D4.4 - Usability Evaluation Report

Summary: This report covers Task 4.3 of the CODE project and deals with the 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. Some of the evaluation results have already been published in international peer-reviewed conference articles.

<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>Know-Center</td>
</tr>
<tr>
<td>Revision</td>
<td>final</td>
</tr>
<tr>
<td>Authors</td>
<td>Patrick Hoefler, Belgin Mutlu, Gerwald Tschinkel, Vedran Sabol</td>
</tr>
</tbody>
</table>

Statement of originality: This document contains original unpublished work except where clearly indicated otherwise.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

<table>
<thead>
<tr>
<th>Role</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>2014-04-18</td>
<td>B. Mutlu</td>
<td>Know-Center</td>
<td>Vis Wizard</td>
</tr>
<tr>
<td>2nd Draft</td>
<td>2014-04-23</td>
<td>P. Hoefler</td>
<td>Know-Center</td>
<td>Query Wizard</td>
</tr>
<tr>
<td>3rd Draft</td>
<td>2014-04-27</td>
<td>V. Sabol</td>
<td>Know-Center</td>
<td>Proof reading, conclusions</td>
</tr>
<tr>
<td>Final Version</td>
<td>2014-04-30</td>
<td>P. Hoefler, B. Mutlu, G. Tschinkel, V. Sabol</td>
<td>Know-Center</td>
<td>Executive Summary</td>
</tr>
</tbody>
</table>
Table of Contents

1 Executive Summary - 4 -

2 Introduction - 5 -

3 Evaluation of the CODE Query Wizard - 7 -
   3.1 STUDY DESIGN - 7 -
   3.2 EVALUATION RESULTS - 11 -

4 Evaluation of the CODE Visualization Wizard - 14 -
   4.1 USER INTERFACE - 14 -
   4.2 EVALUATION METHODOLOGY - 19 -
   4.3 EVALUATION RESULTS - 26 -

5 Conclusion - 30 -

6 References - 31 -
Table of Figures

Figure 1: Overview of WP4 of the CODE project ................................................................. - 6 -
Figure 2: Quantitative results based on the NASA Task Load Index .................................. - 11 -
Figure 3: Five suggested charts, shown as enabled buttons, for a RDF Data Cube with three components .................................................................................................................. - 14 -
Figure 4: Mapping combinations changed for scatter plot .................................................... - 15 -
Figure 5: Interactive Pie and Bar Chart (PAN 2009 challenge results) ................................. - 15 -
Figure 6: Brushing in multiple coordinated views with jQuery table as the base chart .......... - 16 -
Figure 7: Filtering a cube using the numeric filter .................................................................. - 18 -
Figure 8: Base cube (left), aggregated cube, grouped by year, displaying the max value (right) ..... - 19 -
Figure 9: Query Wizard displaying a dataset about how many children were born by women between the ages of 35 and 50 years in the year 2011. ........................................................................ - 20 -
Figure 10: Visualization Wizard: Coordinated View of the visualized dataset (Total value of EC funding to participants in FP7-ICT projects). The brushed data range shows the funding for Germany for the years between 2007 and 2011 ........................................................................ - 21 -
Figure 11: Quantitative results based on the NASA Task Load Index ................................. - 27 -
1 Executive Summary

This report covers the in-depth usability evaluation of the CODE Query Wizard and the CODE Visualization Wizard, which are the two main prototypes of WP4 of the CODE project.

CODE Query Wizard

14 people participated in the usability study (7 with and 7 without a background in computer science), and 2 people took part in the related pre-study. The results of the study showed that the tool was in general very usable, both for people with and without a background in computer science.

The main point for improvement was the missing possibility to add URI filters through a menu in the table header. Also the total number of results could be displayed more prominently, and the implementation of an “infinite scroll” mechanism that automatically displays more data as soon as the users scroll to the bottom of the screen could circumvent the problem that users need to load more data manually in order to find what they are looking for.

CODE Visualization Wizard

8 people participated in the usability study. The general feedback was that the prototype worked very well and bug-free, that it was easy to use, and that the users liked the interactivity of the displayed charts.

The main point for improvement was the fact that it was not always clear to the users that they were working with a filtered dataset when they had previously set a filter through the CODE Query Wizard. The colour coding of the geo chart could also be improved. And finally, some people did not immediately understand why, for a given dataset, certain charts were suggested and others were not. Once they understood the underlying mechanisms, they appreciated the time and effort that the feature had saved them.
2 Introduction

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.

WP4 of CODE therefore deals with the development of web-based visual analytics user interfaces for Linked Data. This means that it should be possible for people who don’t know anything about RDF and SPARQL to access information that is already available but difficult to access in the Linked Open Data Cloud.

WP4 consists 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 [http://code.know-center.tugraz.at/search](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 [http://code.know-center.tugraz.at/vis](http://code.know-center.tugraz.at/vis).

The CODE Query Wizard and the CODE Vis Wizard have been integrated into the CODE Question & Answer Portal 42-data.org and have become a crucial components of this data marketplace.

Figure 1 shows a conceptual overview of how the CODE Query Wizard and the CODE Visualization Wizard interact.

This report covers Task 4.3 of the CODE project and deals with the 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 Query Wizard and the Visualization Wizard. Some of the evaluation results have already been published in international peer-reviewed conference articles (see Section 5).
Figure 1: Overview of WP4 of the CODE project
3 Evaluation of the CODE Query Wizard

During the development of the Linked Data Query Wizard we followed the ‘release early, release often’ principle. This means that as soon as a feature was complete and ready for testing, it immediately rolled out to our staging server and, if no major problems were found, a short time later (usually within hours, sometimes days) is was publicly available at our production server. This also means that the Linked Data Query Wizard has been under permanent scrutiny of fellow researchers from the CODE project as well as other interested colleagues for several months now. They regularly provided valuable feedback on stability and usability issues as well as helpful feature requests.

Additionally, the Linked Data Query Wizard was used in a workshop setting with around 20 students of the Semantic Technologies course at Graz University of Technologies. There it proved helpful in evaluating the quality of the Linked Open Data the students had previously created during the workshop.

To conclude the first development cycle of the Linked Data Query Wizard, an in-depth evaluation was performed. In the remainder of this section we will present the study design, both the quantitative as well as the qualitative study results, and a discussion of the findings.

3.1 Study Design

Our study followed the principles of the Retrospective Think Aloud protocol, combined with the NASA Task Load Index. In the following we will describe the details of the study.

In total, 14 people participated in this study, and 2 people took part in the related pre-study. Each session started with a short explanation of the study and the signing of the declaration of consent. The session was then guided by a survey that was filled out by the participants themselves. The first page of the survey consisted of background questions about the participant:

How’s your English?

The study was conducted in English with participants from different countries, but no English native speakers. 10 participants declared their English skills as fluent, 3 as okay and 1 as basic.

Have you used the Linked Data Query Wizard before?

For this study, only participants with no prior experience with the Linked Data Query Wizard were selected. Accordingly, all 14 participants answered with no.

How frequently do you use spreadsheet applications?

The Linked Data Query Wizard is mainly intended for people that have prior experience with spreadsheet applications. Although this was not checked during the participant selection phase, all participants indeed had at least some experience: 2 of them used spreadsheet applications every workday, 4 of them several times a week, 6 of them several times a month and 2 of them once a month or less often.
How frequently do you look up information on the Internet?
This question aimed to probe the level of the participants' web experience, which turned out to be quite high: 13 of the participants answered every workday, one of them several times a week.

How frequently do you write SPARQL queries?
This question was intended to find out if there were any Semantic Web experts among the participants. Only one of the 14 participants answered once a month or less often, whereas 6 of them answered never and 7 of them What's SPARQL?

What's your age?
The final background question provided information about the age ranges of the participants: 4 of the participants were between 18 and 27 years old, 9 of them between 28 and 37, and 1 between 58 and 67.

After the initial background questions, the participants had to solve 4 tasks using the Linked Data Query Wizard. These were as follows:

**Task 1: Service Data**
There is an available dataset called '% of basic public services for citizens, which are fully available online' provided by EU Open Data.
We are interested only in the data from the year 2010, please filter it accordingly.
After that, please visualize the results.
You have 3 minutes to complete this task.

**Task 2: Data Overview**
This task deals with the same data as before, the dataset called '% of basic public services for citizens, which are fully available online' provided by EU Open Data.
However, this time we are interested in an overview of the data.
Therefore, please aggregate the dataset and display the average values, grouped by year.
After that, please visualize the results.
You have 3 minutes to complete this task.

**Task 3: Pulp Data**
Before you start, please select the data source called 'Wikipedia (CODE Edition)' (5th from the top) in the 'Search Linked Data' section.
There is a film called 'Pulp Fiction'.
1. Was Bruce Willis a cast member of this film?
2. Who was the director of this film?
3. Are there any other films by the same director where Bruce Willis was a cast member? If so, which ones and how many?
You have 5 minutes to complete this task.
Task 4: More Data

Once again, before you start, please select the data source called 'Wikidata (CODE Edition)' (5th from the top) in the 'Search Linked Data' section.

There is a music album that has the word 'Antidote' in it.

1. Who is the performer / musical artist of this album? The one you are looking for starts with 'Mor...'.
2. Which other albums has this artist released? Please make sure that only albums (and no singles) are displayed.
3. Make a MindMap containing all the information that you just looked up.
   You have 5 minutes to complete this task.

After each task was finished — either by the participant successfully completing it, or by reaching the respective time limit — the participants filled out a NASA Task Load Index form and subjectively judged several aspects of the task they had just worked on. The form consisted of the following questions:

Mental Demand
   How mentally demanding was the task?

Physical Demand
   How physically demanding was the task?

Temporal Demand
   How hurried or rushed was the pace of the task?

Performance
   How successful were you in accomplishing what you were asked to do?

Effort
   How hard did you have to work to accomplish your level of performance?

Frustration
   How insecure, discouraged, irritated, stressed, and annoyed were you?

Additionally after each task the participants were asked the following question:

Any comments? What was good / bad / unexpected / difficult?

Firstly they were asked to write down what came to their minds. After that the study conductor asked about specific observations that he had made during the task. After a usually short, sometimes a little longer discussion, the participants added written remarks that came up during the discussion. The final page of the survey consisted of four questions that gave the participants the opportunity to provide additional qualitative feedback. These questions were:

- What did you like about the Linked Data Query Wizard?
- What did you hate about the Linked Data Query Wizard?
- For which tasks would you personally use the Linked Data Query Wizard?
- If you could have solved the tasks with other tools of your choice, which ones would you have used?
After the participant had answered these final questions, the session was concluded.

The four tasks that the participants had to solve were basically divided into two groups:

- **Tasks 1 and 2** concentrated on the Show Available Datasets mode and were intentionally of moderate complexity. Since all participants had, except for a short guided tour, no prior experience with the Linked Data Query Wizard, these tasks were intended to ease them into the system and to discover potential problems with the user experience at the same time.

- **Tasks 3 and 4** concentrated on the Search Linked Data mode and were of significantly higher complexity. By then the participants had become more familiar with the Linked Data Query Wizard, since the learning effect should have already started to kick in.

It was clear that the lack of randomization of the tasks would lead to an uncompensated learning effect. This did not pose a problem under the circumstances, since it was not the goal of this study to compare the different tasks with each other, but rather to evaluate the general usefulness of the system and find its weak points with regard to user experience.

Apart from the quantitative feedback (NASA Task Load Index) and the qualitative feedback (Retrospective Think Aloud), the study conductor also measured the task completion rate.

Task completion time was not measured for several reasons. For example, the study was conducted on the live system and not on a lab setup, so fluctuations in (external) server response times were to be expected. Also, since the Linked Data Query Wizard offers a rather novel interface to access Linked Data, the main goal of the study was to show if users are able to use it at all and where potential problems in understanding arise.

Another idea was to go with a conventional Think Aloud study instead of using the Retrospective Think Aloud protocol. This would have had a negative impact on completion time and, due to the given time limits for completing each task, could have resulted in a lower task completion rate.

Also, competing the Linked Data Query Wizard against other tools was not really an option, since there are currently no tools that could come close enough in functionality to make a direct comparison feasible.

Another possibility would have been to compare against Google searches or SPARQL queries written by Semantic Web experts. However, in the first case, the test cases could have been constructed in a way where the Linked Data Query Wizard would have always won by a landslide, which would have defeated the purpose of a direct comparison. In the latter case, competing against manually created SPARQL queries was not ideal either, since the focus of this evaluation was on regular web users and not on Semantic Web experts.
3.2 Evaluation Results

Due to the combination of the Restrospective Think Aloud protocol with the NASA Task Load Index, it was possible to generated four different kinds of results from the study:

- Quantitative results (NASA Task Load Index)
- Quantitative results (Task Completion Rate)
- Comparison between participants with and without a background in computer science
- Qualitative results (Retrospective Think Aloud)


Quantitative Results (NASA Task Load Index)

The quantitative results of the NASA Task Load Index can be seen in the box plots of Figure 2. The six plots represent the results for the six different aspects of the NASA Task Load Index. Those were mental demand, physical demand, temporal demand, performance, effort, and frustration.

![Figure 2: Quantitative results based on the NASA Task Load Index](image)

- The mental demand was rather low for the first two tasks and increased only slightly for the more complex last two tasks. The variance between the participants was quite high.
- The physical demand was, as expected, very low throughout the study.
- The temporal demand — with respect to the time limits of the tasks — basically corresponded with the results from the mental demand, showing a generally low demand with a high degree of variance between the participants.
The performance scores were very high with a median of 10 out of 10 for all four tasks. Out of the 56 tasks performed in total by the 14 participants, 49 were successfully completed, 6 were not completed entirely in time, and only 1 was not completed at all.

The subjective effort of the participants showed a high variance between the participants, however it also showed the learning effect very nicely: The effort necessary by the participants decreased after the first task, since the second task was similar to the first one. The third task was completely different, which raised the level of necessary effort again. The fourth task was similar to the third task, which again resulted in lower effort.

The frustration level was rather low throughout the study, but again with a very high variance between the participants.

Quantitative Results (Task Completion Rate)
In addition to the subjective quantitative results measured via the NASA Task Load Index, the study conductor also measured the task completion rate objectively. There was, however, no significant difference between the subjective performances as judged by the participants themselves and the objectively measured task completion rates. In detail, this means:

- 13 out of 14 participants were able to solve task 1 completely. 1 participant only received 2 out of 10 points.
- All 14 participants were able to solve task 2 completely in time.
- Task 3 turned out to be the most difficult one: 10 out of the 14 participants were able to solve it completely, the other 4 participants only received 5 out of 10 points.
- Task 4 was completely solved by 12 out of the 14 participants. 1 participant only received 5 out of 10 points, and 1 participant received 0 points.

Background in Computer Science
An interesting research question came up in the preparation of the study: Would there be a significant difference in the results of participants with and participants without a background in computer science? For this reason, 7 of the study participants had a background in computer science, whereas the other 7 did not.

To determine if there was indeed a difference between these two groups, independent two-sample t-tests with equal sample size were performed, comparing all 24 results of the NASA Task Load Index (6 aspects * 4 tasks) as well as the objectively measured task completion rates. The result was that all calculated p-values were larger than 0.1. This means that for our study, no significant difference regarding the results of the subjective NASA Task Load Index or the objective task completion rate between participants with and without a background in computer science could be measured.

Qualitative Results
The qualitative results of the evaluation were based on the statements of the participants collected during the Retrospective Think Aloud phase of the study.
Regarding task 1, the main problem that the participants encountered concerned the filtering: To set a URI filter, the participants had to click on the respective entity (in task 1, it was the year 2010) and select *Add as filter* in its context menu. However, 10 of the 14 participants had problems with setting the filter because they expected it to be set through a context menu item in the table header, as it is the case for other filters (text, number and date) in the Linked Data Query Wizard and in most current spreadsheet applications.

Regarding task 2, the feedback was much more positive, the use of the *Aggregate dataset* feature did not cause any major problems.

Regarding task 3, the combination of multiple filters turned out to be quite a challenge for the participants. Also, the fact that all cast members of the matching movies were displayed even after *Bruce Willis* had been set as a filter confused some users. Because all of the cast members were displayed for each movie, this also meant that the rows became quite high, which several participants found irritating.

The majority of task 4 did not pose a problem for the participants after having completed the similar third task. The fact that the relevant search result did not appear on the first, but on the second result page, caused huge confusion for almost all of the participants, even though the number of total results and the *Load all results* button were visible to all participants all of the time.

When asked about what they liked about the Linked Data Query Wizard, they general opinion was that once they had worked out how the filtering worked, the interface was easy enough to use. Additionally they liked how they could create a useful list of results from a huge database with only a few simple steps.

When asked about what they didn't like about the Linked Data Query Wizard, there was no general theme. Four of the participants mentioned that it would have been hard for them to choose a data source if they had not been told which one to use.

When asked about what they would personally use the Linked Data Query Wizard for, no clear trend could be recognized. The answers ranged from *don't know yet* and *statistical data* to *newspaper entries* and *exploration of data sources*.

Finally, when asked about which other tools they would have used to solve similar tasks if the Linked Data Query Wizard had not been available, it became very clear what the direct competitors for the Linked Data Query Wizard were: Almost every participant immediately mentioned that they would use Google to search for data or information. The majority of participants mentioned that they would also use specialized portals to look for certain information, e.g. IMDB for data about movies. When it came to working with data, analyzing and visualizing it, almost all participants mentioned that they would use Microsoft Excel and that they would probably manually collect and copy the data into the spreadsheet.
4 Evaluation of the CODE Visualization Wizard

4.1 User Interface
The CODE Visualization Wizard is implemented as a web-based tool for visually analyzing Linked Data in the form of RDF Data Cubes. There are 10 different visualizations integrated in Visualization Wizard for different kind of dataset. For the representations of the visualizations we use icons that indicate if the respective visualization is suitable for a given dataset (see Figure 3).

![Possible Charts](image)

**Figure 3**: Five suggested charts, shown as enabled buttons, for a RDF Data Cube with three components

Once the user has chosen the visualization, the Wizard automatically maps the data to the available visual attributes of the visualization. The user sees on the website both the created visualization and as well as the mapping.

Since there are different mapping combinations possible for visualizations with more than three visual attributes, the user has the opportunity to change between different mapping combinations (see Figure 4).
Furthermore, the user is supported by the Visualization Wizard to organize, manage, refine and inspect the visualized data with the following methods:

- Mouse-over inspections
- Brushing in multiple coordinated views
- Filtering
- Aggregation

Data manipulations techniques like aggregation and filtering, and interactive techniques such as coordinated brushing and mouse-over, allow the user to refine the visual representation of the data providing powerful visual analysis functionality.

**Mouse-Over**

Mouse over inspection highlights a visual element, such as a bar or a line, and displays additional information (typically in a tool tip). In the following, we present some visualizations with mouse-over functionality (see Figure 5).

![Figure 4: Mapping combinations changed for scatter plot](image)

**Figure 4: Mapping combinations changed for scatter plot**

![Figure 5: Interactive Pie and Bar Chart (PAN 2009 challenge results)](image)

**Figure 5: Interactive Pie and Bar Chart (PAN 2009 challenge results)**
Brushing
The idea behind brushing is to select a subset of data in one visualization (the “base chart”) and see the distribution of the same data in one or more other visualizations. All integrated visualizations in the Visualization Wizard support coordinated brushing over multiple views. Visually, the data elements selected by the brush retain their original color, while data elements not selected by the brush are shown in gray (see Figure 6).

Figure 6: Brushing in multiple coordinated views with jQuery table as the base chart
Filtering
The Vis Wizard supports two different filter functions:

- Filtering of measures
- Filtering of the data range

Filtering of measures
Once a measure is removed the base RDF Data Cube has a measure less. Hence the mapping algorithm suggests appropriate mapping and visualizations for this “new” RDF Data Cube.

Filtering of the data range (URI, numeric, date)
The Vis Wizard supports the following filtering functions:

- Filtering by URI
- Filtering by numeric entities
- Filtering by date

We use the CODE Query Wizard to set these filters functions.

Filtering by URI
Consider the following example: We have a RDF Data Cube, whose content is the funding amounts to participants (countries) in FP7-ICT projects in the EU. In case the user is only interested in the amounts for Austria, the user only has to set the URI filter in the Query Wizard on Austria.

Filtering numeric entities
In case we have numeric data, we can use the numeric filter in Query Wizard to select a particular data interval. Only the selected data interval from the RDF Data Cube is visualized, whereby either the dimensions or the measures will be removed or changed (see Figure 7).
Filtering date entities
In case we have date information, we can use the date filter in the Query Wizard to select a particular date interval.

Aggregation
The aim of the aggregate function is to group the values of multiple rows together to form a single value. The following aggregation functions are currently included in the Vis Wizard: Average, Count, Maximum, Minimum, and Sum.

Since the numbers of the RDF Data Cube components will be changed after the aggregation, the mapping algorithm determines new visualizations and mapping suggestions for the aggregated RDF Data Cube. After the aggregation, the original and the aggregate cube are displayed together. Additionally the Visualization Wizard provides coordinated brush functionality between the visualization for base cube and the visualizations for aggregated cube (see Figure 8).
4.2 Evaluation Methodology

The main objective of the CODE approach is to simplify the accessing, exploring and analyzing Linked Data. In order to prove this hypothesis, it is necessary to evaluate the whole CODE Workflow, i.e. the process starting from the table extraction to its final visualization. For this purpose, we have prepared statistical data stored in RDF Data Cube format, which allows us to simulate whole CODE workflow without having to integrate all components. Hence we performed an evaluation only with Query and Visualization Wizard, where our focus lay on the latter. We evaluated the Visualization Wizard principally to prove our aforementioned statement but also to identify usability gaps in this tool.

Evaluation Design

For the evaluation we used predefined RDF Data Cubes, which we have stored in our own Linked Data Endpoint CODEV. This endpoint was accessible by the CODE Query Wizard, which provides the interface to Linked Data.

Before the evaluation started the moderator provided a short guided tour through the Query and Visualization Wizard.

For the tour through the Query Wizard the moderator selected a RDF Data Cube from the Vienna Linked Open Data, displayed it on the Query Wizard and performed a numeric filtering. The dataset contend data about the birthrate in Vienna in a particular year associated with the age of the women. The goal of the demonstration was to show participants, how the graph structure of Linked Data will represent in the form of a table and how to refine data on Query Wizard e.g. by using filtering function (see Figure 9).
The second demo targeted the Visualization Wizard. The moderator selected an RDF Data Cube from EU Open Data and visualized it using the Visualization Wizard. The dataset was about the total value of EC funding to participants in FTP-ICT project in a particular year. The goal of the demonstration was to show participants how the Visualization Wizard can simplify the visual analysis of statistical data (which is provided as Linked Data).

The participants have been informed during the demo about how the visualizations and mapping suggestions work and about the automated creation of visualization on the Visualization Wizard. The moderator also demonstrated to the users how to organize, manage, refine and inspect visualized data by using:

- Multiple coordinated views
- Brushing
- Aggregation
- Brushing between base and aggregated data (see Figure 10)
The evaluation has been performed as a combination of the Retrospective Think Aloud protocol with the NASA Task Load Index. In total, 8 people participated in this evaluation. The evaluation started with background questions about the participants:

- **How is your English?**
  The evaluation was conducted in English. From 8 participants 5 declared their English skills as “Fluent” and 3 as “Okay”.

- **Have you used the CODE Visualization Wizard before?**
  We selected for this evaluation only participants with no prior experience with the CODE Visualization Wizard. Accordingly, all 8 participants answered with “no”.

- **Have you used the CODE Linked Data Query Wizard before?**
  We selected for this evaluation only participants with no prior experience with the CODE Linked Data Query Wizard. Accordingly, all 8 participants answered with “no”.

- **How frequently do you use spreadsheet applications?**
  Both CODE Visualization Wizard and the CODE Linked Data Query Wizard are mainly intended for people with prior experience with spreadsheet applications. All participants reported to have at least some experience: 5 of 8 participants use spreadsheet applications several times a month, 2 of them once a month or less often, 1 of them uses spreadsheet applications several times a week and 1 of them every (work) day.

- **How frequently do you use visualization tools?**
  We wanted to find out, how many participants have experience with the visualization tools. 5 of 8 participants use visualizations tools once a month or less often, 3 of them several times a month and 1 of them never uses visualization tools.

- **If you use visualization tools: which tools you have experience with and what do you use them for?**
  This question aimed to check the participant’s experiences with visual analysis tools. The answers for this question were:
  
  - Excel
  - Gnuplot
  - Google charts
o Visio
o SPSS
o Visual Paradigm

- **How frequently do you look up information on the web?**
  This question aimed to probe the level of the participant’s web experience. All participants indicated to look up information on the web every (work) day.

- **How often do you work with Linked (Open) Data?**
  This question was intended to find out if there were any Linked Data experts among the participants. Only 1 of 8 participants answered “Once a month or less often”, 6 of them answered “Never” and 1 of them answered “What’s Linked Open Data”.

- **How frequently do you write SPARQL queries?**
  This question was intended to find out if there were any Semantic Web experts among the participants. Only 2 of 8 participants answered “Once a month or less often”, 6 of them answered “Never”.

- **What’s your age?**
  The final background question provides information about the age range of the participants: 7 participants were between 28 and 37 and 1 of them between 18 and 27.

After the initial background questions, the participants had to complete six tasks using the CODE Visualization and CODE Query Wizard. These were the questions:

**Task 1: Filtering in the Query Wizard**
To begin the task, please click on the following link:
http://codev.knowcenter.tugraz.at/search#?endpoint=http%3A%2F%2Fcode.knowcenter.tugraz.at%3A8890%2Fsparql
- **Data:** This task deals with the dataset called “G:20-Countries: Co2 (Tons per Person), Life Expectancy, Population; 2000-2009”.
- **Task:** Please show the data set in Query Wizard. We are interested only in the countries which have a CO2 Emission over 13 Tons per persons. After that, please visualize the results.
- **Time:** You have 3 minutes to complete this task.

**Task 2: Filtering in the Visualization Wizard**
To begin the task, please click on the following link:
http://codev.knowcenter.tugraz.at/search#?endpoint=http%3A%2F%2Fcode.knowcenter.tugraz.at%3A8890%2Fsparql
- **Data:** This task deals with the same data as before, the dataset called "G:20-Countries: Co2 (Tons per Person), Life Expectancy, Population; 2000-2009".
- **Task:** We are still interested only in the countries which have a CO2 emission over 13 Tons per persons. Please solve this task by using one of the suggested visualizations. Use brushing. Which countries have CO2 emission over 13 Tons per persons?
- **Time:** You have 3 minutes to complete this task.
Task 3: Aggregation
To begin the task, please click on the following link:
http://codev.knowcenter.tugraz.at/search#?endpoint=http%3A%2F%2Fcode.knowcenter.tugraz.at%3A8890%2Fsparql
- Data: This task deals with the same data set as before: “G 20-Countries: Co2(Tons per Person), Life Expectancy, Population; 2000-2009”.
- Task: We are interested in the (average) life expectancy for each country.
  1) Visualize the data set
  2) Aggregate this dataset in such a way to obtain the (average) life expectancy for each country:
     o group by countries
     o select average as aggregation function for life expectancy
  3) Please show this data in a geo chart.
Which countries have the smallest and the largest life expectancy in G20 countries?
- Time: You have 5 minutes to complete this task

Task 4: Aggregation - Multiple Categories
To begin the task, please click on the following link:
http://codev.knowcenter.tugraz.at/search#?endpoint=http%3A%2F%2Fcode.knowcenter.tugraz.at%3A8890%2Fsparql
- Data: This task deals with the dataset called "Countries: Co2 (Tons per person), Life expectancy, Population".
- Task: We want to find out how the average rate of the CO2 emission developed over the years and for different regions.
  1) Visualize the data set.
  2) Please show, for different regions, development of the average rate of the CO2 emission over the years.
     o Use aggregation (tip: you can group over multiple categories)
     o Select the appropriate visualization (tip: change the mapping of Available Categories, if needed)
In which regions are CO2 emissions rising and where are they falling?
- Time: You have 10 minutes to complete this task.

Task 5: Aggregating Multiple Values
To begin the task, please click on the following link:
http://codev.knowcenter.tugraz.at/search#?endpoint=http%3A%2F%2Fcode.knowcenter.tugraz.at%3A8890%2Fsparql
- Data: This task deals with the same dataset called "Countries: Co2 (Tons per person), Life expectancy, Population".
- Task: We want to find out whether there is a correlation between CO2 emissions and life expectancy for regions.
  1) Please visualize the data set.
  2) Please find out whether there is a relation between average CO2 emissions and the average life expectancy.
Tip: you can add multiple aggregate values.
Do you see a correlation between CO2 emissions and life expectancy (for region)?

- Time: You have 12 minutes to complete this task.

**Task 6: Brushing in Multiple Views**

To begin the task, please go to prepared visualization

- Data: This task deals with the same dataset called "Countries: Co2 (Tons per person), Life expectancy, Population". You see visualizations showing the correlation between average CO2 emissions and the average life expectancy for countries
- Task: We want to find out on which continents the countries with
  - the lowest CO2 emissions and life expectancy are located?
  - the highest CO2 emissions and life expectancy are located?

  Use brushing

- Time: You have 5 minutes to complete this task.

After each task was finished (either the participants successfully completed the task or exceed the time limit) the participants filled out a NASA Task Load Index form which allows participants to subjectively judge the level of workload for each task. These were the questions consisted of the form:

- **Mental Demand**: How mentally demanding was the task?
- **Physical Demand**: How physically demanding was the task?
- **Temporal Demand**: How hurried or rushed was the pace of the task?
- **Performance**: How successful were you in accomplishing what you were asked to do?
- **Effort**: How hard did you have to work to accomplish your level of performance?
- **Frustration**: How insecure, discouraged, irritated, stressed, and annoyed were you?

The participants had also the possibility to give any comments about what was good/bad/unexpected/difficult during each task.

The final page of the form consisted of additionally questions that gave the participants the opportunity to provide more qualitative feedback. The following questions were asked:

- What did you like about the CODE Visualization Wizard?
- What did you dislike about the CODE Visualization Wizard?
- How did you like the visual design of the Visualization Wizard?
- Do you have suggestions for improving the design of the Visualization Wizard?
- Was the Visualization Wizard easy to use and interact with?
- Do you have suggestions for improving the usability of the Visualization Wizard?
- For which tasks would you personally use the Visualization Wizard?
- Comment shortly the CODE Query Wizard. What did you like or dislike, what would you use it for?
- If you could have solved any of the tasks with other tools of your choice, which ones would you have used?
- Was filtering (in the Query Wizard) hard or easy to use?
- Was aggregation hard or easy to use?
- Was brushing hard or easy to use?
- Did you find automatic generation of visualizations helpful?
- Was choosing and configuring visualization hard or easy to achieve?

The evaluation was concluded after the participant had answered this final set of questions.
4.3 Evaluation Results

The evaluation has been performed as a combination of the Retrospective Think Aloud protocol with the NASA Task Load Index. Based on this combination, we generated two different kinds of results of our evaluation.

- Quantitative results (NASA Task Load Index)
- Qualitative results (Retrospective Think Aloud)

The quantitative evaluation data is also available for download at http://code-research.eu/wp-content/uploads/2014/04/vis_wizard_eval_data.zip

Quantitative Results (NASA Task Load Index)

The quantitative results of the NASA Task Load Index are presented as box plots of Figure 11. The six plots represent the results for the six different aspects of the NASA Task Load Index: mental demand, physical demand, temporal demand, performance, effort, and frustration.

- The mental demand of the user was rather low for all the tasks. We detect that only by task four the mental demand increased slightly. However, the variance between the participants was quite high.
- The physical demand was constantly low for the first three tasks and increased slightly for the more complex last three tasks.
- The temporal demand remained generally low with a high degree of variance between the participants. From 54 tasks performed in total by 8 participants, 39 were successfully completed in time, 15 were not completed in time.
- The performance scores were generally high but again with a high degree of variance. From 8 participants and 54 tasks in total, the participants successfully completed 39 tasks, 13 tasks could be completed only with moderator’s intervention. 2 tasks could not be completed at all.
- The subjective effort of the participants was varied between the tasks depending on the learning effect. As the user received a task covering a new topic or new functionality, it demanded more effort from the user. Tasks similar to the previous one showed a comparably lower effort.
- The frustration level was rather low throughout the evaluation but with a high variance between the participants.
Qualitative Results (Retrospective Think Aloud)

The qualitative results of the evaluation represent the statements of the participants collected during the Retrospective Think Aloud phase of the evaluation.

Regarding task 1, the task description included a link to the list of the predefined RDF Data Cubes on the Query Wizard. To complete the task the participants should only select the requested dataset on the list. The main problem in this task was that 4 of 8 participants did not expect that the link would lead them directly to the list of predefined RDF Data Cubes. Hence 2 of these 4 participants generated a search in DBpedia and 2 of them browsed the other endpoints searching requested data. The moderator intervened in this case and referred user to the requested endpoint with the requested dataset. Once the dataset was selected, it was very easy for all participants to use the filter and visualize the filtered dataset finally. However, this issue demonstrates that selection of the data set (or of the endpoint to search in) is not trivial and should be given due attention in the UI design.

Regarding task 2, the main problem was that 6 of 8 participants set the filter on the Query Wizard before visualizing. Since the visualizations in the Visualization Wizard already showed the filtered data, it was not necessary to use the brush in Parallel coordinates. This situation caused confusion for the participants, since they did not know how and where they should use the brush to set a filter. Hence 1 of these 6 participants removed the filtering on Query Wizard visualized the original data, and used then the brushing on Parallel coordinates to complete the task. 5 of them concluded the task only after the moderator indicated them that the recently suggested visualizations already show the filtered dataset. In the end effect users completed the task successfully using means other than intended (filtering instead of brushing).

Nevertheless, it is clear that having different functionality (filtering, brushing) serving similar purposes may be confusing to new users.
Regarding task 3, the feedback was much more positive, the participants could use the aggregation function without any major problems. However, for the color encoding in the geo chart we have received a critical review from 7 of the 8 participants: we use red for the color encoding of the countries and depending on the value each country receive a nuance of red. In case of similar values the nuances are also similar to each other. In such a case participants can only recognize very high and very low values but not the values in between. Another interesting result for this task was that 1 of the 8 participants aggregated data using the Query Wizard before visualizing them. Since the user did not have much experience in visualization, the tabular representation was always chosen for transforming the data (e.g. filtering, aggregation) before visualizing them.

Regarding task 4, only 2 participants out of 8 could complete the task without any major issues. The fact that the appropriate visualization (line chart) was not initially suggested, but had to be selected by the user, caused confusion for almost all of the participants. Only 1 participant did select the line chart but tried to complete the task without changing the mapping combination, leading to a failure in completing the task. We learn from this that enabling meaningful visualizations and mappings is not enough, as users have major difficulties in choosing and configuring the appropriate visualization when several are available. Instead, the automatic visualization shall include intelligent ranking of visualizations (and mappings) to initially display the most appropriate visualization to the user. Also, 5 participants of 8 were overwhelmed at the beginning by having to group multiple categories, so they did not group by year. As they noticed the problem they performed the aggregation function again (grouped by year and region) but were confronted with the first issue we mentioned above. These 6 participants complete the task only with the moderator’s intervention. The moderator supported them by selecting the relevant visualization (line chart) and also by setting the relevant mapping combination.

Regarding task 5, we could nicely observe the learning effects. The participants could gather experiences with previous tasks so 5 participants out of 8 could complete the task without any intervention and issues. 3 of them used Parallel Coordinates and 2 selected the Scatterplot for this task. Using multiple aggregated values for this task caused a major problem only for 1 participant. This participant needed a while to detect the missing aggregated value. 2 participants grouped by country instead of region.

Task 6 dealt with the coordinated brushing between Scatterplot chart and Geo chart. 2 participants were confused by the fact that setting the brush interaction in the Scatterplot differs from the one in the Geo chart (the former uses a rectangular brush the later a selection brush through mouse clicks on the countries). They could handle with the coordinated brushing between these two visualizations after moderator explained the differently brushing technique. 1 participant criticized the brushing in the Geo chart which is limited on only one country, and also the non-existent interaction between the legend and the displayed entries on the Geo chart. 1 participant suggested having the opportunity to limit the number of displayed data on Scatterplot chart without the switching on the Query Wizard. We conclude that in coordinated multiple view settings, especially such where the number of available visualizations is large (10 in our case), unified interaction mechanisms for brushing should be provided.

As we asked the participant what they liked about the Visualization Wizard, their general opinion was that the Wizard was easy to use, intuitive, fast and well designed. Additionally they liked the spectrum of available visualizations, the interaction between the visualization and the automated suggestion of meaningful visualizations. As we asked the participants what they did not like about the Visualization Wizard, the most common answers were the color encoding of the Geo chart and the
different brushing techniques for the visualizations. Another issue is the, for performance reasons, the Wizard creates visualizations for 250 data items at the most. When aggregating a larger data set, the user ends up with a wider data range, since original data items are combined together (average/summed/etc.). 1 participant realized the problem when the data range between the original and aggregated RDF Data Cube differed. As future work we want to optimize our mapping algorithm and increase the performance in order to handle large data sets.

Some participants delivered suggestions for improving the Visualization Wizard, for example:

- using rainbow colors for the Geo chart,
- adding additional information to the visualizations to inform user about the visualization itself (for which kind of data is this visualization more suitable) and also about interaction opportunities
- using mouse over tool tips
- supporting other formats than Linked Open Data

As we asked the participants for which tasks they would personally use the Visualization Wizard, the answers were:

- For visualization of any kind of statistical data
- Server log analysis
- Project tracking
- To quickly answer questions which involve data analysis

Finally we asked participants if they could have solved any of the tasks with other tools of their choice, and if yes which ones would they have used. Almost every participant mentioned that they would use Google for the search, and manually collect and copy the data into any spreadsheet application for the visual analysis.
5 Conclusion

This report covered the usability evaluation of the CODE Query Wizard and the CODE Visualization Wizard, two main prototypes of the CODE WP4.

The results of the **CODE Query Wizard** study showed that the tool was generally very usable, both for users with and without a computer science background. The main point of critiques was a possibility to add URI filters through a menu in the table header was missing. Additional suggestions for improvements were to show the total number of results more prominently, and to implementation an “infinite scroll” mechanism that automatically load and displays more data as the users scroll to the bottom of the screen. The general feedback about the **CODE Visualization Wizard** was that the prototype worked well, was generally easy to use, and that the users liked the interactive analysis possibilities offered by the charts. The main point of critiques was the fact that it was not always clear to the users that they were working with a filtered dataset when they had previously set a filter through the CODE Query Wizard. Also, a number of people did not immediately understand why, for a given dataset, certain charts were suggested and others were disabled. Once these users understood the underlying mechanisms, they appreciated the time and effort that the feature had saved them.

For future exploitation plans, the results of the conducted usability studies provide valuable starting points for usability improvement as well as for adding additional features. We will continue research on selected technologies from the two prototypes, such as the automatic suggestion and generation of visualisations, in other EU projects.
6 References

[Hoefler et al. 2013]

[Hoefler 2013]
Patrick Hoefler (2013).
Linked Data Interfaces for Non-expert Users.
The Semantic Web: Semantics and Big Data. 10th International Conference, ESWC 2013, LNCS 7882, P. Cimiano et al. (ed.), Montpellier, France, May 2013.

[Hoefler et al. 2014]
Linked Data Query Wizard: A Novel Interface for Accessing SPARQL Endpoints.

[Mutlu et al. 2013]
Automated Visualization Support for Linked Research Data.
Proceedings of i-SEMANICS 2013 Conference, Graz, Austria, September 2013.

[Mutlu et al. 2014]
Suggesting Visualisations for Published Data.