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

D4.1.1 – The Dicode Collaboration Support Services (initial version)

Due date of deliverable:	31 August 2011
Actual submission date:	31 August 2011
Responsible Partner:	CTI
Contributing Partners:	CTI, NEO

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: Data-intensive collaboration support services, rich interactive search and analysis mechanism, abstract service description, interfaces, operations, development.

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

The Dicode Consortium

Research Academic Computer Technology Institute (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	05-08-2011	First draft version	Manolis Tzagarakis (CTI)
2	19-08-2011	Second draft version (sent to internal reviewers)	Doris Maassen (NEO) Manolis Tzagarakis (CTI)
3	26-08-2011	Third draft version; reviewers' comments incorporated (sent to SC)	Doris Maassen (NEO) Manolis Tzagarakis (CTI)
4	31-08-2011	Final version (approved by SC, sent to the Project Officer)	Manolis Tzagarakis (CTI)

Deliverable managers

- Manolis Tzagarakis (CTI)

List of Contributors

- Manolis Tzagarakis (CTI)
- Spyros Christodoulou (CTI)
- Nikos Karacapilidis (CTI)
- Doris Maassen, (NEO)

List of Evaluators

- Stephan Rueping (FHG)
- Guillermo de la Calle Velasco (UPM)

Summary

This deliverable is to be considered as a progress report on the initial version of the Dicode collaboration services, which are designed and implemented in the context of WP4. Collaboration services developed in Dicode concern the exploitation of the reasoning abilities of humans to capitalize on the outcomes of the data mining services developed in the project. In this deliverable, the technical specifications of the collaboration services related to Tasks 4.1 and 4.4 of WP4 are presented. The intended audience of this document are designers and developers of the Dicode project. The document informs them on which collaboration services have been developed and how they can be used, thus providing a first framework for discussing how to proceed with the full development of the envisioned services. The initial version of the services is presented using a formalized and project-wide adopted service description template.

Table of Contents

1 Introduction 5
1.1 Context 5
1.2 Objectives 5

2 Dicode Collaboration Support Services 5
2.1 Rich Interactive Search and Analysis Mechanism 6
2.1.1 Collaboration Search Service 6
2.2 Collaboration Support Services 9
2.2.1 Collaboration Service 10

3 Development Process 17

4 Future Work 17

5 References 18

Appendix A: Collaboration Support Services Demonstration 19

1 Introduction

1.1 Context

This deliverable presents the initial version of collaboration support services that have been designed and developed in the context of WP4 ("Data-Intensive Collaboration & Decision Support Services") of the Dicode project. More specifically, it reports on the progress of work being carried out, by describing the related initial version of services that have been developed in Tasks 4.1 and 4.4.

This is the first of a series of deliverables reporting on the progress of work of collaboration support services in the context of WP4. While the focus of this deliverable (D4.1.1) is on the initial version of collaboration support services, deliverable D4.1.2 (due in month 24) reports on their enhanced version. The final version of the collaboration support services will be reported in deliverable D4.1.3 (due in month 33).

1.2 Objectives

The purpose of this document is to present the initial version of the developed services related to collaboration support in the context of Tasks 4.1 and 4.4, as they originated from the functional specifications outlined in deliverable D2.2 ("The Dicode Approach – User requirements, conceptual integrative architecture, agile methodology and functional specifications") in order to frame and start the implementation related discussion on how to realize the vision of the Dicode project.

The developed services are presented from a technical perspective, broken down to the level of individual operations, in order to show their role and use and facilitate their assessment with respect to the derived functional specifications. The operations presented are those which are available to clients to be invoked and executed without going into detail about how exactly these can be invoked or executed. In particular, the presented operations can be executed by various technologies such as REST (Fielding, 2000) or Web Services (Web Service Architecture, 2004), but such issues are not the focus of their description. The description of services takes an operation-oriented approach listing the available operations and detailing their aim and purpose.

The initial version of the services are presented using a service description template, which has been derived and used in the context of deliverable D3.1.1 ("The Dicode Data Mining Framework"), called the Abstract Service Description. The Abstract Service Description template provides a technical specification of services by providing an overview of the supported interfaces and the relevant operations. For each operation, a description along with major input and output information is presented.

2 Dicode Collaboration Support Services

The aim of the collaboration support services is to exploit the reasoning abilities of humans to facilitate sense-making of the results of the Dicode data mining services, which are the focus of WP3, thus capitalizing on their outcomes. Towards this, a number of relevant

services will be developed that can fully address the user requirements of Dicode as outlined in deliverable D2.2.

This documents reports on the initial version of the services that have been developed in the context of the following tasks:

- Task 4.1 - Rich interactive search and analysis mechanism, which is concerned with providing full-text and meta-data search of collaborative spaces and interaction of users with the analytical processes, and
- Task 4.4 - Collaboration Support Services, which is concerned with the creation, management and use of innovative workspaces that augment synchronous and asynchronous collaboration.

In the following, we present the service description of the initial version of the services being developed in the abovementioned tasks of WP4. For each service, a short description related to its aim and purpose is given, followed by the abstract service description template of its initial version..

2.1 Rich Interactive Search and Analysis Mechanism

Services related to rich interactive search and analysis mechanisms are the focus of Task 4.1. A workspace in Dicode contains collaboratively generated content of various types: on the one hand, there are dedicated item types such as notes, ideas and comments. On the other hand, there is a wide array of uploaded file types such as word documents, PDF, RTF and images. Both are described using standardized meta-data. Additionally, the user can add his/her own meta-data types. Workspaces serve as structure for collaborative content creation. The Search Service uses workspace meta-data like topic, workspace creator and creation date for result selection and ranking.

Task 4.1 is also associated with the provision of services related to supporting analysis enabling the navigation through the various data processing layers. The description of these services will be presented in future deliverables.

2.1.1 Collaboration Search Service

A prototypical implementation of the "Collaboration Search Service" will be provided. This service allows for indexing and searching of standard documents and helps users explore the opportunities of full-text and meta-data search. A search service tailored to Dicode's requirements can only be implemented after the document and meta-data specification for Dicode's Collaboration Support application is defined. In any case, the following, rather coarse, schema has been elaborated:

Name	Description	Search Specification
content	Element text, text content of PDF or RTF	Full text search
title	Title of the element	Full text search
user	Creator of the element	Field search, Full text search

date	Creation date	Field search
rating	Rating on an ascending scale from 1 to 5 indicating how important the item is	Field search
type	Type of the element such as: Note, Idea, Comment	Field search
Dynamic meta-data[1]	arbitrary key-value pairs which can be defined by the users	Field search

The schema defines which fields are indexed and how they can be searched. In a full-text search, the results are usually ranked by relevance. This relevance is computed as the distance between query and document, where both are represented as a term vector. Additionally, results can be sorted or grouped by date or by natural order. Faceted search groups all search hits by a certain feature, e.g. by user, rating or type. Fields can also be used for result filtering in full-text search. Filtering search results based on particular collaboration workspaces is also possible by introducing a field "workspace id" and return only content workspaces to which the user has access.

The abstract service description of the initial version of the "Collaboration Search Service" is presented below.

Name	Collaboration Search Service
Standards	• REST (Fielding, 2000) • XML (W3C – Extensible Markup Language, 2006)
Description	The Dicode Collaboration Search Service The Collaboration Search Service provides its functionality through the following interfaces: • Indexer: The Indexer indexes full-text and meta-data of documents[2]. • Searcher: The Searcher allows for full-text and field search in the index[3].
Interface	*Indexer*
insertDocument	For indexing, an XML representation of the document is posted to the Collaboration Indexing Service URL. The following example shows a simple document:

[1] The search technology which will be used, Apache Solr, supports dynamic fields, so arbitrary key-value pairs can be added to the meta-data.

[2] The service is provided by a Solr server instance. Indexing is performed via the URL http://${server}:${port}/solr/update

[3] Searching is performed by passing the query string to the following URL http://${server}:${port}/solr/select/?query=myQuery

	<add> <doc> <field name="id">42</field> <field name="title">Recent findings</field> <field name="user">nora</field> <field name="date">210-12-31T23:59:59Z</field> <field name="type">1</field> <field name="rating">5</field> <field name="full-text">At the recent conference ...</field> </doc> </add> The document has to conform to the schema definition of the indexer.
updateDocument	Updates an existing document in the index
deleteDocument	Deletes a document from the index
Interface	*Search*
fullTextSearch	Search for all documents matching a full text query. Depending on the server configuration a set of additional request parameters can be used to determine the return format, to specify the preferred ranking algorithm etc. The developer documentation (Apache Solr Tutorial, 2011) gives the following example which shows a query for the term "video". The parameters determine both the fields returned, the result format and the ranking algorithm to be applied: • q=video&fl=name,id (return only name and id fields) • q=video&fl=name,id,score (return relevancy score as well) • q=video&fl=*,score (return all stored fields, as well as relevancy score) • q=video&sort=price desc&fl=name,id,price (add sort specification: sort by price descending) • q=video&wt=json (return response in JSON format)
facetedSearch	Provides a dynamic clustering of search results according to a set of categories like in the following example where the results are grouped by author and by rating (Faceted Search with Solr, 2009): <lst name="facet_fields"> <lst name="author"> <int name="Ada Albert">17</int> <int name="Berta Brenner">12</int> <int name="Cecilia Cartwright">12</int> </lst>

	<lst name="rating"> <int name="1">5</int> <int name="2">11</int> <int name="3">17</int> </lst>
Example usage	The Collaboration Search Services is accessed by Dicode's Collaboration Support application. The application will trigger indexing of a set of selected elements like full text and meta-data. This process is considered the "export" of a document. The service allows for field updates – only the relevant fields are re-indexed in this case. As a rule of thumb full text indexing of entire documents should be performed only if a significant change of the document occurred.
Comments	The Collaboration Search Services will be implemented based on the Open Source Framework Apache Solr (Apache Solr, 2011) which uses the quasi standard of Open Source Search, Apache Lucene (Apache Lucene, 2011). Solr offers a REST interface for indexing and searching. Alternatively the indexer and the searcher can be accessed via the Java API or one of the libraries for other programming languages like http://code.google.com/p/solrnet/ for asp.net/.NET.
Conformance classes	Not available
Implementation rules	Not available
Implementation status	Under development
UML model	Not available

2.2 Collaboration Support Services

Dicode's Collaboration Support Services are the focus of Task 4.4; these services aim at developing innovative virtual workspaces which support collaboration towards sense-making in data intensive settings.

From a technical perspective, an initial version of a service has been developed, called "Collaboration Service", through which collaboration related functionalities are available to clients[4]. The initial version of the "Collaboration Service" implements the view-based collaboration workspaces presented in deliverable D2.2. In particular, the service provides operations to create and manage workspaces and supports collaboration using the following views:

- **Forum-view,** where the collaboration workspace is presented as a Web-based discussion forum, and

[4] By the term "client" we refer to any program (including Web browsers) able to request functionalities by executing specific operations.

- **Mind-map view,** where the collaboration workspace is presented based on a spatial paradigm.

Each view provides a different set of operations which originate from the purpose of the view. The "Collaboration Service" provides its functionalities via a set of interfaces, each of which defines a set of operations that the service implements. These interfaces and related operations are presented in the next section.

2.2.1 Collaboration Service

The "Collaboration Service" provides the implemented operations via interfaces, each of which focuses on a particular aspect of collaboration-related activities. In general, the available interfaces can be classified into three distinct categories, based on their role and purpose:

- **Core Collaboration Interfaces:** These are interfaces that define operations aiming at facilitating all necessary collaborative actions within Dicode.
- **Integration Interfaces:** These are interfaces that define operations related to integrating collaboration services with other services developed in the context of Dicode.
- **Interoperability Interfaces:** These are interfaces that define operations related to achieving interoperability between the Dicode collaboration support services and third party tools, i.e. tools that have not been developed in the context of the project (such as Web-based forums, standalone applications as Compendium (Buckingham Shum et al 2006) etc). In particular, they permit importing resources or entire collaborations (such as threads in Web-based discussion forums) into the Dicode collaboration workspaces or exporting workspaces to a variety of formats.

Although the integration and interoperability interfaces do not provide core collaboration operations, they nevertheless constitute an important and critical mechanism towards integrating collaboration support services and implementing the "Dicode Workbench" as described in the deliverable D2.2. For this reason, we include their presentation in the following abstract service description.

Name	Collaboration Service
Standards	• REST (Fielding, 2000)
Description	The Collaboration service provides all necessary operations that permit users to create virtual workspaces in which the collaboration takes place and to conduct collaborative activities within workspace. In addition, the collaboration service provides operations to facilitate integration with tools developed in the context of Dicode and interoperability with third party tools (i.e. tools not developed in the context of the project). The Collaboration Service provides its functionality through the

	following interfaces: • *Workspace Interface:* The purpose of the workspace interface is to provide all operations related to the creation and management of virtual workspaces where the collaboration is taking place. • *Forum Interface:* The purpose of the forum interface is to provide all operations related to enabling collaboration between users in the context of Dicode when a workspace is operated in forum view. • *MindMap Interface:* The purpose of the MindMap interface is to provide all operations related to enabling collaboration between users in the context of Dicode when a workspace is operated in mind-map view. • *UserManagement Interface*: The purpose of the UserManagement interface is to provide operations enabling the management of user profiles which are maintained by collaboration services. • *Authentication Interface*: The purpose of the Authentication interface is to provide user authentication and connection operations to the collaboration services. Collaboration services require users to authenticate in order to participate in collaborative sessions. • *External Interface*: The purpose of the external interface is to provide operations that enable interoperation (in terms of import and export) with third party tools. With regard to their role and purpose, the Workspace, Forum and MindMap interface provide operations which are necessary to support collaboration and belong to the Core Collaboration Interface category. The UserManagement, and Authentication interfaces provide operations primarily aimed at achieving integration with tools developed in the context of Dicode, belonging to the Integration Interface category. Finally, the External interface provides operations to facilitate exchange of discourses with third party tools consists an Interoperability Interface.
Interface	*Workspace*
createWorkspace	Creates a new workspace where the collaboration takes place. The CreateWorkspace operation receives as inputs the title of the workspace –indicating the purpose of the workspace - along with an arbitrary number of attribute-value pairs that the creating user may supply. This information constitutes the metadata of the workspace. Upon successful completion of the operation, a new workspace is created and permanently stored into the database that users may open and interact with.
openWorkspace	Opens an existing workspace and returns the items that it

	contains. The operation takes as input the ID of the workspace. Upon successful completion, the items contained in the workspace are returned. The OpenWorkspace operation opens a workspace in mind-map view.
updateWorkspace	Updates the access rights of a workspace. Takes as input the id of the workspace as well as the list of users who have access to the workspace. Upon successful completion, the designated user list becomes access (read and write) to the workspace.
copyWorkspace	Creates a copy (i.e. a clone) of an existing workspace, duplicating its metadata, access rights and all items that the original workspace contains. Takes as input the ID of the workspace that is to be copied (the "source"). Upon successful completion a new workspace is created having the same metadata and access rights with the source workspace. In addition, the new workspace contains a copy of each item that belongs to the source workspace. After a successful copyWorkspace operation, the two workspaces evolve independently.
transformWorkspace	Transforms the workspace in a different view. The operation takes as input the ID of the workspace being transformed and the desired target view (can take one of the values: forum, mind-map and formal). Upon successful completion, the workspace items are returned according to the selected view.
createKnowledgeType	Creates a new knowledge type, active in specific workspaces when the workspace is operated in mind-map view. Once a new knowledge type is define/created, instances of the type can be uploaded into the associated workspaces.
Interface	*Forum*
postItem	Posts a new item (reply) on a workspace in the forum view. The operation takes as input the ID of the workspace, the ID of the topic the post will appear, the content of the post as well as an optional list of files which are attachments to the post. Upon successful completion, the new post will appear on the specified topic.
Interface	*Mind-map*
createItem	Creates a new item of a specific type in a workspace. The operation takes as input the ID of the workspace, where the item will be created, the knowledge type of the item, the title of the item, the content, the mime-type of the content as well as the coordinates within the workspace where the item will be created. Upon successful completion, the new item with the specified input data is created. The newly created item appears on the workspace in mind-map view at the specified coordinates.

uploadItem	Uploads an existing file into a workspace. The file can be of any type such as PDF, PS, Word documents etc. The operation takes as input the ID of the workspace where the file will appear, the content of the file as well as a title. Upon successful completion, the file is permanently stored and appears on the workspace when it is opened.
createRelationship	Creates a relationship between two items that reside in the same workspace. The relationship created with this operation is visible only in mind-map view of workspaces and appear as a directed, colored arrow from a source to a destination item that reside on the workspace. The operation takes as input the ID of the workspace, the ID of the item which is the source of the relationship, the ID of the item which is the destination of the relationship, an optional title which will appear as the relationship's caption and the color of the arrow which represents the relationship. Upon successful completion, a relationship between these two items is created, which is depicted as a directed arrow when the workspace is opened in mind-map view.
updateItem	Updates the metadata of an existing item. The operation takes as input the ID of the item, all metadata fields of the item (such as title, type) as well as an arbitrary list of attribute value pairs. Upon successful completion, the metadata of the specified item is updated.
updateRelationship	Updates the title and visual appearance of a relationship on a workspace. The operation takes as input the ID of the relationship, the relationship's title and the visual appearance of the relationship which refers to the thickness of the arrow representing the relationship and the arrow's color. Upon successful completion, the title and/or visual appearance of the relationship is updated.
deleteItem	Deletes an item from a workspace. Takes as input the ID of the workspace that contains the item and the ID of the item being deleted. Upon successful completion, the specified item is removed from the workspace. In case the deleted item is either the source or the destination of a relationship, the corresponding relationships are also deleted.
deleteRelationship	Deletes a relationship from a workspace. The operation takes as input the ID of the workspace and the ID of the relationship to be deleted. Upon successful completion, the relationship (i.e. arrow) is removed from the workspace.
moveItem	Moves an item from one position to another on the workspace. The operations takes as input the source and destination coordinates of the move.

createGroup	Creates a visual grouping of items on a workspace. A visual grouping on a workspace is a colored rectangle, which encloses all items that belong to the same group. In mind-map view, such grouping of items is only for visualization purposes. This means that at the storage layer (i.e. the database) no information is kept related to which item belong to which group. The operation takes as input the ID of the workspace where the rectangle (group) will be created, the title for the rectangle, the coordinates of the upper left corner of the rectangle representing the group, the rectangle's height and width and its color. Upon successful completion, a new rectangle with the specified characteristics is created and stored into the database.
updateGroup	Updates the title and visual characteristics of a grouping (rectangles) on a workspace. The operation takes as input the ID of the workspace, the ID of the rectangle representing the grouping, the title of the rectangle and the rectangle's color. Upon successful completion, the attributes of the rectangle are updated.
deleteGroup	Deletes a rectangle (grouping) from a workspace. The operation takes as input the ID of the workspace where the rectangle belongs and the ID of the rectangle that is to be deleted. Upon successful completion, the rectangle (grouping) is removed from the space. When deleting a rectangle which group items on a workspace, the contained items are not deleted and remain on the workspace. The DeleteGroup operation only removes the rectangle that encapsulates the items, not the items themselves.
Interface	*UserManagement*
registerUser	Registers a new user within the collaboration services by creating a new user profile. In order to use collaboration services, users have to register and authenticate themselves before they can gain access to existing collaboration workspaces. The operation takes as input a name, which will be displayed whenever the users interact with the collaboration services (such as creating items on workspaces), a desired password, the user's first and last name, the user's email address along with additional information such as his/her postal address, website, home and work phone number as well as a short description. In addition, a list of arbitrary attribute value pairs can be provided. Out of those input variables, name, password and email are mandatory while the other variables (first name, last name, postal address, website, home/work phone, description and arbitrary attribute value pairs) are optional. Upon successful completion, a new user profile is created and the ID of the created user profile is returned.
updateProfile	Updates the profile of an existing user. The operation takes as input all fields comprising the user profile (specifically name, password, email, first name, last name, postal address, website,

	home/work phone, description and arbitrary attribute value pairs). Upon successful completion, the user profile is updated with the new values.
Interface	*Authentication*
loginUser	Authenticates a user to the collaboration services and initializes the all data structures to allow users to use the collaboration services. The operation takes as input the name of the user, the password and a timestamp. The input variables name and password are mandatory while the timestamp is mandatory when requests originate from outside hosts. The value of the timestamp represents the time the LoginUser operation was invoked, expressed in GMT. The timestamp variable is used to check the validity of requests. In particular, a LoginUser operation is only valid if it has been received by collaboration services within a short timeframe from its invocation (configured to be 10 minutes). LoginUser requests that exceed this period (i.e. the difference between current time in GMT and the timestamp is greater than 10 minutes) are considered invalid and rejected. Upon successful completion, the operation prepares all necessary data structures to allow users to use collaboration services.
logoutUser	Disconnects a user from the collaboration services. The operation takes as input the unique ID representing the user. Upon successful completion, the user is disconnected and all related data structures (including the session) are cleared.
Interface	*External*
exportWorkspace	Allows exporting a collaboration workspace, including all the items it contains, to an external file system file. The operation takes as input the ID of the workspace which is to be exported as well as options that permit the configuration of the export process. The options allow the caller to specify the format (XML, CSV, zip file) as well as whether files which reside in the workspace should be exported. Upon successful completion, a file – containing the contents of the workspace - with the specified format is created in the file system.
importWorkspace	Allows importing of files, which have been exported using the ExportWorkspace operation. The operation takes as input the ID workspace where the contents should be imported, the file containing the content of the exported workspace and the type of the file. If a negative value for the workspace ID is given, a new workspace is created. Upon successful completion, the contents of the file will appear either in an existing or new workspace.
exportToCompendium	Allows exporting an existing collaboration workspace to file, suitable for importing it into the tool Compendium (Buckingham

	Shum et al 2006). The operation takes as input the ID of the workspace which is to be exported and the export options. Upon successful completion, a file in a format that can be imported into the Compendium tool is created.
importFromCompendium	Allows importing files into the collaboration services that have been exported using the Compendium tool into a workspace. The operation takes as input the ID of the workspace, where the items should be imported as well as the file which contains an export from Compendium. If a negative value for the workspace ID is given, a new workspace is created. Upon successful completion, the items contained in the Compendium export file will appear in a new or existing workspace.
importFromForum	Allows importing the contents of an external web-based discussion forum into a workspace. The operation takes as input the ID of the workspace, where the external forum should be imported, the URL which identifies a topic of a forum on the Web and the type of the forum (whether it is phpBB-based etc). The operation crawls the topic starting at the specified URL and identifies - using regular expressions - the individual posts. These posts are transformed into a knowledge type (e.g. note) and added to a workspace whose content is the content of the post. If a negative value for the workspace ID is given, a new workspace is created. Upon successful completion, the posts of the specified forums will appear as items in a new or existing collaboration workspace.
Example usage	A web application will provide the necessary user interface through which the previously mentioned operations can be executed by end users. In general, all Dicode use cases that require support for collaboration towards sense-making will be able to use the above operations. In particular, users of the Dicode use cases can use the operations provided by the workspace interface to create and configure new workspaces where the collaboration will take place. Depending on their needs, they may deploy the workspace either in forum or mind-map view, where they are able to upload and process the available collaboration items via the respective interfaces. Operations belonging to the integration interfaces are intended to be used by other Dicode services. In particular, the User Management and Authentication interfaces will be used by the Dicode integration services (which are the focus of WP5) to implement the Dicode workbench
Comments	The initial version of the service reported provides a proof-of-concept implementation of all the presented operations, which nevertheless have bugs and malfunctions. Such issues will be addressed in future versions of the services, along with a more

	thorough evaluation of their usefulness.
Conformance classes	Not available
Implementation rules	Not available
Implementation status	Prototypical version implemented.
UML model	Not available

The **source code** of the initial version of the collaboration service, implementing the above operations, can be found at the following Subversion repository:
https://anivas.cti.gr/websvn/listing.php?repname=ftel&path=/dicode/trunk/src/dll/

In Appendix A, an **example** of how the above operations of the collaboration service can be used to support the needs of the Dicode use cases is presented.

3 Development Process

The initial version of the previously described services have been developed using the guidelines outlined in deliverable D5.1.1 - Standards and guidelines for development. In particular, the Java Guidelines coding convention has been used to format the authored code. In addition, design patterns have been deployed to solve common software design problems.

To manage code changes, the involved project partners have installed on their site the Subversion code repository, which hosts the code of the initial version of the services they develop.

4 Future Work

This deliverable presents the initial version of collaboration support services that have been designed and developed in the context of Tasks 4.1 and 4.4, representing a first implementation of the functional specifications outlined in deliverable D2.2. Future work will concentrate on improving and enriching the above services in order to fully address the needs of the Dicode project. Specifically, with respect to the implementation of the collaboration support service, future work will focus on a number of issues which include:

- Correcting bugs and malfunctions of the initial version of the implemented services.
- Assessing the initial version of the developed services against the functional specifications. The purpose of this action is to see how the initial version of the services must be changed in order to properly support the Dicode use cases. This includes identifying which operations must change their functionality as well as which operations must be added in order to fully address the needs of the use cases.

5 References

Apache Lucene (published 28/07/2011). http://lucene.apache.org/

Apache Solr. (published 28/07/2011). http://lucene.apache.org/solr/

Apache Solr Tutorial. (updated 07/07/2011). http://lucene.apache.org/solr/tutorial.html

Buckingham Shum, S.J., Selvin, A.M., Sierhuis, M., Conklin, J., Haley, C.B. , Nuseibeh, B. (2006). Hypermedia Support for Argumentation-Based Rationale: 15 Years on from gIBIS and QOC. In: Rationale Management in Software Engineering, (Eds.) Allen H. Dutoit, Raymond McCall, Ivan Mistrik, and Barbara Paech. Springer-Verlag.

Dicode Deliverable D2.2 . (2011). The Dicode Approach – User requirements, conceptual integrative architecture, agile methodology and functional specifications.

Dicode Deliverable D3.1.1 (2011). The Dicode Data Mining framework (initial version).

Dicode Deliverable D5.1.1 . (2011). Standards and guidelines for development (initial version).

Faceted Search with Solr. (2009). http://www.lucidimagination.com/devzone/technical-articles/faceted-search-solr

Fielding, R. T. (2000). Architectural styles and the design of network-based software architectures, PhD thesis, University of California, Irvine, Chair-Richard N. Taylor, accessible from http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm .

Huttenhower, C., Hibbs, M., Myers, C., Troyanskaya, O.G. (2006) A scalable method for integration and functional analysis of multiple microarray. Bioinformatics, 22(23):2890-2897.

Mrowka, R., Lieberneister, W., Holste D. (2003). Does mapping reveal correlation between gene expression and protein-protein interaction? Nat Genetics, 33:15-16.

Shabalin, A.A., Tjelmeland, H., Fan, C., Perou, C.M., Nobel, A.B. (2008). Merging two gene-expression studies via cross-platform normalization. Bioinformatics, 24(9): 1154-1160.

Tusher, V.G., Tibshirani, R., Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response, in Proc. Natl. Acad. Sci. USA 2001, 98(9): 5116-5121.

W3C - Extensible Markup Language (XML) 1.0. (2006). http://www.w3.org/TR/2006/REC-xml-20060816

W3C – Web Services Architecture. (2004). http://www.w3.org/TR/ws-arch/

Appendix A: Collaboration Support Services Demonstration

This appendix presents an example of how the previously described collaboration support services (and their operations) are used in a specific context. In particular, we present and discuss their use in the context of a scenario belonging to Dicode's use case 1, which is relevant to clinico-genomic research. First, we present a motivational scenario outlining a typical collaboration example, when researchers engage in joint clinic-genomic research work. Based on this scenario, we then present how the collaborative aspects of the scenario are conceived in terms of the services and operations presented in this deliverable.

Motivating Scenario

Consider two researchers, Jim and Alice, aiming to investigate which genes or groups of genes are associated with breast cancer disease. Initially, they create a new collaboration session (logbook), where they exchange ideas related to which data sources to use, based on their own data analysis experience and literature knowledge. They search relevant literature via PubMed (http://www.ncbi.nlm.nih.gov/pubmed) using the appropriate search services. Jim has conducted an initial analysis with some in-house gene-expression datasets; however, his findings were not very encouraging, which was attributed to the small sample size (i.e. number of patients) available. He informs Alice about it and suggests potential solutions. The discussion proceeds and finally, in order to overcome the limited sample size problem, they decide to augment their samples with publicly available gene-expression data derived from the GEO (http://www.ncbi.nlm.nih.gov/geo/) and SMD (http://smd.stanford.edu/) databases.

After deciding what data to use, they keep collaborating in order to discuss how the data will be processed. Both suggest solutions, comment on them, and finally decide to use the normalized data for each platform and the UniGene annotation database (http://www.ncbi.nlm.nih.gov/unigene) to uniformly map all genes. Jim knows that there are particular confounding effects in such kind of analysis and for that reason suggests a specific strategy that would account for these effects. Particularly, they decide to first analyze the integrated dataset using the well-known Significant Analysis of Microarrays (SAM) methodology (Tusher et al., 2001). This analysis will serve as a baseline to any further analysis they attempt. Jim is also offering to provide all the necessary R scripts (http://www.r-project.org/) for this initial statistical analysis. In addition, they decide to employ model-based data integration methodologies (Huttenhower et al., 2006) (Shabalin et al. 2008) that have been recently published and claim to perform better than simple data integration techniques (Mrowka et al., 2003).

Some of the models are readily available; however, others need to be coded. Jim offers to write the relevant scripts. Alice, being an experienced programmer, offers to hard code them using parallel programming and various servers available at her department. Parallel and cloud computing will ensure fast results, since they have both agreed that they should apply the selected methodologies to numerous datasets. Their goal is to identify novel or already reported groups of genes associated with breast cancer disease. In addition, they are interested in comparing the findings of the chosen methodologies to those of the simple analysis conducted by Jim. They decided to quantify and check the statistical and biological significance of their results via the DAVID tool (http://david.abcc.ncifcrf.gov/) and the KEGG database (http://www.genome.jp/kegg/pathway.html).

Both researchers can execute the available services and retrieve the results of the invoked tool (e.g. a scatter plot or heatmap plot). Once the results are available, they engage into interpreting the results in terms of the initial research question.

The scenario presented above indicates particular situations where biomedical researchers need to collaborate. Their goals may vary during a collaboration session, ranging from merely sharing resources to brainstorming on available solutions and fostering sense-making to making crucial decisions. Support for collaboration in the biomedical domain require solutions that may easily enable researchers create and maintain private or public workspaces, where the most pertinent information about the problem at hand can be gathered, linked, synthesized and assessed. Through such workspaces, they need to carry out synchronous or asynchronous collaboration to accommodate and elaborate the outcomes of biomedical data mining, get recommendations, identify inconsistencies, spot and repair information gaps, reason about actions etc.

Addressing the scenario with the collaboration support services

Revisiting our motivating scenario, Jim and Alice would initially create a new workspace to support their collaboration. As their initial goal is to accumulate a critical mass of relevant resources, they create a new collaboration workspace (logbook) and may start using it in the 'forum-view' (Figure 1) by using the *createWorkspace* operation.

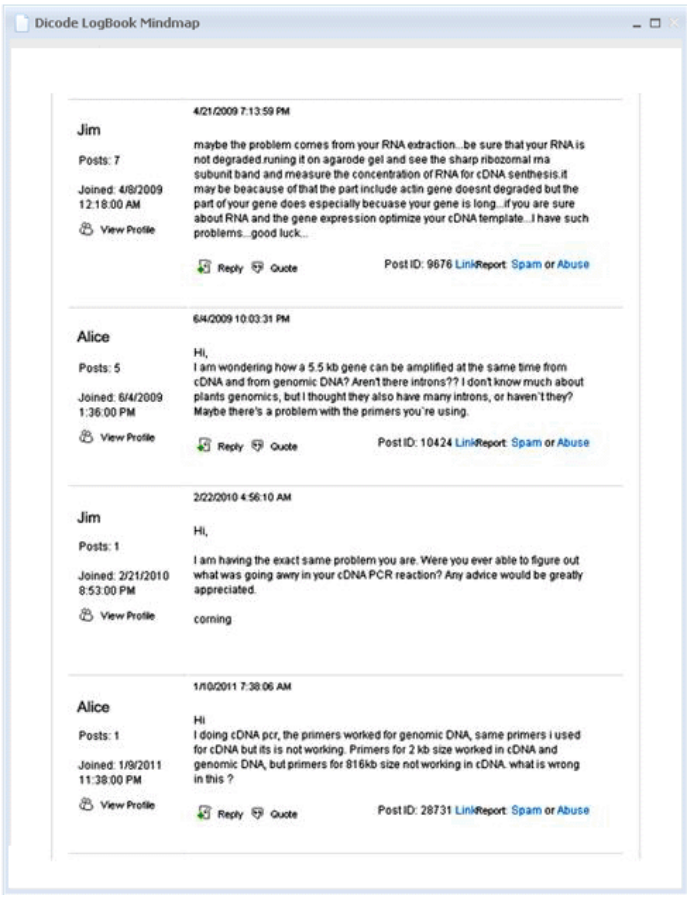

Figure 1: A forum view of the collaboration workspace.

The forum view primarily aims to effortlessly collect and share the available resources (without the need to interrelate them). During this collaboration phase, Jim and Alice upload available resources and assess them informally, by briefly commenting on them by executing the *postItem* operation.

When many resources start appearing in the forum view, Jim and Alice may decide to switch to a mind-map view (where they can better manage the numerous resources). They can perform this task by deploying the *transformWorkspace* operation transforming the workspace from forum view into the mind-map view. Once the *transformWorkspace* operation has been executed, the same workspace appears in mind-map view, as shown in Figure 2. In this view, Jim and Alice may organize the available items in more advanced ways and exploit dedicated item types such as ideas, notes and comments. Ideas stand for items that deserve further exploitation; they may correspond to an alternative solution to the issue under consideration and they usually trigger the evolution of the collaboration. Notes are generally considered as items expressing one's knowledge about the overall issue, an already asserted idea or note. Finally, comments are items that usually express less strong statements and are uploaded to express some explanatory text or point to some potentially useful. Multimedia resources can be also uploaded. Instances of the abovementioned item types can be uploaded into the workspace by utilizing the *createItem* operation, supplying the appropriate input parameters. Removal of items is possible by using the *deleteItem* operation while updating an item is done by executing the *updateItem* operation. All items on the workspace can be freely moved around via the *moveItem* operation.

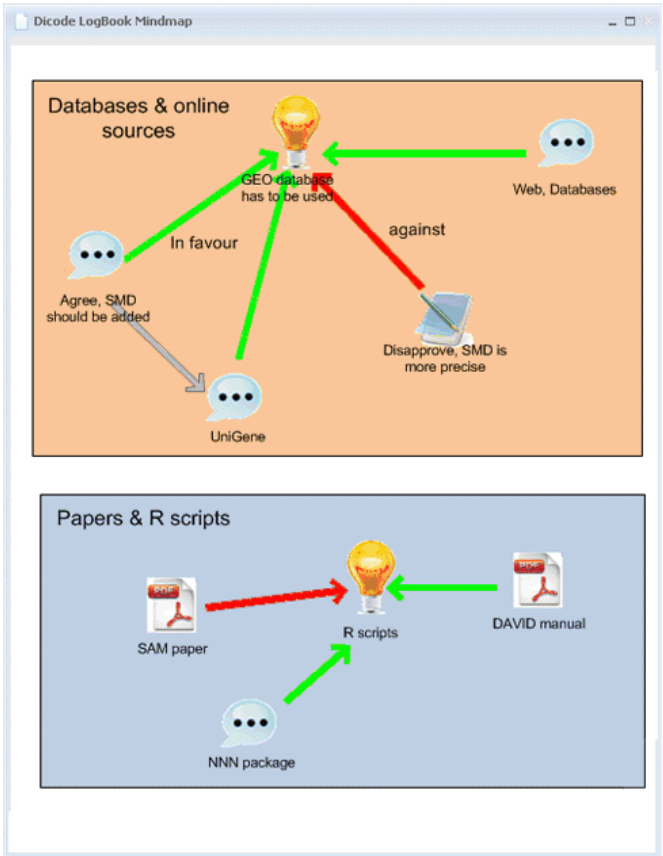

Figure 2: A mind-map view of the collaboration workspace shown in Figure 1.

All the above items can be interrelated by trouble-free actions (see the coloured arrows in Figure 2). When interrelating items, Jim and Alice may select the color of the connecting arrow and provide a legend describing the interrelationship they conceive. These legends are intentionally arbitrary. The visual cues of the arrows bear well-defined semantics: green arrows declare support, whereas red ones declare opposition. Such relationships can be created by calling the *createRelationship* operation with the appropriate input parameters indicating the label and the arrow's visual appearance. Relationships can be deleted and updated by using the *deleteRelationship* and *updateRelationship* operations respectively. Another visual cue that appears in Figure 2 concerns the coloured rectangles that have been created by Jim and Alice to group/cluster related items which can be created via the *createGroup* operation.

By using the mind-map view, Jim and Alice can transform the resources from a mere collection of items into coherent knowledge structures that facilitating sense making on the available resources. By using the search facilities of the workbench, they are also able to search for relevant literature or data sets, which can be also uploaded on the collaboration workspace. For this, they may execute the *fullTextSearch* operation of the collaboration search service.

A more comprehensive presentation of the above in the context of the collaboration service can be found in the following publications:

- Karacapilidis, N., Tzagarakis, M., Christodoulou, S., Tsiliki, G. (2011a). Facilitating Scientific Collaboration in Data-Intensive Biomedical Settings. In: Proceedings of the 10th International Workshop on Biomedical Engineering, Kos Island, Greece, October 5-7 (to appear).
- Ammari, A., Dimitrova, V., Lau, L., Tzagarakis, M., Karacapilidis, N. (2011). Augmented Collaborative Spaces for Collective Sense Making: The Dicode Approach. In: A. Paramythis, L. Lau, S. Demetriadis, M. Tzagarakis, & S. Kleanthous (Eds.), Proceedings of the International Workshop on Adaptive Support for Team Collaboration 2011 (ASTC 2011), Girona, Spain, July 15, CEUR Workshop Proceedings, Vol. 743, pp. 3-13.
- Karacapilidis, N., Tzagarakis, M., Christodoulou, S., Tsiliki, G. (2011b). Facilitating and Augmenting Collaboration in the Biomedical Domain. International Journal of Systems Biology and Biomedical Technologies (IJSBBT) (submitted).