



Mastering Data-Intensive Collaboration and Decision Making

FP7 - Information and Communication Technologies

Grant Agreement no: 257184

Collaborative Project

Project start: 1 September 2010, Duration: 36 months

D6.2.2 – Report from the evaluation of use case #1 (second version)

Due date of deliverable: 30 April 2013
Actual submission date: 26 April 2013
Responsible Partner: BRF
Contributing Partners: UPM, FHG, UOL, CTI

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: evaluation process, Dicode integrated services, integrative analysis, biomedical data, correlation coefficient



The Dicode project (dicode-project.eu) is funded by the European Commission, Information Society and Media Directorate General, under the FP7 Cooperation programme (ICT/SO 4.3: Intelligent Information Management).

The Dicode Consortium

	<p>Computer Technology Institute & Press “Diophantus” (CTI) (coordinator), Greece</p>
	<p>University of Leeds (UOL), UK</p>
	<p>Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e.V. (FHG), Germany</p>
	<p>Universidad Politécnica de Madrid (UPM), Spain</p>
	<p>Neofonie GmbH (NEO), Germany</p>
	<p>Image Analysis Limited (IMA), UK</p>
	<p>Biomedical Research Foundation, Academy of Athens (BRF), Greece</p>
	<p>Publicis Frankfurt Zweigniederlassung der PWW GmbH (PUB), Germany</p>

Document history			
Version	Date	Status	Modifications made by
1	15-03-2013	First draft	Georgia Tsiliki, BRF
2	31-03-2013	Second draft	Georgia Tsiliki, BRF
3	15-04-2013	Third draft (sent to internal reviewers)	Manolis Tzagarakis, CTI Guillermo de la Calle, UPM
4	18-04-2013	Reviewers' comments incorporated, sent to SC	Georgia Tsiliki, BRF Sophia Kossida, BRF
5	26-04-2013	Final (approved by SC, sent to the Project Officer)	Georgia Tsiliki, BRF Sophia Kossida, BRF

Deliverable manager

- Georgia Tsiliki, BRF

List of Contributors

- Georgia Tsiliki, Sophia Kossida, BRF
- Eduardo Alonso Martínez, Martha E. Rojas Vera, UPM
- Spyros Christodoulou, Nikos Karacapilidis, CTI
- Natalja Frieman, FHG
- Fan Yang-Turner, Lydia Lau, UOL

List of Evaluators

- Manolis Tzagarakis, CTI
- Guillermo de la Calle, UPM

Summary

This deliverable reports on the second evaluation round of Use Case 1 “Clinico-Genomic Research Assimilator” and particularly on the evaluation of the Dicode workbench and the enhanced services integrated into it. The evaluation process was performed by using properly formulated metrics and instruments which are described in D6.1 based on the specifications of D2.2. Collaboration services, data mining and decision making support services, as they are developed and accommodated in the Dicode workbench, are evaluated by senior real users under real Use Case 1 related settings. The document also comments on how the feedback provided in the first evaluation round, as described in D6.2.1, and the second review meeting of the project, were taken into account. Particular emphasis is given to the usability and acceptability of the Dicode services, to guide final revisions and improvements.

Table of Contents

1. Introduction.....	5
1.1 Context	5
1.2 Objectives.....	5
1.3 Structure.....	5
2 First evaluation round: overview of Use Case #1.....	6
2.1 Outcome and related actions.....	7
2.2 Reviewers' comments.....	8
3 Second evaluation round: evaluation process & feedback	9
3.1 Evaluation process.....	10
3.1.1 2 nd evaluation round: sample recruited	10
3.1.2 2 nd evaluation round: goal	10
3.2 Detailed evaluation feedback.....	12
3.2.1 Overall impression of the Dicode Workbench	12
3.2.2 Support of data analysis, collaboration and decision making.....	14
3.2.3 Potential benefit to current work practices	17
4 Discussion.....	20
References.....	22
Appendix: Dicode questionnaire - second round	23

1. Introduction

1.1 Context

This deliverable reports on the evaluation of the enhanced version of the Dicode services relative to Use Case #1, “Clinico-Genomic Research Assimilator” (CGRA), as these have been developed in the context of WP2 - WP5. Specifically, it presents the evaluation outcome of the integrated Dicode services (through the Dicode workbench) by summarizing the answers of senior real users given to the questionnaire shown in the Appendix.

This is the second of two deliverables reporting on the validation and assessment of the developed Dicode services through the CGRA use case. While the focus of the first deliverable (D6.2.1) was on the first version of the CGRA Dicode services, this deliverable (D6.2.2) presents the evaluation of their enhanced version, and also comments on how the feedback provided in the first evaluation round was taken into account. Dicode project’s evaluation activities will be summarized in deliverable D6.5 (due in month 36), together with an assessment of the project’s outcome from external experts/ bodies or communities.

1.2 Objectives

The purpose of this document is to present the second evaluation round of the integrated Dicode services related to the CGRA use case in the context of Tasks 6.2, 6.3, as these originate from the functional specifications outlined in the deliverable D2.3 and described in the deliverables D3.1.2, D3.2.2, D4.1.2, D4.2.2, and D5.4.2, in order to assure their usability and accessibility. The cost-effectiveness and readiness for the market of Dicode services are also assessed.

The Dicode workbench and corresponding innovative work methodologies are evaluated based on how they can improve the everyday practise of real senior users (e.g. bioinformaticians, biologists) during the processing of voluminous and heterogeneous biomedical data, whilst enhancing collaboration, decision making, and collective sense making in the biomedical domain.

The evaluation of the Dicode services is performed based on a series of Key Success Indicators (KSIs) presented in D6.1, overall aiming to assess:

- maturity of the technology used through Technology Readiness Level (TRL);
- dissemination and exploitation activities;
- usability and acceptability of Dicode services in the CGRA use case.

In this deliverable emphasis is given to those KSIs measuring usability and acceptability of the Dicode services. The evaluation results and feedback collected and critically presented in this document will then serve as a guideline for the service providers to develop the final Dicode suite of services, emphasizing on the improvement of their usability and overall quality.

1.3 Structure

In Section 2, we present an overview of the first evaluation round and briefly explain its outcome whilst highlighting the key issues raised by the evaluators. The project reviewers’ comments are also discussed in relation to the evaluation’s outcome. In Section 3, information about the process of the second evaluation round of CGRA are provided,

emphasizing on the characteristics of the enhanced services now integrated into the Dicode workbench, along with detailed evaluation feedback and summary statistics for the evaluation results. Section 4 discusses the benefits of the evaluation procedure and outlines possible directions for the final revision of Dicode services.

2 First evaluation round: overview of Use Case #1

During the first evaluation round of the CGRA use case, the Dicode workbench and five services of the project were evaluated (see Table 2.1). At the time of the first evaluation round, two of the services were integrated into the Dicode workbench whereas the remaining three were evaluated as standalone software.

Services	Characteristics	Indication of integration in the Workbench
Workbench	The Dicode integration platform for all data analysis and collaboration.	-
Collaboration & Decision Making Support	Exploits user's reasoning abilities to facilitate sense-making, decision making and collaboration.	integrated
Storage	Share and exchange of information (data sets, reports etc.).	integrated
Forum Summarization	Identify topics in discussion threads of particular public forums.	Stand-alone
Subgroup Discovery	Estimates subgroups in user specified data set.	Stand-alone
Recommendation	Recommend similar users or documents.	Stand-alone

Table 2.1: Dicode services evaluated through the first evaluation round.

The first evaluation round was conducted according to the Dicode Evaluation Framework presented in D6.1 and to some extent updated in D6.2.1. In what follows, we are highlighting the main characteristics of the evaluation procedure:

1. Use case partners and technical partners invited participants to join the user trials.
2. Due to the fact that the majority of developed RESTful Dicode Services produced output in Json format (<https://tools.ietf.org/html/rfc4627>; (Crockford, 2006)), the target population was defined to be mainly computer scientists/informatics scientists with knowledge in Web Programming/Web Engineering.
3. The Dicode Services were evaluated by 61 volunteers from the four participant countries of the project (Greece, Spain, Germany, United Kingdom).
4. The sampling method used was Snowball sampling (Kitchenham & Pfleeger, 2002).
5. Five questionnaires (three taken from D6.1 and another two that were designed and tested for the Forum Summarization service) were distributed to evaluators together with the Dicode Evaluators' Guide that aimed at gathering mainly quantitative but also qualitative feedback when appropriate.
6. Evaluators were asked to carefully read the instructions, have a 'hands-on' session for each service and finally fill in the questionnaires.
7. Summaries of the evaluation analysis results were produced using the Statistical Package for the Social Sciences (SPSS 17v) routines.

8. Summary statistics were presented in D6.2.1, whereas detailed statistics relevant to the completed questionnaires collected for the three Dicode Use Cases' evaluation were uploaded in the Dicode wiki:
<https://wiki.dicode-project.eu/display/DIC/WP6+Evaluation+results>.

2.1 Outcome and related actions

Based on the feedback received from the first evaluation phase of the Dicode project, the Dicode workbench was reported to be a promising tool, which facilitates users to set their research objectives and better understand the data and methodologies used in their research. The vast majority of the evaluators appreciated the potential of exploiting the synergy of machine and human reasoning through data mining and collaborative decision making services. The innovative approaches on the text-mining services seemed to be appreciated by evaluators, who generally agreed on the usefulness and acceptability of the provided services. Nevertheless, additional work towards the improvement of Dicode services in terms of their documentation, user interfaces and performance seemed to be essential. Another issue raised relates to testing these services in various data-intensive contexts, in order to further assess their applicability and potential, and gradually build their generic nature.

In Table 2.2 we summarize the main points raised by the reviewers per service evaluated, as well as the actions that took place to address them. The issues marked with an asterisk (*) are addressed in the second evaluation round.

Services	Issues raised	Action taken
Workbench and integrated Storage service	Technical issues involving its performance and interface*.	UPM, CTI: improved performance & user-friendliness
Collaboration & Decision Making Support	Technical issues involving its performance and interface. More detailed or differently structured scenarios needed*.	CTI: improved performance
Forum Summarization	Relatively limited functionality of the service to extract topics and search for relations.	UOL: improved service's usefulness
Subgroup Discovery	Technical issues involving its interface and presentation of results. Increase flexibility of the service by applying it to additional data*.	FHG: improved service's functionality, usefulness and flexibility
Recommendation	Technical issues involving its interface and outcome.	FHG: improved service's functionality, and usefulness

Table 2.2: Dicode first evaluation round: issues raised and actions taken. Issues marked with an asterisk are addressed in the second evaluation round.

Additionally, other major issues were raised which involved all services, i.e. user interface and design, data-intensiveness, general purpose statistical analysis and additional guidance

throughout the process. As seen in Table 2.2, the improvement of the user interface and design were specially addressed by all technical partners. Data-intensiveness is closely related to the data-mining services provided for CGRA, namely Subgroup Discovery and GEORecommender (called Recommender in D6.2.1). FHG worked towards this goal by allowing extra and differently structured data for Subgroup Discovery service. More specifically, the Subgroup Discovery service can now handle sequencing data (RNASeq), and in that way addresses the relatively new and tackling issue of integrating data from microarray and sequencing technologies. A great assistance is given by the GEORecommender service which can now give access to any kind of biomedical data currently stored in the Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) database. Data sets (or any other type of document) can be stored in the Workbench, using the Storage service, hence addressing the challenging storage issue of sequencing data. UPM is working on the challenging data-intensiveness issue by updating the PubMed Mobile service; the up-to-date version of the service focuses on searching across the approximately 2.7 million full text articles published in PubMed Central (<http://www.ncbi.nlm.nih.gov/pmc/>) database, in order to suggest interesting relations between biomedical entities based filtered articles.

Finally, CTI worked on the user-friendliness and performance of the Dicode collaboration and decision making services through the alternative views offered (see deliverables D4.1.2 and D4.2.2).

2.2 Reviewers' comments

The results of the second year Dicode's review were positive with respect to WP6 and the evaluation procedure followed for the initial set of services of the project. Figure 2.1 shows the relevant quotation taken from the reviewers report. A special note is made for the lack of real users as evaluators of Use Cases 2 and 3. With respect to Use Case 1 and deliverable D6.2.1, the following comments were made:

1. 'Goal- or task-dependent service recommendation to the user can contribute to increased usability.'
2. 'The GEO recommender needs to consider strategies for dealing with 'cold start' problems.'
3. 'The GO categories identified in the discovered subgroups need to be stratified for statistical relevance.'

With respect to the first point we have been working toward increasing the flexibility and performance of the GEORecommender service as reported in D3.2.2. Furthermore, UPM have been working on increasing the functionalities of the PubMed Mobile service aiming to estimate and recommend strong relationships between biological entities given users' specified criteria. This is a promising service, auxiliary to data analysis options offered by data mining services. The second point is also addressed by the enhanced version of GEORecommender service, which has been presented in D3.2.2. Finally, the third point has been addressed and now the outcome of the Subgroup Discovery service includes groups of GO categories along with their ranking statistics. Extra stratification techniques are considered.

<p>- WP 6: This workpackage is about the evaluation and validation of the services along the three use-cases. The reviewers comments from last year were addressed appropriately. Unfortunately, use-cases 2 and 3 were not evaluated by real users. This remains to be done in the third year. However, the applied methodology is according to high scientific standards of using statistical methods. Everything is reported on wikis. The WP was delayed by 3 weeks in order to incorporate feedback from the stakeholders. Everything else was on time and according to plan. This is just a very minor deviation.</p>

Figure 2.1: Progress notes for WP6 as reported in the second review meeting report.

3 Second evaluation round: evaluation process & feedback

For the second evaluation round, all technical partners have worked towards finalizing the enhanced version of all the CGRA related services, in close proximity with BRF. The services now available are integrated into the Dicode workbench and display all the features presented in D3.1.2, D3.2.2, D4.1.2, D4.2.2, and D5.4.2. As shown in Section 2.1, all issues raised by the evaluators of the first round as well as the reviewers remarks raised in the second review meeting of the project are addressed. As explained in Section 2, the update of CGRA was based on both the enhancement of the already developed and tested services, and on suggestions for the development of additional services to address the evaluators' comments and feedback.

Rather than evaluate the usability of tools for each service, in the second evaluation round we were interested in evaluating the usability of tools to facilitate collaborative processing of the task at hand. Given the fact that all services were integrated into the Dicode workbench and most of their foreseen functionalities are already implemented, the second evaluation round was conducted by recruiting senior members of the biomedical community who are able to assess the project's outcomes with the use of video-casts. This is in accordance to Jadhav and Sonar (2009) and Sun et al. (2010), who have emphasized the importance of video-casts through the evaluation process. Video-casts have the advantage that they can be used as often as the individual user finds appropriate. Moreover, video-casts are suitable for assessing collaborative behaviours that facilitate mutual participation (Cheng et al., 2010). Similarly to the approach described in Sun et al. (2010), an overview of the suite of services now available is presented to real expert users through a concrete everyday usage scenario, who are then requested to imprint their thoughts about services' functionalities as well as their overall opinion of the Dicode workbench. Each individual was also given an overview of the Dicode project and the functionalities and goals of the Dicode workbench. Based on the video-casts provided and the short description of the Dicode project, individuals were asked to provide their general impression on the Workbench as well as critique specific features (as demonstrated in the walkthrough video-casts) by providing quantitative and qualitative remarks (Sun et al., 2010).

In the second evaluation round, we have focused on collecting the opinion of real senior users in the biomedical field. We decided to combine the two evaluation methodologies, namely scenario-formed video-casts and questionnaire, based on the suggestions of Cugini et al. (1997) in order to capture experts' judgements about tool usage and overall user's ratings.

Emphasis is given to measuring usability, acceptability and functionality of the Dicode workbench and integrated services, as most technical issues were analytically explored in D6.2.1. Thus, questionnaires aimed to capture the users' opinion based on scenario-formed video-casts. Summaries of the evaluation analysis results were produced using the Statistical

Programming language R (<http://www.r-project.org/>, R version 2.15.3) and routines supplied via Google Docs platform (<https://www.google.com/intl/en/drive/start/apps.html#product=sheets>). The evaluation results and feedback collected and critically presented in this document will then serve as a guideline for technical partners to finalize and improve services' usability and overall quality, and for the use case partner as constructive dissemination guidance. In the following section, we discuss more details of the second evaluation round.

3.1 Evaluation process

3.1.1 2nd evaluation round: sample recruited

We recruited ten participants from the biomedical domain, similarly to Cheng et al. (2010), using the Snowball sampling method (introduced in D6.2.1). We find that this is an adequate sample size which will allow us to focus on qualitative remarks of senior users. Participants came from different disciplines/departments and even though they are all experts in the biomedical domain, their area of expertise is different. In Table 3.1 we present their scientific background, area of expertise, position held and seniority. Their distribution is in line with the main description of CGRA use case, and it was planned to cover important and currently evolving sub-areas of biomedical research and bioinformatics. Additionally, as already mentioned above, special care was given to their seniority distribution (fourth column in Table 3.1). We clearly targeted senior members of the biomedical community, however two special groups were considered: researchers with 5-10 years (indicated with 1 in Table 3.1), and 10-20 years experience after they acquired their PhD degree (indicated with 2 in Table 3.1). Thus, in the evaluators group we have included postdoctoral fellows, associate professors, professors, and head of departments. This seniority distribution is essential because we wanted to combine significant research experience with deep and up-to-date knowledge of the field.

3.1.2 2nd evaluation round: goal

The goal of the second evaluation round is to simulate a multi-tasking environment where users need to work on diverse CGRA tasks. Thus, our hypothesis can be summarized as following:

H₁: using the Dicode Workbench to discuss and analyze biomedical data will lead to more productive and effective work practices compared to current practices.

Based on the above hypothesis, a real working scenario was presented to the evaluators via two video-casts. Here we include a brief summary of the scenario presented:

'You are a scientist working at the clinico-genomics field as part of your everyday practice; consider the case that you want to analyze two sets of data and biologically interpret your findings. You would also like to discuss with your peers the workflow of your analysis, the methodologies applied and the results produced, since you find that their different scientific backgrounds and experience will add value to your research. For that reason you decide to launch the Dicode workbench and create a collaborative session where you can analyze your data, and discuss the results with your colleagues (possibly from other disciplines).'

Evaluator No	Area of expertise	Position hold	Seniority (in years)
1	Bioinformatics	Associate Professor	2
2	Bioinformatics/ Structural Biology	Postdoc Fellow	1
3	Bioinformatics/ Mathematical Modelling	Postdoc Fellow	1
4	Bioinformatics/ Structural Biology	Head of Department	2
5	Bioinformatics/ Structural Biology	Professor	2
6	Bioinformatics	Head of Department	2
7	Proteomics	Associate Professor	2
8	Bioinformatics/ Mathematical Modelling	Professor	2
9	Bioinformatics/ Computing	Postdoc Fellow	1
10	Bioinformatics/ Computing	Postdoc Fellow	1

Table 3.1: Characteristics of the evaluators' sample. In the seniority column 2 denotes senior members and 1 less senior members (5-10 years after PhD).

For CGRA, the main task is to understand and analyze the data available under a particular working framework, that's why we based our working scenario on the Subgroup Discovery service, which can be applied to large data derived by microarray and sequencing technologies. Particularly, the integrated services considered are:

- Collaboration services,
- Decision-making services,
- Data-mining services and in particular the Subgroup Discovery service,

whereas an overview of auxiliary services and functionalities were also shown in the video-casts, i.e. the Document viewer and the PubMed services, as well as the Dicode Workbench log-in process. The questionnaire (see Appendix) provided links to the two accompanied video-casts.

Feedback collected by online questionnaires was uploaded via Google Docs (<https://www.google.com/intl/en/drive/start/apps.html#product=docs>). Quantitative and qualitative feedback was collected by evaluators aiming to explore:

- the ease of use of the enhanced version of the services, and
- the functionalities of the enhanced and integrated version of the services.

3.2 Detailed evaluation feedback

This section includes summary statistics for CGRA use case evaluation as well as representative qualitative feedback from the questionnaires. The table and figures included summarize the quantitative evaluators' responses relative to ease of use, accessibility & acceptability, the scenario provided, the acceptability and the overall quality of the Dicode services. Answers to the quantitative questions of the questionnaires are given for ordinal data in a 1-5 scale, where 1 stands for 'I strongly disagree' and 5 for 'I strongly agree' (Nielsen, 1991; Norman, 1998). Note that missing data were not imputed.

The detailed statistics relevant to the completed questionnaires collected for all three Dicode Use Cases' evaluation are uploaded in the Dicode wiki:

<https://wiki.dicode-project.eu/display/DIC/WP6+Evaluation+results>.

3.2.1 Overall impression of the Dicode Workbench

CGRA evaluators watched the first video-cast provided and responded to the first set of questions (Section A) of the Dicode Evaluation Questionnaire. Figure 3.1 summarizes the responses to questions of Section A, and particularly the mode and quartile trend (minimum, median and maximum values) of the responses relative to the evaluators overall impression of the Dicode Workbench. The Section A questions are:

1. Question 1: The information provided on the video is clear.
2. Question 2: The design of the DICODE workbench is very pleasant.
3. Question 3: The use of DICODE workbench is easy.
4. Question 4: The user interface of DICODE workbench is intuitive.

Overall mode/median values range from 3 to 5, where the minimum value is 3 except from Question 1 which is 2. Evaluators were satisfied with the information provided by the video (only one value is 2), and its pleasant design (mode and median values equal 3). Particularly for the video quality some suggestions were given:

'1-Include voice guide.

2-Power point presentation (allow interaction)'

Suggestions for the Dicode interface were also given:

'There is t[o]o much void space. Used "blank" white space. Main workspace could be larger'

'I didn't see any help files. A nice idea is to include an extended tutorial showing all capabilities.'

'The windows seem squeezed in, but generally user friendly.'

These are all issues being currently addressed by the project's technical partners (especially UPM – few amendments have been already reported in D5.4.2).

Ease of use is ranked with 3 and 3.5 (mode and median values, respectively), which can be partly explained from the lack of hands-on experience. Specifically, evaluators needed extra time to familiarize with the Workbench:

'It appears [straightforward], yet some training of the user and a presentation by an expert would be helpful. In general some wider dissemination of the platform to the community would be very useful.'

This is a very important point for the remaining of the project, and indeed one of BRF's goals for further exploitation and dissemination.

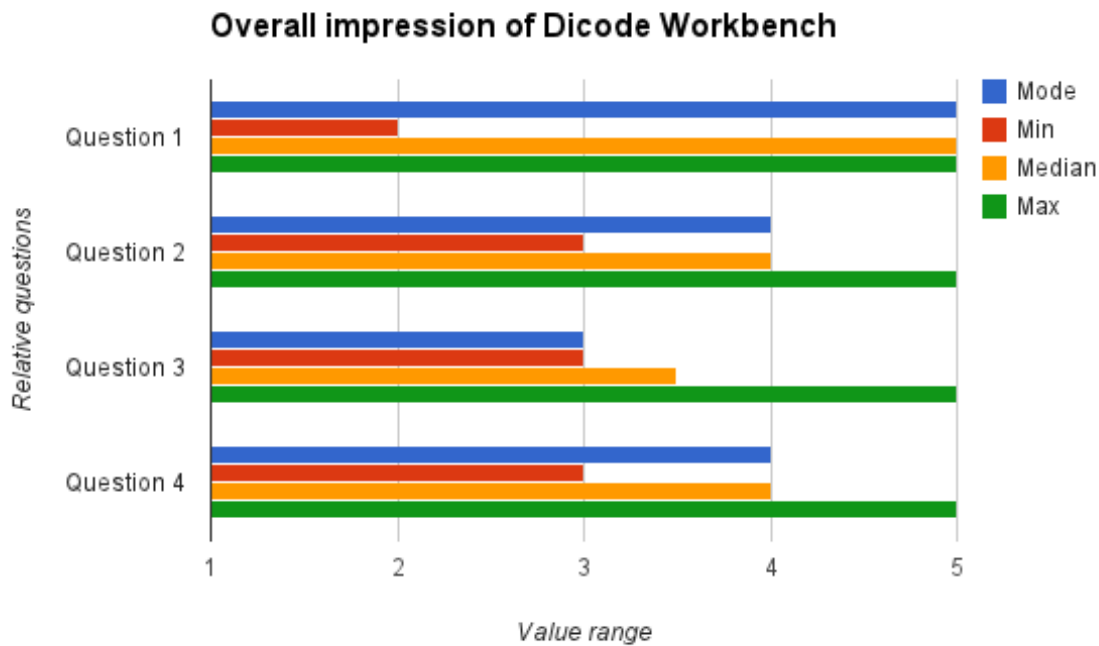


Figure 3.1: The overall impression of Dicode Workbench: mode, minimum, median, and maximum values are presented for Questions 1-4. The range of original values is 1-5, where 1 denotes strong disagreement and 5 strong agreement.

The Dicode workbench interface is reported to be intuitive (mode and median values equal 4). The positive intuition can be seen in the following remark:

'It looks like it needs some getting used to, but after 1-2 tries it should be straight-forward.'

Additionally, we found that for Questions 1-4 the seniority of evaluators does not significantly affect the bad or good impression of the evaluators towards the Dicode Workbench (chi-squared test $p\text{-value} > 0.05$). In Figure 3.2 we can see that seniority does not affect the responses of the reviewers for Section A questions, since values are roughly within the $[-0.5, 0.5]$ interval.

Lastly, in Table 3.2 we present the marginal distribution of evaluators' responses relative to their overall impression of the Dicode Workbench in Questions 1-4. We can observe that the majority of their responses are '3' and '4', marks with their marginal frequency being 0.3 (12/40) and 0.45 (18/40) respectively, whereas the highest rate '5' has marginal frequency 0.225. This constitutes an overall good impression of evaluators for the Dicode Workbench, also indicating some technical issues of interface finalization that need to be considered.

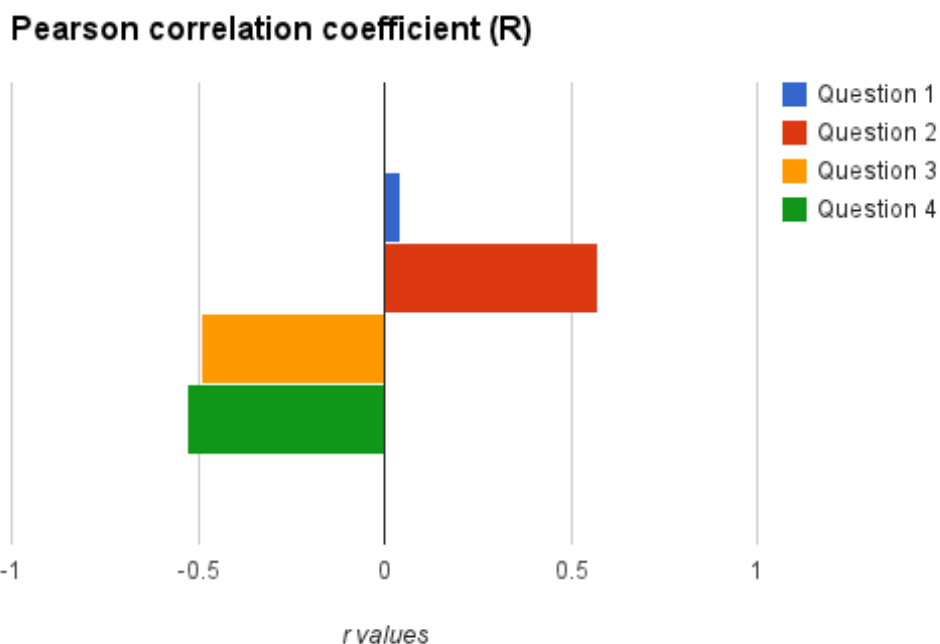


Figure 3.2: The overall impression of Dicode Workbench: the Pearson correlation coefficient is reported for Questions 1-4. Values close to 1 or -1 indicate high positive or negative correlation with seniority of users, respectively.

Questions \ Marks	1	2	3	4	5	N/A
Question 1	0	1	1	2	6	-
Question 2	0	0	4	5	1	-
Question 3	0	0	5	4	1	-
Question 4	0	0	2	7	1	-
SUM	0	1	12	18	9	-

Table 3.2: The overall impression of Dicode Workbench: marginal distribution of responses for Questions 1-4.

3.2.2 Support of data analysis, collaboration and decision making

CGRA evaluators watched the second video-cast provided and completed the second and third set of questions (Sections B, C) of the Dicode Evaluation Questionnaire. We present summarized responses and remarks for questions of Section B in this section and for Section C in the following section (Section 3.2.3). The summarized responses of the evaluators for questions of Section B of the questionnaire provided can be seen in Figure 3.3. Particularly, we present the mode and quartile trend (minimum, median and maximum values) of the

responses relative to the evaluators' opinion for the support of data analysis, collaboration and decision making via the Dicode Workbench. The original answers were given in a 1-5 scale (ranging from 1: strongly disagree to 5: strongly agree). The questions of Section B are:

1. Question 5: DICODE services can help me to deal with data-intensive issues.
2. Question 6: DICODE services can help me to deal with cognitive-complex issues.
3. Question 7: The DICODE workbench can facilitate collaboration.
4. Question 8: The DICODE workbench can enhance decision making.
5. Question 9: The services of DICODE workbench are very well integrated.
6. Question 10: The DICODE workbench can help me be more productive and concentrate on creative activities.

Thus, Questions 5-10 refer to the usefulness and assessment of the integrated into the Dicode Workbench services, i.e. the collaboration and decision-making services as well as to the capabilities of the Subgroup Discovery data-mining service.

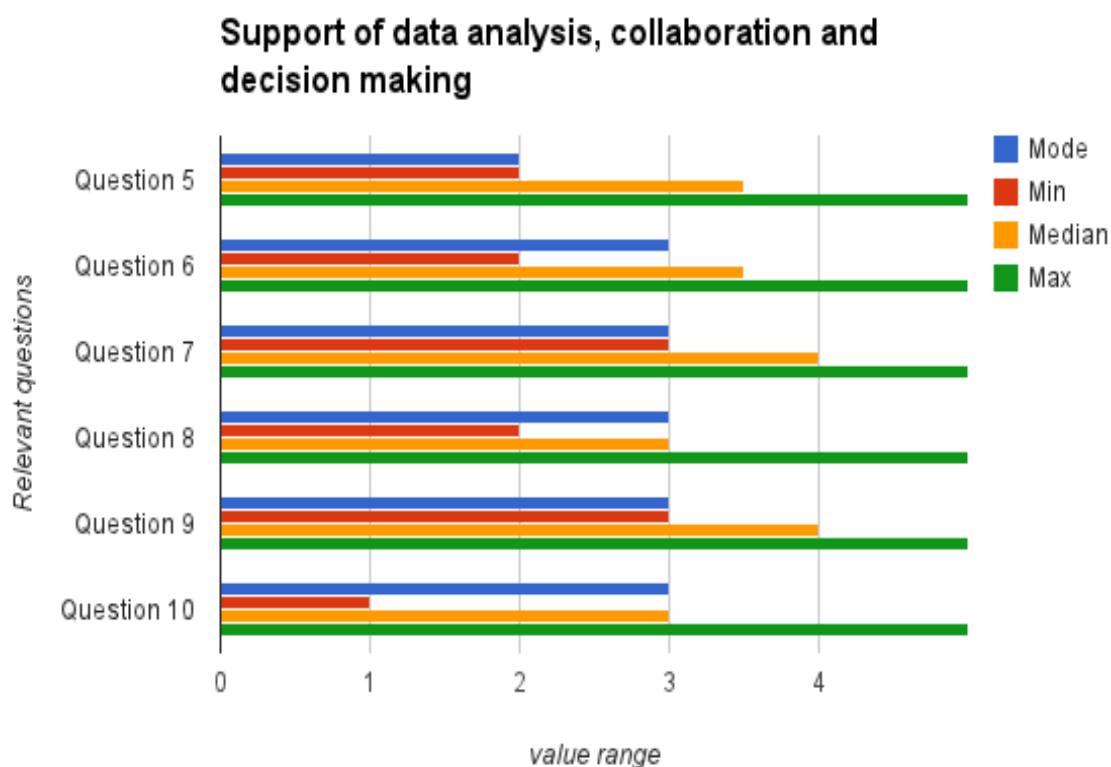


Figure 3.3: Support of data analysis, collaboration and decision-making: mode, minimum, median, and maximum values are presented for Questions 5-10. The range of original values is 1-5, where 1 denotes strong disagreement and 5 denotes strong agreement.

Overall mode/median values range from 2 to 4, where the mode values are the lowest ranging from 2 to 3. Evaluators were more reluctant towards the second set of questions, which is partly explained by the fact that these questions enquire their impression over scientifically significant matters involving data analysis and decision-making. Researchers are always cautious towards the data analysis methods they use, especially given the complexity of the biomedical data. Another reason of their reluctance should be attributed to the sample distribution, as those senior researchers who are confident with their analysis skills are less willing to adopt other analysis routines. Evaluators are less sceptical for the

ability of the Dicode Workbench integrated services to deal with cognitive-complex issues, as well as to enhance collaboration between their peers and in that respect help them to be more productive and concentrate on creative activities. Indicative is a remark given in Question 5:

'If users can upload big data, it could give an indication of data patterns without much effort.'

Also, some positive remarks were given for Question 6 and 7 respectively:

'By discussing with my peers I suppose.'

'I like that it keeps track of ideas, and a history of the project.'

which emphasize on the importance of collaboration between biomedical researchers and in particular for issues which require extended background knowledge. Among the research stages that is supported by the above features is decision-making, maybe the most important step in the research procedure, which is marked with mode and median values equal 3 (nevertheless three responses were above 4).

Evaluators were satisfied with how the services were integrated in the Workbench, however they need more feedback in order to decide whether the Dicode Workbench will help them concentrate on more creative and productive activities:

'More detailed video is required for that answer. Using the workbench could save me time.'

'Productive yes, in terms of ease of facilitating sharing data/results with colleagues.'

Suggestions for further implementation of the services are also given:

'the platform is not yet applicable for proteomics data analysis which is my focus of research, nevertheless, it appears to largely facilitate data comparison and integration'

which is a direction we are currently looking at in collaboration with FHG.

In Figure 3.4, we can observe that seniority significantly affects the responses of the reviewers for Question 9, $r \approx -0.783$, indicating that the degree to which evaluators are satisfied with how well services are integrated in the Workbench is inversely related to seniority. The remaining correlation coefficient values are within the $[-0.531, 0.447]$ interval and do not indicate significant degree of correlation between seniority and questions' outcomes. The above conclusions are supported by chi-squared test p-values, which is also reporting a marginally statistically significant p-value for Question 10 (p-value=0.04043) at a 0.05 significance level. In particular, it seems that for Question 10 less senior evaluators marked with the two far ends, i.e. 1 and 5.

In Table 3.3 we report the marginal distribution of evaluators' responses relative to Questions 5-10 enquiring about data analysis, collaboration and decision making issues of the Workbench integrated services. From the 57 valid answers (3 N/As were reported) we can observe that most of them range in the scale from 3 to 5, with the frequency of $p(\text{mark} \geq 3) \approx 0.719$ and $p(\text{mark} \geq 4) \approx 0.526$. This is in accordance with the mode and median trend values presented above.

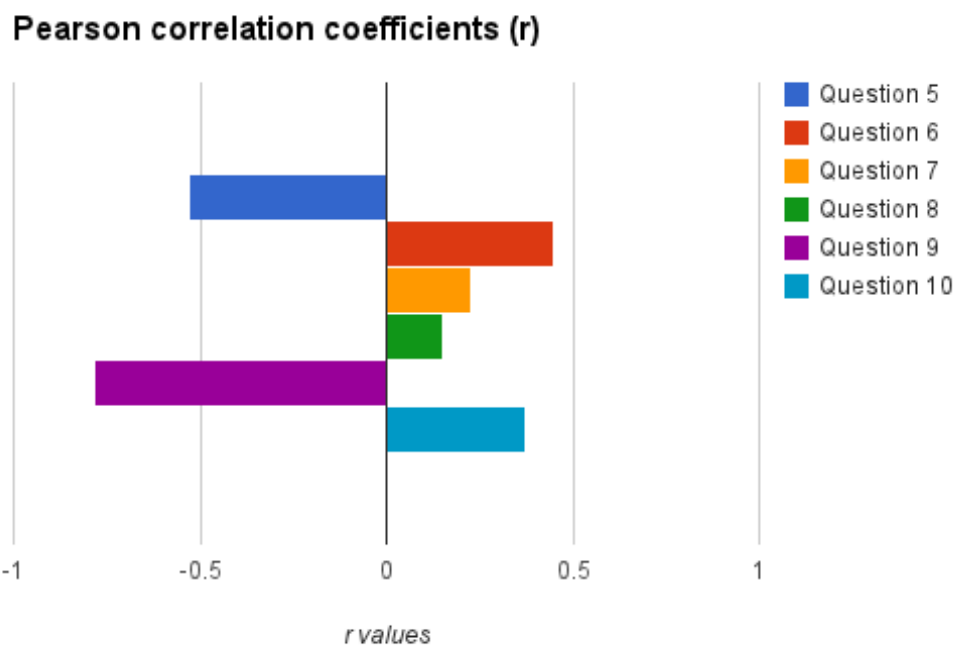


Figure 3.4: Support of data analysis, collaboration and decision-making: the Pearson correlation coefficient is reported for Questions 5-10. Values close to 1 or -1 indicate high positive or negative correlation with seniority of users, respectively.

Questions\ Marks	1	2	3	4	5	N/A
Question 5	0	0	2	7	1	-
Question 6	0	2	2	2	2	2
Question 7	0	0	4	2	4	-
Question 8	0	1	5	3	1	-
Question 9	0	0	4	4	1	1
Question 10	1	2	4	2	1	-
SUM	1	5	21	20	10	3

Table 3.3: Support data analysis, collaboration and decision-making: marginal distribution of responses for Questions 5-10.

3.2.3 Potential benefit to current work practices

CGRA evaluators answered the last set of questions (Section C) based on the two video-casts shown. In this section, we summarize their responses to Questions 11-16 which explore the potential benefit of Dicode to evaluators current work practices. Namely, the relative questions are:

1. Question 11: I can see the potential benefit of using DICODE workbench in my work.
2. Question 12: DICODE provides sufficient services to support my work.
3. Question 13: I would consider using the DICODE workbench in the near future.
4. Question 14: The use of DICODE will have positive impact on my current work practices.
5. Question 15: The use of DICODE will change my current work practices.
6. Question 16: I will recommend the DICODE workbench to my peers/ community.

In Figure 3.5, we present the mode and quartile trend (minimum, median and maximum values) of their responses. The original answers were given in a 1-5 scale (ranging from 1: strongly disagree to 5: strongly agree). We can observe that mode and median values range from 2 to 3. Evaluators are keen to change their current work practices, although they are instances of very low or very high marks for all six questions considered here. Overall, they believe that the Dicode Workbench brings potential benefit to their work (mode, median equal 3) and provides sufficient services to support their work (mode, median values equal 3), whereas they are willing to recommend it to their community (mode, median equal 3). They are more reluctant to use the Dicode integrated services in the near future (mode, median equal 2), *'It will definitely take some getting used to it time...'*, or in changing their current work practices (mode, median values equal 3 and 2 respectively), *'maybe facilitate, but not change'*. This can be attributed partly to the reasons mentioned in the previous section, i.e. there are numerous data analysis techniques in the biomedical field which indicates the need to disseminate Dicode's progress to multi-disciplinary environments.

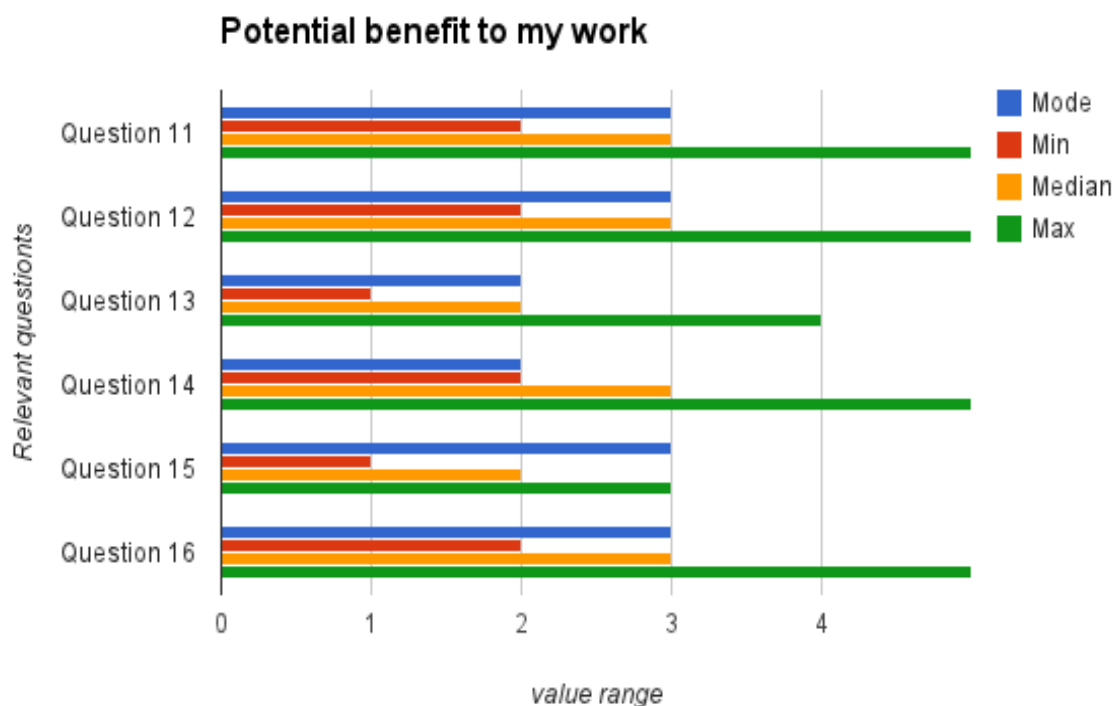


Figure 3.5: Potential benefit to my work: mode, minimum, median, and maximum values are presented for Questions 11-16. The range of original values is 1-5, where 1 denotes strong disagreement and 5 denotes strong agreement.

Specific expected/ foreseen benefits of the use of Dicode Workbench are:

'Information exchange'

'track publications exchange information with collaborators'

whereas an evaluator mentions an area of biomedical research we are currently considering for possible expansion:

'... I hope that the platform will be soon expanded to accommodate additional -omics data. If proteomics data input is established, then I would foresee many beneficial applications, as DICODE integrates various useful tools [.]'

Relative to the sufficiency of the Dicode presented services, remarks collected include the following:

'I can't claim that by the videos- however, it seems promising.'

'I would need to use it to have an opinion. It looks like it needs some getting used to, but after 1-2 tries it should be straight-forward...'

'Extra data mining services will be needed.'

When evaluators were asked to name specific aspects of their current work practices that they foresee/ expect a positive impact with the use of Dicode, they replied:

'Ease of communication between colleagues'

'literature archiving review writing up'

'Archiving'

again emphasizing the need to strengthen exploitation and dissemination activities.

In Figure 3.6, we can observe that the seniority of the evaluators does not significantly affect their responses for questions relative to the benefit of the Dicode Workbench and integrated services; all Pearson correlation coefficients range in [-0.055,0.491]. This is in accordance with the chi-squared test of association examined for all Questions 11:16.

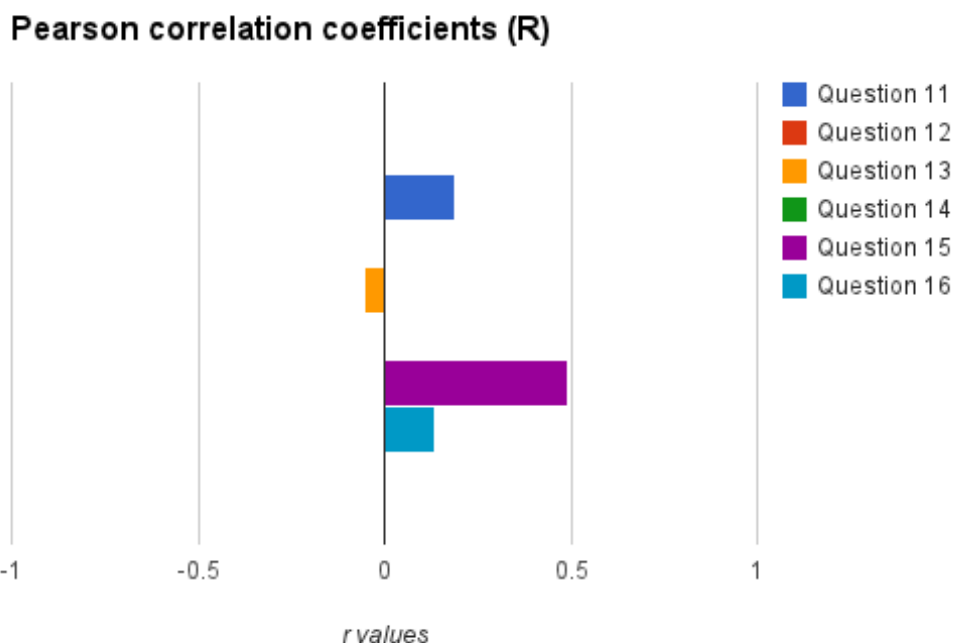


Figure 3.6: Potential benefit to my work: the Pearson correlation coefficient is reported for Questions 11-16. Values close to 1 or -1 indicate high positive or negative correlation with seniority of users, respectively.

In Table 3.4, we report the marginal distribution of evaluators' responses relative to Questions 11-16 enquiring about potential benefit to researchers' work by the Dicode Workbench and integrated services presented in the video-casts. From the 58 valid answers (2 N/As were reported) we can observe that most of them range in the scale from 2 to 3, with the frequency of $p(\text{mark} \geq 2) \approx 0.948$ but $p(\text{mark} \geq 4) \approx 0.19$. This is in accordance with the mode and median trend values presented above.

Questions\ Marks	1	2	3	4	5	N/A
Question 11	0	2	5	2	1	-
Question 12	0	2	6	0	1	1
Question 13	1	7	1	1	0	-
Question 14	0	4	2	2	1	1
Question 15	2	4	4	0	0	-
Question 16	0	1	6	1	2	-
SUM	3	20	24	6	5	2

Table 3.4: Potential benefit to my work: marginal distribution of responses for Questions 11-16.

4 Discussion

Based on the feedback received from the first evaluation phase of the Dicode project, the Dicode workbench was reported to be a promising tool, which facilitates users to set their research objectives and better understand the data and methodologies used in their research. Additionally, actions have been taken to address all comments of the reviewers with respect to goal or task-dependent 'cold-start' problems for the data-mining services and GEORecommender in particular, and also the ranking of SD results based on the importance of the GO categories reported.

On the second evaluation round, emphasis was given to senior real users from the biomedical and bioinformatics domain. We have combined two evaluation methodologies, namely scenario-formed video-casts and questionnaire, based on the suggestions of Cugini et al. (1997), in order to capture experts' judgements about the usage of Dicode services and their overall ratings in an effortless way.

Our goal was to measure the usability, acceptability and functionality of the Dicode workbench and data-mining integrated services, since they are directly facing some important issues the biomedical community needs to cope with, and particularly:

- Data analysis issues,
- Understanding and assessing data analysis findings based on collaboration and decision-making in a multi-disciplinary environment.

The Dicode workbench interface was reported to be intuitive with well integrated services. Evaluators were satisfied with the information provided by the video-cast, although they

reported that extra time was needed to familiarize with the Workbench. They were sceptical about adopting new practices but less sceptical for 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, they reported that the platform offers ease of communication, and strong data/ information archiving features.

Overall, we found that the seniority of the evaluators did not affect their responses to the questionnaire. Evaluators deemed that the Dicode workbench brings potential benefit to their work and provides sufficient services to support their work. Nevertheless, they were reluctant to use the Dicode integrated services.

Future work will concentrate on enriching and finalizing the Dicode services that are related to Use Case 1 by incorporating the detailed feedback of the evaluators. The update of CGRA will be based on the evaluators' suggestions, and particularly:

- Include user-friendly help files;
- Finalize specific technical issues of the Workbench interface, i.e. workspace design and outcome presentation;
- Extend the functionalities of the services to other data sets/types (e.g. developing a more generic version of Subgroup Discovery service to also handle, for instance, proteomics data);
- Further improve the usability of the data-mining services;
- Ensure a wider dissemination of the Dicode workbench.

The approach presented here has potential for adaptation and application to professional evaluators from a wide range of different organisations. In that way, during the remaining period of the project, we aim to present the Dicode Workbench in its finalized form to professionals from different biological backgrounds and investigate how generic it can be, i.e. we aim to identify additional 'application domains' and target users, as also suggested by Spilsbury et al. (2007). Results and feedback will be presented in D6.5 (for all use cases).

References

- [1] Cheng, J., & Cosley, D. (2010) kultagg: Lucid Design for Tagging Interfaces, Proceedings on 16th ACM Intern. Conf. on Supporting group work, GROUP'10, November 7-10, Sanibel Island, Florida, USA, p. 169-172.
- [2] Crockford, D. (2006). The fat-free alternative to XML. *Proc of XML*. Boston.
- [3] Cugini, J., Damianos, L., Hirschman, L., Kozierok, R., Kurtz, J., Laskowski, S., & Scholtz, J. (1997) Methodology for Evaluation of Collaboration Systems, The Evaluation Working Group of The DARPA Intelligent Collaboration and Visualization Program (ICV).
- [4] Jadhav, A.S., & Sonar, R.M. (2009) Evaluating and selecting software packages: A review, *Information and software technology*, 51, 555-563.
- [5] Kirigin, A., Klein, G., & Adelman, L. (2005) Heuristic Evaluation Techniques for Collaborative Software, Technical Report, George Mason University.
- [6] Kitchenham, B., & Pfleeger, S. L. (2002). Principles of Survey Research Part 5: Populations and Samples. *Software Engineering Notes*, 27 (5), 17-20.
- [7] Nielsen, J. (1991). *Designing Web Usability: The Practice of Simplicity*. Indianapolis: New Riders Publishing.
- [8] Norman, D. A. (1998). *The Design of Everyday Things*. The MIT Press.
- [9] Spilsbury, M.J., Perch, C., Norgbey, S., Rauniyar, G., & Battaglino, C. (2007) Lessons Learned from Evaluation: A platform for sharing knowledge, Technical report (specila study paper no 2), United Nations Environment Programme.
- [10] Sun, Y., & Greenberg, S. (2010) Places for Lightweight Group Meetings: The Design of Come Together, Proceedings on 16th ACM Intern. Conf. on Supporting group work, GROUP'10, November 7-10, Sanibel Island, Florida, USA, p. 235-244.

Appendix: Dicode questionnaire - second round

The live form of Dicode's Use Case #1 CGRA was uploaded via Google Docs platform and can be found in <https://docs.google.com/forms/d/1XzRR0ealwUbYnoCmDp8m-4Uf8VAqrDXbakwlCrsTZqU/viewform>. The Questionnaire includes:

- an introduction to the Dicode project,
- a description of Dicode project's objectives,
- a description of Dicode project's evaluation objectives,
- an introduction to the accompanied videos explaining the walkthrough video scenario presented,
- links to the two accompanied videos,
(<http://hodgkin.dia.fi.upm.es:8080/dicode/workbenchVideo2.avi>,
<http://dicodedev.cti.gr/screencast/screencast.html>),
- a set of 16 questions aiming to explore the functionality of the Dicode Workbench, the functionality of the integrated services and their future potential under Use Case #1. The services considered are:
 - 1 Workbench – collaboration service
 - 2 Workbench decision-making service
 - 3 Data mining – Subgroup Discovery service

We hereby include the whole body of the questionnaire as presented to the evaluators of the 2nd evaluation round of the Dicoed project – Use case #1.



Tell us what you think!

We appreciate your input on the DICODE EU FP7 project. The following questionnaire will help us collect your feedback and understand what aspects of the DICODE workbench and integrated services satisfy you.

Before filling in this questionnaire please watch the accompanied screen casts, which show on a step-by-step basis the log-in process to the DICODE workbench and the options offered for intuitive collaboration and data analysis via the workbench integrated services.

Think about all the tasks that were presented at the screen casts while you answer these questions. Please read each statement and indicate how strongly you agree or disagree with the statement by ticking a number on the scale. If a statement does not apply to you, please tick N/A.

Whenever it is appropriate, please write comments to explain your answers.

Thank you for your participation!



Introduction to DICODE: [DICODE](#) aims to facilitate and augment collaboration and decision making in data-intensive and cognitively-complex settings. A special use case of the project focuses on supporting multidisciplinary biomedical communities ranging from biologists to bio-informaticians. In this context, the DICODE workbench provides certain tools as well as an overall environment to collaboratively explore, evaluate, disseminate and diffuse relative scientific findings, results and research objectives in the clinico-genomics field.

DICODE Objectives: The DICODE workbench aims to support interdisciplinary collaboration and decision-making by facilitating meaningful data mining and integration, under a common workbench.

DICODE's evaluation: This task aims to assess how the DICODE workbench can improve the processing of voluminous and complex clinico-genomics data in collaboration and decision making settings. Through this process the DICODE workbench and its integrated work methodologies will be evaluated towards their usability and acceptability from the scientific community.

Screen casts walkthrough: You are a scientist working at the clinico-genomics field as part of your everyday practice; consider the case that you want to analyze two sets of data and biologically interpret your findings. You would also like to discuss with your peers the workflow of your analysis, the methodologies applied and the results produced, since you find that their different scientific backgrounds and experience will add value to your research. For that reason you decide to launch the DICODE workbench and create a collaborative session where you can analyze your data, and discuss the results with your colleagues (possibly from other disciplines).

The following screen casts will help you go through some of the steps that need to be followed when using the DICODE workbench to analyze genomics data.

Please answer the first set of questions (questions 1-4) after watching screen cast 1 and the two remaining sets (questions 5-16) after watching and screen cast 2.

Screen casts:

1. The [first screen cast](#) presents the log-in process along with the functionalities of the DICODE workbench

<http://hodgkin.dia.fi.upm.es:8080/dicode/workbenchVideo2.avi>

2. The [second screen cast](#) presents an overview of the data mining service functionalities as employed in the DICODE workbench. Gene expression data are analyzed while emphasis is given on the collaboration and decision-making procedure which can be followed whilst launched on the DICODE workbench.

<http://dicodedev.cti.gr/screencast/screencast.html>





Mastering Data-Intensive Collaboration and Decision Making

A. Overall impression of DICODE workbench

1. The information provided on the video is clear.

Strongly disagree 1 2 3 4 5 Strongly agree

Add comments if any.

2. The design of the DICODE workbench is very pleasant.

Strongly disagree 1 2 3 4 5 Strongly agree

Please name any specific features you did not like so much.

3. The use of DICODE workbench is easy.

Strongly disagree 1 2 3 4 5 Strongly agree

Add comments if any.

4. The user interface of DICODE workbench is intuitive.

Strongly disagree 1 2 3 4 5 Strongly agree

Please indicate any alterations you would suggest.

B. Support of data analysis, collaboration and decision making

5. DICODE services can help me to deal with data-intensive issues.

Strongly disagree 1 2 3 4 5 Strongly agree

Add comments if any.

6. DICODE services can help me to deal with cognitive-complex issues.

Strongly disagree 1 2 3 4 5 Strongly agree

Add comments if any.

7. The DICODE workbench can facilitate collaboration.

Strongly disagree 1 2 3 4 5 Strongly agree

Add comments if any.

8. The DICODE workbench can enhance decision making.

Strongly disagree 1 2 3 4 5 Strongly agree

Add comments if any.

9. The services of DICODE workbench are very well integrated.

Strongly disagree 1 2 3 4 5 Strongly agree

Add comments if any.

10. The DICODE workbench can help me be more productive and concentrate on creative activities.

Strongly disagree 1 2 3 4 5 Strongly agree

Please name specific activities you foresee to perform using DICODE workbench.

C. Potential benefit to my work

11. I can see the potential benefit of using DICODE workbench in my work.

Strongly disagree 1 2 3 4 5 Strongly agree

Please name specific foreseen/ expected benefits:

12. DICODE provides sufficient services to support my work.

Strongly disagree 1 2 3 4 5 Strongly agree

Add comments if any

13. I would consider using the DICODE workbench in the near future.

Strongly disagree 1 2 3 4 5 Strongly agree

Add comments if any

14. The use of DICODE will have positive impact on my current work practices.

Strongly disagree 1 2 3 4 5 Strongly agree

Please name specific aspects of your current work practices where you foresee/ expect a positive impact:

15. The use of DICODE will change my current work practices.

Strongly disagree 1 2 3 4 5 Strongly agree

Please provide specific examples of the foreseen/ expected changes:

16. I will recommend the DICODE workbench to my peers/ community.

Strongly disagree 1 2 3 4 5 Strongly agree

Add comments if any

Thank you for your participation!