



## 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**

## D6.2.1 – Report from the evaluation of use case #1 (first version)

**Due date of deliverable:** 29 February 2012  
**Actual submission date:** 20 March 2012  
**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:** biomedical research, bioinformatics, statistical datasets, databases, data management, collaboration, decision making



The Dicode project ([dicode-project.eu](http://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



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

Document history			
Version	Date	Status	Modifications made by
1	10-02-2012	First draft	Anastasia Kastania, BRF Georgia Tsiliki, BRF
2	12-02-2012	Corrections	Sophia Kossida, BRF
3	20-02-2012	Incorporate evaluation results	Anastasia Kastania, BRF Georgia Tsiliki, BRF
4	29-02-2012	Sent to internal reviewers	Manolis Tzagarakis, CTI Guillermo de la Calle, UPM Anastasia Kastania, BRF
5	10-03-2012	Input from UOL incorporated	Fan Yang-Turner, UOL Lydia Lau, UOL Vania Dimitrova, UOL Anastasia Kastania, BRF
6	13-03-2012	Reviewers' comments incorporated, sent to SC	Anastasia Kastania, BRF Georgia Tsiliki, BRF
7	16-03-2012	SC's comments incorporated	Georgia Tsiliki, BRF Anastasia Kastania, BRF Sophia Kossida, BRF
8	20-03-2012	Approved by SC, sent to the Project Officer	Georgia Tsiliki, BRF Anastasia Kastania, BRF Sophia Kossida, BRF

### Deliverable managers

- Anastasia Kastania, Sophia Kossida, Georgia Tsiliki, BRF

### List of Contributors

- Anastasia Kastania, BRF ( (i) Dicode services user trials organization, (ii) Dicode services user trials implementation in Greece, (iii) Redesign of questionnaires of D6.1. according to EC Reviewers comments, (iv) Design of the Dicode user trials protocol, (v) Construction of the Dicode evaluators manual, (vi) Production of the descriptive statistics evaluation results (statistics for deliverable body and appendices both for UC1 & UC3), (vii) Implementation of the reliability analysis (both for UC1 & UC3))
- Georgia Tsiliki, Sophia Kossida, BRF
- Spyros Christodoulou, Nikos Karacapilidis, CTI
- Guillermo de la Calle, Miguel García-Remesal, UPM
- Daniel Trabold, FHG
- Doris Maassen, NEO
- Fan Yang-Turner, Lydia Lau, Vania Dimitrova, UOL ( (i) Dicode services user trials organization, (ii) Redesign of questionnaire of D6.1, (iii) Production of the descriptive statistics evaluation results for Forum Summarization service)

### List of Evaluators

- Manolis Tzagarakis, CTI
- Guillermo de la Calle, UPM
- Anastasia Kastania, BRF

## Summary

This deliverable is to be considered as a first version of the evaluation of Use Case 1 Clinico-Genomic Research Assimilator and particularly the evaluation of the services already developed in the Dicode platform. The evaluation process is performed by using properly formulated metrics and instruments which are described in D6.1 based on the specifications of D2.2. The deliverable describes the evaluation results of the Use Case 1 related collaboration services, data mining services and decision making support, as they are currently developed and accommodated under the Dicode platform. The document also includes an updated description of Use Case 1. Although the project's evaluation is an ongoing progress, the assessment of the already developed services raised interesting technical and methodological issues.

## Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>6</b>
1.1	Context .....	6
1.2	Objectives.....	6
1.3	Structure.....	7
<b>2</b>	<b>Overview of Dicode Use case #1 .....</b>	<b>7</b>
2.1	Description of Dicode Use case #1 .....	8
2.2	Input data.....	10
2.3	Output data.....	10
2.4	Dicode’s implemented services .....	10
<b>3</b>	<b>Evaluation process and feedback.....</b>	<b>11</b>
3.1	Evaluation process.....	11
3.2	Detailed evaluation feedback.....	12
3.2.1	Workbench service (includes the Storage service) .....	12
3.2.2	Collaboration and Decision Making Support services (integrated in the Workbench service).....	16
3.2.3	Forum Summarization service .....	19
3.2.4	SubgroupDiscovery service.....	22
3.2.5	Recommender service .....	26
3.3	Reliability analysis.....	30
<b>4</b>	<b>Future work directions .....</b>	<b>31</b>
	<b>References.....</b>	<b>33</b>
	<b>Appendix: Dicode Evaluator’s Guide.....</b>	<b>35</b>

# 1 Introduction

## 1.1 Context

This deliverable presents the evaluation of the initial version of the developed Dicode services relative to Use Case #1 Clinico-Genomic Research Assimilator (CGRA), as these have been designed and developed in the context of WP2 - WP5 of the Dicode project. Specifically, it reports the evaluation outcome of the Dicode services by summarizing, in a statistical manner, the answers of the internal reviewers and CGRA evaluators given to the evaluation forms and questionnaires shown in the Appendix (up-to-date version of D6.1 Appendices C (part II), G, H).

This is the first of two deliverables reporting on the validation and assessment of the developed Dicode services through the CGRA use case. While the focus of this deliverable (D6.2.1) is on the first round of evaluation of the CGRA Dicode services, deliverable D6.2.2 (due in month 32) will present the evaluation of their enhanced version and also comment on how the feedback provided in the first evaluation round was taken into account. The project's evaluation activities will be summarized in deliverable D6.5 (due in month 36), together with an assessment of the project from external experts/ bodies.

## 1.2 Objectives

The purpose of this document is to present the first evaluation round of the developed Dicode services related to the CGRA use case in the context of Task 6.3, as these originate from the functional specifications outlined in the deliverables D2.2 and implemented in the deliverables D3.1.1, D3.2.1, D4.1.1, D4.2.1, D5.3.1 and D5.4.1, in order to assure their usability and accessibility. The cost-effectiveness and readiness for the market of Dicode services are also assessed.

With the CGRA use case Dicode aims to support clinico-genomic scientific research in the today post-genomic epoch. An enhanced description of CGRA is presented in this document to emphasize the current practice and collaboration options now available to the biomedical scientific community. The Dicode suite of services and innovative work methodologies are evaluated based on how they can improve the processing of voluminous and heterogeneous biomedical data in collaboration and decision making settings, so as to assist social sharing and collective sense making.

The evaluation of the Dicode services is performed based on a series of Key Success Indicators (KSIs) implemented 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 revise the Dicode suite of services and specifications as well as to improve their quality.

### 1.3 Structure

Section 2 starts with an enhanced description of CGRA, followed by a detailed description of possible options for the input and output data. In Section 3, information about the evaluation process of CGRA is provided, along with detailed evaluation feedback and summary statistics for the evaluation results. Section 4 discusses the benefits of the evaluation procedure and possible directions for the revised Dicode services.

## 2 Overview of Dicode Use case #1

The field of biomedical research has recently seen a vast growth in publicly available biomedical resources, including multiple types of datasets and databases. A major advance is that now researchers have access to complementary views of a single organism by analyzing multiple types of data, including whole genome sequencing, expression profiling and other high-throughput experiments. Those data which are often called ‘-omics’ data, include the genome sequencing data (genomics), the complete set of RNA transcripts produced by the genome and analysed via microarray, Real-time PCR or Next-Generation Sequencing platforms (transcriptomics), protein structures and function (proteomics), or any other data available for the organism under study, and provide novel views of cellular components in the biological systems (Tsiliki & Kossida, 2011). As a consequence, an enormous amount of digital content is produced everyday (i.e. information that is created, captured, or replicated in digital form as well as hundreds of analysis systems), resulting in high rates of new information being distributed and demanding attention (Karacapilidis, Tzagarakis, Christodoulou, & Tsiliki, 2012). This enables biomedical researchers to have instant access to a great wealth of resources that are critical for supporting their research.

Most of those data sets are well organised in publicly available databases, although there are existing limitations in accessing, storing, mapping and managing the increasing amount of data available (Sullivan, Gabbard, Shukla, & Sobral, 2010). Managing the amount and diversity of data that such experiments produce is a task that must be supported by appropriate algorithmic analysis and software tools (Koschmieder, Zimmermann, Tribl, Stoltmann, & Leser, 2011). Moreover, there have been attempts to algorithmically unify the data above (Lukk, et al., 2010) and their supplementary views (Joyce & Palsson, 2006). However, choosing the right datasets, databases and tools for a given project is difficult even for an expert, which increases the importance of the ability to handle the data (Pennisi, 2011). To address such limitations cloud and distributed computing, schema-free solutions, domain-specific and process-oriented programming languages or special algorithmic solutions are applied (Huttenhower, Schroeder, Chikina, & Troyanskaya, 2008; Pennisi, 2011; Baker, 2012). Overall there is a growing need for data and computing resources to be readily reused, repurposed and extended by other scientists (Baker, 2012). For instance, the well known Galaxy Project (<http://galaxy.psu.edu/>) offers a web-based platform allowing researchers to perform and share complete analyses, and the GenePattern platform (<http://www.broadinstitute.org/cancer/software/genepattern/>) provides access to more than 180 tools for genomic analysis to enable reproducible *in silico* research. Other attempts, such as BioMart (<http://www.biomart.org/>) or its cancer specialized version IntOGen (<http://www.intogen.org/home>), focus on linking biological databases (Guberman, et al., 2011; Perez-Llamas, Gundem, & Lopez-Bigas, 2011).

Within this environment, biomedical research has become increasingly interdisciplinary and collaborative in nature (Lee, 2007; Baker, 2012). The increasingly specialized resources show

that the way forward is to form biomedical research collaboration teams to address complex research questions. Such interdisciplinary teams would better meet challenges relative to various problems such as how to store, access, analyze and integrate multiple types of data (Pennisi, 2011); or, how to work with multiple databases simultaneously (Finholt, 2003); or even, how to make data accessible and usable to life sciences researchers (Sullivan, Gabbard, Shukla, & Sobral, 2010). In addition, tools facilitating sense- and decision-making by appropriately capturing the collective intelligence that emerges during such collaboration are lacking. Biomedical researchers need such tools to efficiently and effectively collaborate and make decisions by appropriately assembling and analyzing enormous volumes of complex multi-faceted data residing in different sources. Supporting team collaboration under such circumstances is still considered a challenging task (Spencer, Zimmerman, & Abramson, 2011).

To address those issues, Dicode aims to support the entire life cycle of biomedical collaboration by supporting clinical researchers and bio-scientists allowing them to easily examine, reuse and interpret heterogeneous clinico-genomic data and information sources, as well as to reach decisions for the production of new insightful conclusions, without having to worry about the method of locating and assembling huge quantities of data. Currently, data considered are genomic, clinical, gene ontology data (analytically shown in Table 2.2.1), nevertheless, the project aims to include transcriptomic/ proteomic data and facilitate collaboration specific needs for any sub-field of biomedical science. The present evaluation procedure will assess the degree of Dicode's flexibility to serve all members of the collaborating group by creating or capturing and propelling conquered and reliable biomedical knowledge.

## 2.1 Description of Dicode Use case #1

The CGRA Dicode Use Case #1 aims to provide certain tools as well as an overall environment to assimilate clinico-genomic research information, achievements and findings, in order to support a multidisciplinary biomedical community ranging from biologists to bio-informaticians. The intersection of the practices, methodologies and findings of the clinical, biological, statistical and information sciences could bring innovative biomedical research findings to the everyday practice of clinics (Kawamoto, Lobach, Willard, & Ginsburg, 2009). In this context, CGRA aims to collaboratively explore, evaluate, disseminate and diffuse relative scientific findings, results and research objectives in the clinico-genomics field.

CGRA is built as a general research assimilator environment which will be able to handle large data such as plant data, although currently the data and practices considered involve breast cancer data disease. The developed CGRA services aim to underline interdisciplinary collaboration and decision-making by facilitating information integration under a common platform in the following ways:

- identify data and annotation databases – data repositories and annotation data to augment the data size and relative information,
- social sharing – exchange valuable experiences and tacit knowledge,
- data sharing – exchange 'similar' or important data for augmenting data size,
- manipulate large datasets – provide means to efficiently handle large amounts of data and avoid 'out-of-memory' errors,
- tacit knowledge – provide suitable means to establish and record expert's knowledge and experience through a trustworthy evaluation framework,

- analysis practices/tools - building and sharing predictive models for data analysis, as well as integration models or templates for merging heterogeneous data/sources,
- efficient use of expertise and results – provide support in decision-making if possible throughout the entire biomedical task in a collaborative setting,
- efficient presentation and grouping of tools/results – suggest ranking of tools, services or users based on relevancy to the issue at hand,
- biomedical resources sharing and interconnection – in summary, provide a social collaborative network for interconnection and interaction between biomedical researchers, tools, data and resources.

Towards the accomplishment of this goal, a number of services will be further developed in order to fully address the relevant requirements of CGRA Dicode as outlined in D2.2. This document aims to demonstrate how the initial version of the developed Dicode services can support clinico-genomics research. Particularly, this document reports on the initial version of CGRA assimilator developed based on the context of Task 6.3 which envisages to

- plan an integrated clinico-genomic knowledge discovery and decision making use case,
- accomplish integration of both heterogeneous data clinico-genomic data sources and advance analytical techniques.

Data type	Data description	Web sites	Data in numbers	Data in size
<b>Gene-Expression profiles (GEP)</b>	Gene-expression data (normalized or raw data)	Gene Expression Omnibus ( <a href="#">GEO</a> )	86 datasets – 7,607 samples	~ 500Kb per sample ~ 32Mb per dataset
		<a href="#">ArrayExpress</a>	978 experiments – 69,483 samples	
		Stanford Microarray Database ( <a href="#">SMD</a> )	508 experiments	
<b>Phenotypic data</b>	Supplementary, clinical or phenotypic data available	Same as GEP or from original publication's sup. files	2 files on average per dataset	~10Kb per dataset
<b>Molecular Pathways (MP)</b>	Data from known and established molecular networks	Kyoto Encyclopedia of Genes and Genomes ( <a href="#">KEGG</a> )	416 pathway maps (153,758 total)	
		<a href="#">Reactome</a>	3,931,211 data entries	
<b>Annotation data</b>	Reference databases for biomedical & genomic information	Gene Ontology ( <a href="#">GO</a> )	~ 30,000 terms ~ 50,000 relationships	
		National Center of Biotechnology Information (NCBI) : 1. <a href="#">RefSeq</a> 2. <a href="#">UniGene</a> 3. <a href="#">GenBank</a> 4. <a href="#">PubMed</a>	1. 26,473 annotated coding regions, 2. 129,493 homo sapiens entries, 3. ~ 126,551,501,141 bases, 4. > 21 million citations for biomedical literature	

**Table 2.2.1** : Input data considered for CGRA. Note that GEO is an NCBI repository which is separately presented here since it is one of the main CGRA's data sources.

## 2.2 Input data

CGRA is planned and materialized to seamlessly link and mine disparate clinico-genomic data sources. Currently the data resources incorporated into the Dicode services are provided in four different types of data shown in Table 2.2.1. For the initial version of the Dicode suite of services the breast-cancer (BrCa) case is targeted but the scenario could be generalized to other diseases (e.g. cardiovascular disease) or organisms (e.g. plant data). Another database we intend to incorporate into CGRA is PubMed citations (<http://www.ncbi.nlm.nih.gov/pubmed/>) for access to scientific journals and online books.

## 2.3 Output data

CGRA output options are directly related to the extent of the use of the Dicode platform. The use case is founded on an integrated knowledge discovery scenario that amalgamates and maps gene-expression profiles onto gene-regulatory networks in an aim to uncover molecular regulatory mechanisms that govern target phenotypes. The Dicode platform provides the means to access and critically assess the essential resources and tools in a common interface which bundles all functionalities together. Having said that, the output of the developed Dicode services covers the following tasks:

- facilitate and augment the collaboration and decision-making in data-intensive and cognitively-complex settings,
- facilitate in sharing knowledge on data, analysis methodologies or web tools,
- apply methodologies to derive predictive disease models (diagnostic and prognostic) such as feature selection and classification analysis,
- utilize mining methodologies in the context of clinico-genomic research field.

## 2.4 Dicode's implemented services

CGRA aims to integrate information and knowledge in a clinico-genomics setting by taking advantage of the collaboration options offered, where the current working example is to uncover molecular regulatory mechanisms that govern target phenotypes. Towards this goal, initial versions of the following Dicode services were developed, based on their full description outlined in D2.2:

- **Workbench service** – a web application built for all Dicode use cases which provide a common graphical interface to access and use heterogeneous services. The Dicode workbench is the integration platform for all Dicode data analysis and collaboration services (D4.1.1, D4.2.1, D5.4.1). Within the Workbench the **Storage service** is included, an integrated service built for all Dicode use cases to comfort the sharing and exchange of information (files, reports, etc.) in data-intensive and cognitively-complex settings. The service provides all functionalities needed to allow permanent and reliable storing of files as well as their accessibility (D5.4.1).
- **Collaboration & Decision Making Support services** – integrated services built for all Dicode use cases. The **Collaboration Support** service exploits the reasoning abilities of humans to facilitate sense-making of the Dicode data mining services results and capitalize on their outcomes (D4.1.1, D4.2.1). The **Decision Making Support** service turns information and knowledge, currently only for a limited set of fixed discourse moves and message types, machine interpretable in order to allow active participation of the system in collaborative activities (D4.1.1, D4.2.1).
- **Forum Summarization service** – a standalone service built for the needs of CGRA and Dicode Use Case #3. The service takes a cluster of discussion threads as input

from relevant public forums and identifies the most prominent terms (topics) in the discussion threads that belong to the cluster (D2.1, D4.2.1). The identified topics can be used by the users to identify what the main theme and the discussion threads in that cluster are about.

- **SubGroupDiscovery service** – a standalone service built for the needs of CGRA and Dicode Use case #3. The service searches for subgroups in any user provided data by searching the rules that cover many target value examples and few non-target value examples. The service supports two different Subgroup Discovery algorithms (D3.2.1, D4.1.1, D4.2.1).
- **Recommendation service** – a standalone service built for all Dicode use cases. The service recommends similar users or documents from log file data (D3.2.1, D4.1.1, D4.2.1) based on similarity models learned by using the Dicode Similarity Learning Service.

During the evaluation process the Workbench service was assessed as an integrated service together with the Storage, Collaboration Support and Decision Making Support services, however particular emphasis was given to the Collaboration Support service of the Workbench since the evaluation was conducted through a CGRA specific collaboration scenario. The following section includes a description of the evaluation process, statistical summaries of the evaluators' answers and representative free text comments.

## 3 Evaluation process and feedback

### 3.1 Evaluation process

The Dicode user trials were conducted up to now successfully in the four participating countries of the project using the Dicode Evaluation Framework. Use case partners and technical partners invited participants to join the user trials. Due to the fact that the majority of the up to now developed RESTful Dicode Services produce 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. The sampling method used was Snowball sampling (Kitchenham & Pfleeger, 2002).

The Dicode Services Evaluators volunteer sample finally consists of 44 evaluators from Greece, 13 evaluators from Spain, 3 evaluators from Germany and 1 evaluator from the United Kingdom. To assure the fairness of the Dicode Services trials, as pointed by the EC Reviewers of the Project the following strategy was applied:

- Ethical issues were applied in trials design with human participants (the right not to suffer harm, the right to full transparency, the right of self decision, the rights of privacy, anonymity and confidentiality).
- The three questionnaires of the D6.1. that were applied in the conducted survey research were adapted including evaluator's confidence measurement in most of the qualitative questions including comments for the optimization of the Dicode Services. Two more questionnaires were designed and tested for the Forum Summarization Service.

The summaries of the evaluation analysis results presented in the sections 3.2- 3.3 have been produced using the Statistical Package for the Social Sciences (SPSS 17v) routines.

The Dicode Evaluators' Guide with instruction on how to fulfil their task is presented in the Appendix.

### 3.2 Detailed evaluation feedback

This section includes summary statistics for all CGRA use case evaluations as well as representative qualitative feedback from the questionnaires. The table and figures included summarize the quantitative evaluators' responses relative to Use Case Validation/Service evaluation instrument, ease of use, accessibility & acceptability of Dicode suite of services. When specific service evaluation is considered, evaluators assess the scenario provided, the acceptability and the overall quality of the service. Answers to the quantitative questions of the questionnaires are given for ordinal data in a 1-5 scale (Quality, Acceptability and Accessibility questionnaires), where 1 stands for 'I strongly disagree' and 5 for 'I strongly agree', and for continuous numerical data (scale data) in a 0-10 scale (Usability questionnaire), where 0 stands for 'none' and 10 for 'excellent' (Nielsen, 1991; Norman, 1998). The evaluation outcome is presented for each service separately. Note that missing data were not imputed.

The detailed statistics relevant to the completed questionnaires collected for the three Dicode Use Cases' evaluation are uploaded in the Dicode wiki:  
<https://wiki.dicode-project.eu/display/DIC/WP6+Evaluation+results>.

#### 3.2.1 Workbench service (includes the Storage service)

CGRA evaluators used the service and completed the related Dicode Evaluation Framework Questionnaires. Figures 3.2.1.1, 3.2.1.2, 3.2.1.3, summarize the evaluators' responses relative to the quality, acceptability, accessibility and usability of the Workbench Integrated services. In these figures, the median and mode values of the evaluators' responses are shown; the original answers were given in a 1-5 scale (ranging from 1: strongly disagree to 5: strongly agree). For the overall quality of the service, the evaluators agree that the objectives of the service are met (median and mode values equal 4), that the service is novel to their knowledge (median=4, mode=5), they are satisfied with the performance of the service (median=4, mode=4) and they are overall satisfied with this service (median=4, mode=4). It is worth noticing that the highest mode value is reported for the novelty of the service (mode=5). The evaluators seem to be to some extent sceptical as to whether the service addressed the data intensive decision making issues (median and mode values equal 3). A particular comment relatively to the data-intensiveness addressed by the service is the following: *'In theory. Practical there is not so much to do at the moment'* (DE01).

As long as the acceptability of the service is concerned, the evaluators overall agree in that the service has all the functions they expected (median=4, mode=4), the interface of the service is pleasant (median=4, the highest mode value of 5 is reported here) and that they will recommend this service to their peers/community (median=4, mode=4): *'The workbench shows promising functionalities and well-presented all accessible options, but seems to be a bit green at the moment.'* (ES12). Evaluators' responses are just above the average (3 median and mode values) concerning finding the particular service very helpful to their work and also regarding their intention to use this service in their work. Evaluators' responses vary for the usefulness of the service to their work ranging from *'I'm not really sure about the possibilities of this service to my work.'* (ES11), to *'Yes, I found the service very helpful to my work'* (GR19) or *'It could be adopted by many inter-/multidisciplinary communities'* (GR34). Nevertheless they found the interface *'user friendly'* (GR16) with some flaws: *'The many IFrame are not very nice, since scrolling in such small areas is not comfortable.'* [sic] (DE3).

For Figure 3.2.1.3 we can observe that the evaluators are satisfied with the access given to the service, the instructions given and the time spent to learn to use the provided solution (in all three cases median=4, mode=4). The lowest mode scores for accessibility and other issues, i.e. 3, were reported for the ease of use and the confidence regarding the evaluators' confidence for their answers for this service.

In Table 3.2.1.1, the mean values of the responses relative to usability are given; the scale in this case is 0-10. Particularly, the following usability measures are assessed: visibility, tolerance, physical mapping, restrictions, conceptual models, feedback, error prevention, flexibility, ease of recognition, flexibility of the use efficiency, provision of clear error messages, aesthetic of the minimalist design, help and documentation facilities, visibility, user control capabilities, consistency and presentation standards, fault prevention, ease of detection. Overall the mean values are above 6.26 across categories, with relatively low variance values (1.926 for the particular case). Low values are reported for the error prevention, flexibility regarding the variety of operation modes and in the provision of clear error messages. The highest scores were reported for the visibility, tolerance, physical mapping and maintain the same presentation of the interface, with low standard deviation values indicating that evaluators generally agreed on the assessment of those usability aspects.



**Figure 3.2.1.1:** Workbench services' quality. The quality evaluation for the Workbench service is shown.

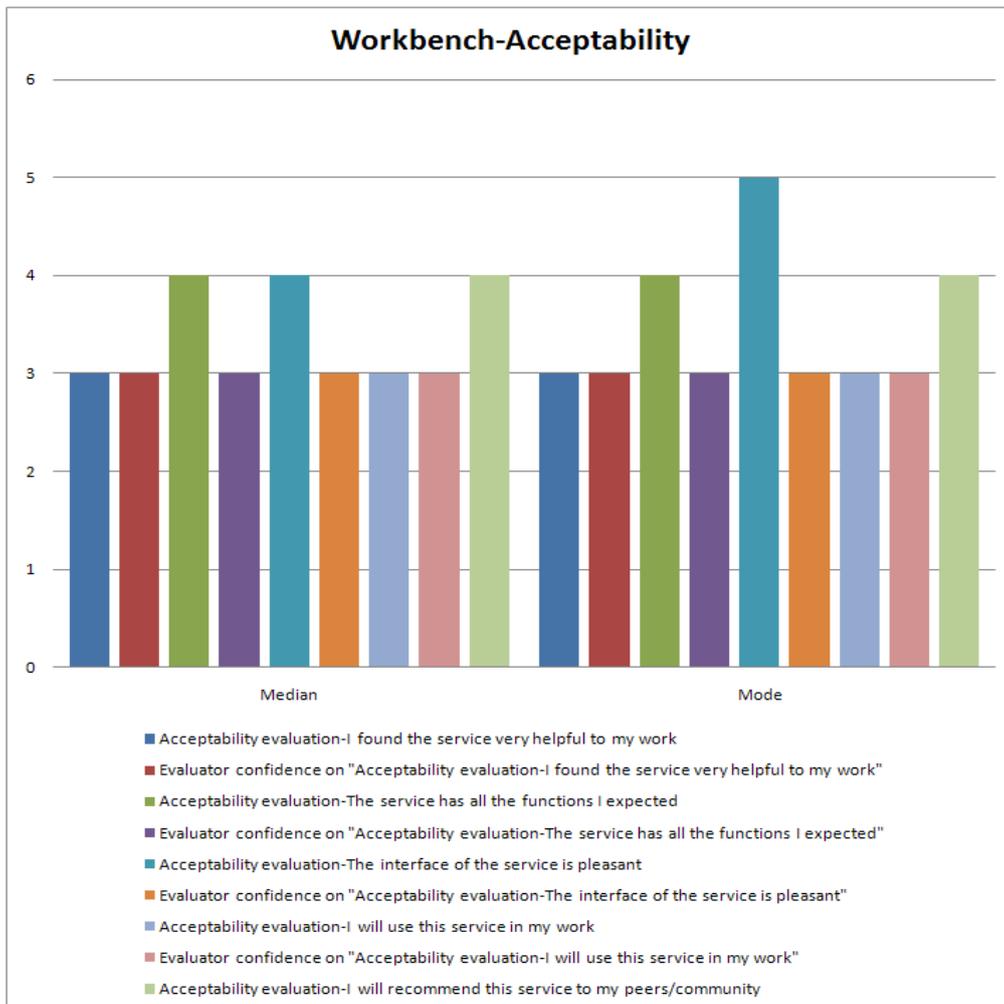


Figure 3.2.1.2: Workbench services' acceptability.

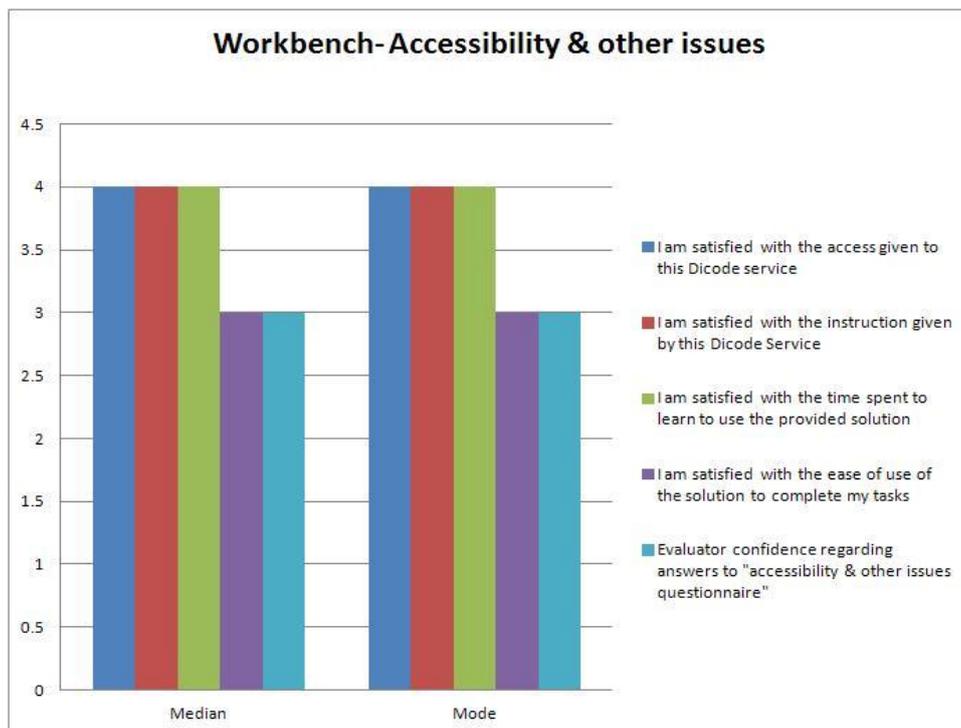


Figure 3.2.1.3: Workbench services' accessibility & other issues.

Questions	N	Minimum	Maximum	Mean	Std. Deviation
Usability principles (Norman, Nielsen)-Rate the visibility: optical information to facilitate the user	42	3	10	7.31	1.774
Usability principles (Norman, Nielsen)-Rate the tolerance: behavior similar to expectations	42	3	10	7.48	1.565
Usability principles (Norman, Nielsen)-Rate the physical mapping: conceptual correspondence between commands and functions	42	3	10	7.50	1.596
Usability principles (Norman, Nielsen)-Rate the restrictions: incomplete ways in the actions performed and in the design of operations in understandable manner	42	2	10	6.48	2.233
Usability principles (Norman, Nielsen)-Rate the conceptual models: the operation of the proposed actions according to the perception of user for these actions	42	3	10	7.12	1.783
Usability principles (Norman, Nielsen)-Rate the feedback: notification regarding the user's position	42	2	10	6.95	1.912
Usability principles (Norman, Nielsen)-Rate the error prevention: restrict user errors & support for their solution	42	2	9	6.26	1.926
Usability principles (Norman, Nielsen)-Rate the flexibility: variety of operation modes	42	2	9	6.33	2.056
Usability principles (Norman, Nielsen)-Rate the ease of recognition: easy identification of the required actions	42	2	10	6.67	1.921
Usability principles (Norman, Nielsen)-Rate the flexibility of the use efficiency: shortcuts provision, capabilities for the user to affect the configuration of the system	42	0	10	6.52	2.371
Usability principles (Norman, Nielsen)-Rate the provision of clear error messages: simple language in error messages and a proposal to resolve them	42	0	10	6.33	2.515
Usability principles (Norman, Nielsen)-Rate the aesthetics of the minimalist design: messages with the necessary information	42	3	10	7.00	1.808
Usability principles (Norman, Nielsen)-Rate the help facilities and the documentation facilities: help facilities related to user action	42	1	9	6.36	2.034
Usability principles (Norman, Nielsen)-Rate the visibility of the system status: user briefing regarding the work progress	42	1	9	6.81	1.903
Usability principles (Norman, Nielsen)-Rate the system matching with the real world: comprehensive and explanatory language	42	2	10	7.10	1.819
Usability principles (Norman, Nielsen)-Rate the user control capabilities and the freedom of action: understandable and direct processes as undo and redo	42	2	10	6.76	1.948
Usability principles (Norman, Nielsen)-Rate the consistency and presentation standards: maintain the same presentation of the interface	42	3	10	7.62	1.724
Usability principles (Norman, Nielsen)-Rate the fault prevention: mechanism for preventing error occurrence	42	2	10	6.40	1.939
Usability principles (Norman, Nielsen)-Rate the ease of detection: obvious & easy actions required during service use	42	2	10	6.86	1.842

**Table 3.2.1.1:** Workbench – Usability principles Descriptive Statistics

### 3.2.2 Collaboration and Decision Making Support services (integrated in the Workbench service)

CGRA evaluators used the service and completed the related Dicode Evaluation Framework Questionnaires. Figures 3.2.2.1, 3.2.2.2, 3.2.2.3, summarize the evaluators' responses relative to the quality, acceptability, accessibility and usability of the Workbench Collaboration and Decision Making services. In these figures, the median and mode values of the evaluators' responses are shown, originally given in a 1-5 scale. The highest values are observed for the quality and accessibility of the Workbench Collaboration and Decision Making services. Regarding the overall quality, evaluators agree that the objectives of the service are met (median=4, mode=3), the service is novel to their knowledge (median=4, mode=4), they are satisfied with the performance of the service as well as overall with this service (median=4, mode=4). The lowest mode value is given for the data intensive decision making issues addressed by the service. In the acceptability graph, Figure 3.2.2.2, we can observe that evaluators generally found all the functions they expected (median=4, mode=3), the interface of the service was pleasant (median=4, mode=4), and subsequently they would most probably recommend the service to their peers/ community (median=4, mode=3). Also, apart for the pleasant interface, a high mode value of 4 can be seen for the usefulness of the service (helpful to their work).

Regarding the accessibility and other issues shown in Figure 3.2.2.3, the evaluators generally agree in that they are satisfied with the access given to this Dicode service (median=4, mode=4), they are satisfied with the instruction given by this Dicode service (median=4, mode=4) and that they are satisfied with the time spend to learn to use the provided solution (median=4, mode=4). The evaluators' opinion regarding their satisfaction with the ease of use of the solution to complete their tasks appears to exceed the average (median and mode values of 3). Particularly, evaluator's comments ranged from *'it was not such easy'* [sic] (GR33) to *'This solution is very easy to use and complete complex tasks'* (GR32).

The suggested scenario seemed to have been understood well by the evaluators, although some extra instructions were thought to be necessary by some reviewers. Overall reviewers found the service *'promising'* (ES12). Some technical issues were raised: *'It's very useful for a complex use case. I'd also like to move with mouse holding left click'* (GR32), *'A bit slow loading time both for Workspaces list and Mind-map view'* (ES12), *'Overlapping adornments. Resizing does not work smoothly in Firefox for Linux'* [sic] (ES12), *'The arrow's graphic wasn't very pleasant for me. It starts from the middle of the icon and not from the begging of the square. The overall idea however, was quite good.'* (GR7), *'Web browsing is interesting and "universal" but the way it is presented and the usability of the interface given could be much better.'* (ES12).

In Table 3.2.2.1 the mean values of the responses relative to usability are given in a 0-10 scale. The mean values are above 5.53 for all questions. The standard deviation values vary in [1.579, 2.481]. Low score values were reported for the restrictions in actions and design, flexibility related to variety of operation modes, ease of recognition, help/ documentation facilities, and user control capabilities of the service. The highest reported mean values with the lowest variability refer to the visibility and interface presentation maintenance.

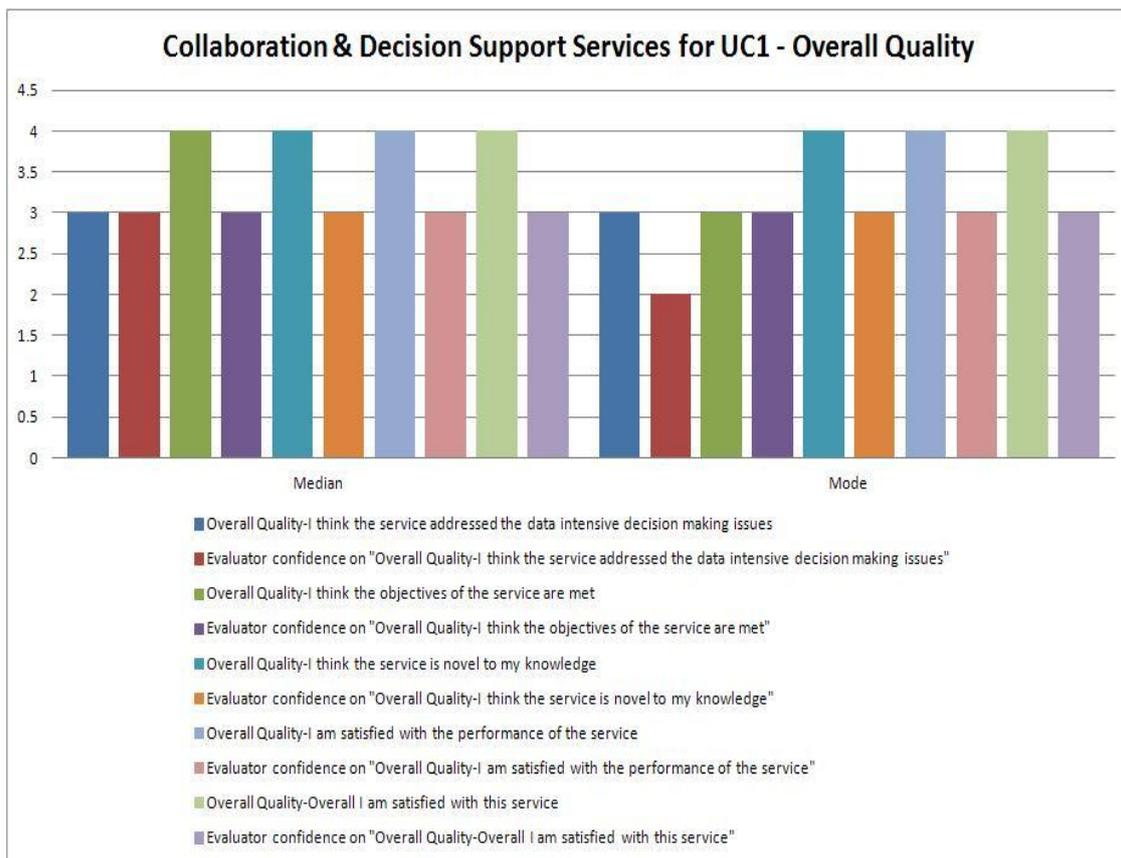


Figure 3.2.2.1: Workbench Collaboration and Decision Making Support services’ quality.

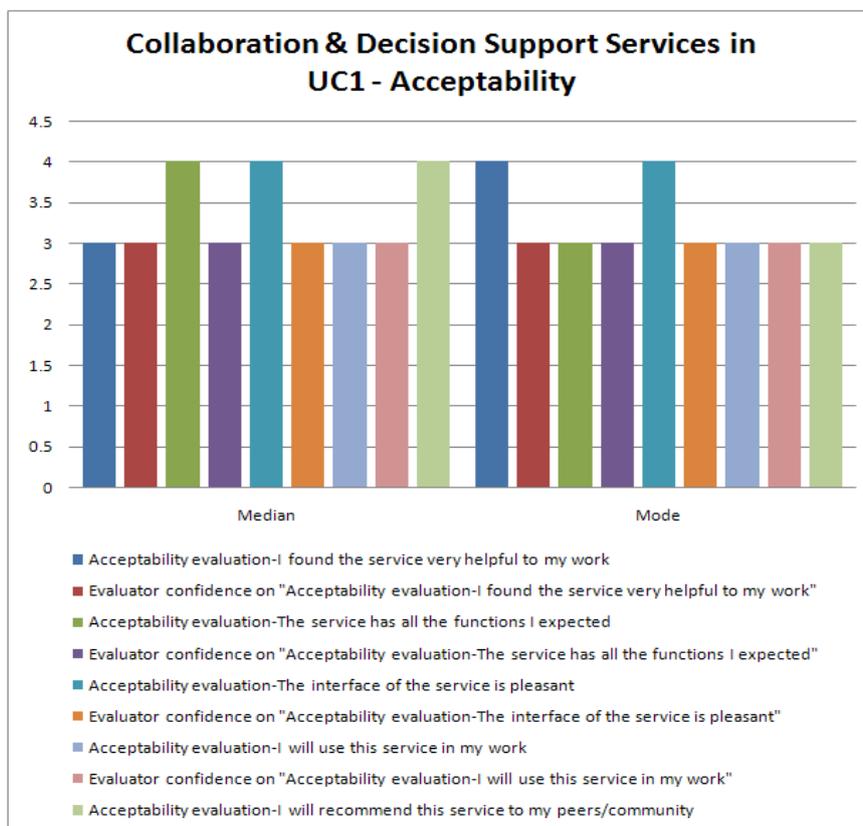
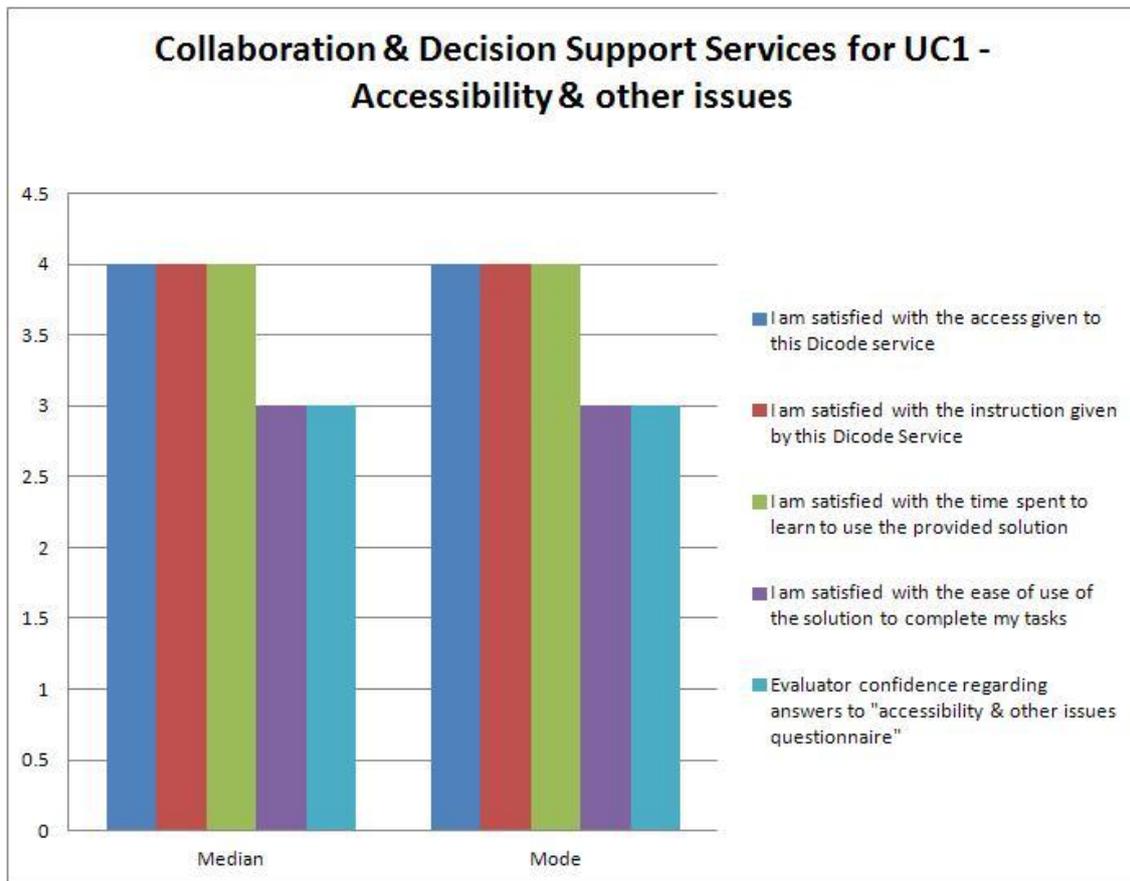


Figure 3.2.2.2: Workbench Collaboration and Decision Making Support services’ acceptability.

Questions	N	Minimum	Maximum	Mean	Std. Deviation
Usability principles (Norman, Nielsen)-Rate the visibility: optical information to facilitate the user	17	4	10	7.41	1.661
Usability principles (Norman, Nielsen)-Rate the tolerance: behavior similar to expectations	17	2	10	6.71	1.929
Usability principles (Norman, Nielsen)-Rate the physical mapping: conceptual correspondence between commands and functions	17	3	10	6.94	1.952
Usability principles (Norman, Nielsen)-Rate the restrictions: incomplete ways in the actions performed and in the design of operations in understandable manner	17	0	8	5.82	2.069
Usability principles (Norman, Nielsen)-Rate the conceptual models: the operation of the proposed actions according to the perception of user for these actions	17	3	10	6.88	1.764
Usability principles (Norman, Nielsen)-Rate the feedback: notification regarding the user's position	17	3	9	6.12	1.764
Usability principles (Norman, Nielsen)-Rate the error prevention: restrict user errors & support for their solution	17	4	10	6.06	1.676
Usability principles (Norman, Nielsen)-Rate the flexibility: variety of operation modes	17	0	10	5.82	2.481
Usability principles (Norman, Nielsen)-Rate the ease of recognition: easy identification of the required actions	17	2	10	5.82	2.243
Usability principles (Norman, Nielsen)-Rate the flexibility of the use efficiency: shortcuts provision, capabilities for the user to affect the configuration of the system	17	2	10	6.12	2.176
Usability principles (Norman, Nielsen)-Rate the provision of clear error messages: simple language in error messages and a proposal to resolve them	17	4	10	6.71	1.649
Usability principles (Norman, Nielsen)-Rate the aesthetics of the minimalist design: messages with the necessary information	17	1	10	6.35	2.370
Usability principles (Norman, Nielsen)-Rate the help facilities and the documentation facilities: help facilities related to user action	17	2	10	5.53	2.095
Usability principles (Norman, Nielsen)-Rate the visibility of the system status: user briefing regarding the work progress	17	3	9	6.35	1.693
Usability principles (Norman, Nielsen)-Rate the system matching with the real world: comprehensive and explanatory language	17	3	10	6.71	1.961
Usability principles (Norman, Nielsen)-Rate the user control capabilities and the freedom of action: understandable and direct processes as undo and redo	17	2	10	5.88	2.342
Usability principles (Norman, Nielsen)-Rate the consistency and presentation standards: maintain the same presentation of the interface	17	3	10	7.53	2.035
Usability principles (Norman, Nielsen)-Rate the fault prevention: mechanism for preventing error occurrence	17	4	10	6.24	1.786
Usability principles (Norman, Nielsen)-Rate the ease of detection: obvious & easy actions required during service use	17	3	9	6.35	1.579

**Table 3.2.2.1:** Collaboration & Decision Support Services for UC1 – Usability principles Descriptive Statistics



**Figure 3.2.2.3:** Workbench Collaboration and Decision Making Support services' accessibility & acceptability.

### 3.2.3 Forum Summarization service

Forum Summarization Service (FSS) combines techniques from data mining and distributed systems to provide effective support for users to quickly make sense of the huge and dynamic pool of information emerging from Internet Forums (Ammari, Lau, & Dimitrova, 2012).

For evaluation purpose, a web-based interface is implemented to show the output of FSS. As FSS is still in its early stage of design especially in its user interface, we believe a formative evaluation (Bhola, 1990) would be appropriate and valuable because with more qualitatively rich data, we can build a more emergent, structured picture of what is happening with this design in a rich interactive context, and to indicate what might be done for improvement.

To simulate Dicode use cases, we select two forums from [stackexchange.com](http://stackexchange.com), a fast-growing forum family web site which has over 1.5 million users on diverse topics from programming to cooking. The two selected forums are: *Web application forums* - to simulate Use Case 3: data analysts in consultant company; *Statistical analysis forum* - to simulate Use Case 1 researchers using statistics in their work.

#### *Time spent on each task*

During the study, we asked the participants to record the time they spent on each slide (task), the data can be used to find out if participants have spent more time on one task than

the other. As there are not enough observations to confidently test whether the differences between tasks are normally distributed, we analysed the data using a signed-rank test on two groups of pairs on one-sided hypothesis: task 1 < task 3 and task 2 < task 3. For those two groups of pairs, there are two nominal variables: task and participant, and one measurement variable (time of completion). The results show that the participants for Web Application forum has spent significantly more time on task 3 than task 1 ( $p=0.014$ ) or Task 2 ( $p=0.002$ ).

Wilcoxon Signed-ranked test	Task 1 < Task 3	Task 2 < Task 3
Web Application	$p=0.014$	$p=0.002$
Statistical Analysis	$p=0.070$	$p=0.052$

**Table 3.2.3.1:** Wilcoxon Signed-ranked test

The participants who felt task 3 is either difficult or very difficult have provided some reasons:

- Information incomplete  
Neither topic clouds nor list of questions alone could provide an immediate answer like task 1 or task 2.  
*"It's a little bit complicated to know what thread links Yahoo and Gmail."*  
*"It was a bit difficult to find statements for this task."*  
*"I did not found the topics that I was looking for in the topic clouds."*
- Information conflict  
Some information gained from topic clouds could not be confirmed by list of questions.  
*"When I search a term and in the detail topics the original word is not in. For example, I clicked in Yahoo inside Email Cloud and in the first discussion Yahoo doesn't match"*
- Search function is not available  
There is no search functionality available in either topic clouds or list of questions to conduct a combined two-word search so they can find some relationships.  
*"I did not found a way to merge the two concepts "yahoo" and "Gmail", because I cannot execute queries with AND, only OR between the words. I have used double-quotes, and the word AND, but it does not work. My idea was to find these two concepts so: I have searched for "yahoo" in the threads and then in the results I have used CTRL+F with the word "Gmail". I have analysed the highlighted results manually and I have decided my responses."*  
*"Sorry, but after 15 minutes of searching, I can't say I found something to type. I have tried searching with terms "Sample Size Hypothesis testing" and just "size Hypothesis". I also opened the topic clouds Sample (size) and Test (hypothesis). Maybe it is due to my lacking knowledge of the field, but it seemed to me that I couldn't find appropriate results."*

### **Confidence and Difficulty**

Most of the participants (more than 70% on both forums) showed they were either very confident or confident in completing all three tasks although they spent more time on task 3 from previous analysis. More than 60% of the participants of Web Application forum claimed that task 3 is neutral or difficult. While 40% the participants in statistical analysis forum claimed that task 3 is neutral or difficult. Participants of web application showed a slightly drop in confidence level for task 3 (70%) compared to task 2 or task 1 (both are nearly 90%) but a big drop in difficulty level for task 3 (nearly 30% as easy or very easy) compared

to task 2 or task 1 (both over 80% as easy or very easy). There is almost no obvious change in confidence level for Statistical Analysis forum but a similar drop in difficulty level from task 2 to task 3 (90% to 60% of participants claimed the task is easy or very easy.) Surprisingly, the difficulty level (easy or very easy) in task 1 and task 3 claimed by the participants of Statistical Analysis forum are almost the same (60%).

The subjective feedbacks of confident participants for task 1 and task 2 have shown that they all have a clear reason on interpreting the questions: “most discussed issues”. Participants who believed that the most discussed issues are shown in the topic clouds went to exam the topic clouds. Participants who believed the most discussed issues are “most viewed” or “most replied” questions went to examine the lists of questions. The participants who look at both topic clouds and list of questions, which did not seem to provide consistent answers showed lower confidence level.

### *Useful functionalities*

For web application forum, from task 1 to task 2 and then task 3, more participants who believed topic clouds are more helpful (3 in task 1; 4 in task 2; 7 in task 3).

For statistical analysis, the participants show different opinions on different tasks. For task 1, half of the participants believed the list is more useful while the other half believed the topic clouds are more useful. For task 2, more participants (60%) believed topic clouds are more useful and among the rest of the participants, half (20%) believed the list is more helpful and half (20%) believed both functions are helpful. The similar pattern is also shown for task 3, half participants (40%) believed topic clouds are more useful among the rest of the participants, half (30%) believed the list is more helpful and (20%) believed both functions are helpful.

As all the participants shows they have similar experience level of the application domain (moderate experience in Web application and basic level experience of Statistical Analysis), it is difficult to link their feedbacks with their domain experience. However, as we have different levels of students (Undergraduates, Masters and PhDs), we have looked into whether the functionality preference have any relationship with their current role.

For Web Application forum, we have 5 Undergraduates, 4 Masters and 5 PhDs. All 5 Undergraduates have clear preference for all three tasks, either List or Clouds. All of them completed the tasks using only one function. Some of the subjective feedbacks about how they completed the tasks are listed here:

*“I took the list of threads and selected the most commented.”*

*“I search for Google Search in the text search field. Then you can see that these issues are the issues with the most number or replies.”*

*“I look at the topic clouds and search the topic email. Then I look for the most replied issues and selected what could link Gmail and Yahoo.”*

For both forums, there was only one participant (a Masters student) who looked at both functions on all tasks. All others complete most of the tasks with only one function, regardless of their roles.

### *Comments on Topic Clouds*

Participants believed the topic clouds are good at:

- Visualizing the essence of the content

*“Hierarchical structure Better visualization of results”*

*“Clear layout Easy to see which topic is being used. A good number of related topics shown”*

- Organizing the topic in groups and easy for browsing the topics  
*“1) Quick 2) easy to use because of the groups that the topics are organized into groups”*  
*“It is easy and fast access to information 2. Reasonable clustering of the topics”*  
*“Simple Clear Easy to locate main topics”*
- Easy for presenting simple summary  
*“Overall view of the forum - Establish two kind of relationships: topic-term and term-term - See the importance of each term and topic”*  
*“In general, it provides easy information and a summary of the forums rather than reading through the thread discussions.”*

Participants believed the topic clouds are not good enough because:

- No searching or sorting functions  
*“It doesn’t have search function”*  
*“1. It lacks of any search and sort feature. 2. It would be good to have a combination feature, to be able to select multiple topics and see the AND or OR results. (I think this would be useful in questions like the one at slide 3).”*
- Not clear in relationships, No hierarchy (i.e. clouds within clouds)  
*“1) It would have helped if clicking the clouds within clouds would have provided their clouds as well rather just giving the discussions they were in. 2) Better to sort clouds in clouds in descending order for items with the same font size.”*  
  
*“It could be more interesting to have another kind of representation as overlapped circles with common terms in intersections - When you click in a term you cannot look up over the list of discussions you get.”*
- Only based on one-word terms  
*“Based mainly on single keywords”*
- Visual clues are not very clear: colouring, font, layout etc.  
*“When you click on a term you cannot search over the retrieved list of answers - A representation based on circles with a size depending on the relevance of the topic and intersections where common terms are located could be more graphical and representative”*
- No details about the topic  
*“Sometimes you don't know exactly what the topic is about”*  
*“Repeated similar topics More information would be appreciated”*

A full description of the service’s results can be found under the Forum Summarization section in: <https://wiki.dicode-project.eu/display/DIC/Evaluators%27+responses+UC1+-+quantitative+results>.

### 3.2.4 SubgroupDiscovery service

CGRA evaluators used the service and completed the related Dicode Evaluation Framework Questionnaires. Figures 3.2.4.1, 3.2.4.2, 3.2.4.3, summarize the evaluators’ responses relative to the quality, acceptability, accessibility and usability of the SubgroupDiscovery service. In

these figures, the median and mode values of the evaluators' responses are shown in a 1-5 scale. Regarding the overall quality of the service, the evaluators agree that the objectives of the service are met, the service is novel to their knowledge and that they are satisfied with the performance of the service. The evaluators responses gave a median and mode value of 3 concerning other overall quality issues, e.g. that the service addressed the data intensive decision making issues and their overall satisfaction with the service. With respect to addressing data intensive decision making issues evaluators noted that *'It certainly can be very helpful for discovery of interesting inter-dependencies in data.'* [sic] (DE03) and that *'I think that if the input files have correct information the service could help in making such decisions.'* (GR12). Regarding the acceptability of the service, it can be seen in Figure 3.2.4.2 that the evaluators responses' median and mode values equal 3, when they were asked if they found the service very helpful to their work, if the service has all the functions they expected, if the interface of the service is pleasant, if they will use this service in their work and if they will recommend this service to their peers/community. Particularly low mode values are reported for the interface of the service and its usefulness to the evaluators' work.

The service was not found to be particularly relative to work of the reviewers who generally were not very happy with the information provided by the service *'Although the program responded correctly the messages were not the expected'* (GR6), *'Really poor documentation of the problem, grammatical errors and missing the gaps between the words.'* (GR24). Relative low values were given for the acceptability of the service, and particularly relevant to how pleasant the interface was. Instructions were found to be adequate, although difficult to follow at times: *'I don't need a specification about how this system is used. What I need is a clear example that indicates and explains what I get for a concrete request and what is the purpose of using this service.'* (ES09), agreeing with *'I didn't really understand the objectives.'* (GR36). Regarding accessibility and other related issues, as can be seen in Figure 3.2.4.3, the evaluators agree that they are satisfied with the access given, the instruction given by the service and the time spend to learn to use the provided solution, with median and mode values equal 4.

In Table 3.2.4.1 the minimum, maximum, mean and standard deviation values of the responses relative to usability are given; the scale in this case is 0-10. Overall the mean values are above 4, whereas most of them exceed the mean 5 (the highest mean value observed equals 6.47) with standard deviation values varying in [1.914, 2.932]. The above results suggest that there is room for improvement concerning the usability of the SubgroupDiscovery service. Particularly low mean score values were given for the visibility, flexibility, and user control capabilities, whereas high scores were given to the tolerance and conceptual models (proposed actions relative to perception of users for actions) of the service, with standard deviation values 1.919 and 2.189 respectively. A particularly favourable comment involved the data-intensiveness of the service: *'It certainly can be very helpful for discovery of interesting inter-dependencies in data.'* (DE3). Reviewers agree on the service being *'fast'* (GR2) and of *'satisfactory performance'* (GR8). Also, another useful comment involved the last session retrieval results: *'It is strange that there is no ID for my Task. There is no information about how results can be retrieved over time. Are task and result connected by a user session (should I use the same browser for result access?'* (DE1).

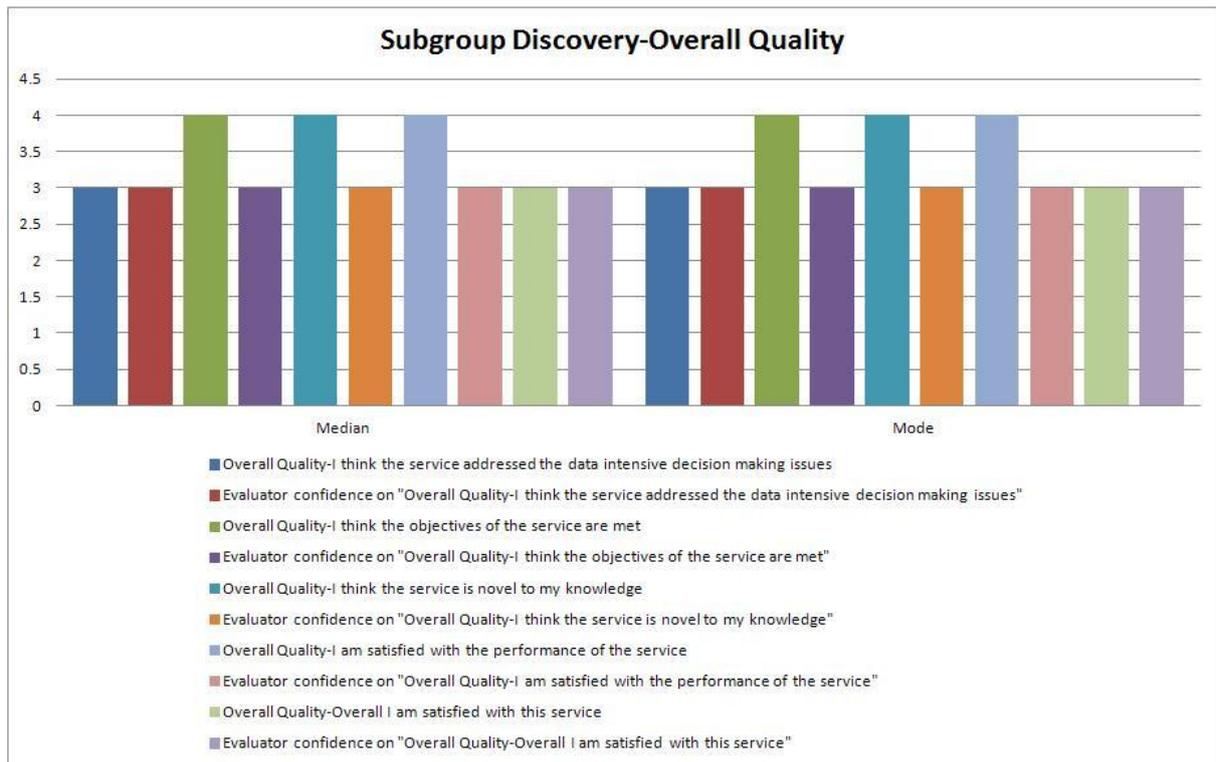


Figure 3.2.4.1: SubgroupDiscovery service quality.

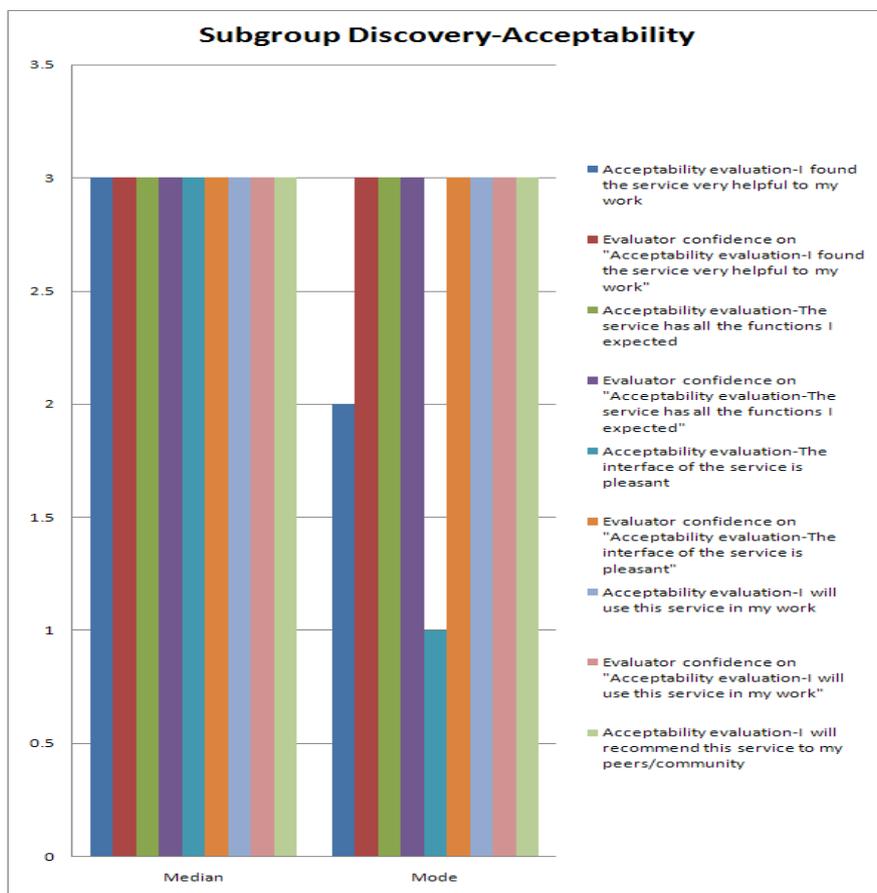


Figure 3.2.4.2: SubgroupDiscovery service acceptability.

Questions	N	Minimum	Maximum	Mean	Std. Deviation
Usability principles (Norman, Nielsen)-Rate the visibility: optical information to facilitate the user	43	0	10	4.98	2.932
Usability principles (Norman, Nielsen)-Rate the tolerance: behavior similar to expectations	43	3	10	6.47	1.919
Usability principles (Norman, Nielsen)-Rate the physical mapping: conceptual correspondence between commands and functions	43	0	10	6.14	2.189
Usability principles (Norman, Nielsen)-Rate the restrictions: incomplete ways in the actions performed and in the design of operations in understandable manner	43	0	10	5.77	2.103
Usability principles (Norman, Nielsen)-Rate the conceptual models: the operation of the proposed actions according to the perception of user for these actions	43	3	9	6.05	1.914
Usability principles (Norman, Nielsen)-Rate the feedback: notification regarding the user's position	43	0	10	5.53	2.208
Usability principles (Norman, Nielsen)-Rate the error prevention: restrict user errors & support for their solution	43	0	9	5.14	2.386
Usability principles (Norman, Nielsen)-Rate the flexibility: variety of operation modes	43	0	10	5.26	2.583
Usability principles (Norman, Nielsen)-Rate the ease of recognition: easy identification of the required actions	43	0	9	5.53	2.240
Usability principles (Norman, Nielsen)-Rate the flexibility of the use efficiency: shortcuts provision, capabilities for the user to affect the configuration of the system	43	0	9	4.77	2.419
Usability principles (Norman, Nielsen)-Rate the provision of clear error messages: simple language in error messages and a proposal to resolve them	43	0	10	5.07	2.649
Usability principles (Norman, Nielsen)-Rate the aesthetics of the minimalist design: messages with the necessary information	43	0	10	5.47	2.261
Usability principles (Norman, Nielsen)-Rate the help facilities and the documentation facilities: help facilities related to user action	43	0	10	5.28	2.292
Usability principles (Norman, Nielsen)-Rate the visibility of the system status: user briefing regarding the work progress	43	0	9	5.12	2.509
Usability principles (Norman, Nielsen)-Rate the system matching with the real world: comprehensive and explanatory language	43	0	9	5.88	2.332
Usability principles (Norman, Nielsen)-Rate the user control capabilities and the freedom of action: understandable and direct processes as undo and redo	43	0	10	4.79	2.704
Usability principles (Norman, Nielsen)-Rate the consistency and presentation standards: maintain the same presentation of the interface	43	0	10	5.51	2.576
Usability principles (Norman, Nielsen)-Rate the fault prevention: mechanism for preventing error occurrence	43	0	10	5.33	2.408
Usability principles (Norman, Nielsen)-Rate the ease of detection: obvious & easy actions required during service use	43	0	10	5.65	2.525

**Table 3.2.4.1:** Subgroup Discovery – Usability Principles Descriptive Statistics

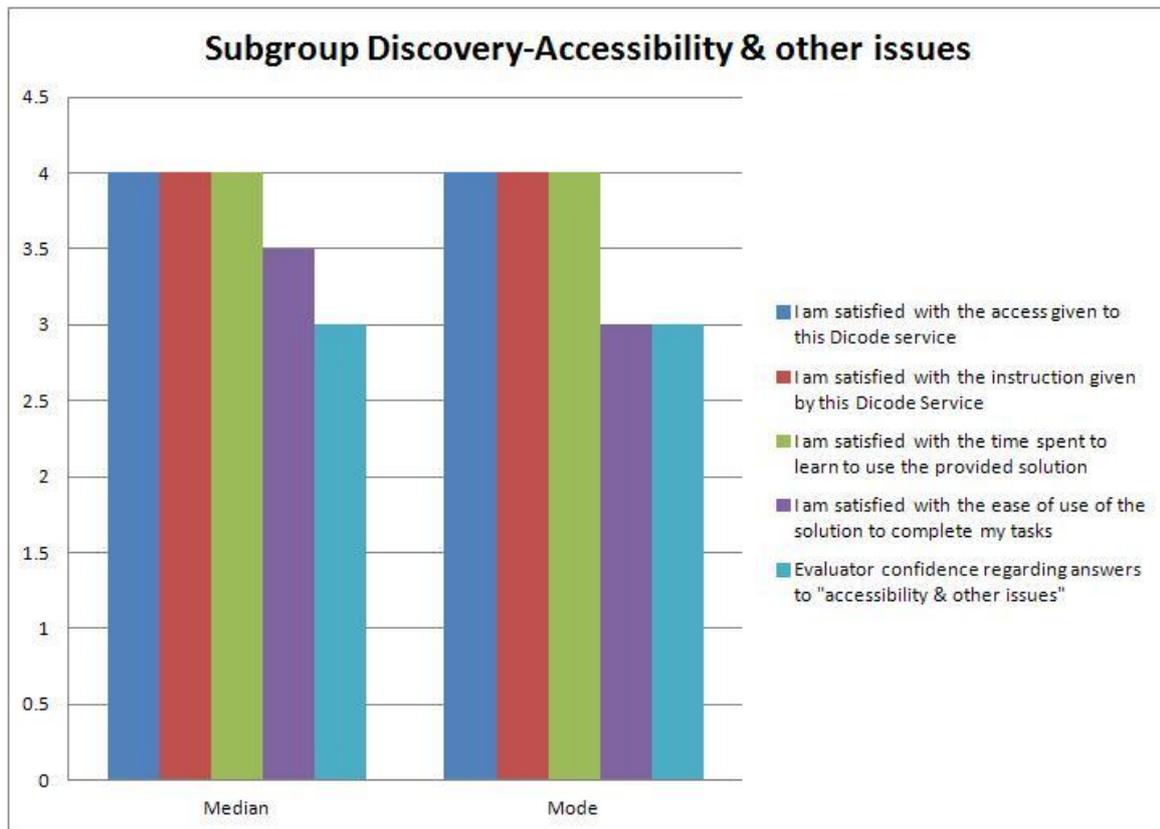


Figure 3.2.4.3: SubgroupDiscovery service accessibility and other issues.

### 3.2.5 Recommender service

CGRA evaluators used the service and completed the related Dicode Evaluation Framework Questionnaires. Figures 3.2.5.1, 3.2.5.2, 3.2.5.3 summarize the evaluators' responses relative to the quality, acceptability, accessibility and usability of the Recommender service. In these figures, the median and mode values of the evaluators' responses are shown, where those were given in a 1-5 scale. Regarding the overall quality of the service the evaluators' responses are neutral concerning whether the service addressed the data intensive decision making issues, if the objectives of the service are met, if the service is novel to their knowledge, and if overall they are satisfied with the service. Comments such as '*seems to help but depends on the application and the framework.*' (GR16), '*I don't know the objectives in detail but I think they are met*' [sic] (GR32) and '*Its performance was very good, its response time too, the only thing missing was a better user interface.*' [sic] (GR12), are helpful in explaining the just above the average median and mode score values in the overall quality of the service. The evaluators agree that they are satisfied with the performance of the service (median and mode values of 4).

Regarding the acceptability of the service, as can be seen in Figure 3.2.5.2, the evaluators are neutral in that the service is helpful to their work, it has all the functions they expected, the interface of the service was pleasant, they will use this service in their work and that they will recommend this service to their peers/community. The acceptability of the service perhaps can be explained by the following comment: '*The function of the service is very interesting, I cannot see how it could be extended.*' (GR31).

In Figure 3.2.5.3 we can observe the accessibility and other related issues median and mode score values given for the Recommender service. The evaluators agree with the access given, the instruction given by the service, they are satisfied with the time spend to learn to use the provided solution and they are satisfied with the ease of use of the solution to complete their tasks. Evaluators agreed that the service was easy to use and fast, although a noticeable comment refers to its output: *'Wasn't easy due to the complex outputs.'* (UK1). Reviewers also agree in the lack of interface. The system's messages provided were thought to be insufficient: *'I wanted more system messages/help in order to understand the service's use and errors.'* (GR15), *'Informations were not enough to understand what I had to do for the service.'* (GR7). An interesting comment about the information provided was the following: *'The information provided by the service was not good because of the following reasons : -The parser error has grammatical flaws and lacks space between the words. -As a result of the first reason it is quite hard to understand what the problem is.'* (GR24).

In Table 3.2.5.1 the mean values of the responses relative to usability are given in 0-10 scale, where overall values are reported to be above 4.63 ranging in [4.63, 6.21], whereas standard deviation values range in [2.310, 3.052]. Particularly low values were given for the feedback, flexibility of use efficiency, flexibility related to variety of operation modes, error prevention, help/documentation facilities, visibility of system's status, and the users' control capabilities of the service. High mean score values were given for the tolerance (6.21) and the physical mapping (6.13) of the service, but with relatively high standard deviation values.



Figure 3.2.5.1: Recommender service quality.



Figure 3.2.5.2: Recommender service acceptability.

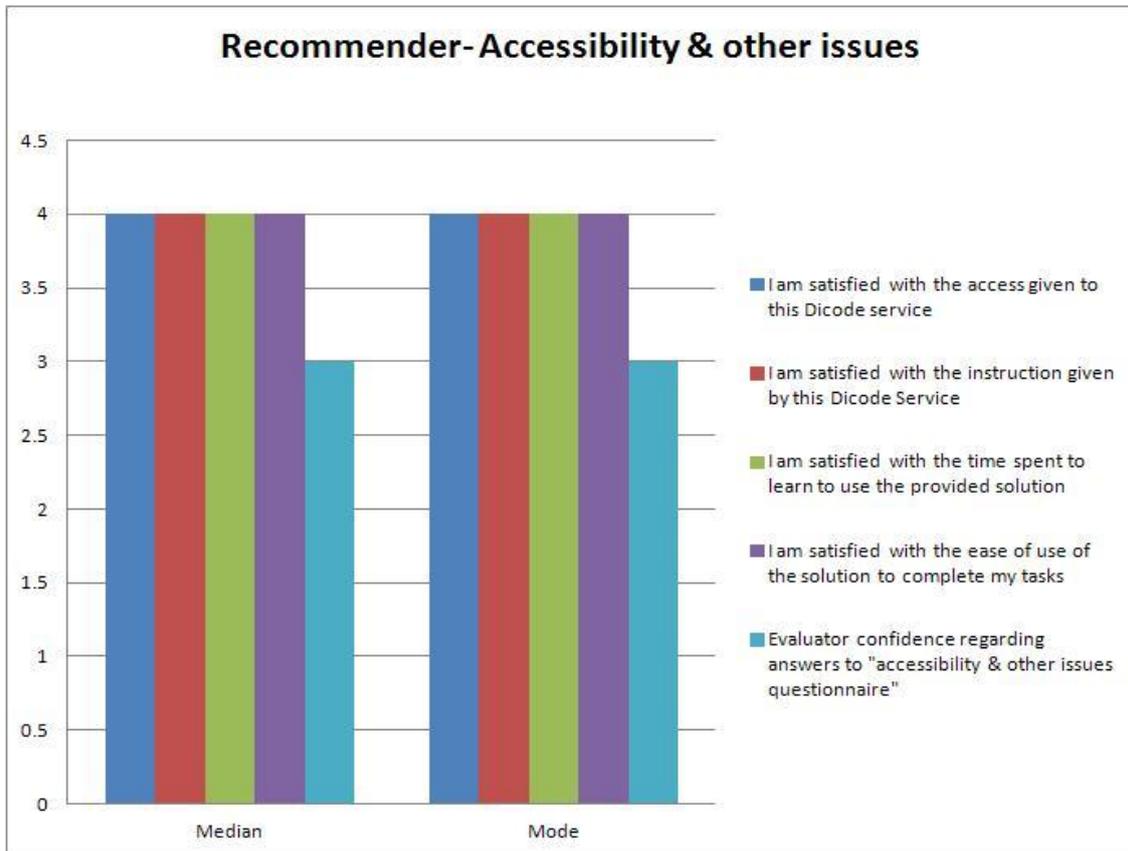


Figure 3.2.5.3: Recommender service accessibility and other issues.

Questions	N	Minimum	Maximum	Mean	Std. Deviation
Usability principles (Norman, Nielsen)-Rate the visibility: optical information to facilitate the user	48	0	10	5.52	2.961
Usability principles (Norman, Nielsen)-Rate the tolerance: behavior similar to expectations	48	0	10	6.21	2.697
Usability principles (Norman, Nielsen)-Rate the physical mapping: conceptual correspondence between commands and functions	48	0	10	6.13	2.749
Usability principles (Norman, Nielsen)-Rate the restrictions: incomplete ways in the actions performed and in the design of operations in understandable manner	48	0	10	5.33	2.538
Usability principles (Norman, Nielsen)-Rate the conceptual models: the operation of the proposed actions according to the perception of user for these actions	48	0	10	5.69	2.477
Usability principles (Norman, Nielsen)-Rate the feedback: notification regarding the user's position	48	0	10	5.21	2.752
Usability principles (Norman, Nielsen)-Rate the error prevention: restrict user errors & support for their solution	48	0	10	5.33	2.470
Usability principles (Norman, Nielsen)-Rate the flexibility: variety of operation modes	48	0	10	5.29	2.568
Usability principles (Norman, Nielsen)-Rate the ease of recognition: easy identification of the required actions	48	0	10	5.77	2.452
Usability principles (Norman, Nielsen)-Rate the flexibility of the use efficiency: shortcuts provision, capabilities for the user to affect the configuration of the system	48	0	10	5.02	2.522
Usability principles (Norman, Nielsen)-Rate the provision of clear error messages: simple language in error messages and a proposal to resolve them	48	0	10	5.44	2.600
Usability principles (Norman, Nielsen)-Rate the aesthetics of the minimalist design: messages with the necessary information	48	0	10	5.52	2.458
Usability principles (Norman, Nielsen)-Rate the help facilities and the documentation facilities: help facilities related to user action	48	0	9	4.63	2.702
Usability principles (Norman, Nielsen)-Rate the visibility of the system status: user briefing regarding the work progress	48	0	9	4.79	2.617
Usability principles (Norman, Nielsen)-Rate the system matching with the real world: comprehensive and explanatory language	48	0	10	5.56	2.835
Usability principles (Norman, Nielsen)-Rate the user control capabilities and the freedom of action: understandable and direct processes as undo and redo	48	0	10	4.96	3.052
Usability principles (Norman, Nielsen)-Rate the consistency and presentation standards: maintain the same presentation of the interface	48	0	10	5.52	2.705
Usability principles (Norman, Nielsen)-Rate the fault prevention: mechanism for preventing error occurrence	48	0	10	5.06	2.310
Usability principles (Norman, Nielsen)-Rate the ease of detection: obvious & easy actions required during service use	48	0	10	5.79	2.689

**Table 3.2.5.1:** Recommender – Usability principles Descriptive Statistics

### 3.3 Reliability analysis

Reliability analysis is an assessment of the degree of consistency between multiple measurements of a variable (Kastania & Kossida, 2011). The most commonly used measure of reliability is internal consistency (i.e., consistency among the variables in a summated scale). Reliability analysis was performed using the Cronbach's  $\alpha$  reliability coefficient (Cronbach, 1951). The Cronbach's  $\alpha$  reliability coefficient measures the consistency among individual items in a single scale, i.e. how close related are a set of items as a group (Norman, 1998). For the questionnaires examined, it indicates how well the different items (questionnaire questions expecting a quantitative answer) complement each other in their measurement of different aspects of the same variable, i.e. reliability of service. As a rule of thumb, a value of 0.7 or higher is obtained on a substantial sample (Lewis, 1995). It is worth noting that the reliability coefficient depends on the number of users that completed the questionnaires in our case, and the number of questions which expect a quantitative answer (number of items).

In Table 3.3.1, Cronbach  $\alpha$  values range from 0.959 to 0.693. It is evident that the three tested Dicode Evaluation Framework questionnaires for each service show high internal consistency as indicated by high alpha coefficients, which exceed the recommended level of 0.7. The Forum summarization service is not included in Table 3.3.1 since it was derived from a differently structured questionnaire compared to the standard three Dicode Evaluation Framework questionnaires used for the other services.

Dicode Service	Dicode Evaluation Framework Questionnaires	Up to now Users	Reliability Statistics	
			Cronbach's Alpha	Number of Items
Workbench Integrated Services	Use Case/Workbench Integrated Services Validation Instrument	45	0.938	30
	Workbench Integrated Services Ease of Use Assessment	42	0.957	25
	Workbench Integrated Services Accessibility & Acceptability Assessment	43	0.770	5
Collaboration Services for UC1	Use Case/Collaboration Services for UC1 Validation Instrument	17	0.930	30
	Collaboration Services for UC1 Ease of Use Assessment	17	0.896	25
	Collaboration Services for UC1 Accessibility & Acceptability Assessment	16	0.693	5

Subgroup Discovery	Use Case/Subgroup Discovery Service Validation Instrument	43	0.939	30
	Subgroup Discovery Service Ease of Use Assessment	43	0.949	25
	Subgroup Discovery Service Accessibility & Acceptability Assessment	42	0.844	5
Recommender	Use Case/Recommender Service Validation Instrument	48	0.944	30
	Recommender Service Ease of Use Assessment	48	0.959	25
	Recommender Service Accessibility & Acceptability Assessment	50	0.750	5

**Table 3.3.1:** Reliability Analysis Table: Table shows the Cronbach  $\alpha$  reliability statistic measures for all the CGRA Dicode services.

## 4 Future work directions

This deliverable presents the initial version of the evaluations of the Dicode suite of services and innovative work methodologies for CGRA that have been designed and developed in the context of Task 6.3, representing a first assessment of their usability and acceptability. Taking into account the feedback received from the first evaluation phase of the Dicode project, the Dicode platform seems to be a promising tool which facilitates users to set their research objectives, and better understand the data and methodologies used in their research. Overall, the project provides a suite of innovative, adaptive and interoperable services that satisfy the requirements reported in Section 2.1. The big 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 should be directed towards the improvement of Dicode services in terms of their documentation, user interfaces and performance. Another concern relates to testing these services in various data-intensive contexts, as indicated in Section 2, in order to further assess their applicability and potential, and gradually build their generic nature.

Future work will concentrate on improving and enriching the above services by incorporating the detailed feedback of the evaluators. Careful summary and statistical estimation of the evaluators' quantitative and qualitative ranking of the services should result in an update of the CGRA description in order to fully address the needs of the Dicode project and of its future users. The update of CGRA will be based on the evaluators' suggestions, and particularly:

- on the enhancement of the already developed and tested services,

- on suggestions for the development of additional services to address the evaluators' comments and feedback.

With respect to enhancement of already developed services, issues raised from the first evaluation process are:

- **Workbench integrated services:** technical issues were raised by the reviewers relative to the Workbench performance and interface. Many of their suggestions point to a more user-friendly interface (more colours, distinguishable schematics), with well structured help files and messaging system.
- **Workbench Collaboration and Decision Making Support services:** Similarly with above, some Workbench flaws were spotted. Reviewers also suggested that more detailed or differently structured scenarios should be supplied, since many times the user could not understand the purposes/ objectives of the task performed.
- **Forum Summarization service:** Our results partially confirm the service's usefulness to facilitate the two sense-making processes (Read & Extract and Search for Relations), which could be due to the limitation of both functionalities and the interpretation of the tasks by the participants. Results clearly show that participants tend to use one function to complete the task irrespectively of their roles.
- **SubgroupDiscovery service:** Reviewers emphasized the poor documentation of the service, as well as the unpleasant interface or lack of aesthetics, which forced them in characterizing this service inflexible.
- **Recommender service:** Here again reviewers emphasized the need for a more user-friendly service which should produce less complex outputs. The help files and messages were thought to be insufficient.

Based on the valuable evaluators' comments, a number of future directions are considered towards the improvement of CGRA:

- **Data-intensiveness issues:**
  - GEO recommendation service aiming to suggest datasets to the user based on a text-mining scan of his descriptive request.
  - PubMed recommendation service aiming to suggest scientific articles relevant to the user's request. The service envisages alerts for new articles, publications, consortiums or innovations relevant to the main interests of the user.
- **Design issues:** Provide meta-data and enhanced search mechanisms
- **User interface issues:** Better visualization, provide search and sorting functions.
- **General purpose statistical analysis:** R/Bioconductor service aiming to provide access to the R statistical programming environment (<http://www.r-project.org/>) through the Dicode platform. Users would be able to upload their r code scripts in the platform not only to share and discuss them with their collaborators but to run their code and discuss the outputs. The service aims to provide r scripts for some popular in biomedical research methodologies.
- **Additional guidance throughout the processes:** A 'log book' utility accommodated to CGRA aiming to record and audit all the steps in the process of the data collection, manipulation, analysis and output interpretation, to assists or even establish future practices and transfer tacit knowledge.

The above actions are aiming to assure the functionality and acceptability of the Dicode CGRA platform as an exchange and collaboration environment to support researchers across disciplines. Most importantly, the CGRA platform is aiming to accommodate future multidisciplinary research efforts in the biomedical field.

## References

- [1] Ammari, A., Lau, L., & Dimitrova, V. (2012). Distributed Data Mining for User Sensemaking in Online Collaborative Spaces. *Mastering Data-Intensive Collaboration through the Synergy of Human and Machine Reasoning (dicoSyn 2012)*, CSCW.
- [2] Baker, M. (2012). Quantitative data: learning to share. *Nat Meth* , 9, 39-41.
- [3] Bholra, H. S. (1990). Evaluation planning, design and implementation, and utilization of evaluation results. In *Evaluating "Literacy for development" projects, programs and campaigns*. Hamburg Germany.
- [4] Crockford, D. (2006). The fat-free alternative to XML. *Proc of XML*. Boston.
- [5] Cronbach, L. J. (1951). Coefficient alpha and internal structure of tests. *Psychometrika* ,16(2),297-334.
- [6] Dicode Deliverable D2.2. (2011). *The Dicode Approach User requirements, conceptual integrative architecture, agile methodology and functional specifications*.
- [7] Dicode Deliverable D3.1.1. (2011). *The Dicode Data Mining Framework (initial version)*.
- [8] Dicode Deliverable D3.2.1. (2011). *The Dicode Data Mining Services (initial version)*.
- [9] Dicode Deliverable D4.1.1. (2011). *The Dicode Collaboration Support Services (initial version)*.
- [10] Dicode Deliverable D4.2.1. (2011). *The Dicode Decision Making Support Services (initial version)*.
- [11] Dicode Deliverable D5.4.1. (2011). *Integrated Dicode Services (initial version)*.
- [12] Dicode Deliverable D6.1. (2011). *The Dicode Evaluation Framework*.
- [13] Finholt, T. A. (2003). Collaboratories as a new form of scientific organization. *Econ Innov New Technol* , 12 (1), 5-25.
- [14] Guberman, J. M., Ai, J., Arnaiz, O., Baran, J., Blake, A., Baldock, R., et al. (2011). BioMart Central Portal: an open database network for the biological community. *Database* , 2011.
- [15] Huttenhower, C., Schroeder, M., Chikina, M. D., & Troyanskaya, O. G. (2008). The Sleipnir library for computational functional genomics. *Bioinformatics* , 24 (13).
- [16] Joyce, A. R., & Palsson, B. Ø. (2006). The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* , 7, 198-210.
- [17] Karacapilidis, N., Tzagarakis, M., Christodoulou, S., & Tsiliki, G. (2012). Facilitating and Augmenting Collaboration in the Biomedical Domain. *Int J Syst Biol Biomed Technol* .
- [18] Kastania, A. N., & Kossida, S. (2011). Advances in Telemedicine: Technologies, Enabling Factors and Scenarios. In G. Grasczew, & T. A. Roelofs (Eds.), *Statistics in Telemedicine* (pp. 191-210). INTECH Open.
- [19] Kawamoto, K., Lobach, D. F., Willard, H. F., & Ginsburg, G. S. (2009). A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. *BMC Med Inform Decis Mak* , 9.
- [20] Kitchenham, B., & Pflieger, S. L. (2002). Principles of Survey Research Part 5: Populations and Samples. *Software Engineering Notes* , 27 (5), 17-20.
- [21] Koschmieder, A., Zimmermann, K., Tribl, S., Stoltmann, T., & Leser, U. (2011). Tools for managing and analyzing microarray data. *Brief Bioinformatics* , 13.
- [22] Lee, S. E. (2007). Facilitating Collaborative Biomedical Research. *GROUPE '07*. Sanibel Island, Chicago, USA.
- [23] Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *Hum-Comput Interact* , 7 (1), 57-78.
- [24] Lusk, M., Kapushesky, M., Nikkilä, J., Parkison, H., Goncalves, A., Huber, W., et al. (2010). A global map of human gene expression. *Nat Biotechnol* , 28, 322-324.
- [25] Nielsen, J. (1991). *Designing Web Usability: The Practice of Simplicity*. Indianapolis: New Riders Publishing.
- [26] Norman, D. A. (1998). *The Design of Everyday Things*. The MIT Press.
- [27] Pennisi, E. (2011). Will Computers Crash Genomics? *Science* , 331, 666-668.
- [28] Perez-Llamas, C., Gundem, G., & Lopez-Bigas, N. (2011). Integrative Cancer Genomics (IntoGen) in Biomart. *Database* , 2011.
- [29] Spencer, D., Zimmerman, A., & Abramson, D. (2011). Special Theme: Project Management in E-Science: Challenges and Opportunities. *Comput Sup Coop Work* , 20, 155-163.
- [30] Sullivan, D. E., Gabbard, J. J., Shukla, M., & Sobral, B. (2010). Data integration for dynamic and sustainable systems biology resources: challenges and lessons learned. *Chem Biodivers* , 7 (5), 1124-1141.

[31] Tsiliki, G., & Kossida, S. (2011). Fusion methodologies for biomedical data. *J Proteom* , 74, 2774-2785.

## Appendix: Dicode Evaluator's Guide

We hereby include the parts of the Dicode Evaluator's Instructions relevant to CGRA use case, as introduced in section 2.4. The Instructions include general information about the service, usage scenarios which emphasize the settings under which the service is applied, and a detailed description of the actions the evaluator needs to perform during evaluation. Note that the Forum Summarization service questionnaire is included, as well as links to all other services' questionnaires. Instructions are given in *italics*.

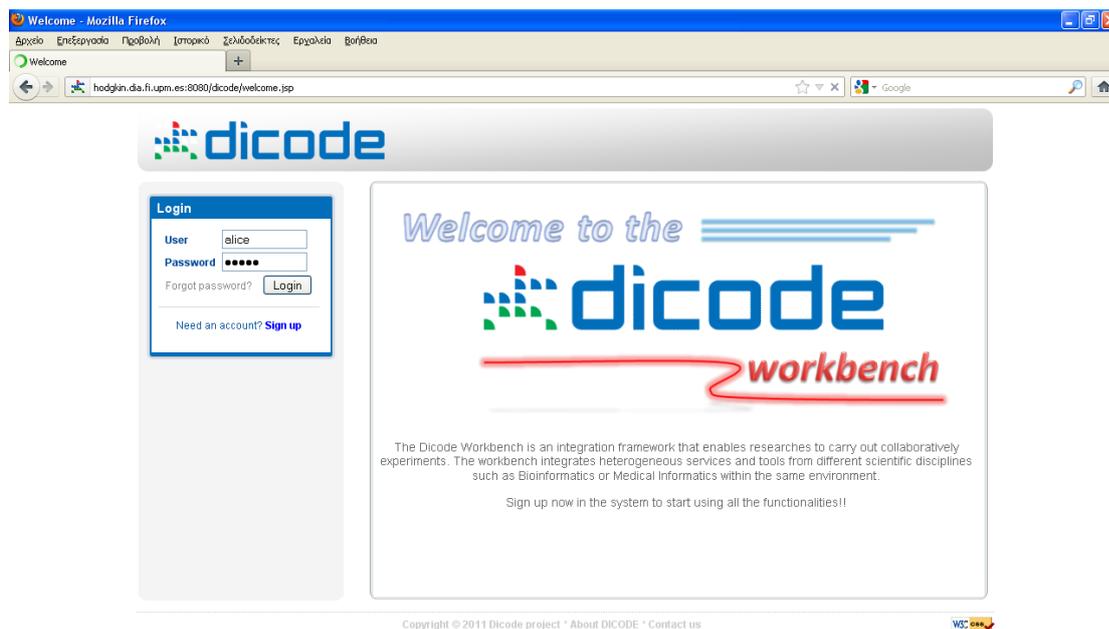
### 1. The Dicode Workbench service (refers to the Workbench, Storage, Collaboration Support and Decision Making integrated services)

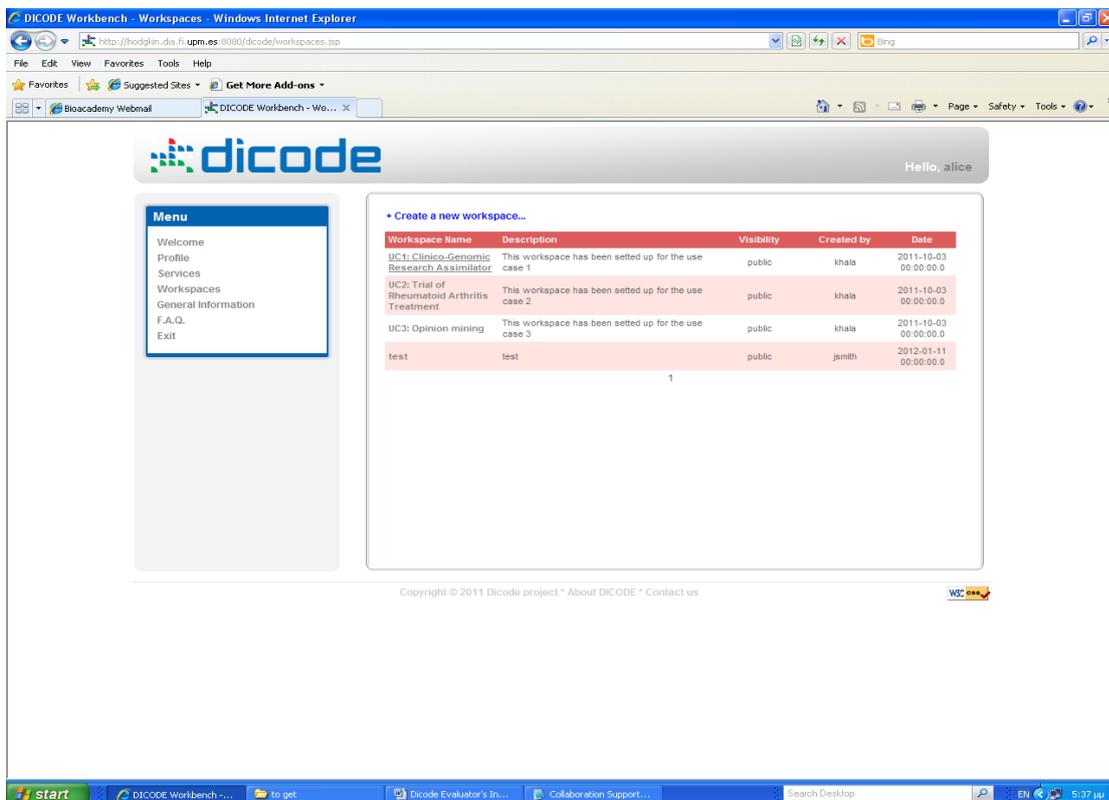
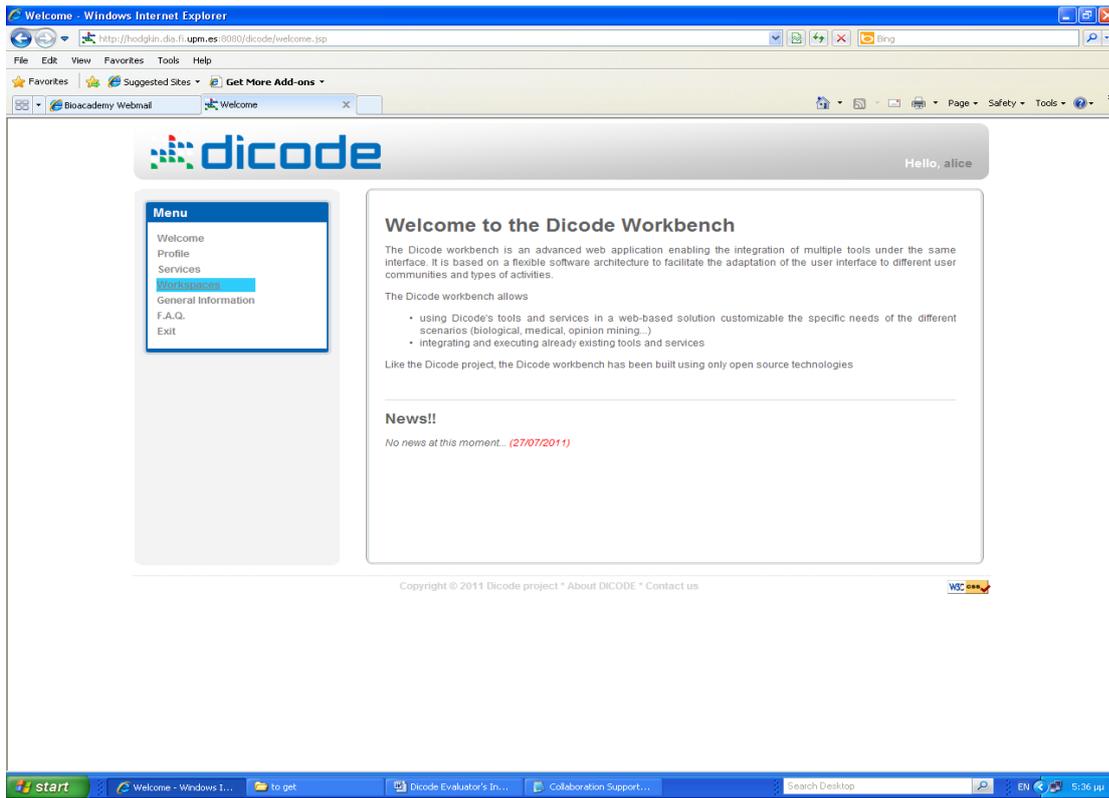
*Service's url:* <http://hodgkin.dia.fi.upm.es:8080/dicode/>

*Feel free to exploit the existing functionalities.*

*User name:* *alice*

*Password:* *alice*





Copyright © 2011 Dicode project \* About DICODE \* Contact

Copyright © 2011 Dicode project \* About DICODE \* Contact

Copyright © 2011 Dicode project \* About DICODE \* Contact



Workspace Name	Description	Visibility	Created by	Date
UC1: Clinico-Genomic Research Assimilator case 1	This workspace has been set up for the use	public	khala	2011-10-03 00:00:00
UC2: Trial of Rheumatoid Arthritis Treatment case 2	This workspace has been set up for the use	public	khala	2011-10-03 00:00:00
UC3: Opinion mining case 3	This workspace has been set up for the use	public	khala	2011-10-03 00:00:00
test	test	public	jemth	2012-01-11 00:00:00

Copyright © 2011 Dicode project \* About DICODE \* Contact us



### Scenario walkthrough – Use Case 1

**Overview:** Two researchers, Jim (bioinformatician) and Alice (biologists), aim to investigate which genes or groups of genes are associated with breast cancer (BC) disease. Jim & Alice have independently conducted similar analysis with in-house gene-expression datasets; however, their findings were not very encouraging, which was attributed to the small sample size (i.e. number of patients) available. They decide to search the available public repositories and download suitable data to augment their data in order to produce more reliable results. Neal (data manager) suggests the following alternatives (alphabetic order):

- i. Array Express
- ii. Gene Expression Omnibus Datasets (GEO)
- iii. Stanford Microarray Database (SMD)

All three of them are good candidates being public repositories that archive and freely distribute high-throughput multi-platform gene-expression data submitted by and under the standards of the scientific community. Jim and Alice are aware of all three databases, however Alice prefers using database (i), whereas Jim has previously used database (iii) without any problems. Neal has used all three databases and clearly prefers GEO-(ii), because:

- 1) it is highly populated (a billion individual gene expression measurements from 100 organisms),
- 2) each dataset is accompanied with supplementary files such as platform information and clinical data,
- 3) there is an R/Bioconductor package to retrieve the data via R,
- 4) all standard blast results display 'E' icons that link directly to GEO-Profiles expression data,
- 5) it includes tools for a first analysis of the data.

Alice points out that the Blast information will be very important for her research, nevertheless she claims that the ArrayExpress-(i) database is a transcriptomics repository and also presents the 1), 2), 3), 5) advantages. Particularly concerning the 2nd point raised by Neal, ArrayExpress offers extra options via the Gene Expression Atlas, namely it provides information about gene-expression in different cell types, organism parts, developmental stages, disease states, sample treatments and other biological experimental conditions.

Jim argues that from his point of view a very important advantage is how well populated and up-to-date the database is which also constitute the disadvantages for the SMD database. Nevertheless, when using SMD he found that it is compatible with GEO and he also found particularly useful the 'data repository' where users can save and share the results of data analysis with other users or groups. Neal also finds SMD's data repository very helpful and adds that it also provides a jumping-off point for various analysis tools (e.g. hierarchical clustering, singular value decomposition). But another disadvantage of SMD Neal finds to be important is that there is a submission cost involved in case Jim or Alice are willing to submit their data. Jim thinks that the submission cost is relevant to the volume of data submitted which in his case is not very high, but Alice argues that given this is an ongoing collaboration the cost is something they should have in mind. Also Alice notes that the SMD data are also submitted to Array Express, as well as some data from GEO. Jim and Neal agree that Alice's last point is sufficient to delete SMD from their list. Jim also adds that in that case he wouldn't mind using either ArrayExpress or GEO, in fact since Alice prefers Array Express but also likes the Blast option of GEO she should decide which one of the two to choose. Finally, Alice suggests that given GEO includes more datasets than Array Express, and it is more cited in scientific journals - Neal adds, they should all agree on GEO database.

**Step-by-step actions:** In the following we describe in a step-by-step manner a part of the previous scenario where Neal, Alice have a collaborative discussion about the ArrayExpress database. Please be creative in summarizing the text of the 'Tasks' and supplying a title in the comments when this is not already done for you.

Neal username/ password: neal

Alice username/password: alice

<b>Tasks</b>	<b>User action</b>	<b>System response</b>
<p>1. Neal, Alice: log-in to the Dicode workbench (<a href="http://hodgkin.dia.fi.upm.es:8080/dicode/">http://hodgkin.dia.fi.upm.es:8080/dicode/</a>, <a href="http://dicodedev.cti.gr/Login/Default.aspx">http://dicodedev.cti.gr/Login/Default.aspx</a>).</p>	<p>Inert username and password. (If using Cope_it! select 'Continue anyway'). In the Welcome page, select the Workspaces from the Menu, and then the Use Case 1: Clinico-genomic research Assimilator. If using Cope_it! Select 'Workspaces', 'Create Workspace', type 'CGRA' in the 'Workspace title', and click the green tick.</p>	<p>Active workspace in the Mind-map view. There is no Mind-map view in the Cope_it!.</p>
<p>2. Neal suggests the following alternatives:</p> <ul style="list-style-type: none"> <li>i) ArrayExpress,</li> <li>ii) GEO,</li> <li>iii) SMD.</li> </ul>	<p>Neal inserts three new collaboration windows with titles: Stanford Microarray Database (SMD), ArrayExpress, Gene Expression Omnibus (GEO). New Collaborative, type in Name the title, and click 'Create collaborative'. If using Cope_it!, from the 'menu' in the bottom left corner of CGRA workspace select 'Add an adornment', which causes a coloured window to appear. Then click on 'Heading', type 'ArrayExpress' and click the green tick. Repeat for ii), iii).</p>	<p>Three windows appear in the mind-map view (or workspace in Cope_it!) with headers ArrayExpress, GEO, SMD.</p>
<p>3. Neal quotes from the ArrayExpress site (<a href="http://www.ebi.ac.uk/arrayexpress/">http://www.ebi.ac.uk/arrayexpress/</a>) that 'The ArrayExpress Archive is a database of functional genomics experiments including gene expression where you can query and download data collected to MIAME and MINSEQE standards. Gene Expression Atlas contains a subset of curated and re-annotated archive data which can be queried for individual gene expression under different biological conditions.'</p>	<p>In the ArrayExpress collaboration window Neal inserts a basic 'idea' object with its 'title' being ArrayExpress and in the content he pastes the text in Task 3. If using Cope_it!, double click anywhere in the workspace, and fill in the gaps in the dialog window as instructed above. Then move the 'idea' icon into the ArrayExpress collaboration window.</p>	<p>In the collaboration window there is now an 'idea' icon with a title, its content, the date it was posted and the name of the user.</p>
<p>4. Alice argues that ArrayExpress offers extra</p>	<p>In the ArrayExpress collaboration window Alice inserts a comment</p>	<p>A 'comment' icon will appear on the</p>

<p>options via the Gene Expression Atlas (<a href="http://www.ebi.ac.uk/gxa/">http://www.ebi.ac.uk/gxa/</a>), namely it provides information about gene-expression in different cell types, organism parts, developmental stages, disease states, sample treatments and other biological experimental conditions.</p>	<p>directly related to the ArrayExpress idea icon. Insert new object – comment – text – title: ‘Gene Expression Atlas’ – Insert the text in Task and click the green tick. If using Cope_it!, double click anywhere in the workspace, and fill in the gaps in the dialog window as instructed above. If you click on the comment, you will see at the top left corner a ‘connect objects’ icon, you click once and connect the comment to the ‘idea’ icon with an arrow. If you click the arrow, you can add text, specify the arrow’s width, its colour (‘In favour’, ‘Against’, ‘Neutral’ comment), or delete the arrow. Select the green colour, i.e. the note you just added is in favour of the idea analyzed in this collaboration.</p>	<p>ArrayExpress collaboration window with Alice’s comment. The comment is pointing with a green arrow to the ‘idea’ of this collaboration.</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------

Please complete the following Questionnaires:

- Questionnaire 1: Dicode Evaluation-First Trial-UC1 Clinicogenomic Research Assimilator Collaboration Services-Integrated Scenario

[http://kwiksurveys.com?s=LJMMJM\\_c8dc2f8d](http://kwiksurveys.com?s=LJMMJM_c8dc2f8d)

- Questionnaire 2: Dicode Evaluation-First Trial-UC1 Clinicogenomic Research Assimilator Collaboration Services-Ease of Use Assessment User Questionnaire

[http://kwiksurveys.com?s=LJMLGJ\\_e2d4ae54](http://kwiksurveys.com?s=LJMLGJ_e2d4ae54)

- Questionnaire 3: Dicode Evaluation-First Trial-UC1 Clinicogenomic Research Assimilator Collaboration Services- Acceptability Assessment User Questionnaire

[http://kwiksurveys.com?s=LJMLGI\\_7bddffee](http://kwiksurveys.com?s=LJMLGI_7bddffee)

## 2. The Dicode Forum Summarization service – Statistical analysis

You are taking part in a small experimental study about summarizing a forum (questions/answers) and finding relationships between topics. For this purpose, you will be provided with a **web-based tool** which has **two functionalities**:

- List of all forum questions;
- Topic cloud (tag/word cloud, a weighted word list). If you are not familiar with topic cloud, here is an example: [http://en.wikipedia.org/wiki/Tag\\_cloud](http://en.wikipedia.org/wiki/Tag_cloud)

The study involves three steps:

- **Step 1:** Providing information about your background;
- **Step 2:** Getting familiar with the tool;
- **Step 3:** Performing tasks and providing feedback.

The study is conducted in *anonymous* manner. No personal information (e.g. name, institution) will be kept. The results of the study will be used solely for research purposes and mainly to inform the improvement of the functionality offered by the tool which is being evaluated.

The study is conducted as part of the EU project Dicode ([www.dicode-project.eu](http://www.dicode-project.eu)). If you have any questions or comments about the study, please do not hesitate to contact Dr Fan Yang-Turner ([F.Yang-Turner@leeds.ac.uk](mailto:F.Yang-Turner@leeds.ac.uk)).

*Thank you very much for taking part in the study. We greatly appreciate your time and your participation.*

### Step 1: Providing Information about Your Background

#### Your research experience

We expect that you have some research experience, please tell us your current role:

- Undergraduate student     
  Masters student     
  PhD student  
 Postdoctoral researcher     
  Faculty (academic staff)     
  Other than above

#### Your skills of statistical analysis

We expect that you have some experience with statistics, please tell us your skills level

- Basic level (I have attended university statistics course.)  
 Experience level (Statistical analysis is part of my work.)  
 Degree Level (I have a university degree related to statistics.)  
 Professional Level (I am a professional statistician.)

#### Your experience with Internet forums

We expect that you have some experience with Internet forum, please tell us that you have

- visited Internet forum, reading the discussions  
 registered with Internet forums, asking questions  
 registered with Internet forums, answering questions  
 registered with Internet forums, asking and answering questions

## Step 2: Getting Familiar with the Tool

*In this study, we have collected 4854 questions from a forum called "Cross Validated", where data analysts or statisticians post questions on statistical analysis. For this study, you are provided with a web tool (URL below) to access these questions:*

<http://imash.leeds.ac.uk/dicode/wp4/Stats2-Forum/> (Main Page)

*From this URL, you have access to two functionalities:*

**View All Collected Discussion Threads** - *the first link on the main page, where all questions are listed sorted by number of views. In the list, you can see the title of the question (Thread Title), number of the views of this question, number of replies of the question and a timestamp showing when the question was posted. By clicking on Thread Title, you can read full detail of the question: the body of the question and its answers, if there are any.*

*You can use key word search to look for a list questions related to a particular topic by typing key words in the search box (at the top of the page). Search result will show the number of questions retrieved.*

**Browse Topic clouds** - *the second link on the main page where all questions are grouped into topic groups by a machine algorithm taking into account the similarity of the text content. Each group is presented with a topic cloud, showing the most frequent words (topics) mentioned in the text content for that group. Each topic cloud has a title, which shows the most discussed topic in that group, other topics listed in the word cloud are presented in different font sizes, which represent the number of discussions.*

*For example, "Regression" is the title in one group as Regression is the most frequent word in the content for that group. Within this group, there are another 10 topics that show other 10 most frequent words in the content for that group. In each topic group, under the title, it also shows the number of discussions in that group and its percentage among all discussions in the forum.*

*By clicking the topic in the topic cloud (group title or words in the group), you can see all discussions from the group which include this word. By clicking on the title of the question, you can read the full detail of the question: the question and its answers, if there are any.*

*Now, please spend 5 minutes to get yourself familiar with this tool and then go to Step 3.*

### Step 3: Performing Tasks and Providing Feedbacks

#### Scenario

*Imaging that you are a member of a research team, which works on analysing a large volume of data to discover meaningful patterns (for example a bioinformatics research team analysing clinico-genomic data). Your team conducts pre-processing of the data, setting the correct sample size for testing hypotheses, analysing the data, and presenting the findings. The team has learnt that forums where people ask relevant questions about statistical analysis can provide helpful information, as it shows recent topics in this area and shares practical experience between people with different expertise. One such forum is "Cross Validated" where people conducting statistical analysis discuss issues they have encountered. You have been asked by your team to check this forum and report back to them.*

You are required to prepare **three presentation slides** to the team using the templates provided. You have **60 minutes at maximum** to do this task. The title and subtitles of the presentation slides have been pre-defined in the template.

*After each slide, you are required to **answer questions** about your experience of completing the slide and your **overall comments** on topic clouds. Please also **record the time** you have spent on completing each slides.*

To complete the task, please use this link **only**:

<http://imash.leeds.ac.uk/dicode/wp4/Stats2-Forum/>

**Do not use any other sources for completing the tasks. We do not evaluate your knowledge but the usefulness of the two functionalities:**

- View All Collected Discussion Threads
- Browse the Topic Clouds

Now, please complete the three presentation slides in the following pages and answer the questions.

Please complete the following Questionnaire:

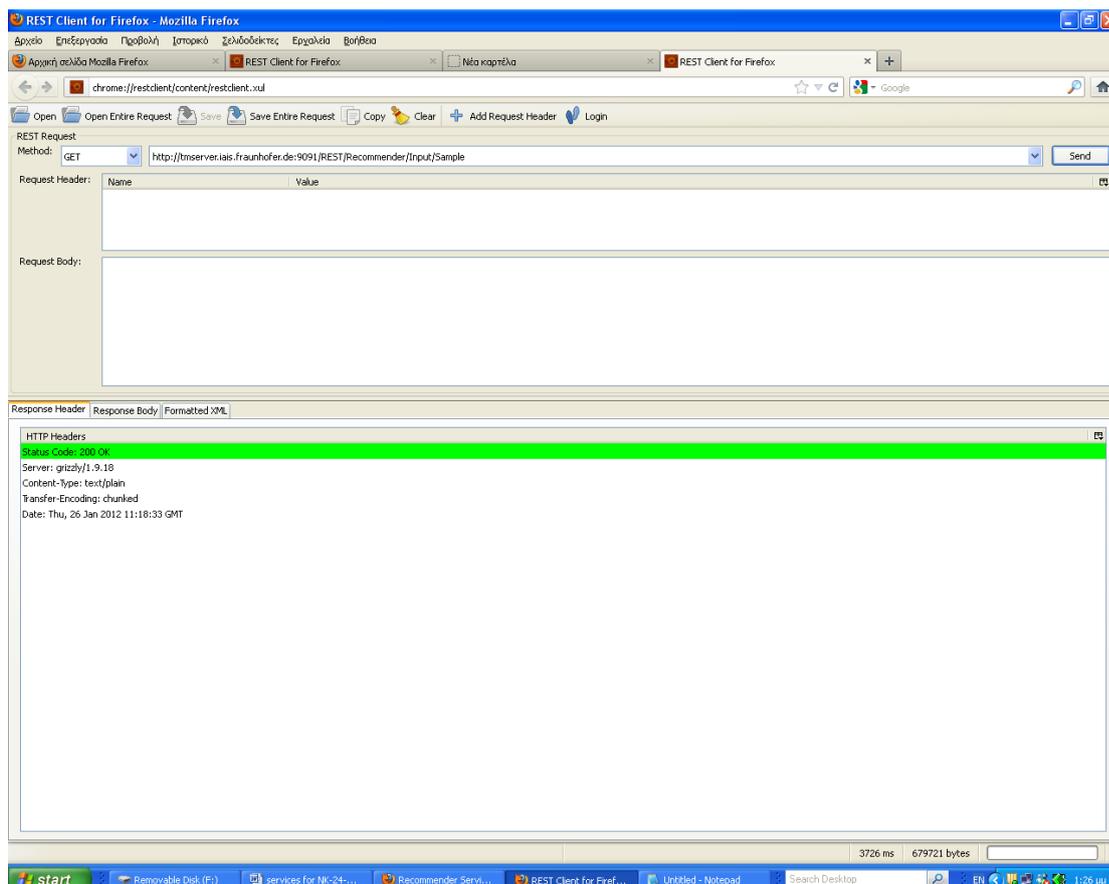
[http://kwiksurveys.com?s=LJHMJJ\\_61664a1c](http://kwiksurveys.com?s=LJHMJJ_61664a1c)

### 3. Recommender service

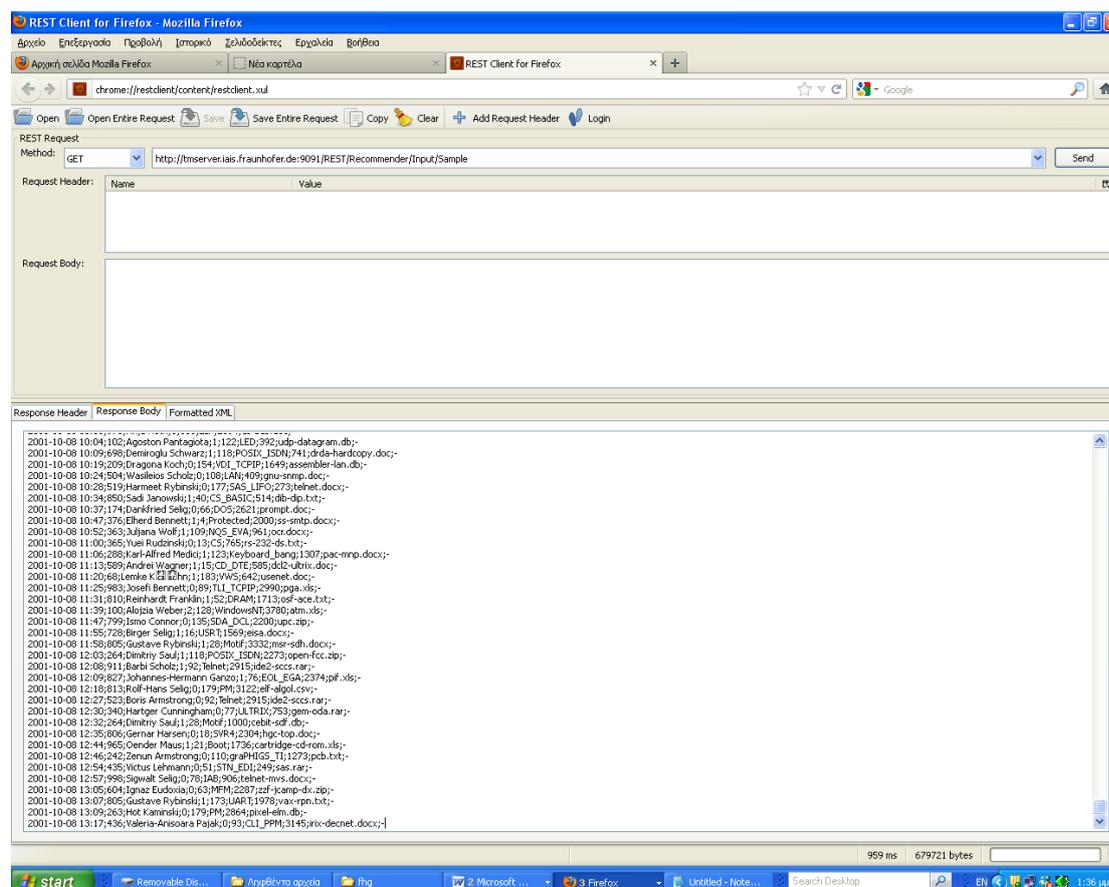
In order to use the service a REST-Client such as RESTClient must be installed (<https://github.com/chao/RESTClient>).

Example Request:

GET <http://tmserver.iais.fraunhofer.de:9091/REST/Recommender/Input/Sample>



### Example Output:



Please complete the following Questionnaires:

- Questionnaire 1: [Dicode Evaluation-First Trial-Standalone Service-Recommender Service, http://kwiksurveys.com?s=LKLJGI\\_fa8ccd89](http://kwiksurveys.com?s=LKLJGI_fa8ccd89)
- Questionnaire 2: [Dicode Evaluation-First Trial-Standalone Service-The Recommender Service -Ease of Use Assessment User Questionnaire http://kwiksurveys.com?s=LKLJGH\\_8d8bfd1f](http://kwiksurveys.com?s=LKLJGH_8d8bfd1f)
- Questionnaire 3: [Dicode Evaluation-First Trial-Standalone Service-The Recommender Service - Acceptability Assessment User Questionnaire http://kwiksurveys.com?s=LKLJGF\\_6a33d018](http://kwiksurveys.com?s=LKLJGF_6a33d018)

Additional instructions in:

<https://wiki.dicode-project.eu/display/DIC/Recommender+Service>

## 4. SubgroupDiscovery service

Example Request:

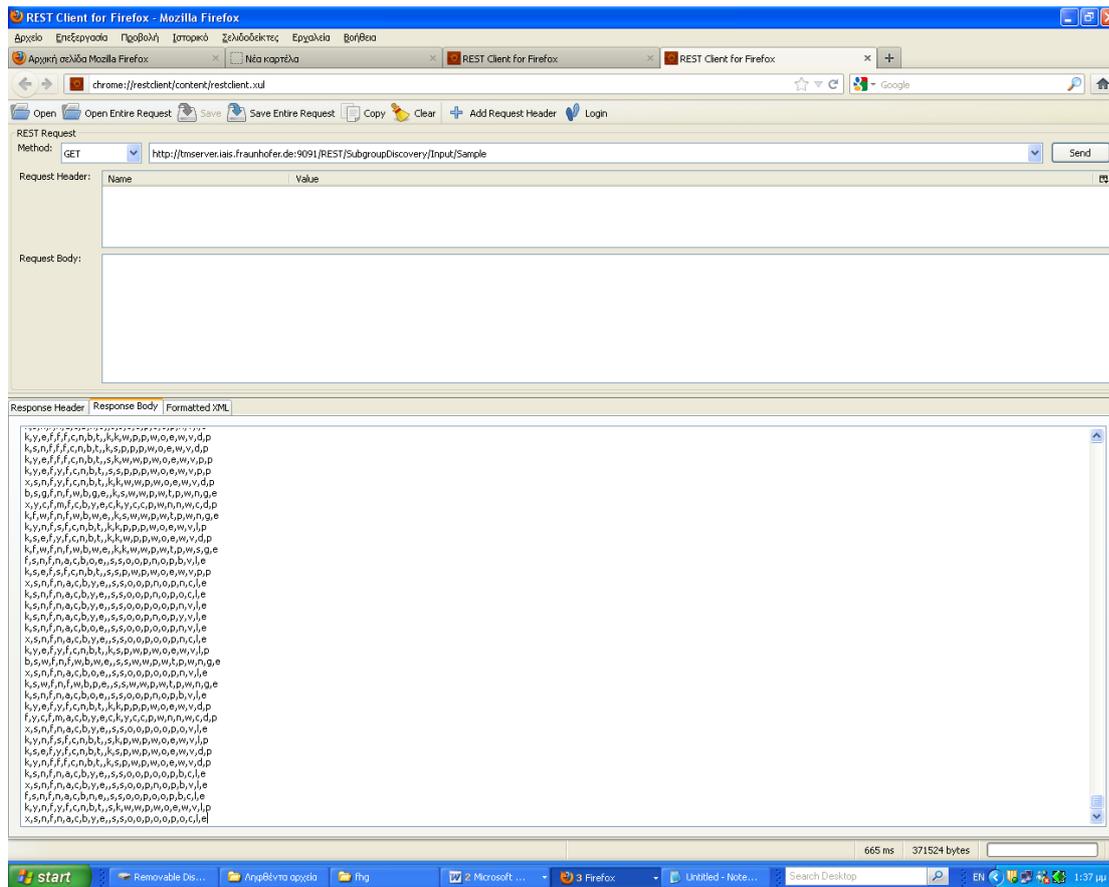
GET <http://tmsserver.iais.fraunhofer.de:9091/REST/SubgroupDiscovery/Input/Sample>

The screenshot displays the REST Client for Firefox interface. The top section shows the REST Request configuration with the Method set to GET and the URL set to <http://tmsserver.iais.fraunhofer.de:9091/REST/SubgroupDiscovery/Input/Sample>. The Request Header and Request Body sections are empty. The bottom section shows the Response Header, which includes the following information:

```
HTTP Headers
Status Code: 200 OK
Server: grizzly/1.9.10
Content-Type: text/plain
Transfer-Encoding: chunked
Date: Thu, 26 Jan 2012 11:14:29 GMT
```

The status bar at the bottom indicates a response time of 3951 ms and a response size of 371524 bytes.

Expected output:



Please complete the following Questionnaires:

- **Questionnaire 1: Dicode Evaluation-First Trial-Subgroup Discovery Service**  
[http://kwiksurveys.com?s=LKLENH\\_5715016b](http://kwiksurveys.com?s=LKLENH_5715016b)
- **Questionnaire 2: Dicode Evaluation-First Trial-Standalone Service-Subgroup Discovery Service -Ease of Use Assessment User Questionnaire**  
[http://kwiksurveys.com?s=LKLENG\\_c7aa1cfa](http://kwiksurveys.com?s=LKLENG_c7aa1cfa)
- **Questionnaire 3: Dicode Evaluation-First Trial-Standalone Service-Subgroup Discovery - Acceptability Assessment User Questionnaire**  
[http://kwiksurveys.com?s=LKLEOO\\_d06aa589](http://kwiksurveys.com?s=LKLEOO_d06aa589)

Additional instructions in:

<https://wiki.dicode-project.eu/display/DIC/SubgroupDiscoveryService>