insights on data. Therefore, we developed a discussion and bookmarking portal around Linked Open Data with a special focus on research data and facts obtained from research papers. 42-data is established as a mix between StackExchange and Delicious, but centred on research data. We identified that the data itself does not bear huge monetization potential. Hence we opted to support financial transaction between participants on the portal in form of donations. Target for donations can be any person registered in the portal, i.e. data provider, users answering questions, users providing resources or users asking question. The portal has been made publicly available by mid of February in an alpha version. After removing critical bugs and after adding important features like the CODE Visualisation Wizard we have announced it to a broader public by end of March. Community building is still ongoing and the most challenging part currently. At end of April we have only attracted 41 users and 205 resources. Our evaluation shows, that the platform lacks content which makes it less attractive to new users. Furthermore, utilizing Open Data in discussions has a steep

\textsuperscript{10} \url{http://blog.mendeley.com/progress-update/desktop-contents-tables-and-figures/}

\textsuperscript{11} \url{http://www.mindmeister.com/help/map/experimental}
learning curve: although the user community is willing to utilize Open Data, they have challenges in developing the necessary skills. Our exploitation plan foresees to develop the platform further, provide more content and reduce the entry barrier for new users.

**CODE Services**

All three ecosystems rely on the following CODE services developed in WP 2-WP 4:

- The CODE PDF Extraction and Annotation services that extracts tables, figures, table of content and entities from PDF documents. The extraction service outputs JSON-LD as semantic web format.
- The CODE Annotator Toolkit (CAT) that allows users creating their own annotation models over research publications. However, the toolkit has not been integrated by Mendeley yet.
- The CODE Disambiguation Service named DoSeR\(^\text{12}\) that links entities in the Linked Open Data cloud and that allows inferring column types for extracted tables. The framework is available under a permissive Open Source license.
- The Balloon\(^\text{13}\) service as a Linked Open Data aggregation and consumption services that provides query federation capabilities and REST based analysis functions over Linked Open Data endpoints. Balloon is available under a permissive Open Source license.
- The Bacon\(^\text{14}\) service providing RDF Data Cube storage and merging capabilities. Bacon is also available under a permissive Open Source license.
- The CODE Query Wizard for easy querying of Linked Open Data through non-IT experts and the CODE Visual Analytics Wizard, which consumes RDF Data Cubes and provides visual analytics capabilities on top of it.

Detailed descriptions are provided in the individual WP 2-WP 4 reports as well as in the corresponding deliverables.

**Exposed Datasets**

The following data sets have been made available by the project through the above services:

- 1.2 Million semantically annotated Open Access publications consumable via Mendeley’s API
- Over 100,000 semantic mind maps represented as RDF SKOS Thesaurus available either as SPARQL endpoint\(^\text{15}\) or via URL dereferencing.
- 10,000 annotated research papers available as JSON-LD dump
- 3,928 automatically extracted and annotated RDF Data Cubes, available via the 42-data endpoint.
- 205 resources available via 42-data.

All datasets will continue to grow due to user uptake in the three ecosystems.

**Success Factors**

CODE has set very ambitious goals where some of them could be met, while others have not been achieved. The ecosystems developed aimed for two routes:

---

\(^{12}\) https://github.com/Quhfus/DoSeR/

\(^{13}\) https://schlegel.github.io/balloon/

\(^{14}\) https://github.com/bayerls/bacon

\(^{15}\) http://datahub.io/dataset/code-mindmeister-endpoint
First, we integrated CODE services into the partner platforms in order to ensure uptake. This required focusing mainly on stability and quality of features. As a result, we achieved good uptake especially for mining research papers. Usage numbers are still growing and will continue to do so in the near future. Hence, CODE services for consuming and producing Linked Open Data will be sustained and extended by the two SME Partners Mendeley and Meisterlabs after the lifetime of the project.

Second, we focused on using CODE services in a new, experimental platform that combines Web 2.0 features with (Linked Open) data. This resulted in a new platform called 42-data. For such a new platform we have not been able to achieve high uptake yet, but are confident to do so in the future. The success factors show that uptake increases, but only at a slow pace. The community is interested in such a platform, but yet the incentives for investing in labour intensive analysis work are still missing.

Especially due to the low update of the 42-data platform, we could not achieve usage based success factors, namely the number of validated annotations and the number of recommendations accepted. However, most importantly the SME partners will continue the usage and development of CODE services which shows uptake and impact of the project. More Details on the success factors are provided in the WP 5 description in Section 2.1.5.

2.1 Work Progress and Achievements

In the following, results achieved in each work package are summarized along tasks defined in the Description of Work [1].

2.1.1 Work Package 1 – Requirements Definition

Work Package 1 aimed at coining a detailed common understanding on the project goals and to develop concrete use-cases in form of user stories as well as their technical realisation. It provides the basis for all other work packages and for researching market place principles around Linked-Science-Data. The work package has been structured in two tasks – the development of the user centric few in Task 1.1. and deriving the underlying technical concepts in Task 1.2. The work package and all associated deliverables have been finished successfully. As a result, MS 1 Project Setup could be achieved. For detail please see the corresponding deliverables and half-year reports.

2.1.2 Work Package 2 – Semantic Enrichment & Integration

WP2 has the following objectives (Objectives WP 2) for Year 2:

- Refinement of the semantic enrichment and integration models and refinements regarding the implemented semantic enrichment workflows (Task 2.4)

In year 1 WP 2 focused on developing basic services and workflows covering the objectives on semantic enrichment and integration. In terms of infrastructure we developed two web-services, namely the Enrichment Service and the Disambiguation Service. Both services have been integrated into clients of Mendeley and MeisterLabs.

In year 2 we focused on improving and refining these services, as well as assessing and evaluating their performance (see D2.3 for the evaluation report of these services). In this process we have
taken feedback into account, which has been gathered during the integration of the services. The feedback includes technical feedback as well as feedback from user interviews.

**Task 2.4 Refinement & Evaluation**

**Refinement of the workflows:** In the first year we have identified 10 workflow use-cases starting with supporting users in defining the terms and ending up in entity marketplaces, for details please see D2.1.1. In the first year six out of these 10 use cases have been completed. These use cases were further refined, primarily by improving the semantic enrichment and integration algorithms.

In the second project year the coverage of the use cases has been further expanded to cover all proposed use cases. Here the use cases “2.8 - Structured Data Extraction from Table” and “2.10 - Article Navigation through References” have been identified to be of highest priority, thus they received the most attention. Finally these workflows were implemented in the second project year and been put into productive use within the Mendeley client and available to all Mendeley users.

The use case “2.4 – Article Navigation and Summarisation through Entities” has not been followed further as it has been found that from a user experience point of view little benefit can be achieved over the already supported “2.3 – Article Navigation and Summarisation through Structure”. The same is true for “2.7 – Entity Marketplace”, where interviews and feedback from the quality assurance team of Mendeley have indicated little interest to support this use case, specifically within work package 2.

The last remaining use case, “2.6 – New Entity Class Creation” has been addressed in the second part of the year 2 by developing the CODE annotator tool, which is described in D7.5 including the workflow on how to create user specific models.

**Refinement of the semantic enrichment algorithms:** In the first year we initially focused on the development of a service infrastructure which satisfies the functional as well as non-functional requirements. From the non-functional requirements the scalability of the system has been identified as a critical issue in order to reach the foreseen success indicators. For the Enrichment Service, as well as for the Disambiguation Service, we achieved the necessary scalability which allows adding more machines in order to cope with the demand. For details on the related success factors see D5.3.

In regard to the semantic enrichment aspects which relate to the extraction of facts from scientific articles we worked on improving the algorithms for table detection and table extraction and the extraction of facts from the article’s text. Table extraction is motivated by the assumption that tables will often list relevant facts. By refining the extraction algorithms we improved the F1 from 69.59% (phase I) to 77.67% (phase II) for correctly identifying the table cells’ content and then further to 83.95% (phase III).

For the extraction of entities from the text we agreed to focus on the concepts “corpus” (typically a data-set) and “algorithm” for articles from the computer science domain. For the biomedical domain the initial results have not been fully satisfying from a user’s perspective. We addressed this issue by replacing the generic named entities of the CALBC corpus by more specialised entities and relations from the data set assembled by the BioNLP initiative, where we outperform the results from the initial shared task (achieving a F1 value of 75% instead of 73%).

In order to arrive at the extraction of facts, the extraction of entities is combined with the extraction of relations between them. In order to assess the performance of our approach, we used a data set from the BioNLP initiative focusing on the detection of protein components and subunit complex relationships. Here our results are in line with the current state-of-the-art in this domain.
Going beyond the biomedical domain, our focus has been on the computer science domain. Here we developed an ontological structure in the first project year. In the second year we refined this structure and added relationships between the concepts. We worked on semi-automatic extraction mechanisms, as well as fully automatic approaches. Here our results progressed from 45% to 81% in terms of F1 for the “algorithm” concept and from 65% to 85% for the “corpus” concept in our evaluation runs.

The extraction of the structure of a scientific article plays an important role as aid for navigating the article, as well as for generating mind maps, that give an at a glance overview of the article. We improved our PDF extraction pipeline by integrating a reference extraction component, which goes beyond state-of-the-art in this field\textsuperscript{16}. Furthermore, in the second project year we focused on improving robustness and also on covering cases that only arise when working with real-world data, in contrast to well defined scientific data-sets. For example, we made sure that the bounding boxes are pixel-perfect, even for PDFs that do not adhere to the standard.

Refinement of the semantic integration algorithms: In the second project year we worked on improving the user perceivable performance. We achieved major performance improvements, and at the same time the quality of the entity disambiguation algorithms have been tweaked and fine-tuned and disseminated.

Furthermore we worked on the disambiguation of table cell and column type inference, were the later feature plays an important role in the data cube creation. Especially for the table cell disambiguation the results are matching and in part surpassing the current state-of-the-art.

2.1.3 Work Package 3 – Data Querying, Aggregation and Provenance

WP3 of the CODE project has been finished as planned in the DoW. There have been no delays in the completion of the prototypes.

Within WP3, three major open source prototypes have been developed:

- **Bacon**\textsuperscript{17} presenting an open source framework that enables interactive and crowd-sourced Data Integration on Linked Data (Linked Data Integration), utilizing die RDF Data Cube Vocabulary and the semantic properties of Linked Open Data.
- **Balloon**\textsuperscript{18} builds a Linked Open Data index of available structural information. The aggregation of this data in one place enables balloon Fusion and balloon Commonalities to ease the access to Linked Data and discovery new knowledge. Further, subprojects tackle visualization of RDF data as well as the exploitation of extracted structures to find shared characteristics of entities and types.
- **Data Extractor** undertakes a triplification processing chain for semantic lifting of primary research data. Here, tabular data is being transformed into a RDF data cube and stored in the 42-data SPARQL endpoint

\textsuperscript{16} http://www.dlib.org/dlib/september13/kern/09kern.html

\textsuperscript{17} https://github.com/bayerls/bacon/

\textsuperscript{18} http://schlegel.github.io/balloon/
Besides those, open accessible mindmaps of MindMeister have been semantically enriched and stored in a triple store. The MindMeister endpoint is a periodically updated triple store storing semantically enriched public mind maps. This triple store is registered at Datahub.io at the following link:

http://datahub.io/dataset/code-mindmeister-endpoint

The last six months of the project were partitioned in two phases:

- **Phase II** (M19–20) focused on the beta deployment of the refined prototypes
- **Phase III** (M21–24) evaluated the interplay of all prototypes within the 42-data platform

With respect to the aforementioned prototypes, the following enhancements have been made within phase II summarized in Task 3.4 “Refinement & Evaluation”:

**Task 3.4 “Refinement & Evaluation”**

- **Task 3.1: Federated Querying**
  
  CODE Balloon can be seen as the major outcome of Task 3.1. In the refinement phase, it has been extended by the following means:
  
  - Integration of Neo4J as storage layer
  - Crawling process relies on an arbitrary set of properties; currently co-reference, type and subclass properties are being indexed
  - Discovery of common super types as well as common predicates of given semantic entities

  The crawled data of CODE Balloon is hosted at the University of Passau and reachable via FTP at:

  ftp://moldau.dimis.fim.uni-passau.de/data/

  Beside CODE Balloon, a SPARQL endpoint for semantically enriched Mindmeister mindmaps has been created. Further, the SPARQL backend of 42-data has been developed and maintained by WP3 holding approx. 4000 data cubes at the moment.

- **Task 3.2: Linked Data Aggregation**
  
  The Bacon prototype is the major outcome of Task 3.2. Its features have been extended in the following means:
  
  - Complete UI revision
  - Guessing algorithm to identify merge-able data cubes due to similar structure
  - Structural changes of a data cube during merging process
  - Semi-automatic cube extension mode as expert mode of the Data Extractor; here, the service of CODE Balloon searching for common super types is integrated.

  The current prototype is hosted at the University of Passau at the following link:

  http://zaire.dimis.fim.uni-passau.de:8383/cube-merging/select

  The prototype of the Data Extractor is hosted at the University of Passau and reachable at the following link:

  http://zaire.dimis.fim.uni-passau.de:8383/code-server/demo/dataextraction

  Please note, that the prototype can be only used with a registered 42-data user.

- **Task 3.3: Provenance Schema**
The outcomes of Task 3.3 were the definition of fundamental provenance workflows inside the CODE project, along with the definition of a subset of the W3C PROV ontology\textsuperscript{19} as well as a specific CODE Data Provenance Vocabulary\textsuperscript{20}. Those concepts are now applied to marketplace concepts to establish trust mechanisms.

Evaluations of the prototypes are mainly part of phase II and are reflected in D3.4 accordingly. In this deliverable, an in-depth evaluation of the Bacon prototype has been made. In addition, a comparison of the most recently crawled data set of Balloon to the data set crawled in early 2013 has been conducted to show its evolution and latest trends.

2.1.4 Work Package 4 – Web-based Visual Analytics Interfaces

WP4 of CODE has been completed as planned according to the plan laid out in the Description of Work and the feedback received at the first review meeting. There are no major deviations.

The focus of this work package has been on the development of two major components:

The CODE Query Wizard helps people select relevant data from the Linked Data Cloud – to be more exact, from publicly available SPARQL endpoints. The resulting data is presented in an easy-to-use web-based interface that looks and feels similar to current spreadsheet applications. For advanced data manipulation, e.g. aggregations, the data is converted to the RDF Data Cube format. The prototype is publicly available at \url{http://code.know-center.tugraz.at/search}

Once the users have selected the relevant data, the CODE Visualization Wizard (or Vis Wizard for short) supports them in creating visualizations to make the data easier to understand. Multiple visualizations can be combined to form a visual analytics interface. In order to facilitate this step, not only the data but also the visualizations have semantic descriptions in the form of OWL ontologies. The prototype is publicly available at \url{http://code.know-center.tugraz.at/vis}

The last 6 months of the project consisted of two phases:

In Phase II (project month 19–20), the work focused on completing the feature set of the CODE Query Wizard and the CODE Vis Wizard. Among others, the following features were added:

**Query Wizard:**
- Display SPARQL queries and statistics (for semantic web experts)
- Export displayed data as JSON-LD
- Advanced filtering functionality

**Vis Wizard:**
- Coordinated brushing for all visualizations
- Multiple coordinated views for cubes and aggregated cubes

\textsuperscript{19} \url{http://www.w3.org/2011/prov/}
\textsuperscript{20} \url{http://code-research.eu/ontology/code-prov-vocabulary}
For a detailed feature description, please refer to Deliverable 4.3 Refined Semantic Enrichment, Semantic Integration and Visual Analytics Interfaces.

Figure 1: SPARQL query and statistics for experts in the Query Wizard

Figure 2: Multiple coordinated view in the Vis Wizard

In Phase III (project months 21–24), the work focused on integrating both prototypes with the CODE question and answer portal 42-data.org as well as on in-depth user experience testing of the developed visual analytics interfaces. Evaluation of the components and user interfaces has been conducted by performing formative web-based usability testing with friendly users as well as summative in-depth evaluation of the user experience of the CODE Query Wizard and the CODE Visualization Wizard. Some of the evaluation results have already been published in international
peer-reviewed conference articles. For details about the conducted usability studies, please refer to Deliverable 4.4 Usability Evaluation Report.

To sum up, the outcomes of the tasks of WP4 are as follows:

- **All tasks have been completed successfully.**

- **Task 4.1 “Semantic Descriptions for VA Components”** (months 2–6) was concerned with the development of semantic descriptions for user interface components to visualize aggregated information from the linked data cloud. The outcome was the Visual Analytics Vocabulary as reported in Deliverable 4.1 Semantic Descriptions for Visual Analytics Components.

- **Task 4.2 “Visual Analytics Interfaces”** (months 2–12) dealt with the development of analysis components to visualize aggregated linked data and semantic enrichment results. The outcomes were the CODE Query Wizard and the CODE Vis Wizard as reported in Deliverable 4.2 Semantic Enrichment, Semantic Integration and Visual Analytics Interfaces.

- **Task 4.3 “Usability Evaluation & Refinement”** (months 15–24) dealt with further development and enhancement of analysis components to visualize aggregated Linked Data and semantic enrichment results (see 4.3 Deliverable Refined Semantic Enrichment, Semantic Integration and Visual Analytics Interfaces) as well as in-depth usability testing of the developed prototypes (see Deliverable 4.4 Usability Evaluation Report).

2.1.5 Work Package 5 – Platform Integration & Deployment

WP5 began in month 4 and completed at the end of month 24. During this period, resources have been split between three tasks: Task 5.1 (Platform Development and Alpha Deployment); Task 5.2 (Platform Refinement and Beta Deployment); and Task 5.3 (Success Factor Monitoring and Sprint Planning). Task 5.1 completed at the end of month 14 with the Alpha deployment. This deployment attempted to integrate CODE features into Mendeley Desktop in a bespoke release. It included features for enhanced data extraction from PDFs, semantic lookups, enhanced creation and merging of data cubes and data visualisation.

Integrating all of these features into Mendeley Desktop, however, didn’t prove to be a user-friendly experience. Linking out to external websites such as the data cube management and data visualisation sites was off-putting to users, breaking up their natural workflow. In addition, very few users downloaded the bespoke version of Mendeley Desktop making it unlikely for us to hit our target success indicators. Finally, there was an interesting request during the review meeting to integrate the CODE features into a single platform based around a question and answering portal.

Task 5.2 began in month 15, following the completion of Task 5.1, and completed at the end of month 20. We took the learnings from the Alpha deployment into account and realised that trying to integrate all of the CODE features into Mendeley Desktop wasn’t going to create a good user experience. Instead, Mendeley Desktop should only integrate the subset of CODE features that fit naturally with researcher workflows. Namely, it would include the features for enhanced data extraction from PDFs. The other features would be showcased in a question and answering portal that, using the enhanced data extracted from PDFs and exposed through Mendeley’s API, could be used to support answers given to questions. In turn, this portal fits very well with experiments for building a marketplace around data, exploiting the data visualisation, mind maps and data cube management that has also been developed. Finally, in order to address the problem of low usage of
CODE features, the enhanced data extraction tools have been integrated into Mendeley’s main Mendeley Desktop build and API, exposing it to a large community of over 2.5 million users. Task 5.3 began in month 4 and completed at the end of the project.

**Task 5.1 Platform Development and Alpha Deployment**

Task 5.1 was completed at the end of month 14 as planned. The features and services that were demonstrated during the annual review meeting were deployed on the platform and made publicly available. Since the deployment, it became clear that meeting the success factor targets for the end of the Beta deployment would likely mean giving the features more exposure. As a result, in order to reach Mendeley’s vast community, the decision was taken to implement the features in Mendeley Desktop’s main build rather than in a bespoke version of it.

**Task 5.2 Platform Refinement and Beta Deployment**

Task 5.2 began in month 15, as planned. It lasted for six months. As previously covered, the main learnings from the Alpha deployment were that the CODE features were not getting enough usage in their current forms and they were being unnaturally squeezed into Mendeley Desktop. As a result, another approach was taken in which the set of features available through Mendeley Desktop were reduced to make them a better fit, they were released to all users and a question and answering portal was developed that integrated together the CODE features into a single platform.

The enhanced data extraction from PDFs tools reached a reasonable level of quality where Mendeley’s internal Quality Assurance team was happy for them to be integrated and released in a main Mendeley Desktop build. This includes extracting the table of contents, tables and figures from PDFs and presenting them to researchers as both summarisation and navigation supports. The tools were run over all Open Access articles in Mendeley’s catalogue resulting in over one million enrichments. All of these enriched documents can be freely downloaded and viewed using Mendeley Desktop and Mendeley’s API.

By providing this enhanced data through the API, third parties are no longer restricted to using just the metadata of articles in their applications but can also use their tables of contents, tables and figures. Feedback mechanisms were also implemented in Mendeley Desktop that help users to see which articles have enriched data and to report errors in them. These errors can then be used to help focus on the areas of the extraction tools that require improvement. This version of Mendeley Desktop with enhanced data extraction from PDFs was made available to all Mendeley users.

MindMeister’s new experimental features on Semantic Mind Maps emphasize 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. While we started with the intention of creating a corresponding marketplace around semantically enriched mind maps, user surveys and uptake analysis showed that it is far beyond the scope of current business users. Users don’t see the benefits of sharing data in a Semantic Web compliant format. Similarly, re-using Semantic mind maps still lacks corresponding enterprise technology. However, consuming Linked Open Data, namely Freebase, showed an interesting possibility to enhance mind mapping with intelligent services. A service called Wunderkind has been introduced that automatically extends mind maps with semantic web concepts. From a commercial point of view the revenue chain is again a Freemium subscription model. However, the most critical point for a wider uptake lies in the data quality itself. Unfortunately we found that the current data

---

quality and services availability of Linked Open Data lacks easy uptake and that services provided by major companies like Google provide a better quality and coverage than Linked Open Data.

As a joint marketplace, we developed 42-data – “A Data Flea market for Research”\(^{22}\). 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 open research data. The concept is similar to StackExchange and Delicious but with data as the object of interest. 42-data aims to involve people interested in data, but which are not necessarily IT experts. The portal has been realised and made public in an early beta at the end of March 2014\(^{23}\). We engaged in several dissemination activities via social media, press articles, talks on conferences and by running a Linked Open Data challenge. However, at the date of writing, we could not achieve significant uptake. We identified that although people are quite interested in discussing and utilizing open data, hardly anybody engages in the process of data preparation and analysis. While we assumed that this labour intensive task would be covered only by a low number of users, we were surprised by the extremely low uptake. Moreover, expert interviews with people from media and communication showed that the value of publishing data as Linked Open Data was not immediately clear to people. The provided tool chains did not yield immediate pay offs and hence have not been pursued by potentially interested people. The learning curve is too steep. Secondly, although we offered discovery features for Linked Open Data, people barely engaged with it. Usability experiments confirmed that the tools provide good usability. So we infer that utilizing data itself remains challenging and that the whole tool chain of conducting analytics over linked data is not well suited for IT laymen. Thirdly, we discovered a lot of technical and data quality issues that hindered uptake. For example, endpoints delivering data do often have very varying service quality and hence data is not displayed reliably.

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.

**Task 5.3 Success Factor Monitoring and Sprint Planning**

With the close of the CODE project, it is now possible to compare the success factor targets with the actual values achieved. In general, most of the success factors were hit, with the exceptions of Number of Recommendations Accepted and Validated Annotations due to delayed public releases of CODE features and low uptake of the 42-data platform. Some of the highlights on the success factors are as follows:

\(^{22}\) http://42-data.org/home

\(^{23}\) http://42-data.org/home
• Developed scalable Enrichment Service which can be adapted to different workloads (low, medium, high), and at peak times could enrich almost 900,000 PDFs using 512 workers (64 machines each with 8 cores).
• Made available 87 LOD endpoints via the Balloon Service, 46 of which are used in the 42-data Platform and the Query and Visualisation Wizard.
• Generated a large number of Aggregated Data Sets consisting of Data Cubes, enriched PDFs and mind maps as SKOS thesauri.
• Achieved good uptake of CODE features deployed in Mendeley receiving over 500,000 API calls by over 8,000 external users in less than two months after launch.
• Made 120,000 public mind maps available in the Linked Open Data cloud
• Developed a new portal, 42-data, for data centric discussions and socializing around data, which has 41 registered users and 205 created resources.
• Provided access to over 10,000 data cubes in the Linked Open Data Cloud and released 4,000 data cubes from Open Access articles via the 42-data SPARQL Endpoint.
• Delivered three new cutting-edge features to MindMeister end users, resulting in very positive user feedback and increased publicity of the server.

A detailed report on success factors is provided in Deliverable 5.3 [8].

Sprint planning continues as a light-weight model that’s appropriate for all partners. As every partner prefers to use their own system this mostly involves coordination through in person meetings every quarter and regular bi-weekly teleconferencing calls. Results are summarized in the project TRAC and in the issue tracking systems of every partner. We will present the details of sprint planning at the final review meeting.

2.1.6 Work Package 6 – Socio-Economic Mechanisms in Data Marketplaces

WP 6 started in month three and ended in month 24. Within the first year the work in this package focused on researching potential models for establishing a successful marketplace based upon the code services and in the context of the application partners. As outcome of Task 6.1, deliverable 6.1 (Report on data marketplace sustainability factors) has been submitted prior to the Y1 review and outlined our analysis on potential marketplace routes for the consortium partners.

Based on the year-1 review feedback, we have invested additional effort into researching and proposing marketplace concepts for the newly established Q & A portal scenario, where users can post data-centric questions and answer other user’s questions, backed up with linked data references. This has been done (and is still ongoing) as an extension to Task 6.1, which was supposed to be concluded at this stage. In Task 6.2 we have been evaluated sustainability factors for marketplace models based on the revisited concepts of D 6.1 and our evaluation of the three implemented scenarios.

Task 6.1 Analysis of Marketplace Sustainability Factors

Until the Y1 review, much work has gone into the D6.1 deliverable, focusing on the scenarios detailed in the respective report. The newly proposed Q & A portal requires different marketplace and monetization concepts and thus an extension of Task 6.1 into year two. The outcomes have been integrated into the final deliverable D 6.2.

CODE Q&A Portal (42-data): As the new scenario is a question & answer community similar to many existing sites, market research was conducted into leading Q&A portals and two of them (stackexchange.com and quora.com) have been analysed further with regard to their positioning,
market success, business models and trust & reputation models, i.e. how users are incentivised and remunerated for posting answers.

Based on these findings, a marketplace concept for the CODE Q&A Portal (42-data) has been proposed and delivered to the CODE partners in October 2013. The cornerstones of the suggested concept are:

**Positioning:** We propose to position the 42-data as a “Question & Answer Community for Data-Centric Problems”, or a “Stackexchange for Data”

**Target audience:** The main target audience of 42-data are researchers, and only those who work with and routinely use (statistical) data.

**Business model:** As the Q&A market clearly doesn’t support direct monetization models, especially for the planned target audience of researchers, we propose and advertisement- and/or donation-based model, possibly extended by software licensing.

**Trust & Reputation Model:** We suggest a simplified version of the Stackoverflow model where users can post questions and answers as well as up- and downvote other people’s questions and answers. Users earn reputation score for upvoted questions and answers. Various levels of reputation score earn users badges, combined with a trusted user provenance scheme from Mendeley.

While the CODE Q&A Portal forms the integrated marketplace platform of CODE, we still pursue the other two scenarios in integrating CODE services in the partner platforms.

In particular, Mendeley integrates CODE services for crowdsourcing the annotation of semantic research papers and MeisterLabs developed the Semantic Mind Map scenario (i.e. representation and publishing of public maps in MindMeister as RDF)

**Semantic Research Paper Mining:** Due to lack of uptake of the CODE Version of Mendeley Desktop, we plan to integrate CODE Services into the main build of Mendeley Desktop. This requires improving especially the quality of our services and is subject to internal Quality Assurance procedures. Users will be enabled to decompose research papers into smaller semantic units and to link those units to LOD repositories. Besides annotations made available by the Mendeley Desktop users, the consortium plans to release over 3 million automatically annotated open access research papers in a semantic web compliant format (i.e. JSON-LD). The data will be available for other research groups and fed into the C-QAP System.

**Semantic Mind Maps:** Based on the remarks from the review meeting we worked on publishing mind maps in RDF. Semantic mind maps serve two scenarios: first, they can be used for giving a presentation based on semantically annotated research papers and data sets accumulated in the CODE Q&A Portal. Second, they can be used as lightweight thesaurus editor and re-use them in very different scenarios ranging from categorizing documents to inputs into semantic enrichment processes. It is planned to integrate semantic mind maps as expressive means for answers in the Q&A portal as well as in investigating potential marketplace scenarios similar to slideshare.

**Task 6.2 “Evaluation of Marketplace Sustainability Factors”**

Task 6.2 covered the evaluation of marketplace sustainability factors along the three realized scenarios. We extended the task to compare existing ecosystems and their trust mechanism to our scenarios. Therefore, we developed a Data Value Chain and identified important roles and trust factors for data-centric ecosystems.

Based on this analysis, we could identify the following major sustainability factors for ecosystems:
• Successful ecosystems are either socially driven with low value commodities, or data and/or technology driven targeting highly valuable verticals like advertisement or finance.

• Domain focus and targeting high value verticals play a major role for data centric ecosystems.

• In case of non-domain specific data-ecosystems there is also a technological component involved. So generic data alone does not constitute a sufficient commodity.

• Most of the successful ecosystems do not take use of semantic web technologies.

The 3 ecosystems developed in CODE have to handle low value commodities and, based on our analysis, must be socially driven. Hence, revenue possibilities are limited. For the two partner systems, Mendeley’s research paper mining and MeisterLabs semantic mind maps, we identified a subscription-based model, in particular a freemium model, for making CODE services sustainable. The prototypes showed success and in the case of Mendeley high uptake. Both partners came to the decision to maintain CODE services after the project lifetime, which can be seen as major success. Our third ecosystem, 42-data, constitutes a flea market for research data. It can be categorized as socio-data system that tries to establish a community and collaboration around open research data. Due to the low value commodity we focused our sustainability model on donation based micro transaction among participants. While we identified interest in the community for such a portal, we could not generate high uptake until the date of this writing. However, University of Passau will continue to host 42-data and try to focus on research and education support through the platform.

Overall the work package has been successful in developing socio economic mechanisms and realising them in 3 scenarios. Although our new platform did not receive high uptake yet, we could establish CODE services in the existing ecosystems of our SME Partner. They committed themselves to continue the services and their future development.

2.1.7 Work Package 7 – Dissemination & Exploitation

WP7 has been completed as outlined in the Description of Work. Dissemination activities have been ongoing throughout the project, with the project partners participating in over 20 events, submitting the same number of publications and engaging in various marketing and exploitation activities over the course of the projects. The team has been updating the inbound marketing channels set up in Task 7.1 (website, fact sheet, and social media) and set up new web-based interfaces for the three developed prototypes. The main efforts of WP7 in phase 2 have been cantered on marketing these prototypes. Within Task 7.2, an exploitation and sustainability plan has been developed, and the standardization proposal from Task 7.3 has been submitted together with the final deliverables.

Task 7.1 Dissemination Activities

As in phase 1, the CODE team has documented all dissemination activities as well as relevant news on its public website at http://www.code-research.eu, as well posting to the corresponding social media channels on Facebook, Twitter and Google Plus.

In terms of public dissemination activities to scientific communities, the CODE project has been presented at 23 events in total. A list of past and planned events is available in Section 2.3. A full list of presentations - including slide sets for the most part - can be found on the CODE website http://code-research.eu/presentations.
In terms of scientific publication, the consortium got 17 publications accepted, two more publications are currently in the process of submitting, and a further two are planned.

List of accepted publications (available via the Mendeley CODE Publication Group24):


24 http://www.mendeley.com/groups/3344951/code-publications/papers/
As indicated in the 2\textsuperscript{nd} 6-months report, the consortium intensified the dissemination activities with regard to the developed prototypes of Uni Passau and the Know-Center, as well as the new Freemium functionalities of the commercial partners.

Activities to publicise the newly available feature included:

- Blog Posts
- Newsletters
- Social Media Marketing (Facebook, Twitter, Google+, LinkedIn)

Here is a list of marketing publications regarding CODE developments that the project partners did:

- **Mendeley**

- **MeisterLabs**

- **Know-Center**
Task 7.2 Exploitation and Sustainability Planning

This task has been started in M 15 as planned and finished with the delivery of D7.2. In this deliverable 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

University of Passau foresees a two-fold exploitation strategy: First, they aim to publish results as Open Source in order to give others the opportunity to uptake and refine prototypes and to re-use them in other research projects. Second, they plan on a start-up around 42-data and Balloon. The exploitation route for 42-data may also involve other partners, in particular the Know-Center in Graz, where the goal is to conduct joint dissemination to academia and to take up 42-data also in educational processes at Graz University of Technology.

Exploitation plans of the Know-Center primarily focus on the two groups of prototypes we have developed within the project: first, the enrichment tools (Enrichment Service, Annotator Tool), and second, the CODE wizards for easy discovery (Query Wizard, Visualisation Wizard). Targeting academia, CODE tools and services will be used for teaching at the Graz University of Technology. Commercially, the exploitation efforts will be based on the Austrian COMET programme which supports applied research projects to promote the transfer of research results from the academia into the industry. Furthermore, the Extraction service will be exploited within cooperation projects with Mendeley and some of the developed technologies will be integrated in the commercial product lines of the Know-Center. All source code produced during the project is or will be released under an open source licence.

Mendeley exploits the data provided by CODE partners (table of contents, tables, figures and entities extracted from Mendeley's collection of Open Access research articles) 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 makes use of for commercial and non-commercial applications. Mendeley will continue to use the tools that have been developed and prototyped during the CODE projects and exploit them through Mendeley Desktop’s premium plan and the Mendeley API.

MeisterLabs will exploit the CODE technologies implemented by extending its product platform with a number of end-user focused features, thereby increasing the attractiveness of MindMeister in general, and its Premium offering in particular. Especially the Wunderkind feature bears the potential to become a substantial upgrade trigger in the future, provided result quality can be improved. All
developed functions can be run with minimal to no additional overhead so sustainability is a non-issue for MindMeister. In addition to extra user signups and upgrades, we expect additional, “soft” benefits from the participation in the CODE project, such as a perceived high innovation level and research-focus of the company.

**Task 7.3 Standardisation**

This task has been started in M 15 as planned and finished with the submission of D7.3. We started our standardisation activities by organising meetings with people from the standardisation community. However, main standardization activities happened in Phase III after completing the CODE prototypes. Our contribution in this respect is twofold: i) RDF Data Cube Vocabulary & PROV Ontology and ii) the Visual Analytics Vocabulary. The former effort was driven by UNI PASSAU where official implementation reports have been provided within the standardization process of the RDF Data Cube Vocabulary and extensions are proposed. Visual Analytics Vocabulary was developed by the Know-Center to semantically describe visualisation components. Given statistical data sets described with the Data Cube Vocabulary, automatic visualisation of such data sets becomes possible by mapping of the data onto suitable visualisation and visual properties.

With respect to the developed prototypes, the following standardization tasks have been identified, initiated and partly finished:

- Implementation report for the RDF Data Cube Vocabulary
- Extensions to the RDF Data Cube Vocabulary and the PROV Ontology
- Specification of the Data Extraction Microformat
- Specification of the Visual Analytics Vocabulary

### 2.2 Project Management

From the project management point of view, the project is running as expected, no major issues were encountered. There have been no changes in the consortium structure, nor in the management structures and responsibilities (as these were defined at the kick-off meeting). Communication channels and procedures, as well as collaboration infrastructure was set up and has been actively used since the beginning of the project. These communication channels include:

- **TRAC** ([https://nil.dimis.fim.uni-passau.de/trac/](https://nil.dimis.fim.uni-passau.de/trac/)) used as
  - Collaborative project documentation Wiki
  - Issue management system
- **Email lists**
  - general purpose: code@know-center.at
  - WP-lead list: wplcode@know-center.at
- Dropbox file exchange account
- Mendeley paper management system and discussion forums
- MindMeister brainstorming platform

More details can be found in the Deliverable 8.1 “Project Handbook”.

Consortium meetings have been held at regular (approx. tri-weekly) intervals, either as in person meetings where into-depth discussions and planning was made, or as (more frequent) teleconferences to track the progress of the project and coordinate the development and research efforts. Additionally, bilateral meeting and discussions were organised by consortium partners as needed. During the following teleconference, all other partners were informed about the conclusions.
of the bilateral meetings. Protocols of all past meetings and teleconferences are available online at https://nil.dimis.fim.uni-passau.de/trac/wiki/Meetings (with restricted access).

The project Web site (http://code-research.eu/) was set up during the first month of the project and is regularly updated, including a news area, where relevant information can be posted by the partners.

Strategic partnerships have been established with the TPC Consortium and the CLEF PAN Challenge organizers. The TPC Consortium is a group around benchmarking database technology with members from Oracle, Microsoft, Sybase etc. and will provide data on database benchmarking results to be analysed with CODE. The CLEF PAN Challenge is an annually challenge on plagiarism detection which will also provide benchmarking data to be analysed and published with CODE.

2.2.1 Consortium

One partner - Menedley – has been acquired by Elsevier (www.elsevier.com). Possible administrative implications have been discussed with the Project Officer. Mendeley has issued an Information Letter (sent to the Project Officer on 17.07.2013) stating that there will be no transfer of assets, no change in the legal name or capacity of Mendeley Limited, and that the integration of Mendeley into Reed Elsevier group of companies will provide for an excellent continuation of collaboration with the project consortium.

This issue has been resolved with the Project Officer and did not necessitate any changes in the project execution or within the consortium.

2.2.2 Project Plan

Changes to the Project Plan

The project plan experienced two changes: the first small change in late Phase I of the project concerned development activities in WPs 2, 3 and 4 during platform integration; the second change concerned the Milestone MS4 and the completion of Phase III which was caused by the introduction of a new marketplace scenario – the 42-data platform.

Change 1: A small change of the work plan concerned the timing of the work packages 2, 3 and 4 at the beginning of the year two of the project. According to the project Gantt chart (see CODE Description of Work, part B, page 21) no activity should take place in these work packages during months 13 and 14, as the main effort should be focused towards WP5 Platform integration and Deployment. However, in order to address quality and compatibility issues some activity continued in WPs 2, 3 and 4 during months 13 and 14.

Change 2: We have revisited the data marketplace scenarios based on the outcome of the review meeting and adapted the integration of the services accordingly. Development of a new, third data marketplace scenario, the 42-data question and answer platform, was started with the beginning of Phase II in month 14. Milestone MS 4 “Beta Deployment” was reached in month 21, one month later than planned, due to the need to stabilise the 42-data platform before releasing it to the public. Also, since the scenarios changed, the engagement of a wider public was started after reaching MS 4. As a consequence of the MS4 delay, the Phase III of the project also started in month 21, one month later than planned.
Sprint Planning

In order to focus our development efforts towards the three marketplace scenarios, our sprints, which were planned separately for each work package during Phase I of the project, were planned for each of the scenarios during the Phases II and III. The sprints were executed in a bi-weekly cadence including joint Hangouts to discuss the completed sprints and adjust the plans for the next ones. Additional scenario-specific Hangouts were scheduled on-demand for detailed scenario discussions.

2.3 Main Events

The following main events can be reported, where CODE has been presented:

1. The European Data Forum, June’12, Copenhagen, Denmark
2. i-Know’12, September, Graz, Austria
3. 3rd Annual Vivo Conference, September 2012, Florida, USA
4. 1st Workshop on the Big Data Benchmarking Community, December 2012, Pune, India
5. ZPID Colloquium, February 2013, Trier, Germany
6. Cloudscape Workshop, February 2013, Brussels, Belgium
7. The European Data Forum, April 2013, Dublin, Ireland
8. 2nd International B2B Software Days, April 2013, Vienna, Austria
9. 2. Wissensmanagement-Tag, April 2013, Krems, Austria
10. ESWC 2013, May 2013 in Montpellier, France
11. SouthCHI’13, July 2013 in Maribor, Slovenia
12. I-Know’13, September 2013, Graz, Austria
13. I-Semantics’13, September 2013, Graz, Austria (co-located with the I-Know’13 conference)
14. 3rd ESWC Semantic Web Summer School
15. Science 2.0 Policy Workshop organised by DG Research & Innovation C2 and ERIAB, 4th November 2013, Brussels
16. ASIS&T 2013 Annual Meeting, 5th November 2013 in Montreal, Canada
17. First Standardization and collaboration meeting with Erik Manners from University Gent, December 17th 2013, University of Passau, Germany
19. European Data Forum (EDF 2014), March 19-20, Athens, Greece
22. 5th International Workshop on Data Engineering Meets the Semantic Web, co-located with ICDE 2014, April 2014, Chicago, USA

Next planned events include:

1. 11th ESWC 2014 conference, 25th to 29th May 2014, Anissaras, Crete, Greece
2. 1st i-KNOW ALOA Challenge on Getting Answers through Linked Open Data, 14th International Conference on Knowledge Technologies and Data-driven Business, 16-19 September 2014, Graz, Austria
2.3.1 Meetings

The following Meetings and Workshops have taken place:

<table>
<thead>
<tr>
<th>Date</th>
<th>Meeting</th>
<th>Location</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-05-15</td>
<td>Kick-off Meeting (2 days)</td>
<td>Graz, Austria</td>
<td>all</td>
</tr>
<tr>
<td>2012-07-12</td>
<td>Use Case and Architecture Meeting (2 days)</td>
<td>London, UK</td>
<td>KC, Uni-Passau, Mendeley</td>
</tr>
<tr>
<td>2012-08-23</td>
<td>WP 6 Meeting (2 days)</td>
<td>London, UK</td>
<td>MeisterLabs, Mendeley</td>
</tr>
<tr>
<td>2012-09-06</td>
<td>WP2/WP 3/WP4 Meeting (2 days)</td>
<td>Graz, Austria</td>
<td>KC, Uni-Passau, MeisterLabs</td>
</tr>
<tr>
<td>2012-10-01</td>
<td>Sprint Planning Meeting (2 days)</td>
<td>Passau, Germany</td>
<td>all</td>
</tr>
<tr>
<td>2012-12-14</td>
<td>Mind map integration and marketplaces (1 day)</td>
<td>Graz, Austria</td>
<td>KC, MeisterLabs</td>
</tr>
<tr>
<td>2013-01-28</td>
<td>Integration Meeting</td>
<td>Vienna, Austria</td>
<td>All</td>
</tr>
<tr>
<td>2013-03-06</td>
<td>Data Marketplace Meeting</td>
<td>Passau, Germany</td>
<td>MeisterLabs, KC, Uni-Passau</td>
</tr>
<tr>
<td>2013-05-02</td>
<td>Review Preparation Meeting (2 days)</td>
<td>Graz, Austria</td>
<td>all</td>
</tr>
<tr>
<td>2013-06-14</td>
<td>Review Meeting (2 days)</td>
<td>Luxembourg</td>
<td>all</td>
</tr>
<tr>
<td>2013-07-08</td>
<td>Post-Review Meeting (2 days)</td>
<td>London, UK</td>
<td>all</td>
</tr>
<tr>
<td>2013-08-27</td>
<td>Data publishing &amp; semantic mindmaps meeting (2 days)</td>
<td>Passau, Germany</td>
<td>Uni-Passau, MindMeister</td>
</tr>
<tr>
<td>2013-09-05</td>
<td>WPs 2-4 planning meeting @ I-Know 2013</td>
<td>Graz, Austria</td>
<td>Uni-Passau, Know-Center</td>
</tr>
<tr>
<td>2013-10-10</td>
<td>Reconciliation and sprint planning meeting (2 days)</td>
<td>Passau, Germany</td>
<td>all</td>
</tr>
<tr>
<td>2014-01-30</td>
<td>Phase III Kick-off meeting (2 days)</td>
<td>Vienna, Austria</td>
<td>all</td>
</tr>
<tr>
<td>2014-04-29</td>
<td>Final project meeting (2 days)</td>
<td>Graz, Austria</td>
<td>all</td>
</tr>
</tbody>
</table>

*Table 1: Meetings and Workshops.*
The following virtual meetings (via Skype or Google Hangout) have taken place:

<table>
<thead>
<tr>
<th>Date</th>
<th>Meeting</th>
<th>Location</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-05-25</td>
<td>Work package Leader Skype Meeting</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2012-06-21</td>
<td>Work package Leader Skype Meeting</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2012-07-06</td>
<td>Work package Leader Skype Meeting</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2012-08-17</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2012-09-19</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2012-11-07</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2012-12-18</td>
<td>Integrated User Interface Discussion</td>
<td>Virtual</td>
<td>Know-Center, Uni Passau, Mendeley</td>
</tr>
<tr>
<td>2012-12-19</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2013-01-15</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2013-01-18</td>
<td>Visual Analytics Discussion</td>
<td>Virtual</td>
<td>Know-Center, Uni Passau</td>
</tr>
<tr>
<td>2013-02-19</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2013-03-13</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2013-04-12</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2013-06-03</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2013-10-03</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2013-10-23</td>
<td>Scenario 3 Q&amp;A portal discussion</td>
<td>Virtual</td>
<td>Know-Centar, Uni Passau</td>
</tr>
<tr>
<td>2013-10-24</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2013-10-31</td>
<td>Marketplace discussion</td>
<td>Virtual</td>
<td>Uni-Passau, MindMeister, Know-Center</td>
</tr>
<tr>
<td>2013-11-07</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2013-11-28</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2013-12-12</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2014-01-14</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2014-02-19</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2014-03-12</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
</tbody>
</table>
The following Meetings (including virtual Meetings) and Workshops are planned:

<table>
<thead>
<tr>
<th>Date</th>
<th>Meeting</th>
<th>Location</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-03-31</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
<tr>
<td>2014-04-08</td>
<td>D6.2 status and TODOs</td>
<td>Virtual</td>
<td>Uni-Passau, MindMeister, Know-Center</td>
</tr>
<tr>
<td>2014-04-15</td>
<td>D6.2 status and finalization</td>
<td>Virtual</td>
<td>Uni-Passau, MindMeister, Know-Center</td>
</tr>
<tr>
<td>2014-04-23</td>
<td>Work package Leader Hangout</td>
<td>Virtual</td>
<td>all</td>
</tr>
</tbody>
</table>

*Table 2: Virtual meetings (teleconferences).*

<table>
<thead>
<tr>
<th>Date</th>
<th>Meeting</th>
<th>Location</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-05-26 &amp; 27</td>
<td>Final Project Review Meeting (2 days)</td>
<td>Luxembourg</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>• Day 1: dry run</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Day 2: review meeting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 3: Planned meetings.*

## 2.4 Deliverables and Milestones

### 2.4.1 Deliverables

As planned, the following deliverables have been created and submitted:

<table>
<thead>
<tr>
<th>Number</th>
<th>Title</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1.1</td>
<td>Joint Collaboration Infrastructure for research and development</td>
<td>This deliverable presents the collaboration infrastructure that has been set up for the project, such as a Wiki, an Issue Tracker, and collaborative paper management and mind-mapping tools. The deliverable also shortly describes communication procedures and tools used in the project.</td>
</tr>
<tr>
<td>D1.2</td>
<td>Requirements Definition and Conceptual Architecture Document</td>
<td>This deliverable outlines user stories, the derived conceptual architecture and resulting requirements/API definitions. Potential marketplace scenarios are outlined based on the derived conceptual architecture. Due to the short two-year time-frame of the project, all decisions taken, aim to allow parallel realization</td>
</tr>
<tr>
<td>Deliverable</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td><strong>D2.1.1</strong></td>
<td>Semantic Enrichment Workflow Service</td>
<td>This deliverable describes the Semantic Enrichment Workflows in order to crowd-source semantic annotation and fact extraction for research papers. For the first prototype two workflows were realised and integrated into a special version of Mendeley Desktop.</td>
</tr>
<tr>
<td><strong>D2.1.2</strong></td>
<td>Semantic Enrichment Algorithms</td>
<td>This deliverable describes a Web service, named the Enrichment Service, and the associated algorithms developed to allow semantic enrichment of scientific articles (typically PDFs) as well as mind maps. The service relies on (domain-specific) pre-computed models to extract the document structures, a set of metadata (e.g. author names) and tables. The deliverables also reports on conducted algorithm evaluation.</td>
</tr>
<tr>
<td><strong>D2.1.3</strong></td>
<td>Semantic Integration Algorithms</td>
<td>This deliverable describes a Web service developed to allow disambiguation of entities in a research article and in extracted tables. The service links the entities in an article to semantic Web URIs. The deliverables also reports on conducted experiments.</td>
</tr>
<tr>
<td><strong>D2.2</strong></td>
<td>Refined Semantic Enrichment &amp; Integration Workflow available as Component/Service</td>
<td>This deliverable describes refinements on two web services: the Enrichment Service and the Disambiguation Service. For scientific articles (PDF documents) the structure (i.e. table of content) and metadata (such as author names), tables and images are extracted. Disambiguation Service disambiguates extracted entities and table columns. Deliverable also reports on conducted evaluations and the resulting publications.</td>
</tr>
<tr>
<td><strong>D2.3</strong></td>
<td>Evaluation and Performance Report</td>
<td>This deliverable describes the evaluation results of two web services: the Enrichment Service and the Disambiguation Service. The algorithms, which have been developed for the services have been evaluated against existing or newly created data-set and. The results of this work have been published.</td>
</tr>
<tr>
<td><strong>D3.1</strong></td>
<td>Federated Querying and Aggregation Service</td>
<td>This deliverable describes prototypes for Federated Querying and Aggregation of Linked Data. For Federated Querying two prototypes are introduced enabling automatic and intelligent query federation. Also described are mechanisms developed to integrate RESTful APIs into the Linked Open Data cloud. For Linked Data Aggregation a refined version of the Data Extractor prototype is described along with first results addressing data cube merging.</td>
</tr>
<tr>
<td>D3.2</td>
<td>Provenance Schema</td>
<td>This deliverable describes our approach to model provenance information, which is present in all workflows involving available project prototypes. The objective is establishing a full revision chain of data, involving statements on the complete history of ownership. This information will provide the foundation for other processes, such as marketplace concepts or query optimization.</td>
</tr>
<tr>
<td>------</td>
<td>------------------</td>
<td>---</td>
</tr>
<tr>
<td>D3.3</td>
<td>Refined Federated Querying and Aggregation</td>
<td>This deliverable introduces the results of “Federated Querying” and “Aggregation of Linked Data” tasks. In For the former two prototypes are introduced supporting intelligent query federation. For the latter a major refined version of the Data Extractor prototype is presented along with first results for the central data cube merging issue.</td>
</tr>
<tr>
<td>D3.4</td>
<td>Evaluation Results</td>
<td>The results and statistics of conducted evaluation are available for the Balloon (indexing and discovery of Linked Open Data) and Bacon (management and aggregation of data cubes) prototypes. Also provided are statistics for the MindMeister endpoint, covering semantically enriched mind maps.</td>
</tr>
<tr>
<td>D4.1</td>
<td>Semantic Descriptions for Visual Analytics Components</td>
<td>This deliverable describes semantic descriptions of user interface components for visualisation of Linked Data. These descriptions serve as mapping between the data and the visual components in order to support and simplify creation of user interfaces for visual analysis of Linked Data.</td>
</tr>
<tr>
<td>D4.2</td>
<td>Semantic Enrichment, Semantic Integration and Visual Analytics Interfaces</td>
<td>This deliverable deals with the development of Web-based analysis components to visualize aggregated Linked Data and semantic enrichment results. It describes a Query Wizard which provides an easy-to-use interface for accessing Linked Data, and a Visualisation Wizard which automatically suggests suitable visualizations for provided data and visualises this data using a variety of visualisation components.</td>
</tr>
<tr>
<td>D4.3</td>
<td>Refined Semantic Enrichment, Semantic Integration and Visual Analytics Interfaces</td>
<td>This deliverable describes results of tasks concerned with ensuring the usability of the developed visual analytics interfaces. Results include enhancement of interfaces and interaction paradigms for analysing, navigating, browsing and visualizing Linked Data. Formative evaluations of the components and user interfaces have been conducted through continuous usability testing with friendly users.</td>
</tr>
<tr>
<td>D4.4</td>
<td>Usability Evaluation Report</td>
<td>This deliverable covers tasks dealing with the</td>
</tr>
<tr>
<td>Deliverable</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>D5.1</td>
<td>CODE Platform Alpha Version</td>
<td></td>
</tr>
<tr>
<td>D5.2</td>
<td>CODE Platform Beta Version</td>
<td></td>
</tr>
<tr>
<td>D5.3</td>
<td>Monitored Success Factors</td>
<td></td>
</tr>
<tr>
<td>D6.1</td>
<td>Report on data marketplace sustainability factors</td>
<td></td>
</tr>
<tr>
<td>D6.2</td>
<td>Report on Trust and Reputation Models for Data Marketplaces</td>
<td></td>
</tr>
</tbody>
</table>

usability testing of the developed visual analytics interfaces. Evaluation of the components and user interfaces has been conducted by performing formative web-based usability testing with friendly users as well as summative in-depth evaluation of the user experience of the CODE Query Wizard and the CODE Visualization Wizard.

This deliverable is a practical getting started guide to the CODE build of Mendeley Desktop and associated services. It exposes the functionalities of PDF data extraction, data visualisation, table extraction, data cube generation and entity meaning lookups.

This deliverables summarizes the CODE Platform that has been build. The platform consists of three prototypes to analyse different value chains for (Linked) Open Data: i) 42-data, a data-centric question and answer portal, ii) the Mendeley Desktop/Client and Server API for semantically enriching research publications, and iii) MindMeister.com web platform for generating semantic web enabled mind-maps and mind-map-based presentations. All three platforms are characterized by their production and consumption of Linked Open Data.

This document summarises the success factors monitored by the three prototypes of the CODE platform. CODE services and tools have been successfully deployed in the three platforms, and users are engaging with them. CODE had set very ambitious success factors where some of them could be met, while others have not been achieved yet. The two SME partners, Mendeley and MindMeister, are committed to continuing development of CODE features as users have found them valuable.

This report analyses roles, revenue models and value chains in data marketplaces as well as corresponding trust and reputation models. Based on secondary research 25 data marketplaces are analysed to derive potential marketplace roles, revenue models, trust measures and value chains. In two large-scale surveys and four expert interviews 3 potential marketplace scenarios for CODE are identified.

For this deliverable we have worked out a generic, role-driven Data Value Chain to analyse existing commercially oriented data ecosystems. Success factors were identified and three trust relationships
are outlined to be established by data-centric ecosystems, namely: the social trust between participants, trust in data quality and trust in actions transforming/providing data. Based on these theoretical foundations we provide insights in the ecosystems developed in the CODE project and insights found by analysing usage and uptake of the ecosystems.

<p>| D7.1.1 | Homepage, Social Media Channels, Fact Sheet | This deliverable presents the project website, gives an overview of the used social media channels – Twitter, Facebook and Google+ – and introduces the Project Fact Sheet. |
| D7.1.2 | Community Engagement Plan | This deliverable describes the plan to engage the MindMeister and Mendeley communities in order to get users involved in CODE as well as on how to reach beyond both communities. We target social media channels, strategic partnerships as well as the organisation of challenges at highly renowned conferences. |
| D7.2 | Exploitation and Sustainability Plan | This deliverable reports on plans for exploiting CODE project results and guaranteeing the sustainability of the developed prototypes. Each partner describes the planned exploitation activities, addressed target groups and the expected relevance for their organisation. We consider exploiting developed prototypes, newly employed technologies, data contributing to the data economy, and contributing source code to the open source community. |
| D7.3 | Standardization Proposal | This deliverable describes standardization activities carried out within the project: i) RDF Data Cube Vocabulary &amp; PROV Ontology, ii) the Visual Analytics Vocabulary. For the former implementation reports were provided within the standardization process of the RDF Data Cube Vocabulary and extensions are proposed. Visual Analytics Vocabulary was developed to semantically describe visualisation components and to enable automatic visualisation of data sets described with the Data Cube Vocabulary. |
| D7.4 | Research Publications | This deliverable gives an overview of scientific publications produced during the CODE project, including Journal papers, publications at international conferences and workshops, and articles in scientific news media. For each paper the title, abstract and the full reference is given. Also included are future publication plans. |</p>
<table>
<thead>
<tr>
<th>Deliverable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>D7.5</td>
<td>Web-enabled Public Showcase</td>
</tr>
<tr>
<td></td>
<td>This deliverable describes Web-enabled public showcases for public dissemination and presentation of the project. The showcase is built around publicly available, Web-based CODE prototypes. The purpose of each prototype is briefly described followed by typical use case descriptions demonstrated along a typical use case.</td>
</tr>
<tr>
<td>D8.1</td>
<td>Project Handbook</td>
</tr>
<tr>
<td></td>
<td>This deliverable presents the Project Handbook, gives an overview of the project structure and provides a description of project processes and resources.</td>
</tr>
<tr>
<td>D8.2.1</td>
<td>6-monthly-Report (M1-M6)</td>
</tr>
<tr>
<td></td>
<td>This deliverable provides an overview of the work completed resp. launched in the first 6 months (1st May to 31st October 2012) of the project. It also provides an overview of the project from the management point of view, including a report on meetings, deliverables, and the overall resource consumption.</td>
</tr>
<tr>
<td>D8.2.2</td>
<td>Progress Report Year 1</td>
</tr>
<tr>
<td></td>
<td>This deliverable provides an overview of the work conducted in the first year. Overall the project is on track by achieving MS 2 “Components Ready”. A set of services usable in three market place scenarios were developed. Phase II will be mostly driven by refinement of the developed services and the realisation of the three scenarios in order to conduct real-user evaluation by M 20.</td>
</tr>
<tr>
<td>D8.2.3</td>
<td>2nd 6-monthly Report (M13-M18)</td>
</tr>
<tr>
<td></td>
<td>This deliverable provides an overview of the work conducted from month 13 to month 18 of the project. Overall the project is on track: milestone MS3 “Alpha Delivery” was achieved in month 14, while milestone MS4 “Beta Delivery” shall be achieved for all services in month 20 as planned. Due to the revised data marketplace scenarios (based on the outcome of the review meeting) and the adapted the integration of the service, engagement of a wider audience is planned after reaching MS 4. Phase III, starting in month 20, will focus on evaluation and refinement of the developed services.</td>
</tr>
<tr>
<td>D8.2.4</td>
<td>Final Report</td>
</tr>
<tr>
<td></td>
<td>This document</td>
</tr>
</tbody>
</table>

Table 4: Descriptions (and major outcomes) of the deliverables submitted during the two years of the project.
<table>
<thead>
<tr>
<th>Del. no.</th>
<th>Deliverable name</th>
<th>Version</th>
<th>WP no.</th>
<th>Lead beneficiary</th>
<th>Nature</th>
<th>Dissemination level</th>
<th>Delivery date from Annex I (project month)</th>
<th>Actual delivery date dd/mm/yyyy</th>
<th>Status</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8.1</td>
<td>Project Handbook</td>
<td>1.0</td>
<td>WP 8</td>
<td>1</td>
<td>R</td>
<td>CO</td>
<td>1</td>
<td>27/07/2012</td>
<td>Submitted</td>
<td>D7.1 has been finished and made available online in Month 1, but submitted later (27/07/2012)</td>
</tr>
<tr>
<td>D7.1.1</td>
<td>Homepage &amp; Social Media Channels, Fact Sheet</td>
<td>1.0</td>
<td>WP 7</td>
<td>1</td>
<td>O</td>
<td>PU</td>
<td>1</td>
<td>27/07/2012</td>
<td>Submitted</td>
<td>D7.1 has been finished in Month 1, but submitted later (27/07/2012)</td>
</tr>
<tr>
<td>D1.1</td>
<td>Collaboration Infrastructure</td>
<td>1.0</td>
<td>WP 1</td>
<td>1</td>
<td>O</td>
<td>CO</td>
<td>1</td>
<td>27/07/2012</td>
<td>Submitted</td>
<td>D1.1. has been finished in Month 1, but submitted later (27/07/2012)</td>
</tr>
<tr>
<td>D1.2</td>
<td>Requirements Definition &amp; Conceptual Architecture</td>
<td>1.0</td>
<td>WP 1</td>
<td>2</td>
<td>R</td>
<td>CO</td>
<td>3</td>
<td>27/07/2012</td>
<td>Submitted</td>
<td></td>
</tr>
<tr>
<td>D4.1</td>
<td>Semantic Descriptions for Visual Analytics interfaces</td>
<td>1.0</td>
<td>WP 4</td>
<td>1</td>
<td>O</td>
<td>PU</td>
<td>6</td>
<td>07/11/2012</td>
<td>Submitted</td>
<td></td>
</tr>
<tr>
<td>D7.1.2</td>
<td>Community Engagement Plan</td>
<td>1.0</td>
<td>WP 7</td>
<td>4</td>
<td>R</td>
<td>RE</td>
<td>6</td>
<td>08/11/2012</td>
<td>Submitted</td>
<td></td>
</tr>
<tr>
<td>D8.2.1</td>
<td>6-Monthly Report</td>
<td>1.0</td>
<td>WP8</td>
<td>1</td>
<td>R</td>
<td>RE</td>
<td>6</td>
<td>15/11/2012</td>
<td>Submitted</td>
<td>Delayed two weeks to allow for collecting resource consumption data from partners. In agreement with the PO</td>
</tr>
<tr>
<td>D 2.1.1</td>
<td>Semantic Enrichment Workflow Service</td>
<td>1.0</td>
<td>WP2</td>
<td>3</td>
<td>P</td>
<td>RE</td>
<td>12</td>
<td>06/05/2013</td>
<td>Submitted</td>
<td></td>
</tr>
<tr>
<td>D 2.1.2</td>
<td>Semantic Enrichment Algorithms</td>
<td>1.0</td>
<td>WP2</td>
<td>1</td>
<td>P</td>
<td>RE</td>
<td>12</td>
<td>06/05/2013</td>
<td>Submitted</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------</td>
<td>-----</td>
<td>-----</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>-----</td>
<td>------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>D 2.1.3</td>
<td>Semantic Integration Algorithms</td>
<td>1.0</td>
<td>WP2</td>
<td>2</td>
<td>P</td>
<td>RE</td>
<td>12</td>
<td>06/05/2013</td>
<td>Submitted</td>
<td></td>
</tr>
<tr>
<td>D 3.1</td>
<td>Federated Querying and Aggregation Service</td>
<td>1.0</td>
<td>WP3</td>
<td>2</td>
<td>P</td>
<td>RE</td>
<td>12</td>
<td>06/05/2013</td>
<td>Submitted</td>
<td></td>
</tr>
<tr>
<td>D 3.2</td>
<td>Provenance Schema</td>
<td>1.0</td>
<td>WP3</td>
<td>2</td>
<td>O</td>
<td>PU</td>
<td>12</td>
<td>06/05/2013</td>
<td>Submitted</td>
<td></td>
</tr>
<tr>
<td>D 4.2</td>
<td>Semantic Enrichment, Semantic Integration and Visual Analytics Interfaces</td>
<td>1.0</td>
<td>WP4</td>
<td>1</td>
<td>P</td>
<td>RE</td>
<td>12</td>
<td>06/05/2013</td>
<td>Submitted</td>
<td></td>
</tr>
<tr>
<td>D 8.2.2</td>
<td>Progress Report Year 1</td>
<td>1.0</td>
<td>WP8</td>
<td>1</td>
<td>R</td>
<td>CO</td>
<td>12</td>
<td>12/05/2013</td>
<td>Submitted</td>
<td>Delayed slightly to allow for collecting resource consumption data from partners. In agreement with the PO</td>
</tr>
<tr>
<td>D 6.1</td>
<td>Report on Data Marketplace Sustainability Factors</td>
<td>1.0</td>
<td>WP6</td>
<td>4</td>
<td>R</td>
<td>RE</td>
<td>12</td>
<td>20/05/2013</td>
<td>Submitted</td>
<td>Delayed 3 weeks due to collecting additional user feedback on data marketplaces and due to illness of the responsible employee.</td>
</tr>
<tr>
<td>D 5.1</td>
<td>CODE Platform Alpha Version</td>
<td>1.0</td>
<td>WP5</td>
<td>3</td>
<td>P</td>
<td>RE</td>
<td>14</td>
<td>11/07/2013</td>
<td>Submitted</td>
<td></td>
</tr>
<tr>
<td>D 8.2.3</td>
<td>2nd 6-monthly Report</td>
<td>1.0</td>
<td>WP8</td>
<td>1</td>
<td>R</td>
<td>CO</td>
<td>18</td>
<td>12/11/2013</td>
<td>Submitted</td>
<td>Delayed slightly to allow for collecting resource consumption data from partners. In agreement with the PO</td>
</tr>
<tr>
<td>D 2.2</td>
<td>Refined Semantic</td>
<td>1.0</td>
<td>WP2</td>
<td>1</td>
<td>P</td>
<td>PU</td>
<td>20</td>
<td>08/01/2014</td>
<td>Submitted</td>
<td></td>
</tr>
<tr>
<td>Deliverable</td>
<td>Description</td>
<td>WP</td>
<td>P</td>
<td>PU</td>
<td>Date</td>
<td>Status</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----</td>
<td>---</td>
<td>----</td>
<td>------------</td>
<td>-------------</td>
<td>-----------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 3.3</td>
<td>Enrichment &amp; Integration Workflow available as Component/Service</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Delayed by 3 weeks due to requirements for the 42-data platform (see deliverable D 5.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 4.3</td>
<td>Refined Federated Querying and Aggregation</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 5.2</td>
<td>CODE Platform Beta Version</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Delayed by 5 weeks due to revised data marketplace plans and implementation of a new platform – 42-data – which went online end of January.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 2.3</td>
<td>Evaluation and Performance Report</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Submitted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 3.4</td>
<td>Evaluation Results</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Submitted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 4.4</td>
<td>Usability Evaluation Report</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Submitted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 5.3</td>
<td>Monitored Success Factors</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Submitted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 6.2</td>
<td>Report on Trust and Reputation Models for Data Marketplaces</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Submitted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 7.2</td>
<td>Exploitation and Sustainability Plan</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Submitted</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 5: Status of the deliverables submitted during the two years of the project.

<table>
<thead>
<tr>
<th>Deliverable</th>
<th>Version</th>
<th>WP</th>
<th>Type</th>
<th>Status</th>
<th>Date Submitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>D 7.3</td>
<td>1.0</td>
<td>WP7</td>
<td>R</td>
<td>PU</td>
<td>24</td>
</tr>
<tr>
<td>D 7.4</td>
<td>1.0</td>
<td>WP7</td>
<td>1</td>
<td>R</td>
<td>24</td>
</tr>
<tr>
<td>D 7.5</td>
<td>1.0</td>
<td>WP7</td>
<td>1</td>
<td>R</td>
<td>24</td>
</tr>
<tr>
<td>D 8.2.4</td>
<td>1.0</td>
<td>WP8</td>
<td>1</td>
<td>R</td>
<td>24</td>
</tr>
</tbody>
</table>

- Delayed to allow for collecting resource consumption data from partners. In agreement with the PO.
Dissemination level:

**PU** = Public

**PP** = Restricted to other programme participants (including the Commission Services).

**RE** = Restricted to a group specified by the consortium (including the Commission Services).

**CO** = Confidential, only for members of the consortium (including the Commission Services).

*Make sure that you are using the correct following label when your project has classified deliverables.*

**EU restricted** = Classified with the mention of the classification level restricted “EU Restricted”

**EU confidential** = Classified with the mention of the classification level confidential "EU Confidential"

**EU secret** = Classified with the mention of the classification level secret "EU Secret"

Nature:

**R** = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other

### 2.4.2 Milestones

All milestones of the project have been achieved. “MS 1 – Project Setup”, “MS2 Components Ready” and “MS3 – Alpha Deployment” have been achieved as planned. “MS4 – Beta Deployment” was achieved one month later than planned due to delayed completion of the 42-data platform (see Section 2.2.2 for details). “MS5 – Evaluation Conducted was achieved as planned for all services, tools and scenarios except for the 42-data platform, where the 4 months planned for community building, dissemination and evaluation have been reduced to 1 month (due to critical bugs and necessity to include additional functionality).

<table>
<thead>
<tr>
<th>Milestone no.</th>
<th>Milestone name</th>
<th>Work package no</th>
<th>Lead beneficiary</th>
<th>Delivery date from Annex I dd/mm/yyyy</th>
<th>Achieved Yes/No</th>
<th>Actual / Forecast achievement date dd/mm/yyyy</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS1</td>
<td>Project Setup</td>
<td>WP1</td>
<td>2</td>
<td>31/07/2012</td>
<td>Yes</td>
<td>27/07/2012</td>
<td>Project infrastructure setup and requirements definition</td>
</tr>
<tr>
<td>MS2</td>
<td>Components Ready</td>
<td>WP 2-4</td>
<td>1</td>
<td>31/04/2013</td>
<td>Yes</td>
<td>31/04/2013</td>
<td>Fist versions of all single components are ready</td>
</tr>
<tr>
<td>MS3</td>
<td>Alpha Deployment</td>
<td>WP5</td>
<td>3</td>
<td>31/06/2013</td>
<td>Yes</td>
<td>31/06/2013</td>
<td>First version of the platform</td>
</tr>
<tr>
<td>MS4</td>
<td>Beta Deployment</td>
<td>WP 2-5</td>
<td>3</td>
<td>31/12/2013</td>
<td>Yes</td>
<td>31/01/2014</td>
<td>Beta deployment of the platform. Results will be Slipstreamed</td>
</tr>
</tbody>
</table>
### Table 6: Project milestones for the full duration of the project.

<table>
<thead>
<tr>
<th>MS5</th>
<th>Evaluation Conducted</th>
<th>WP 2-6</th>
<th>30/04/2014</th>
<th>Yes</th>
<th>30/04/2014</th>
<th>Evaluation of components, platform and marketplace mechanisms has been conducted</th>
</tr>
</thead>
</table>

#### 2.5 Next Steps

The next step in the project is the Final Review Meeting to be held on 27th May 2014 in Luxembourg. Meeting agenda has been approved by the Project Offices and sent to the reviewers (Karima Rafes, Irene Celino). Project Consortium will be represented by:

- Know-Center: Roman Kern, Patrick Höfler, Vedran Sabol
- University of Passau: Michael Granitzer, Florian Stegmaier
- Mendeley: Maya Hristakeva
- MeisterLabs: Michael Hollauf

From the administrative point of view the concluding step of the project will be the completion of the Final Report including the financial reports (Form C). This step will be completed and the forms entered into NEF system within 60 days from the end of the second reporting period (which ends on 30th April 2014).

A major goal to focus on after the completion of the project will be the exploitation of project results. Our exploitation plans 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.

In particular, we identified the following main routes of exploitation after the end of the project:

- 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.

Exploitation of code results also includes publishing of the research results in international Journals and conferences which has already resulted in 17 published papers. As of 30th April 2014, two more papers have been submitted and are awaiting notification of acceptance. Additional 4 publications are planned and shall be submitted by the end of May 2014.

For more details on CODE exploitation plans please consult [5]. More information on future CODE publication plans can be found [6].
3 References

[1] CODE Description of Work, Version Date 20012-01-23
[2] D8.2.1 first 6-monthly Report
[3] D8.2.2 Progress Report Year 1
[4] D8.2.3 second 6-monthly Report
[5] D7.2 Exploitation and Sustainability Plan
[7] D7.5 Web-enabled Public Showcase
[8] D5.3 Monitored Success Factors
4 Appendix - Showcase Materials

Within the project a number of publicly available Web-enabled prototypes were developed. For the purpose of public dissemination and presentation of the project, this deliverable describes these prototypes from the user's point of view. Simple use cases are given in the form of a sequence of steps accompanied by screenshots, illustrating how each prototype is used and applied. Descriptions of following prototypes are included:

- 42 data platform: a data-centric question and answering portal that builds around main CODE services http://www.42-data.org/
- CODE Annotator Tool: a tool for intuitively annotating documents, creating models by applying machine learning techniques, and sharing the created models with others. http://code-annotator.know-center.tugraz.at/
- CODE Query Wizard: an easy to use user interface for searching, refining and transforming Linked Data http://code.know-center.tugraz.at/search
- CODE Visualization Wizard: a tool for visual analysis of Linked Data which supports the user by automating the visualization process http://code.know-center.tugraz.at/vis
- Mendeley Research Paper Mining: 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 http://www.mendeley.com/download-mendeley-desktop/
- MindMeister Semantic Mind Maps: prototype emphasizes the consumption and provision of Linked Open Data through mind maps on the MindMeister platform

For the purpose of public dissemination and presentation of the project, the publicly available Deliverable 7.5 Web-enabled Public Showcase [7] describes the above listed prototypes from the user’s point of view. Simple use cases are given in the form of a sequence of steps accompanied by screenshots, illustrating how each prototype is used and applied.


Following showcase materials are available online:

- CODE Annotator Tool handout:

- CODE Fact Sheet:
- CODE screencasts (YouTube Channel):
  https://www.youtube.com/user/CODEresearchEU