CODE

Commercially Empowered
Linked Open Data Ecosystems in Research

^

D 7.4 - Research Publications

Summary: This deliverable gives an overview of scientific publications produced during the CODE project. Included is a list of 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. An overview of addressed conferences and publication media is also provided. Finally, future publication plans are outlined with a list of papers which are either planned or already submitted (but not yet accepted at the time of writing of this document).

Project Acronym	CODE
Grant Agreement number	296150
Project Title	Commercially Empowered Linked Open Data Ecosystems in Research
Date	2014-04-30
Nature	R (Report)
Dissemination level	PU (Public)
WP Lead Partner	MindMeister
Revision	Final revision
Authors	Vedran Sabol, Lisa Maurer

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Statement of originality: This document contains original unpublished work except where clearly indicated otherwise.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 & Project Coordinators

Project Officer	Stefano Bertolo	
Project Coordinator	Stefanie Lindstaedt	Inffeldgasse 21a, 8010 Graz, Austria +43 (316) 873-9250 (phone) +43 (316) 873- 9254 (fax) slind@know-center.at
Scientific Coordinator	Michael Granitzer	Innstrasse 33, D-94032 Passau +49(0)851-509-3305 michael.granitzer@uni-passau.de

Document Revision History

Revision	Date	Author	Organization	Description
1st draft	2014-04-07	V. Sabol	Know-Center	Document structure, summary, introduction, conclusion
2nd draft	2014-04-17	L. Maurer	Know-Center	Initial list with abstracts
3rd draft	2014-04-24	V. Sabol	Know-Center	Tables, references, submitted and planed papers
Final Version	2014-04-30			

Table of Contents

1 Introduction

A central goal of each research project is the production of scientific output in the form of scientific publications. This deliverable provides an overview of all scientific publications produced in the CODE project over the 24 month period (01.05.2012 to 30.04.2014). We provide detailed information on all accepted publications including the

- Title
- Abstract
- Publication type (Journal publication, conference paper, short paper/demo, poster, scientific news article)
- Full reference

As of 30th April 2014, a total of **17 papers** and articles have been accepted and published: 3 Journal papers, 3 conference papers, 4 workshop papers, 2 short papers, 3 posters and 2 articles in scientific news media. We have contributed to 3 Journals, 10 international conferences and workshops and to 1 scientific news medium. Also provided is our publication plan which currently includes 2 submitted papers and 4 more planed publications.

2 Accepted Publications

This chapter lists all papers and articles which were published during the CODE project, including Journal publications, papers at international conferences and workshops and articles in other media.

2.1 2012 Publications

In 2012, the initial year of the project, 2 papers were published: 1 Journal paper and 1 workshop paper.

2.1.1 Unleashing Semantics of Research Data

Abstract: Research depends to a large degree on the availability and quality of primary research data, i.e., data generated through experiments and evaluations. While the Web in general and Linked Data in particular provide a platform and the necessary technologies for sharing, managing and utilizing research data, an ecosystem supporting those tasks is still missing. The vision of the CODE project is the establishment of a sophisticated ecosystem for Linked Data. Here, the extraction of knowledge encapsulated in scientific research paper along with its public release as Linked Data serves as the major use case. Further, Visual Analytics approaches empower end users to analyse, integrate and organize data. During these tasks, specific Big Data issues are present.

Publication type: workshop paper

Florian Stegmaier, Christin Seifert, Roman Kern, Patrick Hoefler, Sebastian Bayerl, Michael Granitzer, Harald Kosch, Stefanie Lindstaedt, Belgin Mutlu, Vedran Sabol, Kai Schlegel, Stefan Zwicklbauer (2012) **Unleashing Semantics of Research Data**, *The Second Workshop on Big Data Benchmarking (WBDB2012.in),* December 2012, Pune, India.

2.1.2 Exploring different optimization techniques for an External Multimedia Meta-search Engine

Abstract: Along with the tremendous growth of Social Media, the variety of multimedia sharing platforms on the Web is ever growing, whereas unified retrieval issues remain unsolved. Beside unified retrieval languages and metadata interoperability issues, a crucial task in such a retrieval environment is query optimization in federated and distributed retrieval scenarios. This work introduces three different dimensions of query optimization that have been integrated in an external multimedia meta-search engine. The main innovations are query execution planning, various query processing strategies as well as a multimedia perceptual caching system.

Publication type: Journal paper

Kai Schlegel, Florian Stegmaier, Sebastian Bayerl, Harald Kosch, Mario Döller (2012) **Exploring different optimization techniques for an External Multimedia Meta-search Engine**, International Journal of Multimedia Data Engineering and Management, Volume 3, Issue 4, *IJMDEM* 3.4, pages 31-51.

2.2 2013 Publications

In 2013 eight papers were published: 1 Journal paper, 2 conference papers, 1 workshop paper, 2 short papers (incl. demos) and 2 posters. One publication earned the best paper award.

2.2.1 Extraction of References Using Layout and Formatting Information from Scientific Articles

Abstract: The automatic extraction of reference meta-data is an important requirement for the efficient management of collections of scientific literature. An existing powerful state-of-the-art system for extracting references from a scientific article is ParsCit; however, it requires the input document to be converted into plain text, thereby ignoring most of the formatting and layout information. In this paper, we quantify the contribution of this additional information to the reference extraction performance by an improved preprocessing using the information contained in PDF files and retraining sequence classifiers on an enhanced feature set. We found that the detection of columns, reading order, and decorations, as well as the inclusion of layout information improves the retrieval of reference strings, and the classification of reference token types can be improved using additional font information. These results emphasize the importance of layout and formatting information for the extraction of meta-data from scientific articles.

Publication type: Journal paper

Roman Kern, Stefan Klampfl (2013) **Extraction of References Using Layout and Formatting Information from Scientific Articles**, *D-Lib Magazine 19 (9/10),* September/October 2013.

2.2.2 Automated Visualization Support for Linked Research Data

Abstract: Finding, organizing and analyzing research data (i.e. publications) published in various digital libraries are often tedious tasks. Each digital library deploys their own meta-model and technology to query and analyze the knowledge (in further text, scientific facts) contained in research publications. The goal of the EU-funded research project CODE is to provide methods for federated querying and analysis of such data. To achieve this, the CODE project offers a platform, that extracts scientific facts from research data and integrates them within the Linked Data Cloud using a common vocabulary (i.e. meta-model). To support users in analyzing scientific facts, the project provides means for easy-to-use visual analysis. In this paper, we present the web-based CODE Visualization Wizard, which aims to analyze research data visually with an emphasis on automating the

visualization process. The main focus of the paper lies on a mapping strategy, which integrates various vocabularies to facilitate the automated visualization process

Publication type: short paper and demo

Belgin Mutlu, Patrick Hoefler, Vedran Sabol, Gerwald Tschinkel, Michael Granitzer (2013) **Automated Visualization Support for Linked Research Data**, 9th International Conference on Semantic Systems (i-SEMANTICS 2013)*, volume 1026 of CEUR Workshop Proceedings,* pages 40-44, September 2013, Graz, Austria.

2.2.3 Do We Need Entity-Centric Knowledge Bases for Entity Disambiguation?

Abstract: Entity Disambiguation has been studied extensively in the last 10 years with authors reporting increasingly well performing systems. However, most studies focused on general purpose knowledge bases like Wikipedia or DBPedia and left out the question how those results generalize to more specialized domains. This is especially important in the context of Linked Open Data which forms an enormous resource for disambiguation. However, the influence of domain heterogeneity, size and quality of the knowledge base remains largely unanswered. In this paper we present an extensive set of experiments on special purpose knowledge bases from the biomedical domain where we evaluate the disambiguation performance along four variables: (i) the representation of the knowledge base as being either entity-centric or document-centric, (ii) the size of the knowledge base in terms of entities covered, (iii) the semantic heterogeneity of a domain and (iv) the quality and completeness of a knowledge base. Our results show that for special purpose knowledge bases (i) document-centric disambiguation significantly outperforms entity-centric disambiguation, (ii) document-centric disambiguation does not depend on the size of the knowledge-base, while entity-centric approaches do, and (iii) disambiguation performance varies greatly across domains. These results suggest that domain heterogeneity, size and knowledge base quality have to be carefully considered for the design of entity disambiguation systems and that for constructing knowledge bases user-annotated texts are preferable to carefully constructed knowledge bases.

Publication type: conference paper

Stefan Zwicklbauer, Christin Seifert, Michael Granitzer (2013) **Do We Need Entity-Centric Knowledge Bases for Entity Disambiguation?**, *in Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW 2013)*, September 2013, Graz, Austria.

2.2.4 An Unsupervised Machine Learning Approach to Body Text and Table of Contents Extraction from Digital Scientific Articles

Abstract: Scientific articles are predominantly stored in digital document formats, which are optimised for presentation, but lack structural information. This poses challenges to access the documents' content, for example for information retrieval. We have developed a processing pipeline that makes use of unsupervised machine learning techniques and heuristics to detect the logical

structure of a PDF document. Our system uses only information available from the current document and does not require any pre-trained model. Starting from a set of contiguous text blocks extracted from the PDF file, we first determine geometrical relations between these blocks. These relations, together with geometrical and font information, are then used categorize the blocks into different classes. Based on this logical structure we finally extract the body text and the table of contents of a scientific article. We evaluate our pipeline on a number of datasets and compare it with state-of-the-art document structure analysis approaches.

Publication type: conference paper
Awards: TPDL Best Paper Award

Stefan Klampfl, Roman Kern (2013) **An Unsupervised Machine Learning Approach to Body Text and Table of Contents Extraction from Digital Scientific Articles**, 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013), *LNCS 8092*, Springer, pp 144-155, Valletta, Malta September 2013.

2.2.5 Crowdsourcing Fact Extraction from Scientific Literature

Abstract: Scientific publications constitute an extremely valuable body of knowledge and can be seen as the roots of our civilisation. However, with the exponential growth of written publications, comparing facts and findings between different research groups and communities becomes nearly impossible. In this paper, we present a conceptual approach and a first implementation for creating an open knowledge base of scientific knowledge mined from research publications. This requires to extract facts - mostly empirical observations - from unstructured texts (mainly PDF's). Due to the importance of extracting facts with high-accuracy and the impreciseness of automatic methods, human quality control is of utmost importance. In order to establish such quality control mechanisms, we rely on intelligent visual interfaces and on establishing a toolset for crowdsourcing fact extraction, text mining and data integration tasks.

Publication type: workshop paper

Christin Seifert, Michael Granitzer, Patrick Höfler, Belgin Mutlu, Vedran Sabol, Kai Schlegel, Sebastian Bayerl, Florian Stegmaier, Stefan Zwicklbauer, Roman Kern (2013) **Crowdsourcing Fact Extraction from Scientific Literature**, 3rd Workshop on Human-Computer Interaction and Knowledge Discovery (HCI-KDD 2013, held at SouthCHI 2013), *LNCS 7947,* Springer, pp 160-172, Maribor, Slovenia, 2013.

2.2.6 Trusted Facts: Triplifying Primary Research Data Enriched with Provenance Information

Abstract: A crucial task in a researchers' daily work is the analysis of primary research data to estimate the evolution of certain fields or technologies, e.g. tables in publications or tabular benchmark results. Due to a lack of comparability and reliability of published primary research data, this becomes more and more time-consuming leading to contradicting facts, as has been shown for ad-hoc retrieval [1]. The CODE project [2] aims at contributing to a Linked Science Data Cloud by integrating unstructured research information with semantically represented research data. Through crowdsourcing techniques, data centric tasks like data extraction, integration and analysis in combination with sustainable data marketplace concepts will establish a *sustainable, high-impact ecosystem*.

Publication type: poster

Kai Schlegel, Sebastian Bayerl, Stefan Zwicklbauer, Florian Stegmaier, Christin Seifert, Michael Granitzer, Harald Kosch (2013) **Trusted Facts: Triplifying Primary Research Data Enriched with Provenance Information**, 10th Extended Semantic Web Conference (ESWC 2013), p268-270, *LNCS 7955*, Springer, pp 268-270, Montpellier, France, May 2013.

2.2.7 Linked Data Query Wizard: A Tabular Interface for the Semantic Web

Abstract: Linked Data has become an essential part of the Semantic Web. A lot of Linked Data is already available in the Linked Open Data cloud, which keeps growing due to an influx of new data from research and open government activities. However, it is still quite difficult to access this wealth of semantically enriched data directly without having in-depth knowledge about SPARQL and related semantic technologies. In this paper, we present the Linked Data Query Wizard, a prototype that provides a Linked Data interface for non-expert users, focusing on keyword search as an entry point and a tabular interface providing simple functionality for filtering and exploration.

Publication type: short paper and demo

Patrick Hoefler, Michael Granitzer, Vedran Sabol, Stefanie Lindstaedt (2013) **Linked Data Query Wizard: A Tabular Interface for the Semantic Web**, 10th Extended Semantic Web Conference (ESWC 2013), *LNCS 7955*, Springer, pp. 173–177, Montpellier, France, May 2013.

2.2.8 Linked Data Interfaces for Non-expert Users

Abstract: Linked Data has become an essential part of the Semantic Web. A lot of Linked Data is already available in the Linked Open Data cloud, which keeps growing due to an influx of new data from research and open government activities. However, it is still quite difficult to access this wealth of semantically enriched data directly without having in-depth knowledge about SPARQL and related semantic technologies. The presented dissertation explores Linked Data interfaces for non-expert users, especially keyword search as an entry point and tabular interfaces for filtering and exploration. It also looks at the value chain surrounding Linked Data and the possibilities that open up when people without a background in computer science can easily access Linked Data.

Publication type: poster

Patrick Hoefler (2013) **Linked Data Interfaces for Non-expert Users**, PhD Symposium of the 10th Extended Semantic Web Conference (ESWC 2013), *LNCS 7882*, Springer, pp 702-706, Montpellier, France, May 2013.

2.3 2014 Publications

In 2014, the final year of the project, 7 papers have been published or accepted for publication (as of 30th April 2014): 1 Journal paper, 1 conference paper, 2 workshop papers, 1 poster and 2 scientific news articles.

2.3.1 Unsupervised document structure analysis of digital scientific articles

Abstract: Text mining and information retrieval in large collections of scientific literature require automated processing systems that analyse the documents' content. However, the layout of scientific articles is highly varying across publishers, and common digital document formats are optimised for presentation, but lack structural information. To overcome these challenges, we have developed a processing pipeline that analyses the structure a PDF document using a number of unsupervised machine learning techniques and heuristics. Our system uses only information available from the current document and does not require any pre-trained model. First, contiguous text blocks are extracted from the raw character stream. Next, we determine geometrical relations between these blocks, which, together with geometrical and font information, are then used categorize the blocks into different classes. Based on this resulting logical structure we finally extract the body text and the table of contents of a scientific article. We separately evaluate the individual stages of our pipeline on a number of different datasets and compare it with other document structure analysis approaches. We show that it outperforms a state-of-the-art system in terms of the quality of the extracted body text and table of contents. Our unsupervised approach could provide a basis for advanced digital library scenarios that involve diverse and dynamic corpora.

Publication type: Journal paper

Stefan Klampfl, Michael Granitzer, Kris Jack, Roman Kern (2014**) Unsupervised document structure analysis of digital scientific articles,** accepted for publication in *International Journal on Digital Libraries.*

2.3.2 Linked Data Query Wizard: A Novel Interface for Accessing SPARQL Endpoints

Abstract: In an interconnected world, Linked Data is more important than ever before. However, it is still quite difficult to access this new wealth of semantic data directly without having in-depth knowledge about SPARQL and related semantic technologies. Also, most people are currently used to consuming data as 2-dimensional tables. Linked Data is by definition always a graph, and not that many people are used to handle data in graph structures. Therefore we present the Linked Data Query Wizard, a web-based tool for displaying, accessing, filtering, exploring, and navigating Linked Data stored in SPARQL endpoints. The main innovation of the interface is that it turns the graph

structure of Linked Data into a tabular interface and provides easy-to-use interaction possibilities by using metaphors and techniques from current search engines and spreadsheet applications that regular web users are already familiar with.

Publication type: workshop paper

Patrick Hoefler, Michael Granitzer, Eduardo Veas, Christin Seifert (2014) **Linked Data Query Wizard: A Novel Interface for Accessing SPARQL Endpoints**, 7th International Workshop about Linked Data on the Web (LDOW2014) at 23rd International World Wide Web Conference (WWW 2014), April 2014, Seoul, Korea.

2.3.3 Balloon Synopsis: A Modern Node-Centric RDF Viewer and Browser for the Web

Abstract: Nowadays, the RDF data model is a crucial part of the Semantic Web. Especially web developers favour RDF serialization formats like RDFa and JSON-LD. However, the visualization of large portions of RDF data in an appealing way is still a cumbersome task. RDF visualizers in general are not targeting the Web as usage scenario or simply display the complex RDF graph directly rather than applying a human friendly facade. Balloon Synopsis tries to overcome these issues by providing an easy-to-use RDF visualizer based on HTML and JavaScript. For an ease integration, it is implemented as jQuery-plugin offering a node-centric RDF viewer and browser with automatic Linked Data enhancement in a modern tile design.

Publication type: poster

Kai Schlegel, Thomas Weißgerber, Florian Stegmaier, Christin Seifert, Michael Granitzer, Harald Kosch (2014) **Balloon Synopsis: A Modern Node-Centric RDF Viewer and Browser for the Web**, *in Proceedings of the 11th European Semantic Web Conference (ESWC 2014, Poster Session)*, May 2014, Anissaras, Crete, Greece.

2.3.4 Balloon Fusion: SPARQL Rewriting Based on Unified Co-Reference Information

Abstract: While Linked Open Data showed enormous increase in volume, yet there is no single point of access for querying the over 200 SPARQL repositories. In this paper we present Balloon Fusion, a SPARQL 1.1 rewriting and query federation service build on crawling and consolidating co-reference relationships in over 100 reachable Linked Data SPARQL Endpoints. The results of this process are 17.6M co-reference statements that have been clustered to 8.4M distinct semantic entities and are now accessible as download for further analysis. The proposed SPARQL rewriting performs a substitution of all URI occurrences with their synonyms combined with an automatic endpoint selection based on URI origin for a comprehensive query federation. While we show the technical feasibility, we also critically reflect the current status of the Linked Open Data cloud: although it is huge in size, access via SPARQL Endpoints is complicated in most cases due to missing quality of service.

Publication type: workshop paper

Kai Schlegel, Florian Stegmaier, Sebastian Bayerl, Michael Granitzer, Harald Kosch (2014) **Balloon Fusion: SPARQL Rewriting Based on Unified Co-Reference Information**, *in proceedings of the 5th International Workshop on Data Engineering Meets the Semantic Web* (DESWeb 2014), co-located with ICDE 2014, March 2014, Chicago, USA.

2.3.5 Suggesting Visualisations for Published Data

Abstract: Research papers are published in various digital libraries, which deploy their own meta-models and technologies to manage, query, and analyze scientific facts therein. Commonly they only consider the meta-data provided with each article, but not the contents. Hence, reaching into the contents of publications is inherently a tedious task. On top of that, scientific data within publications are hardcoded in a fixed format (e.g. tables). So, even if one manages to get a glimpse of the data published in digital libraries, it is close to impossible to carry out any analysis on them other than what was intended by the authors. More effective querying and analysis methods are required to better understand scientific facts. In this paper, we present the web-based CODE Visualisation Wizard, which provides visual analysis of scientific facts with emphasis on automating the visualisation process, and present an experiment of its application. We also present the entire analytical process and the corresponding tool chain, including components for extraction of scientific data from publications, an easy to use user interface for querying RDF knowledge bases, and a tool for semantic annotation of scientific data sets.

Publication type: conference paper

Belgin Mutlu, Patrick Hoefler, Gerwald Tschinkel, Eduardo Veas, Vedran Sabol, Florian Stegmaier, Michael Granitzer (2014) **Suggesting Visualisations for Published Data**, *in Proceedings of the 5th International Conference on Information Visualization Theory and Applications* (IVAPP 2014), January 2014, Lisbon, Portugal.

2.3.6 CODE Query Wizard and Vis Wizard: Supporting Exploration and Analysis of Linked Data

Article available at: http://ercim-news.ercim.eu/en96/special/code-query-wizard-and-vis-wizard-supporting-exploration-and-analysis-of-linked-data

Publication type: scientific news article

Patrick Hoefler, Belgin Mutlu (2014) **CODE Query Wizard and Vis Wizard: Supporting Exploration and Analysis of Linked Data**, *ERCIM News January 14(96)*, p. 32-33

2.3.7 Lost in Semantics? Ballooning the Web of Data

Article available at: http://ercim-news.ercim.eu/en96/special/lost-in-semantics-ballooning-the-web-of-data

Publication type: scientific news article

Florian Stegmaier, Kai Schlegel, Michael Granitzer (2014) **Lost in Semantics? Ballooning the Web of Data**, *ERCIM News January 14(96)*, p. 18-19

3 Future Publications

This chapter lists all publications which were submitted to international conferences but, by the time of submission of this deliverable (early May 2014), their acceptance (or rejection) was not yet confirmed. Beyond that a plan for future publications is briefly outlined.

3.1 Submitted Papers

As of 30th April 2014, two papers have been submitted with a notification of acceptance expected by the end of May:

Mark Kröll, Stefan Klampfl, Roman Kern, **Towards a Marketplace for the Scientific Community: Accessing Knowledge from the Computer Science Domain**, submitted to International Digital Libraries Conference (DL2014)

Abstract: As scientific output is constantly growing, it is getting more and more important to keep track not only for researchers but also for other scientific stakeholders such as funding agencies or research companies. Each stakeholder values different types of information. A funding agency, for instance, might be rather interested in the number of publications funded by their grants. However, information extraction approaches tend to be rather researcher-centric indicated, for example, by the type of named entities to be recognized. In this paper we account for different perspectives and propose an ontological description of one scientific domain – the computer science domain. We accordingly annotated a set of 22 computer science papers by hand and make this data set publicly available. In addition, we started to apply methods to automatically extract instances and report preliminary results. Automating the process of populating the proposed ontology represents a prerequisite for our vision of a “Marketplace for the Scientific Community” where stakeholders can exchange not only Information but also search concepts or annotated data.

Stefan Klampfl, Kris Jack, Roman Kern, **A Comparison of two Unsupervised Table Recognition Methods from Digital Scientific Articles**, submitted to International Digital Libraries Conference (DL2014)

Abstract: In digital scientific articles tables are a common form of presenting information in a structured way. However, the large variability of table layouts and the lack of structural information in digital document formats pose significant challenges for information retrieval and related tasks. In this paper we present two table recognition methods based on unsupervised learning techniques and heuristics which automatically detect both the location and the structure of tables within a article stored as PDF. For both algorithms the table region detection first identifies the bounding boxes of individual tables from a set of labelled text blocks. In the second step, two different tabular structure detection methods extract a rectangular grid of table cells from the set of words contained in these table regions. We evaluate each stage of the algorithms separately and compare performance values on two data sets from different domains. We find that the table recognition performance is in line with state-of-the-art commercial systems and generalises to the non-scientific domain.

3.2 Publication Plan

A list of planed research publications based on CODE results is as follows.

University of Passau plans to submit following papers:

- A paper titled "Bacon: Linked Data Integration based on the RDF Data Cube Vocabulary" will be submitted to the ISWC 2014 Research Track in early May 2014.
- A paper addressing crowd-sourced creation of taxonomies using Mindmeister semantc MindMaps will be submitted as a full paper at the SEMANTiCS conference (formerly known as I-Semantics) by the end of May 2014.

Know-Center plans to submit the following papers:

- A paper titled "Discovery and Visual Analysis of Linked Data for Humans" will be submitted to the ISWC 2014 Research Track in early May 2014.
- A paper titled “What does it take to get published? A Real-Life Study of Scientific Writing” will be submitted to COLING Workshop on Synchronic and Diachronic Approaches to Analyzing Technical Language (SADAATL 2014) in early May 2014

4 Summary and Conclusion

In the following, a summary of Journals (Table 1), international conferences and workshops (Table 2) and other media (Table 3) is given where CODE papers have been published.

Journal	Publ. count	Year
International Journal of Multimedia Data Engineering and Management (IJMDEM)	1 paper	2012
D-Lib Magazine	1 paper	2013
International Journal on Digital Libraries	1 paper	2014, accepted

Table 1: 3 Journals where CODE papers were published.

Conference	Publ. count	Date	Place
2nd Workshop on Big Data Benchmarking (WBDB2012.in)	1 paper	2012, December 17 - 18	Pune, India
9th International Conference on Semantic Systems (i-SEMANTICS 2013)	1 paper	2013, September 4 - 6	Graz, Austria
13th International Conference on Knowledge Management and Knowledge Computing (i-KNOW 2013)	1 paper	2013, September 4 - 6	Graz, Austria
17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013)	1 paper	2013, September 22-26	Valetta, Malta
3rd Workshop on Human-Computer Interaction and Knowledge Discovery (HCI-KDD 2013), held at SouthCHI 2013	1 paper	2013, July 1 - 3	Maribor, Slovenia
10th Extended Semantic Web Conference (ESWC 2013)	1 paper, 2 posters	2013, May 26 - 30	Montpellier, France
11th Extended Semantic Web Conference (ESWC 2014)	1 poster	2014, May 25 - 29	Anissaras, Greece
7th International Workshop on Linked Data on the Web (LDOW2014) at 23rd International World Wide Web Conference (WWW 2014)	1 paper	2014, April 7 - 11	Seoul, Korea
5th International Workshop on Data Engineering Meets the Semantic Web (DESWeb) , co-located with the 30th IEEE International Conference on Data Engineering	1 paper	2014, March 31	Chicago, USA
5th International Conference on Information Visualization Theory and Applications (IVAPP 2014)	1 paper	2014, January 5-8	Lisbon, Portugal

Table 2: 10 International conferences and workshops where CODE papers were published and presented.

Published in	**Publ. count**	**Date**	Link
ERCIM News	2 scientific news articles	2014, January	http://www.ercim.eu/

Table 3: Other media where CODE papers were published.

As of 30th April 2014, a total of **17 papers** and articles have been accepted and published over the 24 month duration of the CODE project:

- 3 Journal papers
- 3 conference papers
- 4 workshop papers
- 2 short papers
- 3 posters
- 2 articles in scientific news media.

We have contributed to 3 Journals, 10 international conferences and workshops and to 1 scientific news medium. Therefore the project has more than fulfilled the goals defined in the Description of Work [1] (2 Journals, 6 proceedings). Additionally, at least 6 more papers are planned, 2 of them already submitted to international conferences and 4 more in preparation. We conclude that the CODE project has been very successful in producing scientific output and contributing to the research community.

5 References

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