D5.5 – Innovative work methodologies

Due date of deliverable: 31 July 2013
Actual submission date: 26 July 2013
Responsible Partner: CTI
Contributing Partners: CTI, UPM, IMA, BRF, PUB, UOL

Nature:

☒ Report ☐ Prototype ☐ Demonstrator ☐ Other

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)

Keyword List: Innovation, work methodologies, use cases, Dicode Workbench, integrated services, CommBAT, Pinta.

The Dicode project (dicode-project.eu) is funded by the European Commission, Information Society and Media Directorate General, under the FP7 Cooperation programme (ICT/SO4.3: Intelligent Information Management).
The Dicode Consortium

Computer Technology Institute & Press “Diophantus” (CTI) (coordinator), Greece

University of Leeds (UOL), UK

Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e.V. (FHG), Germany

Universidad Politécnica De Madrid (UPM), Spain

Neofonie GmbH (NEO), Germany

Image Analysis Limited (IMA), UK

Biomedical Research Foundation, Academy of Athens (BRF), Greece

Publicis Frankfurt Zweigniederlassung der PWW GmbH (PUB), Germany
Deliverable manager
- Nikos Karacapilidis (CTI)

List of Contributors
- Nikos Karacapilidis, Spyros Christodoulou, Manolis Tzagarakis (CTI)
- Guillermo de la Calle Velasco (UPM)
- Georgia Tsiliki, Sophia Kossida (BRF)
- Mark Hinton (IMA)
- Ralf Löffler (PUB)
- Fan Yang-Turner, Dhaval Thakker, Vania Dimitrova, Lydia Lau (UOL)

List of Evaluators
- Guillermo de la Calle Velasco (UPM)
- Lydia Lau (UOL)

Summary
This deliverable provides a summary of innovative work methodologies shaped in the context of the Dicode project. Exploiting the feedback obtained from the validation of Dicode’s suite of services in the three use cases of the project, we present a series of real scenarios about the use of Dicode solutions, pointing out efficient and cost-effective work methodologies for data-intensive and cognitively complex collaboration and decision making settings. These methodologies reflect our experiences gained from the validation of the project’s results and provide useful suggestions and insights to relevant communities and organizations.
Table of Contents

1 Introduction ........................................................................................................................................... 5

2 Use Case 1: Clinico-Genomic Research Assimilator ................................................................. 5
   2.1 Working with the Dicode platform.............................................................................................. 5
   2.2 Innovative work methodologies and insights.............................................................................. 8

3 Use Case 2: Trial of Clinical Treatment Effects .......................................................................... 10
   3.1 Working with the Dicode platform.......................................................................................... 10
   3.2 Innovative work methodologies and insights ........................................................................... 12

4 Use Case 3: Opinion mining from unstructured Web 2.0 Data .............................................. 13
   4.1 Working with the Dicode platform.......................................................................................... 13
   4.2 Innovative work methodologies and insights ........................................................................... 18

5 The Dicode Workbench .................................................................................................................. 20
   5.1 Introduction.............................................................................................................................. 20
   5.2 Description of the scenario..................................................................................................... 21
   5.3 Innovative functionalities......................................................................................................... 23

6 CommBAT: Community Behaviour Analytic Tool ..................................................................... 25
   6.1 Context for Use ....................................................................................................................... 25
   6.2 Using CommBAT ................................................................................................................... 26
   6.3 Innovation of CommBAT......................................................................................................... 28

7 Pinta: a Tool for Exploratory Search over Linked Data .............................................................. 29
   7.1 Context for Use ....................................................................................................................... 29
   7.2 User interaction with Pinta to perform exploratory search tasks ........................................ 30
   7.3 Key Innovation: Semantic Signposting in Semantic Data Browsers ................................... 32

8 Conclusions ....................................................................................................................................... 32

9 References ......................................................................................................................................... 34
1 Introduction

This deliverable reports on the outcomes of Task 5.5: “Development of innovative work methodologies”. This task exploited the feedback obtained from the validation of Dicode’s suite of services in the three use cases of the project, and elaborated a series of real usage scenarios of the foreseen solutions. The aim is to produce advanced, efficient and cost-effective work methodologies for the problems and settings under consideration. These methodologies take into account the nature and needs of contemporary organisations and communities. They reflect our experiences gained from the validation of the project’s results and provide useful suggestions and insights to relevant communities and organizations. To increase the impact of the Dicode project, such methodologies and related conclusions obtained have been communicated in relevant conferences and high impact journals on the Dicode’s influence areas (a detailed report will appear in deliverable D7.2.2: Dissemination and Exploitation Activities Report - second version, due in August 2013).

The deliverable is structured as follows: As stated in the project’s DoW (Description of Work), Sections 2-4 present the proposed work methodologies through real scenarios from the project’s use cases. In addition, Section 5 provides a scenario of exploiting the Dicode Workbench in a different setting. Finally, Sections 6 and 7 discuss work methodologies related to two stand-alone tools developed in the context of Dicode, namely CommBAT and Pinta, respectively. These tools can be used in parallel to the Dicode’s solutions developed for the three use cases and further augment the proposed work methodologies. The second part of each of the abovementioned sections (i.e. Sections 2-7) focuses on the innovative features of the proposed methodologies and provides insights about their use in diverse data-intensive and/or cognitively complex settings.

2 Use Case 1: Clinico-Genomic Research Assimilator

2.1 Working with the Dicode platform

In this section, we describe a typical multi-tasking biomedical environment, where users need to work on diverse tasks and interpret their research outcome. This is one of the main challenges the biomedical community is facing, due to the ever increasing genomic and transcriptomic data. Our working hypothesis is that using the Dicode platform to discuss and analyze biomedical data leads to more productive and effective work practices compared to current practices (see also deliverable D6.2.2).

Consider a real working scenario for a group of researchers in the clinico-genomics field. As part of their everyday practice, they want to analyze two sets of diverse data and biologically interpret their findings. They would also like to discuss with their peers the workflow of their analysis, the methodologies applied and the results produced. As broadly admitted, collaboration among scientists with different scientific background and experience adds value to their research.

The presented scenario emphasizes on the way researchers collaborate in this field, where research is carried out by multidisciplinary teams consisting of biologists, medical doctors, clinical researchers and statisticians, each of whom contributes from his/her perspective to the problem being discussed. To better illustrate the use of the proposed solution, and in particular how the available functionalities can be used in the biomedical context, we present
a walkthrough of the above scenario with additional details for the discussions and the analysis details that take place.

Figure 2.1 shows an instance of the collaboration workspace (provided by the Dicode Collaboration Service – see deliverables D4.1.3 and D5.4.3) operated in the ‘mind-map view’, where a team of three researchers, namely Alice, Neal and Jim, are discussing an issue related to breast cancer research. In particular, they are collaborating in order to determine how to augment existing datasets to further study how Tamoxifen (Tam) resistant cells modulate global gene expression. Tam is a widely used antagonist of the estrogen receptor, whereas its resistance is a well-known obstacle to successful breast cancer treatment (Huber-Keener et al., 2012). While adjuvant therapy with Tam has been shown to significantly decrease the rate of disease recurrence and mortality, recurrent disease occurs in one third of patients treated with Tam within 5 years of therapy.

Alice, who is a Pharmacology PhD student, initially selected and analyzed gene-expression data from 300 patient samples with the help of Neal, an MD at a collaborating university hospital. These data are derived from whole human genome expression arrays (Affy U133A Plus 2.0, see http://www.affymetrix.com). Although the sample is relatively large, Neal and Jim, a postdoctoral researcher in Bioinformatics, believe that augmenting the data with publicly available data will be a good idea for obtaining statistically significant results. However, the team needs to decide on the sample size and the nature and characteristics of the new data they would incorporate. A practical solution is offered by the Dicode GEORecommender Service (see deliverable D5.4.3), which facilitates researchers to find data sets that resemble their data, given some user specified criteria.

Figure 2.1: Collaboration Workspace (‘mind-map view’) showing collaboration between biomedical researchers. Service items (d) and (e) have finished their execution and are associated with other collaboration items.
All participating researchers may upload items into the collaboration workspace to express their opinion on the issue being discussed. In the instance shown in Figure 2.1, they have uploaded items of type ‘idea’ to propose additional data sets (“Consider also Next Generation Sequencing (NGS) data” (Figure 2.1 - (a)) and “Work with gene-expression data (Figure 2.1 - (b))). Participants have responded to the proposed alternatives (ideas) by uploading items and connecting them via arrows to other items to which they refer. Participants may also change an arrow’s colour to indicate the semantics of the relationship: green-coloured arrows express arguments in favour, red-coloured arrows express arguments against, while grey-coloured arrows indicate neutrality. Furthermore, they can aggregate items on the workspace by drawing coloured rectangles around them and give a title to the groupings. For example, the orange-coloured rectangle with title “Supplementary information” (Figure 2.1 - (c)) groups together bibliographic resources the team has obtained from external repositories and are relevant to their research. Supplementary information may include clinical information, and general information concerning the biological background and the technical difficulties relative to the production or analysis of the data.

As the discussion evolves, the team thinks about exploiting the Subgroup Discovery (SD) data mining algorithm (Atzmueller et al., 2005) using both data sets, microarray gene-expression and NGS RNA-Seq gene-expression, as input. In that way, they will be able to compare the results obtained from the two data sets, and thus argue on the importance of data augmentation. The SD algorithm has been exploited by (and offered through) Dicode Subgroup Discovery Service (see deliverables D3.2.2 and D5.4.3).

SD is the task of finding patterns which describe subsets of a data set highly correlated relative to a target attribute. This is a popular approach for identifying interesting patterns in the data, since it combines a sound statistical methodology with an understandable representation of patterns. For example, in a group of patients that did or did not respond to specific treatment, an interesting subgroup may be that patients who are older than 60 years and do not suffer from high blood pressure respond much better to the treatment than the average.

To invoke the SD algorithm on the NGS data, they upload the associated service item into the workspace (Figure 2.1 - (e)) and start configuring the Dicode Subgroup Discovery service. Configuring the service includes the specification of the URI for the REST-based SD service and specification of parameters such as input file, number of rules to be used, service ontology, and minimum number of subgroups to be retrieved. After configuring the service, they trigger its execution. As long as the SD service is executing, the icon representing the service appears with a green colour. Upon successful termination of the SD service, the icon changes its colour to orange and the results are automatically uploaded into the collaboration workspace (Figure 2.1 - (f)). To clearly indicate the execution of the SD service on the gene expression data and the results it returned, the team groups together the relevant items and supplies a descriptive title (“Apply data mining to NGS data”). The team can now assess the output of the SD execution by commenting on the results and connecting them to other items in the collaboration workspace (Figure 2.1 - (g)).

The team can follow the same procedure (invoking the SD service and collectively assessing its output) for the gene expression data. The three researchers are able to carefully examine the commonalities between the two SD runs (on gene expression and NGS data) and share their insights. Alice notes that there is a notable agreement between the two datasets. Their findings can be further explored by launching the Dicode PubMed service (see deliverables D5.4.2 and D5.4.3) in order to search for relevant scientific findings.
As the collaboration continues and more items are added to the collaboration space, the team decides to switch to a different view, in order to reach a decision. For this, they decide to convert the ‘mind-map view’ into the ‘formal view’, which provides elaborated scoring and reasoning mechanisms that further facilitate the decision making process. By transforming the ‘mind-map view’ of the collaboration workspace, all available semantic types – including the service items - are transformed into the appropriate types of the ‘formal view’, based on well-specified rules.

Figure 2.2 shows the collaboration space in the ‘formal view’. As noted above, the team can continue the collaboration in this view by adding more items (each time a new item is added, the reasoning mechanism is triggered). Furthermore, it allows the team to see which is the best argumented alternative solution (or ‘winning’ solution) by highlighting it using visual cues. Based on the current state of the collaboration, the currently ‘winning’ solution is the alternative “Consider also NGS data” (item in blue underlined font colour in Figure 2.2).

The above scenario is fully demonstrated via two video-casts:
- http://dicodedev.cti.gr/screencast/screencast.html

2.2 Innovative work methodologies and insights

As shown in the above walkthrough scenario, the Dicode platform offers the following innovative solutions to scientists working in the field of clinico-genomics research (related publications include (Karacapilidis et al., 2013), (Karacapilidis et al., 2012) and (Tsiliki et al., 2012)):
- the means to easily analyze large and diverse data derived by microarray and sequencing technologies (via the Subgroup Discovery service),
• the means to explore, categorize and access, based on user specified criteria, the vast amount of genomic and transcriptomic data now available (via the data-mining GEORecommender service),
• the means to explore PubMed published scientific articles based on user specified criteria.

The above tailor-made data/text-mining solutions, have addressed the subject of intense ongoing research (Tyekucheva et al., 2011), i.e. the similarity between microarray and sequencing technologies and how such data can be integrated to augment data analysis results and reinforce the biological outcome. This is indicative of the capabilities offered via the Dicode platform to the biomedical community, where the main issue involves handling vast amount of diverse data. Additionally, the Dicode Collaboration and Decision-Making services presented offer an innovative work methodology for clinico-genomics research in that:

• large scale multi-institutional and interdisciplinary collaboration are supported and monitored via the sharing and analysis of datasets,
• researchers with different backgrounds can easily exchange and organize their ideas, experiences and supporting data, being able to get quickly an insight of the evolution and the current status of the overall process,
• data and decision provenance are effortlessly preserved,
• successful practices can be stored and reused, since the argumentation of participants (and logic underneath) is always available for future reference (researchers can re-launch and revise past conversations),
• successful practices (methodologies or workflows) can be easily adapted by other researchers in similar disciplines,
• discussions can be accompanied by scientific ‘proofs of action’, i.e. users can exploit the set of Dicode services to load graphs, tables, files and any type of analysis outcomes (as these derive from Dicode services) to augment their arguments and further reinforce an ongoing collaboration.

The Dicode solution has great potential for adaptation and application to professional users from a wide range of different organisations. As reported in the two rounds of evaluation of Use Case #1 (see deliverables D6.2.1 and D6.2.2), evaluators from different backgrounds in the biomedical domain, found the Dicode workbench interface intuitive with well integrated services. Although they were sceptical about adopting new practices, they appeared less sceptical concerning the ability of the Dicode Workbench integrated services to deal with cognitive-complex issues, to enhance collaboration between their peers and in that respect assist exchanging of information and advice. The collaboration and decision-making options of Dicode were highly marked; special mentions include the data management mechanism, and the different manners of sharing or discussing data and results. Furthermore, evaluators reported that the platform offers ease of communication, and strong data/information archiving features.

Based on the above feedback, we are optimistic that the overall Dicode solution can significantly inspire future research. A straightforward direction concerns specialization of the Dicode services and/or extension of their functionalities to other data sets/types (e.g. developing a more generic version of SD service to also handle proteomics data). The Dicode project was always aiming to develop generic services, which would be easily adopted form professionals of different biological backgrounds; thus, we believe that given the nature of the data, the Dicode services could be easily adopted by other related application domains, such as genetics, biotechnology, developmental biology, proteomics, and molecular biology.
3 Use Case 2: Trial of Clinical Treatment Effects

3.1 Working with the Dicode platform

In this section, we describe a typical multi-centre clinical trial environment, where users are in multiple sites spread across the globe. The management of clinical trials often requires a great deal of coordination between sites, reviewers and the operations team to ensure smooth running of the study and quality. This produces a complex system where there are risks of miscommunications and inefficiencies in reporting issues or problems that occur in the process of the trial. We therefore envision Dicode as the central platform for effective communication and collaboration between involved parties of the study.

Consider a real working scenario of an imaging clinical trial where study participants (patients) are enrolled in various sites and treatment effects are determined through MRI scans at baseline and follow up visits. Site coordinators designated for each site ensure quality of scans and anonymisation of patient information before it is transferred to a central database of images. Once images are in the central database, they are assessed for treatment effects by a reviewer who will query this database for images. The operations team helps track data movement and also ensures quality assurance of data. Any problems faced by the sites and reviewer need to be promptly reported and communicated to the operations team and involved parties for issues to be solved quickly.

![Diagram of interaction between relevant parties during a clinical trial]

The presented scenario (Figure 3.1) emphasizes on the way clinical trials are run, where a multidisciplinary team works together. To better illustrate the use of the proposed solution, and in particular how the available functionalities can be used in the clinical trial context, we present a walkthrough of the above scenario with additional details for the discussions and the analysis details that take place.
Figure 3.2 shows an instance of the collaboration workspace operated in the ‘mind-map view’, where a trial manager, radiographer, radiologist and IMA representative are working towards solving an issue. In particular, they are collaborating in order to determine how to improve the quality of scans. A radiologist analysing images gathered as part of a clinical trial spots some problems with the data. Out of the 500 scans he has analysed, 100 were not of sufficient quality to obtain any conclusive results. He would like to discuss this with the radiographer. Also present during the discussion is a representative from IMA who knows how data should be acquired to work best in Dynamika. It is hoped that the discussion can reveal the root cause of the problem and enable it to be corrected before any more poor data is acquired.

![Collaboration Workspace](image)

**Figure 3.2:** Mind map view of collaboration between radiographer, radiologists, trial manager and IMA representative on Dicode. The two highlighted components consist of differing opinions on whether (b) contrast injection before or (a) during the scan is the better solution for improving scan quality.

Participants start the collaboration by uploading items into the collaboration workspace. As shown in the collaboration instance of Figure 3.2, they have uploaded items of type ‘idea’ to propose injecting contrast during the scan (Figure 3.2 – (a)) and before the scan (Figure 3.2 – (b)). They have also responded to the proposed ideas by uploading their positions and connecting them with items to which they refer. Similarly to what is described in Section 2.1, participants indicate the semantics of the relationship through the colouring of the connecting arrows. Moreover, they aggregate items on the workspace by drawing appropriate coloured rectangles. For example, the purple-coloured rectangle with title “Contrast injection during scan” (Figure 3.2 – (b)) groups together relevant opinions and resources relating to the implementation of an idea expressed so far. Supplementary information may include clinical information, external files and technical difficulties.

At a certain collaboration instance, participants decide to switch to the ‘formal view’ (they want to exploit the provided decision making functionalities – see Figure 3.3). Based on the
current state of the collaboration, the currently ‘winning’ solution is the alternative “If you use a power injector, the patient could be injected automatically during the scan” (item in blue underlined font colour in Figure 3.3).

Figure 3.3: Formal view of mind map workspace where the Dicode platform provides elaborated scoring and reasoning mechanisms which further facilitate decision making.

The above scenario is included in the video-cast appearing at: http://dicodedev.cti.gr/screencast/UC2_screencast/UC2.html.

3.2 Innovative work methodologies and insights

As revealed from the above scenario, the Dicode platform offers the following innovative solutions to a multi-disciplinary team working together in a clinical trial:

- the means to easily and efficiently communicate issues centrally,
- the means for effective brainstorming and decision making,
- the means for easily identifying the most favoured solution through grouping of ideas.

The above solution will improve communication and efficiency in problem solving during a study. This is indicative of the capabilities offered via the Dicode platform to the clinical trial community, where the main issue involves miscommunications and delayed response to problems. Additionally, the Dicode Collaboration and Decision-Making services presented an innovative work methodology in that:

- multi-centre collaboration, based on rigid protocols such as those followed in clinical trials, can be easily supported,
- multi-disciplinary clinical trial teams can easily follow and advance the evolution of the overall process,
- clinical trial teams may consult (and build on) past collaboration sessions that led to well-validated practices,
- clinical trial teams can easily exploit the Dicode solution to accompany an ongoing collaboration with the required pieces of evidence.

Similar to what reported for Use Case 1 (see Section 2.2), people involved in the evaluation of the Dicode solution for Use Case 2 have acknowledged its value and potential (see deliverables D6.3.1 and D6.3.2). Although they were reluctant about adopting new practices, which is explained by the conservative protocols that are usually followed in the clinical trials field, they recognized the value of the Dicode solution as far as augmenting collaboration, sense-making and decision making are concerned. Users saw this as an
advantage because it is not time or location specific and it was praised for being able to track written conversations and ideas which usually can get lost during a verbal meeting.

Based on the above feedback, we are optimistic that the overall Dicode solution can positively impact the running of future clinical trials. Because of the project’s aim to connect people easily and facilitate collaborative work, we see Dicode’s applicability not only in the clinical trial setting but generally in any type of scenario where an organization or a team requires quick and easy to use tools for idea generation, argumentation and collaborative decision making.

4 Use Case 3: Opinion mining from unstructured Web 2.0 Data

4.1 Working with the Dicode platform

In this section, we present two scenarios about the use of Dicode Workbench and integrated services for the analysis of the voluminous amount of unstructured information existing on the Web, especially in the highly dynamic social media space. We show how different parties from different locations can effectively work together using the abovementioned Dicode solutions. Three people, logged in the Dicode Workbench, are involved in this use case: Jan (Brand Manager Mercedes Benz, Stuttgart), Alice (Social Media Analyst, Hamburg), and Julia (Social Media Engager, Berlin). The Dicode services incorporated in these scenarios are:

- Collaboration Service;
- Top Entities Service;
- Prominence Graph Service;
- Phrase Extraction Service;
- Topic Detection Service.

First scenario: Opinion mining from unstructured Web 2.0 Data

The scenario provides insights on the following questions?

- What people do to analyse the Web 2.0? How do they do it and why?
- How do people work together from different places?

Jan is the Brand Manager of Mercedes-Benz. Once a week, he is using the Top Entities Service to check some interesting websites for relevant buzz (see Figure 4.1). He discovers heavy discussions for Mercedes Benz in a specific time frame and wants to find out more about them (what is the reason for that buzz?).

Jan opens a new ‘Collaborative Workspace’ (offered through the Dicode Workbench from the Dicode’s Collaboration Service) and contacts Alice. She is asked to investigate where the buzz comes from. Alice is reading Jan’s message in this Collaborative Workspace. She sends Jan a short reply and
Figure 4.2: Using the Prominence Graph Service.

First, Alice uses Prominence Graph Service to identify peaks within a specific keyword and timeframe. There are two interesting peaks for the keyword Mercedes Benz (see Figure 4.2). She analyses the first one.

By clicking on the peak, Alice automatically gets forwarded to the Google results for this specific day. By checking the links, she notices a lot of entries related to the Mercedes Benz Fashion Week. She analyzes a few of them and sees a strong connection to the Fashion Week. Then, she double-checks with the peaks for the main competitors: Audi and BMW (just to exclude the case that the peak is generated from the whole set). She observes that it is a Mercedes Benz Peak. Finally, she writes a message to Jan saying that she found out the reason for the buzz and recommends extending the collaboration with the Fashion Week. Then, she connects her results with Jan’s initial message.

Jan checks the message on the Collaborative Workspace and is very impressed about the results. He really likes the idea to boost the collaboration with the Fashion Week and he writes a comment on the Workspace to Alice and Julia. Julia is the social media engager and she should work out the social media strategy for the collaboration. Again, Jan connects his comment with Alice comment on the workspace. Julia sees the conversation, writes a short note to Jan and starts working on the project.

Second scenario: Deep analysis of complex texts

This scenario provides insights about how the Dicode services can help one to deeply analyze complex texts and collaboratively make decisions.

Jan wants to know how the newly launched A-Class is perceived and discussed in the web to adjust communications. Alice (Social Media Analyst) collects car reviews from the “autobild” as a starting point for her analysis (online car magazine). She arranges the texts in directories which are named after the cars that are reviewed. Then, she zips directories into a single file. Alice uses the Topic Detection Service to quickly understand what the conversations are about and what sub-topics drive the conversations.

Through the Dicode Workbench, Alice executes the service. She then chooses to do a fine-grained analysis (with many topics). She wants to get the graphical results display in a PDF file. She specifies the text language: German. Alice also wants to see the car types in the resulting graph: therefore she tags the “display text categories” button and indicates that she wants to see only the 10 most prominent car types. Alice does not want to see all topics of the fine-grained analysis. She decides to restrict the output graph to display only the 20 most prominent topics (see Figure 4.3). Finally, she triggers the execution of the service.
The analysis uncovers the Audi A3 as a close competitor (see Figure 4.4). In addition, three very competitive topics can be identified (see Figure 4.5).
Alice uploads the two review texts. She wants to detect only positive emotions in the texts. She indicates German as text language. She chooses predefined emotional analyser for automotive texts (see Figure 4.6). Alice also wants to see the two texts with the positive phrases highlighted (Figure 4.7). She can also see a tag cloud that is generated from the positive phrases in the two texts (Figure 4.8).

Figure 4.5: Output of the Topic Detection Service (competitive topics identified).

Figure 4.6: Working with the Phrase Extraction Application.
Figure 4.7: Car reviews with positive phrases highlighted.

Figure 4.8: The Tag Cloud shows the identified positive phrases in the reviews (A-Class and Audi A3).

The results are discussed with the whole team. If the results are not satisfying in terms of relevance or they are not plausible, the model can be improved by giving further input. After seeing the results, Jan gets an idea and contacts all Team-Members at the Collaborative Workplace. He decides to push the new A-Class at the next Mercedes-Benz Fashion Week.

The above scenarios are included in the video-cast appearing at: http://dicodedev.cti.gr/screencast/UC3_screencast/UC3_Opinion_Mining.mov.
4.2 Innovative work methodologies and insights

“Innovation is more than having new ideas: it includes the process of successfully introducing them or making things happen in a new way. It turns ideas into useful, practicable and commercial products or services.” - John Adair

Social-Media-Monitoring (SMM) is a software-based observation of the Internet. By now, it is used by many companies to observe what the Internet “thinks” because target groups talk on the Internet and influence the perception of things. SMM tools collect data for such an analysis. Not using SMM means to blind out a whole set of information. These tools are based on different algorithms. A lot of companies already use such tools. However, a lot of them are not happy with the results and stop working with them. Of course, there is a reason for this dissatisfaction.

Why is Social-Media-Monitoring not only important, but even indispensable?

The Internet is an overcomplex, fast-growing and fast-changing space. It is not possible to overview even a small fraction without any software tools. That is why SMM is necessary to capture big data. There is no other valid possibility. Besides the quantitative data pooling, the tools can capture data with a very specific theme in a specific timeframe. To find out relevant aspects about a theme the Internet is absolutely essential – the first thing people do to be informed about something is to “ask” the Internet – it is THE opinion-maker. Influencers are the most relevant group in terms of evaluation and more important than ever. That is why an observation of them is very needful when it comes to research. SMM tools are able to find the influencers on a quantitative basis. Finding influencers is one of the most difficult parts of the influencer analysis.

The other not less important aspect is the opinion-leader-change by the use of the Internet. This year, the Grimme Online Award in Germany was awarded for the very first time to a Twitter-Hashtag – one word, spread out by one person was caught up by thousands of people on the Internet and even influenced the TV political talk-shows. The Internet is the place where you get genuine opinions from real people without the problem of distorted actions. The Internet is a transparent room of opinion-clusters, which directly concern the perception of things: brands, decisions, news, politicians etc.

Biggest challenges

There are two really big challenges regarding SMM. The tools cannot differentiate same words with different content. An example: A Social Media Analyst wants to find out something about the brand “dove” – most tools are not able to differentiate now between the animal “dove” and the brand “dove”, so one gets a bunch of unnecessary data.

The other even bigger challenge is to evaluate the sentiments. Much of the software is not able to identify positive and/or negative and/or neutral opinions. They collect data, but are unable to show more than rough numbers. This problem is fundamental because strategy is consumer-based. If the analysis of the customers and their opinion on the Internet is wrong then every step after the analysis is going wrong. For example, irony or sarcasm are not easily recognizable by the software. Besides, the software is confronted with a range of different languages in different forums. What is irony on a political forum may be very serious on a cooking platform. As long as the sentiment analyses doesn’t work, the whole SMM makes no sense.
Why Dicode?

Quality
A lot of SMM tools have one big problem: they only catch words and collect data. Dicode is able to differentiate between product and category. The Entity Prominence Service can search only for the brand “dove” without getting distortive data about the white bird. This very plausible aspect seems easy but is groundbreaking in the field of SMM. The Entity Prominence Service saves a lot of time and money because no one has to eliminate useless data.

Ordering
Another big problem by just catching up texts is the amount of work by clustering them. The Topic Detection Service of Dicode analyses clusters on its own. The service does the quantitative part of clustering while you need to name the clusters manually. The Topic Detection Service is based on factor analysis, which means that Dicode discovers unstructured variables and finds the common variables. This service enables a meta-analysis, in that it finds the topic behind the topic.

Differentiation & Autodidactics
While other tools review the collected terms based on predefined lists (i.e. the terms are not assessed in their context), Dicode really learns from what the people teach the software. This means that the service is able to understand differences that are dependent, for instance, on product or category. An example: the language in forums about cars is very different from the language in forums about cosmetics. The same sentence means probably something completely different.

The Phrase Extraction Service can learn to discern. The learning is accomplished by humans. So every company or every person working with Dicode can teach the software exactly the way it is useful for the specific needs. This is a big benefit for an effective and efficient work. The human trains the machine and the machine only collects the relevant data. And on the other hand, the people have a much bigger control about the data capturing; they really know what the software collects and why.

Linkage
Dicode not only shows the results, but even gives you the direct link to them. When one observes a peak on a specific timeframe in the Prominence Graph Service, she can directly click on the peak for a direct forwarding to the relevant websites. The service does not only refer to itself like most of the other SMM tools do, but incorporates Google and its algorithm for searching results.

Connection between software and humans
Many other SMM tools are a lot less flexible and do not give this big opportunity of teamwork between humans and software. Interaction between the software and the people is very important for the examination of software results. People really need to work with the results and the control variable. Otherwise, there is an amount of data but the data will never influence the strategy of the company.

Collaborative Work
Dicode even offers a Collaborative Workspace, where all involved people can work together, share the results, contact others, etc. In this way, people are informed very quickly about the status of the process. Ideas can be shared and tasks can be divided without being even at the
same physical place. This Dicode solution opens a new dimension of working together without being together. It is highly important in cooperative agencies and companies.

5 The Dicode Workbench

5.1 Introduction

The Dicode Workbench is a key component which aimed to integrate all services developed in the Dicode project. Our main objective was to produce a working environment where users could work collaboratively using multiple and heterogeneous applications at a time, under a common interface. Requirements for the Dicode Workbench regarding its functionalities, interface, usability and integration of services have evolved along the project. By the end of the project, a typical working space is as shown in Figure 5.1.

![Current interface of the working space in the Dicode workbench.](image)

From the beginning of the project, the Dicode Workbench has been envisioned as a web application based on widgets since current trends in software development are focused on web interfaces. The purpose of widgets is to manage the visualization of the services integrated in the Dicode Workbench. Two integration strategies were defined to integrate services in the Dicode Workbench, enabling service providers to freely adopt any of them according to their needs. For further details on all these issues, see the series of deliverables D5.4.X (i.e. D5.4.1, D5.4.2 and D5.4.3).

The Dicode Workbench has been thoroughly validated through the three use cases of the project. Different workspaces have been configured to demonstrate and simulate real working scenarios (see deliverables D6.2.2, D6.3.2 and D6.4.2). In these workspaces, users were able to use the services specifically developed in the Dicode project for each use case. The Dicode workbench was designed to allow service providers to integrate dynamically services in an easy manner. This feature enables the use of the Dicode Workbench not only in
the context of the Dicode’s use cases; it can facilitate and augment work performed in other (data-intensive and/or cognitively specific) domains and scenarios.

In the next subsection, we present a new scenario about a software development project, where the Dicode Workbench is used by the developers to implement a new product. This scenario describes how the Dicode Workbench could be used by users in a complete different environment from the project’s use cases.

5.2 Description of the scenario

Company XYZ is a young and novel software company dedicated to the production of ad-hoc web solutions. Its philosophy is to always apply the latest and the most prominent technologies to provide the best solutions to their clients. This philosophy is not only applied for final products but also in the development process. The development team consists of young analysts, designers and programmers who combine work at the office with tele-working from anywhere (home or while travelling to a meeting or conference).

Developing a software product is a complex process that requires multiple discussions and agreements between analysts, designers and programmers (for instance, on the architecture of the system, the best tools and technologies to solve a problem, the user interfaces of the application, etc.). Regarding to the programming work, it also requires a strong collaborative work between developers to write code, validate functionalities and test interfaces. Since the development team work sometimes remotely and sometimes at the office, collaboration and sharing information constitutes a key challenge.

John, CTO of XYZ, has read a paper describing the Dicode Workbench and decides to try it in his company since it seems to provide all the functionalities needed: an application web front-end that makes it available from everywhere and an integration framework that supports the integration of multiple services and applications used in the development process.

![Figure 5.2: List of available workspaces in the Dicode workbench](image-url)
John opens his favorite web browser and goes to the Dicode Workbench web page (http://hodgkin.dia.fi.upm.es:8080/dicode). He creates a new account for him and activates it following the link received by email. Then, he uses his username and password to login into the system. Once inside, he clicks on the label “Workspaces” located at the menu on the left. A list of the existing workspaces is presented as shown in Figure 5.2. John tries some of them to evaluate whether someone fits his needs. Since he cannot locate an appropriate workspace for his purposes, he decides to create a new one to start with the development of a new software tool. So, he clicks on the “+ Create a new workspace…” link located at the upper-left corner of the list of workspaces. Then, a web form appears enabling John to input information about the workspace (name, description, domain, tags and the visibility). John fills the form and click on the “Create Workspace” button.

After that, John enters the new workspace. At the beginning, only the Collaboration Service is included in the Workbench. It appears maximized on the right column. John decides to explore the existing services already published in the Dicode Workbench by using the option “Add Services” of the menu. After evaluating some of them, he adds two services to the workspace that could be useful for his purpose: the Storage Service that will allow users to share files, code or test sets, and the Document Viewer Service to visualize documents. At this moment, the workbench appears as shown in Figure 5.3.

![Figure 5.3: New workspace customized by John.](image)

John comments to his team what he has done with the Dicode Workbench and invites them to create their own accounts to start the collaborative work. The team is composed by one analyst, one system designer and three programmers. After all of them have activated their accounts, John realizes that the workspace he created is visible for any user in the system. This constitutes a great security and confidentiality problem, so he decides to change the visibility of the workspace to “Restricted”. This way, only his team will have access to the workspace. He modifies the visibility attribute and adds his team to the list of granted users.
Now, John and his team can start working on the new tool. Since the environment is completely new for all of them, they start a discussion about which tools they are going to use for programming the new tool. After having a discussion using the Collaborative Service, they finally decide to use a development framework for Java that provides a web interface. During the discussion time, Susan, the system designer, had to attend a conference in the States but using the Dicode Workbench she was able to join and follow the discussion.

To use the selected Java framework, John has to publish the application in the Dicode Workbench. He clicks on the “Services” option of the main menu, and then on the “+ Publish a new service…” link. After filling a web form with the information about the Java framework, it is ready to be included and used in the workspace. John returns to the workspace and adds the Java framework application. He configures all relevant parameters of the Java framework. Since this moment, all members of the team have access to the same development framework with the same configuration. Neal, one of the programmers, suggests that there is a nice on-line development help available at http://www.javadevelopers.com, and he wonders whether this help could be integrated in the Dicode Workbench. John answers that “of course, since it is an on-line web page”, and repeats the process detailed above to publish and use the help.

At this stage, the Dicode Workbench looks as presented in Figure 5.4. Now, the team can start developing the new tool knowing that all changes are available for all persons in the team anywhere, and no extra installations of development tools are needed. They are configured once and this configuration is shared by all members of the team.

![Dicode Workbench](image)

**Figure 5.4:** Another instance of the Dicode Workbench (after including new applications and having conducted some discussion).

### 5.3 Innovative functionalities

The traditional tendency to work in isolated environments is quickly becoming obsolete. Nowadays, sharing data and results almost in real time is a common and key requirement in
multiple domains. As we have seen, the Dicode Workbench has been conceived as an open integration platform, that it is not restricted to the concrete use cases of the Dicode project. It provides users with an innovative working framework applicable to multiple domains and scenarios, taking advantage of the latest information and communication technologies (ICTs).

Innovative features provided by the Dicode Workbench that could contribute to a change of current work practices include:

- **Common integrated interface**: All services and applications are available within the same platform. Users do not need to open multiple applications and change between them since most resources can be integrated in the Dicode Workbench.

- **Web User Interface**: The Dicode Workbench is a web application accessible through the Internet by using a common web browser. Using the platform does not require to install any extra applications or libraries in the computer. This feature contributes to reduce the security hazards for the companies and institutions adopting this technology.

- **High availability**: Since all resources (services, applications and data sets) are available through the Internet in the cloud, it is not needed to download anything locally. Services and applications deal directly with the data and information needed to carry out the associated tasks. It reduces time and costs to organizations since no powerful computers are required to work with the Dicode Workbench. All applications and services are naturally distributed.

- **Shared working environment**: The same working environment (Collaborative Workspace) is available for all users or a restricted list of them depending on the visibility defined by the creator of the workspace. This means that all users granted to access a workspace see the same resources and have the same configuration of resources with the other users. This way, everything is automatically shared by users having the required privileges.

- **High customizable working environment**: Although users see the same set of services in one workspace, each user is able to modify the visualization and distribution of services according to his/her wish, just by moving the widgets inside the browser. Users are able to save that configuration between sessions.

- **Collaborative work**: The Dicode Workbench does not only allow sharing working resources between users, but it enables them to work collaboratively. Specific services such as, for instance, the Collaboration Service, are aimed to this purpose. All workspaces include by default the Collaboration Service, already configured and ready to be used.

- **Dynamic integration of services and applications**: The Dicode workbench has been conceived as a living environment. Currently, it integrates a set of services but it is not restricted to them. Services and applications can be dynamically integrated (published) by any user at any time. Thanks to the integration approach defined, almost any web application or service can be directly published in the Dicode Workbench and used from the associated workspaces. The process of publication of one service has been simplified as much as possible to facilitate this task. It is almost enough for one to just provide the URL where the service or application is available and a name to identify the service in the system to publish it.
Configure once, use many times: Sometimes, software applications require strong ICT skills to be configured before being used, and users have to spend much time to learn how to do it. As presented in the scenario above, the Dicode Workbench enables an ICT expert to setup the working workspace once, and then, all users could use it. This feature also reduces time and costs, since workers could concentrate all their efforts on more creative and productive tasks.

Regarding current limitations of the Dicode Workbench, we can mention a couple of them, derived directly from the web nature of the system:

- **Browser compatibility:** Web applications suffer from this common problem. Each web browser manages and displays the information by applying its own rules. This constitutes a real headache for web developers trying to ensure that applications work properly in most web browsers. In Dicode, we have had special care to this issue and the Workbench works fine in some of the most popular browsers, but not in all of them.

- **Integration of web applications and services:** This feature, despite being one of the major advantages, has a limitation. As we mentioned above, almost all web applications can be directly integrated in the Dicode Workbench, but only web applications. At this moment, the Workbench does not provide direct mechanisms to integrate other type of applications or services. Existing services or third-party applications that are not based on the Web could be integrated if pre-wrapped with a Web interface.

### 6 CommBAT: Community Behaviour Analytic Tool

#### 6.1 Context for Use

Dicode provides a novel collaborative workspace to support knowledge workers on collaboration and decision making in data-intensive tasks. Specifically, the Collaborative Workspace (offered through Dicode’s Collaboration Service) supports users in working with their argumentation via different views: forum view, mind-map view and formal view, which permit an incremental formalization of collaboration. This collaborative approach to problem solving involves a process of brainstorming, sense making and decision making.

Working with such a novel Collaborative Workspace is a new experience for most participants (i.e. the knowledge workers). To reap the benefits of having an environment that captures the digital footprints of a problem solving process and its underpinning rationale, its use needs to become embedded into the future work environment. Three main stakeholders are noted in this transition: (i) **Developers** of Dicode Collaborative Workspace would be interested in how users benefit from the novel features of the system and how to support them better. (ii) **Researchers** in collaboration and decision making studies would be interested in identifying patterns of user and community behaviour with the logging ability of Dicode Collaborative Workspace. (iii) **Community leaders or moderators** would be interested in how to use the tool to improve the process of collaboration and decision making.

To meet the above demand, a tool called CommBAT (Community Behaviour Analytic Tool) has been developed to gain insight of community behaviour from logs in the Dicode Collaborative Workspace. CommBAT is a Windows desktop application, which presents statistics of Dicode Collaborative Workspace log data in a visual form (details are reported in
deliverable D4.2.3). It allows users to dynamically explore the factors related to community behaviour: users, activities, objects and semantic types.

CommBAT combines three features: Dynamic Filtering, Visual Presentation and Multi-perspective views. Firstly, it allows users to dynamically explore activities of Dicode Collaborative Workspace via filtering the factors related to community behaviour. Secondly, it presents the dynamic filtering results in different types of chart: pie, column, bar, bubble etc. Finally, CommBAT provides multiple views for users to look into knowledge items (ideas, comment, notes and services): object type view, user activity view and object timeline view. The combination of these three features provides a flexible way for users to develop insights into how the community members use Dicode Collaborative Workspace and how they work together in terms of their activities, interactions and knowledge items.

6.2 Using CommBAT

CommBAT can read log data of a collaborative workspace in Dicode and present information visually (Figure 6.1). The information presented in CommBAT includes users, sensemaking actions, workspace activities and knowledge items. For example, across the top in Figure 6.1: (i) User information is shown in a column chart, which shows five users with their number of activities sorted in a descendental order; (ii) Sensemaking actions are presented in a pie chart. It summarises what users performed in the workspace: contributing knowledge items, reviewing knowledge items or organizing knowledge items etc. (iii) The chart on workspace activities shows details of sensemaking actions, such as how users contributed to the workspace, in creating new knowledge objects or new relations and so on.

CommBAT has three different views for knowledge items: Object Type View (OTV), User Activity View (UAV) and Activity Timeline View (ATV). The bottom half of Figure 6.1 shows OTV, which presents the objects in their semantic types using different shapes and colors. In OTV, the objects are presented in an order of their creation date (x-axis) with the numbers of user activities (y-axis). In Figure 6.1, it shows the “group” type of objects (in pink squares) were created towards the later part of the workspace’s lifespan while “relation” type of objects (in blue crosses) are constantly created during the whole lifespan.

Figure 6.1: CommBAT View of Users, Sensemaking Actions, Workspace Activities and OTV.
User Activity View (Figure 6.2) shows how users contributed to one object. In this view, the objects are also presented in an order of their creation like OTV. In this way, user can see a consistent presentation of objects but different aspects of the objects. For example, from OTV, the top “Idea” object (ID: 28278) has the a high level of user activities and from UAV, we can see that this object only involved activities of three users, while there were six users in the workspace. This might indicate that an alert is needed to inform the other three users to pay attention to this idea.

![CommBAT User Activity View (UAV)](image)

In Activity Timeline View (ATV), we can see objects in terms of time and sensemaking actions. For example, in Figure 6.3, we can see one object (ID: 28234) has been constantly reviewed by participants and from UAV, we can see this object has attracted all users.

![CommBAT Activity Timeline View (ATV)](image)

CommBAT also provides a Multiple Document Interface (MDI), which supports multiple logs of workspaces presented at the same time. This allows users to compare different analytic results among different workspaces. For example, the screen shot of Figure 6.4 compares two workspaces with ID 29967 and 27231.
6.3 Innovation of CommBAT

CommBAT enables visual, dynamic log analysis to examine the community features of Dicode Collaborative Workspaces such as participants and activities. With the novelty of Dicode Workspace, it also examines the semantic features (semantic types of objects and sensemaking activities) of collaboration and decision making.

CommBAT supports the following stakeholders:

- Community Leaders or moderators, who are interested in how to effectively facilitating the process of collaboration and decision making
- Dicode Developers, who are interested in how Dicode workspace are being used in terms of its novel features
- Collaboration and decision making researchers, who are interested in patterns of user and community behavior.

The dynamic filtering and visual views can help the stakeholders to identify:

- How participants contribute to the workspace in terms of activities;
- How participants contribute to the workspace in terms of their use of the provided objects;
- How an object evolves in terms of its interaction with participants and over a timeline;
- How participants interact with one type of object over a timeline;
- What interactions among participants have taken place surrounding a type of object or an object.

With the tool, user and community behaviour can be discovered so that quality and evolution of collaboration and decision making can be monitored. This is a step forward towards improving the collaborative and decision making process using the Dicode Collaborative Workspace.
7 Pinta: a Tool for Exploratory Search over Linked Data

7.1 Context for Use

Linked Data technologies have received wider acceptance, both in industry and academia (Bizer et al., 2009). One of the major factors for this success has been the availability of large amount of semantic data in various formats and domains (http://lod-cloud.net/state/). In parallel with engineering solutions for seamless generation of semantic data, efforts have been made to facilitate user interaction with such data. There are growing arguments that Linked Data technologies can be utilised to enable user-oriented exploratory search systems for the future Internet (Waitelonis et al., 2010). In contrast to regular search, exploratory search is open-ended, multi-faceted, and iterative in nature, and is commonly used in scientific discovery, learning, and sense making (Marchionini, 2006; White et al., 2006).

As part of the Dicode project, an exploratory search system, called Pinta, has been developed for browsing over linked semantic data. Pinta falls under a class of applications called “Unifocal semantic data browsers”. Such data browsers operate on semantically augmented data (e.g. tagged content) and layout browsing trajectories using relationships in the underpinning ontologies. In such browsers, exploration is often restricted to a single start point in the data and uses ‘a resource at a time’ to navigate anywhere in a dataset (Araújo et al., 2009).

The main goal of Pinta is to enable users to easily tap into resources built from the Web and, in particular, exploring the use of the Linked Data paradigm (Heath and Bizer, 2011). Figure 7.1 depicts the traditional three-layer architecture for Pinta which comprises: (i) the Data Layer, including knowledge sources and content, (ii) the Processing Layer, including modules for semantic augmentation and query, and (ii) the Presentation Layer for content browsing. The implementation of Pinta combines semantic web technologies for semantic augmentation, semantic query and data representation.

The Data Layer contains domain specific ontological knowledge sources and content assembled from the Web (Linked Data and other domain specific sources). It is envisaged that this kind of resources will be available more readily and in bigger volume in the future in all domains.
The Processing Layer has two main services: (i) semantic augmentation of the assembled content, and (ii) semantic queries to retrieve content for the Presentation Layer. Semantic augmentation is a process of attaching semantics (in the form of ontology concepts) to a selected part of text. Semantic Query service takes keywords captured from the Presentation Layer, performs concept/content lookup and returns the relevant ‘network of information’ to the user via the Presentation Layer.

The Presentation Layer provides a front-end for the output of semantic queries from the Processing Layer (the template for a focus entity is shown in Figure 7.2). The interface layout includes three main facets and a description (at the top) extracted from the knowledge datasets for the focus entity (being currently explored): (i) Facet 1 includes facts about the focus entity; (ii) Facet 2 includes terms related to the focus entity; and (iii) Facet 3 shows related content. Hyperlinks are provided to further explore the details.

In a cognitively-complex task environment, the ability of showing a network of linked information for user exploration will be helpful to stimulate ideas and insight into the problem domain.

7.2 User interaction with Pinta to perform exploratory search tasks

This section illustrates user’s interaction with MusicPinta, which is an instantiation of Pinta in the Music domain (see deliverable D4.2.3). In this example, the user was asked to conduct two tasks as presented in Table 7.1. These tasks exhibit the characteristics of exploratory search tasks summarised in (Wildemuth and Freund, 2012): the main goal is to learning and/or to investigate aspects of a musical instrument; there is a low level of specificity about the information needed and how to find it; search is open ended, requires finding several items and involves a degree of uncertainty; tasks are ‘not too easy’ and include multiple facets.

Table 7.1: Sample user tasks in MusicPinta.

<table>
<thead>
<tr>
<th>Task 1: Characteristics of a musical instrument [bouzouki]</th>
<th>Task 2: Usage and features of a musical instrument [electrical guitar]</th>
</tr>
</thead>
</table>
| The music shop is extending its collection of instruments with international musical instruments. You work in an advertising agency, which has been asked to prepare an advertisement script for some of the new instruments that will appear in the shop. A key part of the preparation of the advertisement script is the research of the product. You have been asked to conduct a research of one of the new instruments, called bouzouki, using the information available in MusicPinta. You have to identify:  
- the main characteristics of bouzouki;  
- up to five similar instruments to bouzouki;  
- features that make bouzouki distinctive from the similar ones you have chosen.  
Go to ‘Semantic Search’ in MusicPinta and type ‘bouzouki’. Browse the content and follow links. Complete the provided form. | The music shop wants to increase the sales of its traditional musical instruments, such as electrical guitars. It intends to do this by adding links to creative commons album recordings with electric guitars, together with some interesting information about these albums to inspire customers to play/buy electric guitars or other musical instruments. Furthermore, when displaying its electric guitar items, the shop wants to highlight key features people look for when purchasing electric guitars. You are asked is to conduct the research to address the above requirements by using information provided in MusicPinta. You have to review the information about electric guitar and identify:  
- three interesting album recordings that include electric guitars and specify what is interesting;  
- key features that people look for when purchasing an electric guitar.  
Go to ‘Semantic Search’ in MusicPinta and type ‘electric guitar’. Browse the content and follow links. Complete the provided form. |
Task 1 can be completed mainly through browsing the musical instrument classification and reading descriptions from the underlying ontologies and semantic datasets. The task is analytical in nature, as the user has to perform comparison and identification of distinctive features. Example screen shots from a user’s interaction are shown in Figure 7.3 (the user reviews the information provided on ‘bouzouki’ - the image, description, categorisation, and other terms), and Figure 7.5 (examining the content of similar instruments selected from the facet “Plucked string instrument” available on the ‘bouzouki’ interface).

**Figure 7.3:** ‘Bouzouki’ - facts, related terms, description, image, tagged content (obscured).

**Figure 7.4:** ‘Electrical Guitar’ – facts (obscured), related terms, description (obscured), tagged content.

**Figure 7.5:** ‘Lute’ - facts and related terms, description, image, tagged content (all obscured).

**Figure 7.6:** A performance involving ‘Electrical Guitar’.
In contrast, Task 2 is more ambiguous and involves some creative thinking and imagination, and can be completed by browsing through content about music albums (see Figure 7.6) and artists, and reading through Amazon reviews (Figure 7.4).

7.3 Key Innovation: Semantic Signposting in Semantic Data Browsers

Although the technological platforms for exploring linked data are growing, enabling citizen users to explore the inter-connectable links to make sense of structured data is still a key challenge (Schraefel, 2010). Anticipating these new tools for knowledge workers to exploit Linked Data in the future, emerging research begin to identify major issues with user exploration of linked data, derive requirements for new methods, and engineer solutions to implement these methods utilising semantic technologies and tools.

As part of the Dicode project, MusicPinta is piloted in experimental studies, as reported in (Thakker et al., 2013; Yang-Turner et al., 2013; Dimitrova et al., 2013) with users to elicit requirements for intelligent assistance based on observations of challenges users faced while interacting with MusicPinta. Lessons learned with suggested improvements on user adaptation and usability are presented in deliverable D3.3 (due in August 2013).

From these requirements, we have identified the need for further algorithmic support to realise the exploratory search potential of semantic data browsers. There can be many possible ways to address these requirements. One of such possible novel approach is semantic signposting which we have implemented as part of the Dicode project.

In uni-focal exploration, a user focuses on one entity at a time represented in a page. This entity page contains links to various descriptions, image and also links to other entities.

Such entity page can be treated as a juncture in the journey where the explorer has to make few choices (through the links which takes him to different paths). Semantic Signposting provides different types of signposts guiding the explorer in making a choice about paths she can take. An example is provided in Figure 7.7, where important facts are calculated based on the semantic graph and richness of content and signposted for user to view instead of reviewing all the possible facts, hence improving usability and reducing cognitive overload.

8 Conclusions

As reported in the two evaluation rounds of the project, the overall Dicode solution has great potential for a wide range of diverse organizations and groups working in data-intensive and/or cognitively complex settings. The innovative work methodologies proposed in this
deliverable, as illustrated through real-world scenarios of use, exploit the proposed solution that thoroughly builds on the synergy of human and machine reasoning. Having followed a generic approach in the development of the associated services, we are confident that the Dicode solution can be easily adopted from professional organizations and bodies outside the project (the exploitation strategy of each Dicode partner is reported in detail in deliverable D7.2.2, due in August 2013).

Admittedly, the methodologies proposed in this deliverable mask the overall complexity of the underlying issues (i.e. data mining, sense making, collaboration, decision making), thus allowing stakeholders to easily interact with large and complex data, providing them with meaningful recommendations upon which they can base their decisions and actions. Moreover, machine-tractable knowledge concerning the full lifecycle of collaboration and decision making is easily accumulated and maintained. Consequently, the foreseen solution is expected to augment the productivity of stakeholders.

Adopting open standards, and in accordance with EU’s recent initiatives on Open Systems and Data, the Dicode solution has the potential of forming a rich ecology of domain specific and non-specific extensions. It allows for external data service providers to supply information, as well as for external developers to supply additional modules and applications, which are tailored to evolving market conditions. Finally, it enables diverse public and private entities to aggregate, structure, semantically enrich and analyse vast amounts of information. This turns the problem of information overload into a benefit of structured data, which can be used as the basis for decisions of better quality. Consequently, the Dicode solution enables the turning of information growth into economic growth.


9 References


