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

D6.3.1 – Report from the evaluation of use case #2 (first version)

Due date of deliverable:	29 February 2012
Actual submission date:	20 March 2012
Responsible Partner:	IMA
Contributing Partners:	UOL, CTI

Nature: ■ Report ☐ Prototype ☐ Demonstrator ☐ Other

Dissemination Level:

■ PU : Public
☐ PP : Restricted to other programme participants (including the Commission Services)
☐ RE : Restricted to a group specified by the consortium (including the Commission Services)
☐ CO : Confidential, only for members of the consortium (including the Commission Services)

Keyword List: evaluation, randomised clinical trials, MRI, imaging, statistical data analysis, datasets, databases, data management, collaboration, decision making

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

The Dicode Consortium

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 Politecnica 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	20-02-2012	First draft	Mark Hinton, IMA
2	29-02-2012	Second draft, evaluation results incorporated	Lee Tunnicliffe, IMA
3	05-03-2012	Sent to internal reviewers	Mark Hinton, IMA Fan Yang-Turner, UOL Joerg Kindermann, FHG Anastasia Kastania, BRF
4	14-03-2012	Internal reviewers' comments incorporated, sent to SC	Mark Hinton, IMA Lee Tunnicliffe, IMA
5	20-03-2012	Final version (approved by SC, sent to the Project Officer)	Mark Hinton, IMA Lee Tunnicliffe, IMA

Deliverable managers

- Mark Hinton & Lee Tunnicliffe, IMA

List of Contributors

- Mark Hinton, IMA
- Lee Tunnicliffe, IMA
- Dhaval Thakker, UOL
- Vania Dimitrova, UOL
- Lydia Lau, UOL
- Manolis Tzagarakis, CTI
- Nikos Karacapilidis, CTI

List of Evaluators

- Fan Yang-Turner, UOL
- Joerg Kindermann, FHG
- Anastasia Kastania, BRF

Summary

This deliverable is to be considered as a first version of the evaluation of Use Case 2 Trial of Clinical Treatment and particularly the evaluation of services developed in the Dicode platform applicable to this use case. 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 document includes an updated description of Use Case 2, as well as a description of the Dicode services applied to Use Case 2. There is a section describing future requirements to make the Dicode services suitable for inclusion in Image Analysis' commercial platform for supporting clinical trials.

Table of Contents

1 Introduction

1.1 Context

This deliverable presents the evaluation of the initial version of the developed Dicode services relative to Use Case 2 Trial of Clinical Treatment Effects. It focuses on the use of Augmentor as a tool to aid analysis and sharing of clinical research reports generated by Image Analysis software, Dynamika. Dynamika does post processing and analysis of magnetic resonance images (MRI).

The report also considers the extended services that will be needed to make the Dicode services a viable commercial component of Image Analysis roadmap of platform development.

1.2 Objectives

The objectives for this report are: i) Describe the evaluation process and the results of the evaluation of Augmentor, ii) Update Use Case 2 to reflect current thinking from the development of IMA's business model and iii) Comment on the cost-effectiveness and readiness of the Dicode services for IMA's market.

The deliverable is organized as follows: Section 2 describes Use Case 2 and section 3 looks at ethical issues associated with testing Use Case 2 in the Dicode project and beyond. Section 5 reports on the evaluation of the Augmentor. Section 6 concludes the deliverable discussing future work and the current readiness and suitability of the services.

2 Use Case 2: Trial of Clinical Treatment Effects

The second use case, trial of Clinical Treatment Effects, is selected to represent clinical trial communities that collaborate on clinical decision making issues and take into account multimedia materials. This use case was originally concerned only with trials for Rheumatoid Arthritis. This was then IMA's main business focus. Since then, the use of the IMA platform has been extended to more musculoskeletal diseases and also to oncology, cardiology and neurology. Hence the title of this section has changed to 'Use Case 2: Trial of Clinical Treatment Effects' to reflect this. The details of the use cases have also changed accordingly.

To facilitate the process of making clinical decisions in drug trials is quite challenging in terms of complexity of trial data analysis. However, this aim can be achieved by combining datasets from patient results (blood tests, physical examinations) and the different scan modalities (X-Ray, Static and Dynamic MRI scan images). With this use case, Dicode aims to deliver pertinent information to communities of researchers, doctors and patients by revealing trends within a trial and provide consistent recording. The use case in this section is presented in a very general form indicating that it may support different clinical trials related to musculoskeletal diseases.

2.1 Current work practice scenario

2.1.1 Context

A trial for clinical treatment effects is conducted in an academic establishment on behalf of a pharmaceutical company. The trial is evaluating the effectiveness of treatment, by analysing the condition over the period of treatment.

A range of patients with different degrees of disease and healthy controls will have scans at regular periods (e.g. monthly) to monitor their condition. MRI scans are used to show how effective the treatment is, and can show changes in soft tissue inflammation quickly, E.g. results to debilitating RA can be seen within 5 days. Scans consist of high resolution static images, and a series of lower resolution images taken during 5-10 minutes using a contrast agent that is absorbed by different tissues during the scan period. The Dynamika[1] software is used to highlight the contrast absorbtion patterns using the changes in signal intensity as read from the scan series. A radiologist and a clinician will be required to assess the combined data to make a judgement on the condition. The trial results will determine whether the treatment is permitted for treatment of the condition in general practice within Europe and globally.

2.1.2 Goal

The goal for this use case is to analyse the effectiveness of treatment effects during a clinical trial.

2.1.3 Actors

Joan, *patient*, is under treatment of a RA trial.
Alice, *radiographer*, takes scans of patients.
Chris, *radiologist*, conducts scan/image analysis through Dynamika.
David, *lead clinician*, is responsible for the medical evaluation of the treatment for the patients. He makes decisions about the effectiveness of the treatment based on a report by the radiologist and other information related to the trial.
Frank, representative of other *clinicians*, has experience in disease diagnosis and treatment. He might help David to analyse the image or the case if needed.

2.1.4 Steps

The following use case steps are visually illustrated in use case diagram (Figure 1).

1. MRI Scan
Radiographer Alice takes the scans of a patients checks the anatomy position and the quality of the images captured.

2. Access Image
Radiologist, Chris, is notified by Radiology Information System (RIS) that a new case can be seen in the PACS[2]. He accesses the case images through PACS.

3. Radiologist analyses images in Dynamika
   3.1 Upload images to Dynamika

[1] http://www.imageanalysis.org.uk
[2] http://en.wikipedia.org/wiki/Picture_archiving_and_communication_system

Chris uploads scan images into Dynamika. Dynamika performs motion correction and calculates and displays overlaid coloured maps of the contrast absorbtion patterns in the tissues during the dynamic scans.

3.2. Identify ROIs

Using Dynamika, Chris identifies Regions of Interest (ROIs) which will be used together with the maps in generating statistics about the ROIs. When doing so, Chris also made comments for his analysis.

3.3. Generate Report by Radiologist

Chris stores the maps, ROIs and associated statistics in a report associated with the specific scan image. Chris then sends the electronic report (HTML file) to the responsible clinician, David.

4. Clinician analyses images in Dynamika and reports from Radiologist

David receives the report from Chris. He uploads the images to Dynamika and repeats the steps performed earlier by the radiologist: identifying ROIs and generating maps and statistics. Finally he stores the maps, ROIs and associated statistics in a report, (which might be different from the one the radiologist produced earlier), associated with the specific scan image.

5. Synthesized Analysis

David then consults previous relevant statistics along with treatment information from previous visits of the patient, Joan. The information includes Joan's journal and historical reference data via the Patient Trial Details. He further examines the trial protocol used at the time the scans were taken through the Trial Details, in order to reach a conclusive decision on the effectiveness of the treatment for Joan. He then produces an Analysis Report containing these findings. If the images and clinical findings do not correspond to the radiologist conclusion on the specific patient then this patient's case must be discussed further on a conference.

6. Further Analysis

In this instance, if David has some difficulties in concluding on the effectiveness of the treatment on the specific patient. He might want to consult his colleagues, the radiologist and another clinician, Frank who are located elsewhere, to exchange ideas and collaboratively reach a decision. At present, there is no easy way to share data in real time for joint discussion.

7. Follow-up Analysis

Follow up analysis applies to analysis for the whole trial including multiple patients or even multiple trial sites. The output reports for each patient are also stored in CSV and imported into Excel to aggregate the results. The aggregated data then provides visibility over time and across multiple patients to reveal general trends in conditions. As an overview, this shows the success of particular treatments.

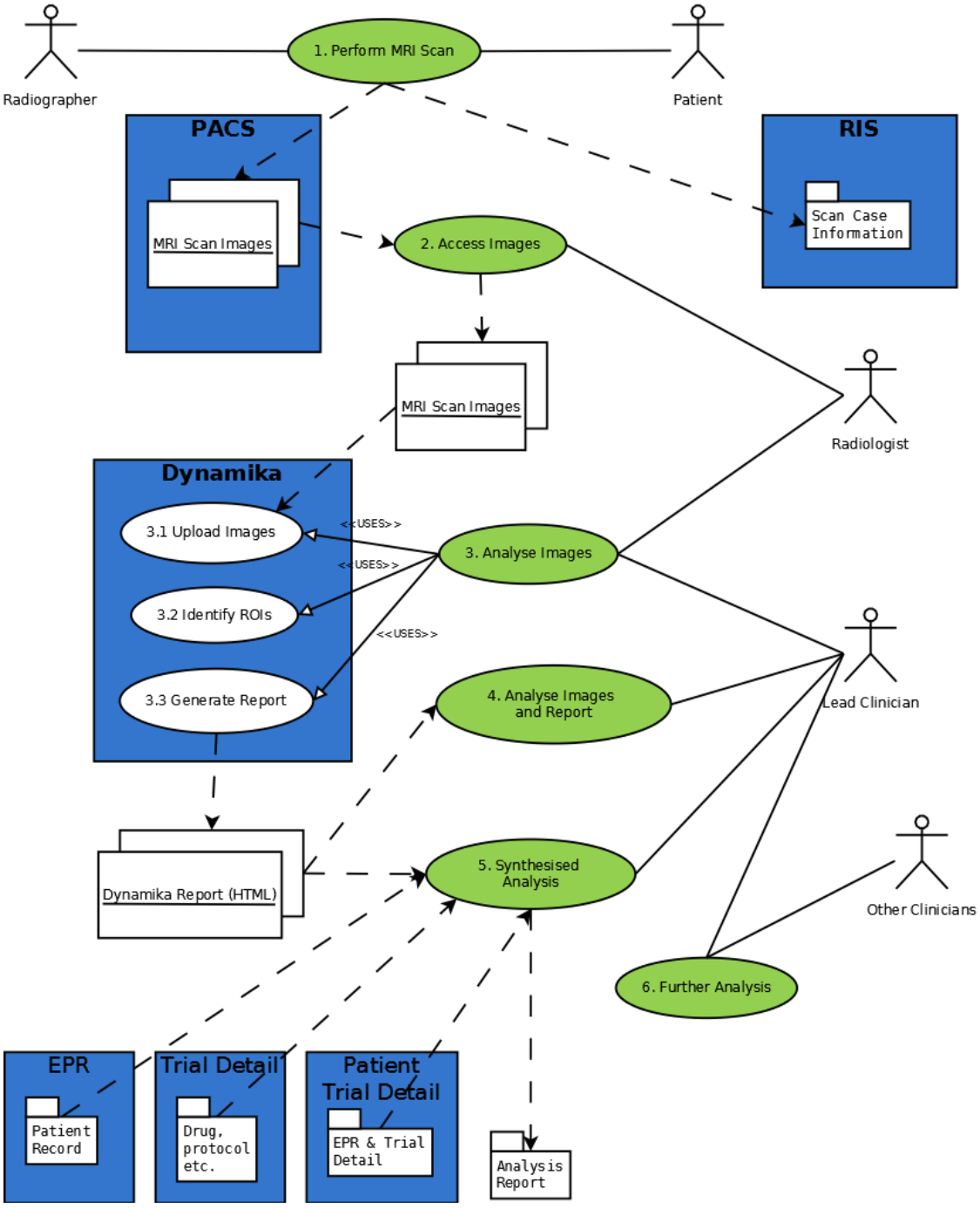

Figure 1: Trial of treatment effects (current work practice)

***Notes*:**

- Only the responsible clinician, in this case, David has full access on the patient data and others can only see the anonymous data or do not have any access to the data at all.
- The radiographers can see all images but they cannot mark that the case have been read - it is only the radiologist who can do that. The clinician can see the images using a web portal that shows the DICOM images in jpeg with a low image quality and slow update, so large datasets are basically impossible to scroll through.

- It requires a PACS environment to assess the image data in a proper way. Several different doctors/user can see the same case, but the changes made to the images by each user cannot be seen and it is only the Radiologist that can see all the changes.

2.2 Analysis of current work practice

2.2.1 Users and communities

From a clinical trial point of view, there is a working team in which everyone takes their role towards the final analysis of the trial result. This working team involves:

Radiographer
The Radiographer takes the scan with the patient, checks the anatomy position, checks the quality of the image and checks that the features of the condition have been captured. The Radiographer is required to adhere to the designated protocol that in this case specifies the scan settings, the timing between scans in the series and the timing and dosage of contrast agent.

Radiologist
The Radiologist uses the output from the radiographer, the static and dynamic scans and highlights the features of the scan and suggests a diagnosis.

Clinician
The Clinician uses the results from the Radiologist and Radiographer to diagnose the condition and analyse the effectiveness of the treatment. The Clinician may need to collaborate on decisions with the radiologist or other clinicians. The Clinician can see the previous statistics and perform additional processing as required. The analysis may consume MRI scan data and treatment information from multiple visits of the same patient, and additionally comparison against historical reference data and other patients' scans and treatment history.

Pharmaceutical company
The Pharmaceutical Company reviews the outputs from the trial. Although, it cannot inform decisions if impartial results are obtained. However, it may recommend changes in treatment during development of drugs.

Patient
The Patient is the most important part of the process, because drugs trials are aimed at improving the quality of life of the patient. E.g. In terms of RA, patient movement can be severely restricted with debilitating consequences on normal day-to-day activity. The patient is involved in the trial as being treated to improve their condition. Diagnosis of the patient includes physical examination, blood tests, various scanning modalities (e.g. X-Ray, MRI, and ultrasound) and patient feedback on the condition, which can be accompanied by a journal.
This closely working team can be considered as a well-formed community.

2.2.2 Data intensiveness

Static Image
Between 1 and 3 static images (~80-160KB) may be taken by radiographer for reference and to establish correct settings on the MRI scanner before commencing the full scan.

Dynamic Scan Images
A series (~12-35) of dynamic scan images (~120KB – 500KB per image) will be produced during which a contrast agent is administered. This is for a single slice dataset and depending on the anatomy there may be between 3 and 120 slices. The scans generate a minimum 6MB of image data per patient, per visit.

A Typical Example
A clinical trial could easily contain 400 patients and each patient could be scanned somewhere around 15 times during the course of the trial. If each scan contains sets of images with 25 time points and 5 slices at a resolution of 512x512 this would translate to 125(25x5) images per scan 1875 images per patient and 750,000 images during the course of the trial. If each of the images is 512kb (512x512) probably closer to 600kb with the DICOM header information we're looking at 430GB (750,000 x 600kb) of data.

As of 2012 there are 1978 open clinical trials worldwide**, generating 830TB** (1978x430GB) of data.

Dynamika Report
Dynamika is used to assist the radiologist in viewing the scan by overlaying coloured maps of the tissues shown in the dynamic scans. ROIs can be marked on the scans and the maps are used to generate statistics about the ROIs. The Radiologist will store the maps, ROIs and associated statistics in a report. This Dynamika output report is a HTML file (size: 3MB) with all the sample images and statistics of the analysis. The report generated by the Radiologist is submitted to the Clinician responsible for the trial. The report is sent in an electronic form within hospitals and to referring clinicians who have electronic communication with the hospital. Otherwise it is printed and sent by mail.

For trials in Phase 3[3], there might be:

- 500 patients
- 10 scan studies per patient
- scan series per study
- 2 reports per scan series

Giving a total of **30,000 reports**, averaging approximately 87,8GB of textual data, generated for the trial over a 2-3 year period. The scans themselves will have been carried out at multiple locations (typically 3 but could be more) and with different manufacturer's scanners.

This is far too much information for any radiologist or clinician to mine and understand without suitable tools. At best only high level trends will be identified.

PACS
PACS handles images storage from various medical imaging instruments, such as MRI scanner. Most PACS have not incorporated post-processing algorithms unless the PACS has a specific plug-in to do this. Usually the post-processing is done on the scanner software and then sent to PACS. The universal format for PACS image storage and transfer is DICOM[4] (Digital Imaging and Communications in Medicine). Non-image data, such as scanned

[3] http://www.nlm.nih.gov/services/ctphases.html
[4] http://en.wikipedia.org/wiki/DICOM

documents, may be incorporated using consumer industry standard formats like PDF, once encapsulated in DICOM.

EPR (Electronic Patient Record)
The EPR includes information about the patient.

Trial Details
Drug details, treatment process and protocols used each time when a scan is taken.

Patient Trial Details
It is important in helping clinician make decision on the effectiveness of the treatment as this connects information about patient (EPR) and trial. It includes previous treatments, previous test results, medical history and basic patient information.

Notes:

- Search through the repositories (PACS, Patient Trial Details, EPR) can be done by patient's name, patient's ID or patient's date of birth. Search can also be done using other keywords and stored characteristics if available for indexing.

2.2.3 Collaboration and decision making activities

Table 1 evaluates if the activities in current work practice involve collaboration or decision making and what tools are used now.

Activities		Collabo-ration	Decision making	Tools	Current practice
MRI scan		No	Selection (Screen, evaluation and choice)	MRI Scanner, Trial Details	Radiographer needs to check the quality of the image and checks that the features of the condition have been captured
Access image		No	No	RIS, PACS	notified by RIS imaged downloaded from PACS
Analyse images in Dynamika	Upload images to Dynamika	No	No	Dynamika	Repeat work by Radiologist and Clinician
	Identify ROIs	No	Selection (Screen, evaluation and choice)	Dynamika	The selection of ROIs very much depends on the knowledge and experience of a radiologist or a clinician
	Generate report	No	Problem Identifi-cation	Dynamika	The report includes comments from radiologist or clinician
Synthesized analysis		No	Selection (Screen, evaluation and choice)	All data sources	With all available information (either in system or on paper), clinician makes final decision.
Further analysis		Yes	Selection	No	No tools to support data

		(Screen, evaluation and choice)		sharing with other clinicians or discussions
Follow-up analysis	Yes	All	No	Further investigation needed for this analysis

Table 1: Activities of rheumatoid arthritis treatment trial (current work practice)

2.3 Users' vision

Users of this use case have given following stories about their vision of Dicode:

"The analysis of image or scans and statistics comparison described at current work practice is done manually at the moment by only one person; this makes the process of decision making difficult and prone to errors."

"One Dynamika report includes rich information about a patient in a trial, not only the images and the DICOM information with the images, but also the comments from radiologist, the ROIs identified by the radiologist and various statistics. This provides a data source for knowledge discovering, for example, the knowledge of identifying ROIs. If all the reports of a trial were combined, the knowledge discovered could be a trend of the trial or a comparison among patients."

"Supporting the involvement of other experts in the process will allow a healthy collaboration between them and will reduce the risk of possible mistakes as opposed to when only one person is manually comparing that amount of data."

"Although PACS, EPR and Patient Trial Details can all be accessed simultaneously, at the moment there is not a collaborative platform that can be used so they can effectively share all the data needed and allow efficient and effective decision making."

"A problem that exists at the moment is that not all people involved can have access to the full patient's data due to privacy issues. So, the lead Clinician might want to control the access rights that people he collaborates with have to the patient's data. At the moment this is not possible."

"Clinician A might request help from another clinician and direct him or her (allowing access) to the patient's data in order for clinician B to access the data. Clinician B may comment on the results and analysis from clinician A, or clinician B may perform his own analysis. The results can be combined in the overall trial results. They might also want to collaborate in real-time in order to make annotations on ROIs and (or) change the text comments in a report together. These features are not available to them in the current practice."

"There is obviously a problem with protecting the patient data; this has to be discussed or even changed by law before such collaboration could take place. The environment has to be secure and shielded from outsiders that are not granted access. Across country borders, this is even more troublesome as each country has its own set of rules; within a country, this setup could work more easily."

"In general, the idea of a central database that could collect all available information is a good idea; this has been implemented in some areas on a regional basis in Denmark within pathology, breast screening, patient reporting but it is not yet a nationwide application, nor do any of these systems integrate yet. In rheumatology in Denmark we have a nationwide patient database that is setup to handle patient data and treatment called DANBIO (Hetland, 2005). The problem is: how to make every clinician report to this database as he has to login and type in the data, which is time consuming and so many clinicians do not use it on a regular basis."

"In radiology we can within the regions (5 in total) see each other's images on the PACS or we can send images to each other and then go through the images, but we cannot always read each other's descriptions of the images because we use different vendors. There is a national bid at the moment to buy the same PACS system which will hopefully solve most of these issues. So in general we lack a proper platform where collaboration 'on the fly' is feasible and where all needed data are available."

Users' vision in this use case is illustrated in Figure 2.

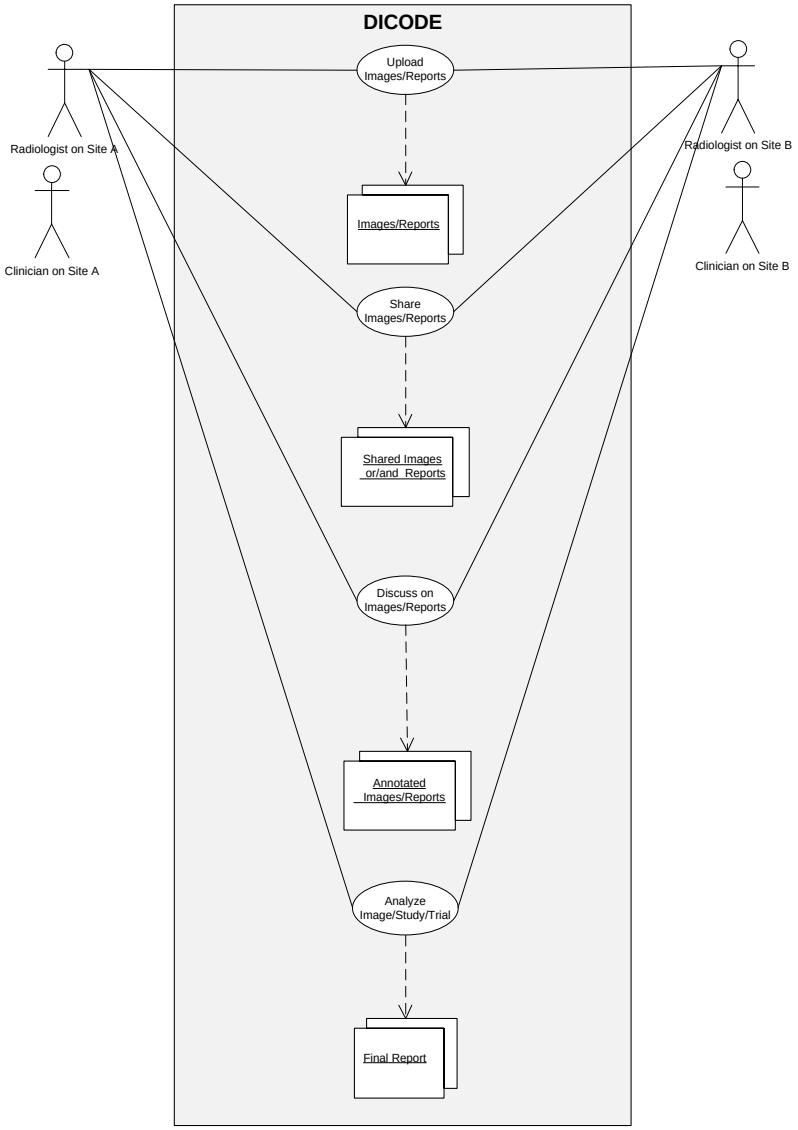

Figure 2: Trial of rheumatoid arthritis Treatment (users' vision)

2.4 Analysis of users' vision

2.4.1 Users and communities

In this scenario, it is expected that collaboration should be encouraged and supported within the team between specialists. This means the team can be expanded from having one radiologist to two or more radiologists; from one or two clinicians to three or more clinicians. If the trial is performed across multiple sites, the team will expand to have all the specialists in different sites.

2.4.2 Data processing

In terms of data processing, expected functions are interpreted from users' statements (Table 2).

Function category	Expected functions	Users' statements
Image analysis	1. Discover ROIs experience Finding the patterns of how radiologists identify ROIs from reports generated from Dynamika. The discovered knowledge can be used to train new radiologist or generate suggestions for radiologist while analysing the images. 2. Discover pattern of a trial Finding the trend of a trial by analysing all the reports. This discovery can help decision makers to stop or adjust the trial in time, consequently save the trial cost. 3 Discover the patients' response pattern. Finding the relationship between patient and his/her response to the treatment. This can be used to explain why patients have same or different response.	'One Dynamika report includes rich information about a patient in a trial, not only the images and the DICOM information with the images, but also the comments from radiologist, the ROIs identified by the radiologist etc.' 'This provides a data source for knowledge discovering, for example, the knowledge of identifying ROIs. ' 'If combine all the reports of a trial, the knowledge discovered can be a trend of the trial or a comparison among patients.'
Large image analysis	4. Finding the most relevant images and most relevant part of the image to scroll through.	'Large datasets are basically impossible to scroll through.'

Table 2: Data processing functions of treatment effects trial (users' vision)

2.4.3 Collaboration and decision making activities

In terms of collaboration and decision making, expected functions are interpreted from users' statements (Table 3).

Activities	Expected functions to support collaboration	Expected functions to support decision making	Users' statements
Image analysis	1. Image analysis can be conducted among radiologists and clinicians, which requires a shared place for image sharing. 2. Users can make comments, annotation and can see others' comments or annotations.	1. Radiologists can get suggestions from the systems based on the knowledge discovered from previous analysis.	'Clinician A might request help from another clinician and directs him or her (allowing access) to the patient's data in order for clinician B to access the data asynchronously.' 'Clinician B may comment on the results and analysis from clinician A, or clinician B may perform his own analysis, where the results can be combined in the overall trial results.' 'They might also want to collaborate in real-time in order to make annotations on ROIs and (or) change the text comments in a report. These features are not available to them in the current practice.'
Report analysis	3. Reports can be created by radiologists and clinicians collaboratively. 4. Users can make comments and can see others' comments in the report.		

Table 3: Activities support functions of treatment effects trial (users' vision)

2.5 Dicode's services

Dicode's team have also proposed services at the creativity workshop for this use case. The summarized services (listed in Table 4) work as an outline for technical partners defining detailed functional specifications.

Services category		Description
Data mining support	Data acquisition	1. DICOM input data contains meta-information usually hidden from the analyst. Consistent acquisition is critical to the accuracy of results, variations in acquisition can be revealed through mining DICOM header. Example fields are time intervals between scans, scanner settings and patient position.
	Data analysis	1. Analyse comments of experts
		2. Mining patient discussion in web forums about diseases, drugs and side effects for relevant information related to the trial. This will benefit for follow-up analysis and monitoring of side effects after drug/treatment is introduced to the market. (ACR[5], ESR[6] forms, name of drugs/treatment/side effects etc.)
		3. Determine trends of a trial, visualize trends of a trial, the result trends against patient population. Collation of results can happen in real-time as results are processed and stored. 4. Determine image analysis process, identify good/bad analysis to help or train others.
		5. Pattern mining using logs, reports: same patterns in selecting ROIs, same ROIs having similar comments.
Collaboration		1. Clinician/radiologist collaborative discussion towards data sources to be used, the data mining algorithm to be invoked as well as the interpretation of data mining outcome. This would build collaborative reports using extracted results from reports and relating them using a discussion platform.
		2. Semantic annotation of comments. 3. Grouping discussions based on detected topics. 4. Use output of data mining in collaborative discussion on trial trends to analyse accuracy of trial, or for use in training.
Decision making		1. Decision making support towards selecting specific course of action.
		2. Guide users with best practice, highlight if anything important has been missed, which can be used in training or to access accuracy of the trial.
Integra-tion	Interface	1. Widget-like approach.
	Ontology/ data	1. Annotation of images/reports using ontology.
		2. Ontologies developed from HL7[7]/DICOM/SNOMED[8].
	System/ services	1. Workflow mapping.
		2. Break out Dynamika to services for Dicode, such as motion correction, ROIs status, image status.

Table 4: Proposed services for treatment effects trial

[5] http://www.acr.org/
[6] http://www.eurorad.org/
[7] http://www.hl7.org/
[8] http://www.ihtsdo.org/snomed-ct/

2.6 Future work practice

Dicode aims to improve time of drugs to market by revealing trends within a trial and provide consistent recording of data to the medical team. This future practice is illustrated in Figure 3; related requirements are summarized in Table 5.

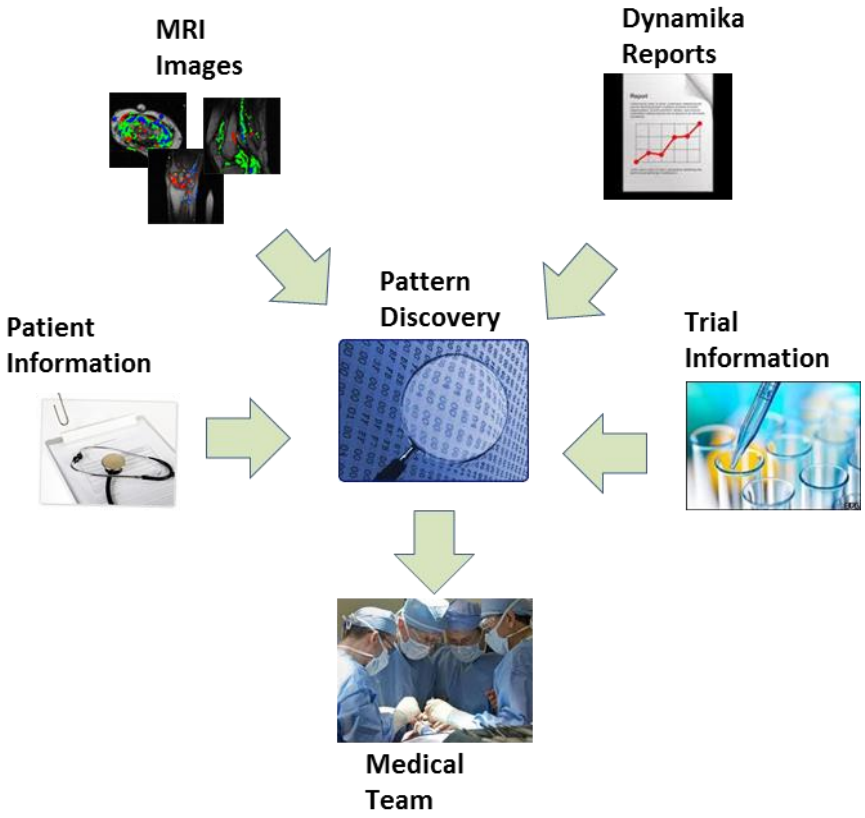

Figure 3: Future work practice of treatment effects trial

This future work practice aims to provide a general understanding of how to discover, access and share image data and drug treatments across health networks, which includes:

- *Consistent provenance of trials:* The process history of the trial is essential when gaining medical approval for a drug or treatment process. This will also be valuable knowledge in improving accuracy and reducing time taken in future trials.
- *Large dataset decision making:* The current datasets are increasing in size and analysis time increases with each time a patient visits. More patient visits improve the accuracy of treatment monitoring and therefore improve the proof of the efficacy or side effects of drugs. Therefore new methods are required to reduce the time needed during collaborative decision making.
- *Monitor overall trial process:* Current protocols and trial practice are carried out by each party. Monitoring the overall process will produce provenance records for the trial and allow workflow governance controls to ensure medical protocols are adhered to.

Front-end requirements	Back-end requirements	Current practice improved	Users' vision addressed
History data collection	Images & reports uploaded to a storage for analysis		Knowledge discovery
Pattern discovery	Analyse comments of experts	Enable knowledge transfer of radiologists or clinicians	
	Mining patient discussion in web forums about diseases, drugs and side effects for relevant information related to the trial. This will benefit for follow-up analysis and monitoring of side effects after drug/treatment is introduced to the market. (ACR, ESR forms, name of drugs/treatment/side effects etc.)	Dicode's innovative idea: bring social media content for trial reference	
	Determine trends of a trial, visualize trends of a trial, the result trends against patient population		Detect trend of a trial
	Pattern mining using logs, reports: same patterns in selecting ROIs, same ROIs having similar comments		Detect patterns of image analysis
Promote collaboration	Clinician/radiologist logbook of contemplation, collaborative discussion towards data sources to be used, the data mining algorithm to be invoked as well as the interpretation of data mining outcome	Support of collaboration among medical team	
	Semantic annotation of comments		
	Grouping discussions based on detected topics		
Decision making support	Decision making support towards selecting specific course of action	Support decision making during the trial	
	Guide users with best practice, highlight if anything important missed		

Table 5: Requirements list of treatment effects trial

3 Ethical Issues

There are several stages in the testing of Dicode services for this use case. In each one IMA has adequate datasets available for testing that can be used by the relevant parties without ethical problems.

3.1 In development testing

IMA has a fully anonymised set of scan studies that are used for testing. This set contains approximately 2.3 GB of data, 12500 images, and some 50 scan studies. This was the raw dataset used in this assessment of Augmentor. A sample of this set has been sent to Fraunhofer for initial analysis, to see if there are patterns and sub groupings that can be found. This dataset has been fully anonymised and has no references that can identify any individual. IMA use this dataset in their own QA stages of development.

3.2 Test Data for integration

The same data set can be used for integration testing with the IMA clinical platform. If required IMA holds a larger dataset of over 19000 scan studies, holding approximately 234,000 images. This dataset is not fully anonymised but IMA has software tools to perform this anonymisation.

3.3 In trial data

This is testing of the Dicode services when they are mature enough to be incorporated in the IMA platform. This will take place with an IMA customer on their live data. IMA is currently engaged with multiple clinical trials in Europe and North America. IMA is the imaging company for these trials. IMA is working with the pharmaceutical company, the research institute and the clinical research organisation. These are multi-site trials using MRI as a biomarker and the IMA platform to analyse and report on scan series. That is, it is a real instance of Use Case 2. This will enable testing of Dicode services in a proper commercial setting. So long as the services are sufficiently mature and the consortium can concentrate on developing them.

4 Augmentor & Semantic Services Evaluation[9]

The main aims of this formative evaluation are:

1. To identify the benefits and limitations of the existing services and underlying framework (technology, techniques, methodology)
2. To identify the future R & D direction in the areas of Dicode semantic services.

The evaluation is carried out in two areas: **a) technical evaluation:** with the focus on evaluating the semantic annotation service which is core to the performance of the other semantic services and Augmentor and **b) user study:** where the focus is on evaluating similarity & relatedness-based browsing functionality and its usability.

[9] Augmentor and Semantic services are described in the Dicode Deliverable 5.3.1

The evaluation exercise was designed to enable the participants to perform various tasks using Augmentor and feedback was collected using questionnaires. The exercise was designed to address the aforementioned two areas and the original exercise document is available in the appendix A. The task 4 from the exercise covers the technical evaluation and tasks 1, 2 & 3 covers the user study.

The evaluation was carried out with 5 Dynamika software developers from the Image Medical Analysis (IMA). The tasks in the evaluation exercise involved 16 Dynamika reports. These reports were prepared by Dynamika users as part of a trial.

4.1 Technical Evaluation

4.1.1 Semantic Annotation service Evaluation

The semantic annotation service is a generic service designed to link content with the concepts from the ontological knowledge bases in order to fully benefit from the reasoning capabilities of semantic technologies. With respect to Dicode, semantic annotation of content using Dicode ONtology (DON, described in deliverable 5.2) makes it possible to enrich unstructured or semi-structured data with domain-specific context. Example content includes comments from a drug-trial report. This service is described in the deliverable 5.3.1 (Thakker, D et al. 2011).

The evaluation of this service was carried out using the task 4 (see Appendix A). The aim of this task was to measure the precision and recall of the service and receive and analyse feedback to form the basis for extending the functionality of the service.

4.1.2 Datasets , Participants & Evaluation Method

The evaluation was carried out with 16 Dynamika reports and the 167 terms detected by the semantic annotation service. The five participants were presented with the terms identified from the reports and were asked to suggest if the service identified correct term by indicating *Yes*, *No* or *Don't know*. The participants were also asked to provide any terms that the service missed.

4.1.3 Results

Precision:
The overall precision is 73% (when considering only *Yes* and *No,* see Figure 4) and 69% (when considering all *three* answers, see Figure 5).

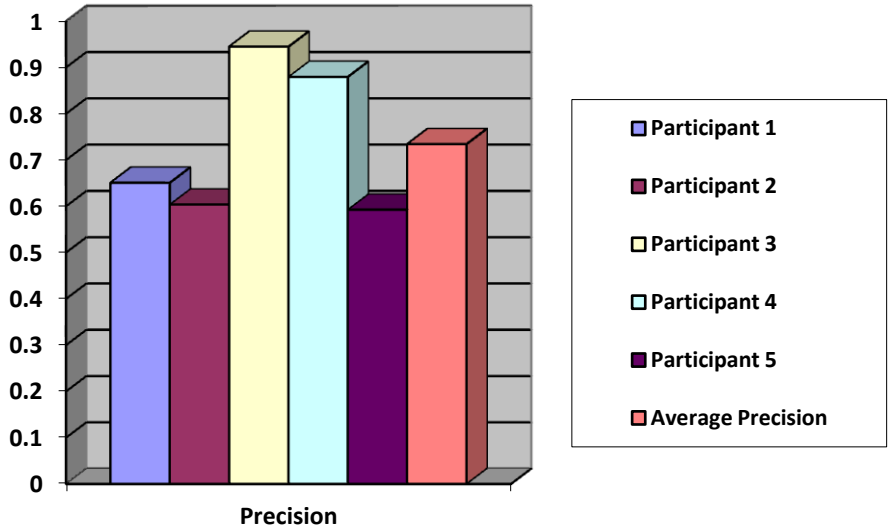

Figure 4: Precision figures for the Semantic Annotation service Precision (considering YES/NO results)

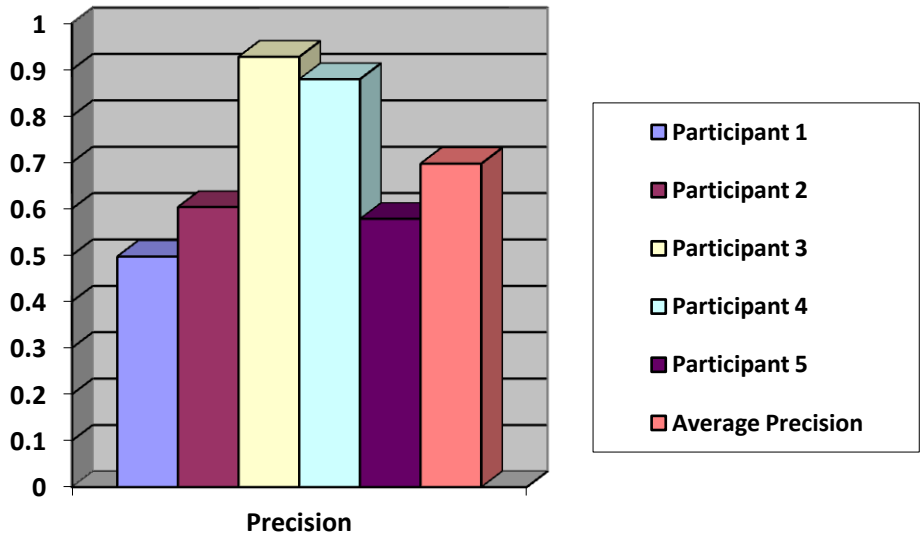

Figure 5: Precision figures for the Semantic Annotation service (considering YES/NO/Don't Know results)

Recall:
The overall recall is 85.7%. See Figure 6.

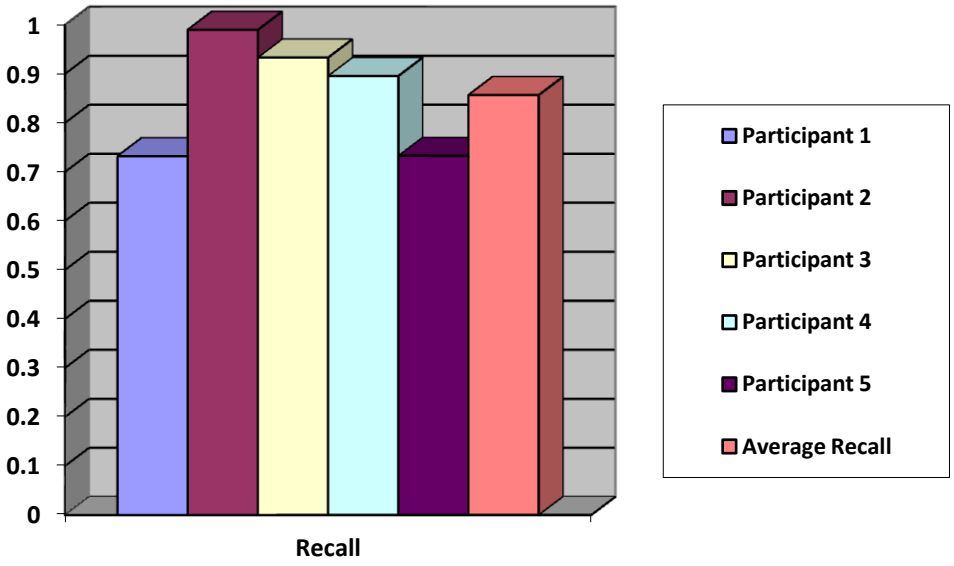

Figure 6: Recall figures for the Semantic Annotation service

The F-Measure using the average precision and recall values is 79%.

4.1.4 Overall agreement between participants:

There are many possible metrics for reporting overall agreement between annotators, such as Cohen's Kappa [Cohen, 1960], bias [Eugenio et al. 2004]. Kappa is considered best metric for IAA when all the annotators have identical exhaustive sets of questions on which they might agree or disagree. In other words, it is a classification task such as task 4 of this evaluation exercise where the terms are presented to the participants and they are required to agree (*yes*) or disagree (*No*).

Kappa is defined as the observed agreements P_o minus the agreement expected by chance P_e and is normalized as a number between -1 and 1[Cohen, 1960].

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

$\kappa = 1$ means perfect agreements, $\kappa = 0$ means the agreement is equal to chance, $\kappa = -1$ means 'perfect' disagreement.

For the participants (or raters) in our exercise, Table 6 outlines pair-wise agreement (Keppa).

	Rater 1		**Rater 2**		**Rater 3**		**Rater 4**		**Rater 5**
Rater 2	0.413299	Rater 1	0.413299	Rater 1	0.056544	Rater 1	0.212383	Rater 1	0.254351
Rater 3	0.056544	Rater 3	0.126343	Rater 2	0.126343	Rater 2	0.175661	Rater 2	0.307704
Rater 4	0.212383	Rater 4	0.175661	Rater 4	0.355274	Rater 3	0.355274	Rater 3	0.085813
Rater 5	0.254351	Rater 5	0.307704	Rater 5	0.085813	Rater 5	0.085813	Rater 4	0.103437
Oveall	0.234144		0.255752		0.155993		0.207283		0.187826

Table 6: Pairwise agreements between participants

The average agreement is 0.2082 which is considered fair agreement (Altman, 1991). As the agreement between raters was not substantial, we utilise alternative method to analyse.
For alternative, we have consider the terms detected by the service and classified minimum acceptance rate (i.e. when certain number of raters said *yes*) in two classes: **higher acceptance rate**, where the term was accepted by at least 3 raters) and **lower level of acceptance,** where the terms were accepted by less than three raters. There were at least 73% (122 out of 167) terms were accepted by at least 3 raters hence received higher level of acceptance. In addition, 32% terms were accepted by all 5 raters. The lower level of acceptance was for 26% of terms and there were 6 terms (3.59%) nobody accepted as important term.

Analysing the source of the most correct and most incorrect terms further (see Figure 7 and Figure 8), RadLex was the source of the most incorrect terms (67%). The most correct terms were from MeSH.

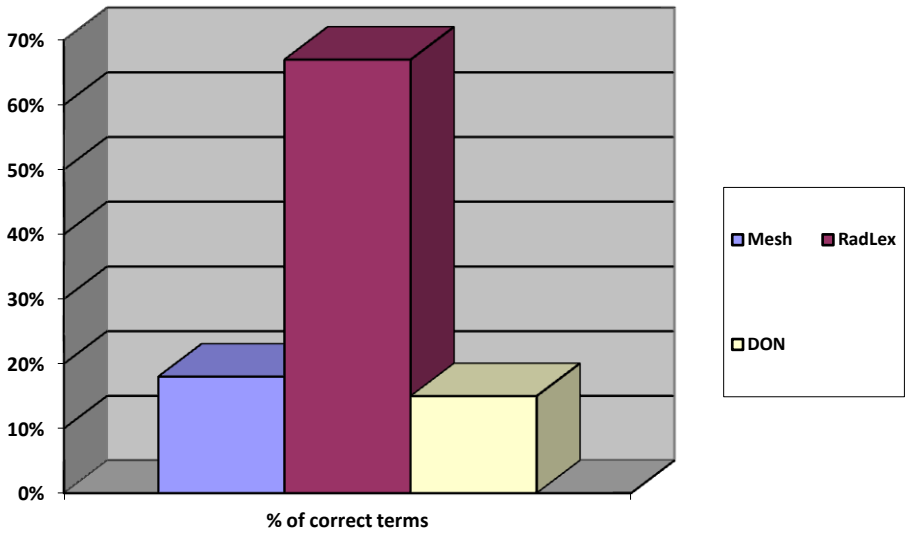

Figure 7: Source of Correct Terms

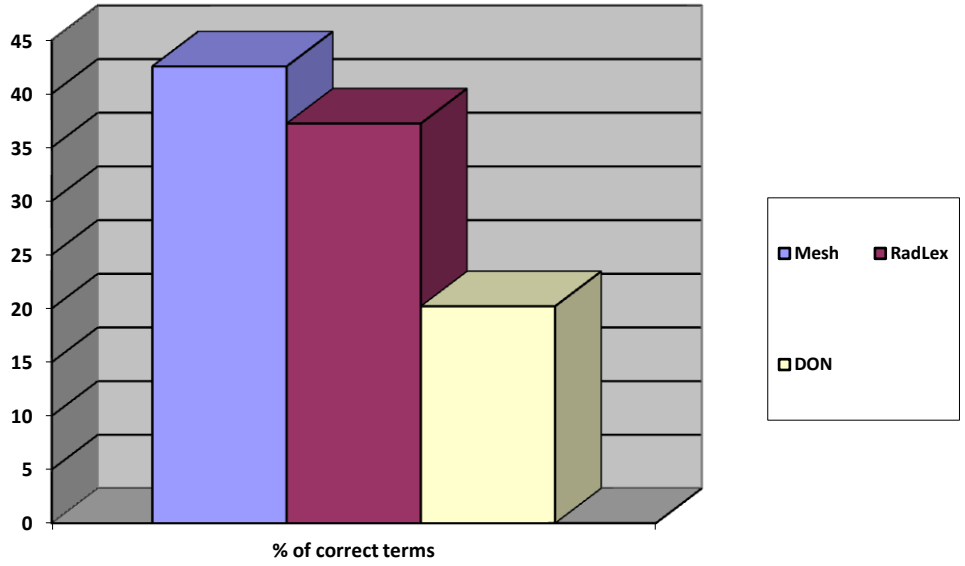

Figure 8: Source of Incorrect Terms

4.2 Analysis of the results

The detailed analysis of the results and feedback from the participants suggest the following areas for consideration while improving the service further:

Preference for specific terms and concepts
There were cases when the participants preferred specific concepts instead of generic concepts. For example, the ontologies offer concepts such as ROI, Wrist, Time, Series and Inflammation, Sub-Patella however participants preferred to have specific terms (e.g. by combining concepts/terms) such as ROI Wrist, instead of ROI, Time Series instead of Time and Series, Anti-inflammatory medicine, Inflammation in Sub-patella.

Considering the quantifiers
In some cases picking the quantifiers was considered important. For example, Little Noise, Small Movement, Very Noisy, Large area of inflammation, Too Much motion, Lot of noise, Erroneous enhancement were suggested when the service respectively detected Noise, Movement, Noisy, Inflammation, Motion, Noise and Enhancement.

Considering the Negation
The precision and recall was affected in the cases where the negation of actual concept was mentioned in the text. For example "Severe" was considered incorrect when the actual surface form was "Not severe".

Controlling concepts from external datasets to use for Information Extraction:
The service relies on the external datasets, such as RadLex, to provide necessary concepts to for information extraction process. As shown in the Figure 7, the precision is affected by certain terms coming from these datasets, for example RadLex is the source for 67% of incorrect terms. One of the reasons for this is that the terms are too generic may not be

considered relevant to the domain in question. Examples of such terms are: "Purpose", "Horizontal", "Measurement", "Patient", "marked". In addition, the datasets also contain abstract/vague terms such as "both", "rest", "work", "acquired", "function", "severe", "small", "large", "probably", "marked".

Incorrect modelling in the Dicode Ontology (DON)
The concepts that were classified under the lower level of acceptance and were originated from DON are listed in the Appendix B. This indirectly evaluates DON and these incorrect concepts shall be marked for further inspection during the development of the next version of DON.

4.3 User Study

4.3.1 Semantic Similarity Service Evaluation

The evaluation of this service was carried out using the task 2 (Appendix A). The aim of this task was to measure the precision and recall of the similarity service and receive feedback to form the basis for extending this service.

Datasets, Participants & Evaluation Method
The evaluation was carried out with 16 reports. The participants were offered a list of reports that semantic similarity service found to be similar to one of the reports. They were asked to suggest whether they agree if they are similar or not and suggest any reports that were similar but the service missed it. They were also asked to comment on the ranking of the reports, where the service ranked the most similar first and the least similar the last. Table 7 presents the results from this exercise. The participants were presented with five reports (report numbers: 6,3,4,8 & 11) similar reports to report number 5 and were asked to suggest if they agree (*yes*, *no*, *not sure*).

Participants	Report 6 (86%)	Report 3 (55%)	Report 4 (51%)	Report 8 (36%)	Report 11 (26%)
Participant 1	1	1	1	3	3
Participant 2	1	1	2	2	2
Participant 3	1	1	1	1	3
Participant 4	1	1	1	1	3
Participant 5	1	1	1	3	2

Table 7: Similarity results and Participants feedback (Yes=1, No = 2, Not sure = 3)

4.3.2 Results

Precision
The overall precision (see Figure 9) for the similarity service is 83% with the least precision (40%) for participant 2 and the most (100%) for the participant 1, 3 & 4.

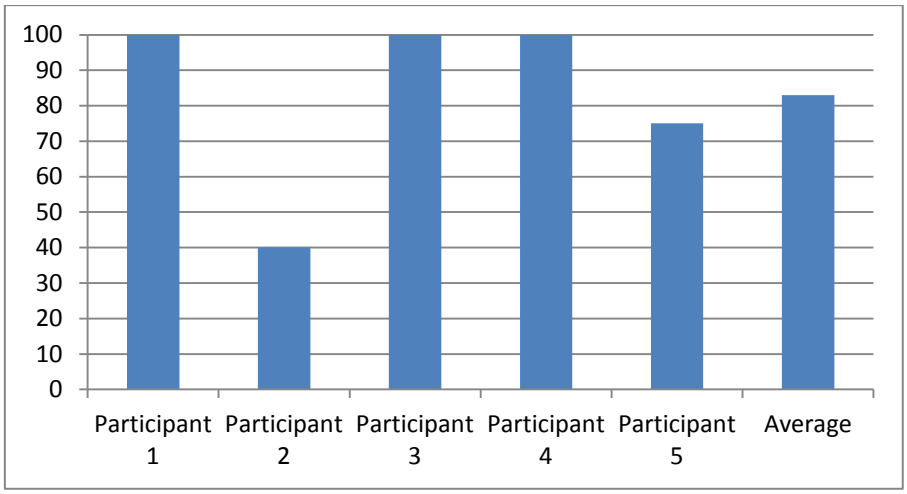

Figure 9: Similarity Precision

Recall

The overall recall (see Figure 10) for the similarity service is 93% with the recall precision (66%) for the participant 4 and the most (100%) for the participant 1, 2, 3 & 5.

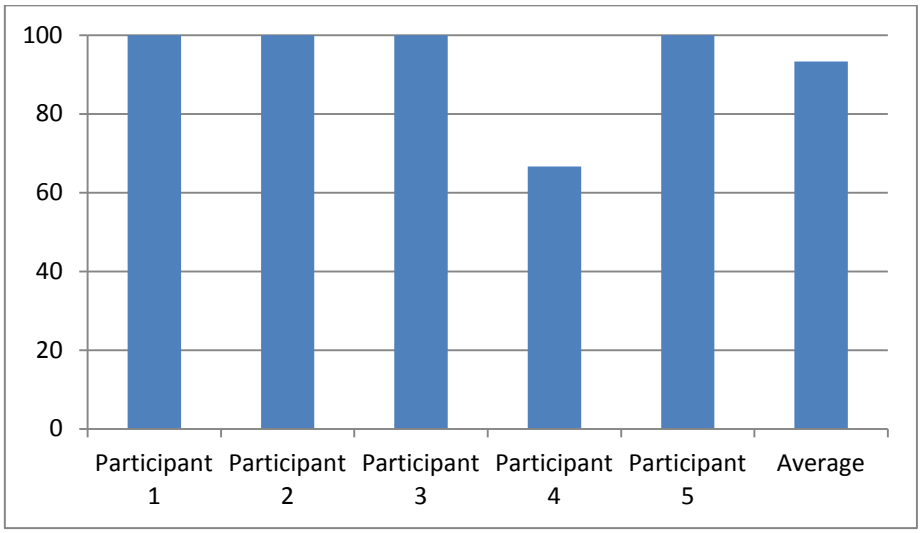

Figure 10: Similarity recall

Precision (Ranking)

The overall precision (see Figure 11) for the similarity service in terms of ranking is 92% with the least precision (60%) for participant 4 and the most (100%) for the participant 1, 2, 3 & 5.

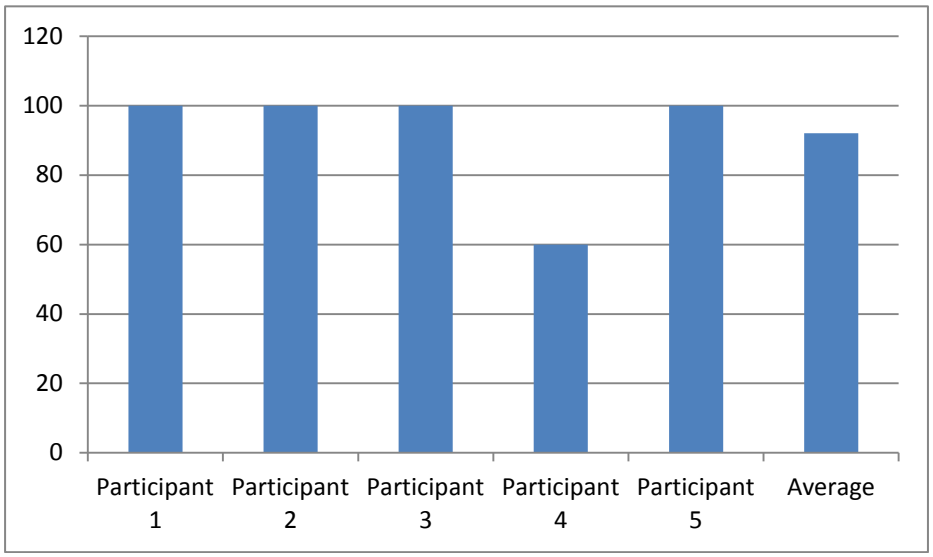

Figure 11: Similarity Ranking Precision

Feedback on the usefulness of the service

In addition to the correctness of the service, feedback was also collected on the perceived usefulness of the service (Figure 12). In terms of functionality, the service was received positively by the participants as three of the five participants found it potentially useful (the remaining two were not sure either way). One participant commented on the usefulness "To automatically consult historic data similar to my patients would facilitate treatment outcome prediction." Other commenting "Identify similar cases, Go through other's opinions about related cases, Refer to an expert's comments." Third participant commenting "This enables me to see that I have not forgotten to do anything. Draw a ROI, Measure and comment on meanME inside the ROIs, Describe the level of inflammation, Reduce noise by performing motion correction, and suggest further monitoring / treatment."

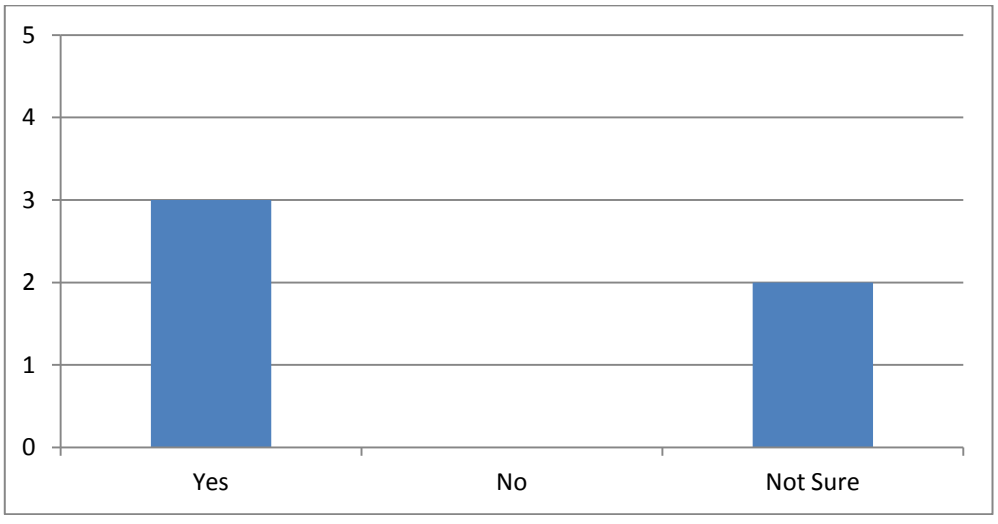

Figure 12: Feedback on service usefulness

Analysis of results

There are two perspectives for the analysis of these results: a) to analyse how the users in this domain consider similarity and b) analyse the cases where the precision/recall is low.

Participants, such as Participant 3, consider similarity by comparing important terms in the reports, as the current similarity service does, have rated precision and recall complete. We are more interested in analysing different interpretation (ways) of considering similarity hence following is the analysis of extreme disagreement (e.g. in the case of the participant 4 or participant 2).

Considering fewer features (terms) for similarity
The current version of the service considers all the terms for similarity. The participants remark that rather than counting all the possible features (i.e. terms) the similarity can be improved by considering only certain aspects (e.g. outcome of the study was the same, or studies where the noise was fairly low).

Considering complimentary & related features
The participant 4 (recall 66% and ranking precision 60%) has focused on the presence of features (e.g. noise) and contrasted against absence of the complimentary or related features (such as motion correction – because the "noise" is corrected with "motion correction"). Hence, the challenge for the similarity service is to consider the other features that might not be present in the text itself but are related or complimentary to the features that are present.

Difference overriding similarity
In some cases the similarity was rejected by participants by citing difference in features that overridden the similar features among the reports. For example, participant 2 rejected 3 reports by contrasting features, quoting from the feedback "for report 8, focus on inflammation and blood ROI gives a different focus hence should score less than 36% and for report 11, Location is good but procedure and findings have little correlation would expect less than 26%".

4.4 Semantic Summarisation Results

The semantic summarisation service is an extension of the semantic query service (Deliverable 5.3.1). The evaluation of this service was carried out with the task 1 (see Appendix A).

Datasets, Participants & Evaluation Method
In the task, a summary of the discussions from a clinical trial were presented to the participants as "Tag Cloud". They were asked to look at the tag cloud and report their findings back to their organisation to see if the discussions during this trial have potential to provide any useful information for their organisation. The aim of this task was to receive feedback on the potential usability of tag cloud generation using semantic query service.

Results
All participants were able to complete the task (i.e. report their findings in a presentation format). Their rating of difficulty in completion is reported in Figure 13 and Table 8. 4 participants either found it "very easy" or "easy".

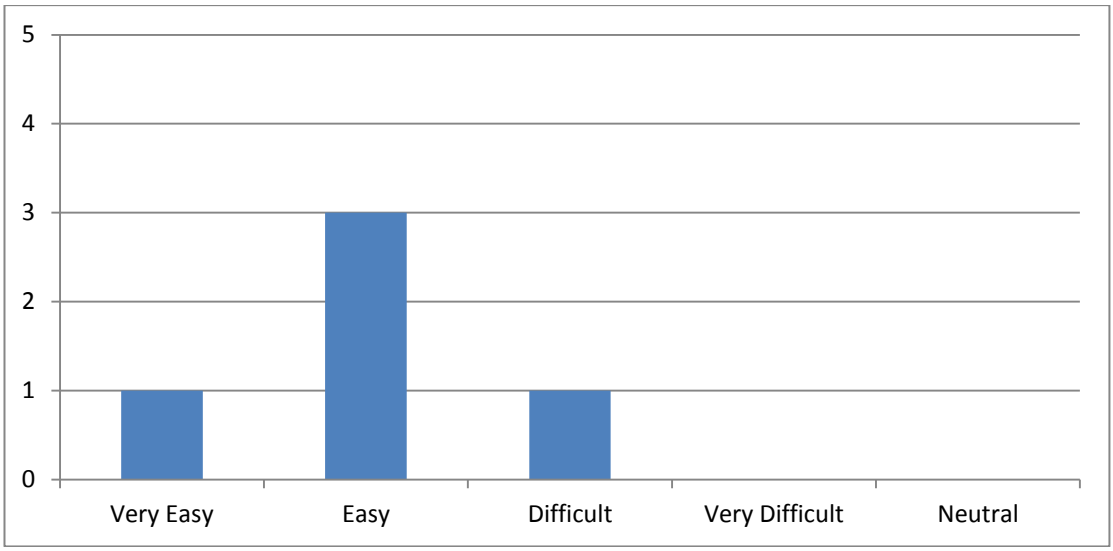

Figure 13: Difficulty in Completing Task

Participants	Comment
Participant 1	It needs extensive knowledge about the topic to make up something meaningful, and even though I have the knowledge e.g. that the patella is related to the knee, based on this cloud It is not clear, why It was so important,(what made It to appear with a larger size) and why other anatomy in the knee was not mentioned...
Participant 2	Size of text and repetition of synonyms are clear indicators
Participant 3	I'm comfortable with the domain's keywords and am a regular Dynamika user.
Participant 4	Tag clouds are a very good representation of this type of data. It's very easy to gain useful information from them and they're easy to look at and understand.
Participant 5	The report has highlighted the most repeated information so it does make a sense for gathering information and wasn't so difficult.

Table 8: Feedback on difficulty level to complete task

Their rating of confidence level in findings is reported in Figure 14 and Table 9. Three participants were "confident" and two users were "somewhat confident" in their findings.

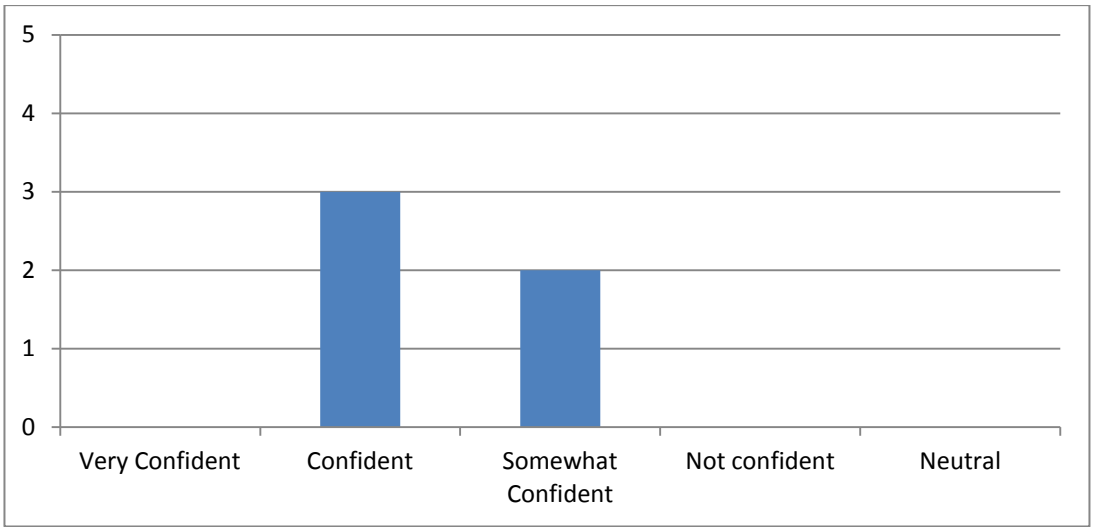

Figure 14: Confidence in findings

Participants	Comment
Participant 1	Definitely the size of the letters gives a clue about the importance of the words, but I really would be interested about the relations between the words appear in the cloud. E.g. if the word severe would be close (or right next to) the word inflammation (and I know that there is a correlation between the distance of the words and how often they appear together), I could guess that in some cases the patient had severe inflammation...
Participant 2	Tag clouds are used extensively in professional publications such as The Economist and FT as well as in social media.
Participant 3	Tag cloud is quite suggestive. This fact, 'augmented' with my domain knowledge makes me feel confident.
Participant 4	The tag cloud shows what I would expect. It picks out the key features of Dynamika. It clearly shows subject of the clinical and it picks out the features of Dynamika which are useful in this type of trial and on the anatomy concerned. It shows some of the key issues Dynamika addresses such as noise and patient motion.
Participant 5	I wasn't much confident in separating information related to Dynamika and clinical trial. It was difficult to separate into two areas through one report.

Table 9: Feedback on confidence in finding

Analysis of the results

Difficulty Level

The comment by participant 1 (Table 8) relating to the explanation of why certain terms are important ("why It was so important, what made It to appear with a larger size and why other anatomy in the knee was not mentioned...") is relevant to highlight the benefit of the Augmentor's semantically-enabled tag cloud. The tag cloud in the current version already

allows browsing background information and the content for a "Tag". However the participants were not required to use it for completing this task. Such information and browsing allows to see other terms related to the "Tag" from the cloud and the list of content tagged with this "Tag" addressing the need highlighted by this participant.
Participants 2,3 and 4 cited their familiarity with the tag cloud and the domain while rating the difficult level.

Confidence in Findings
Participant 1 makes a very interesting suggestion where the message of the summary can be made more effective and intuitive with better placement/layout of the tags to reflect correlation between the distances of the tags.

Participant 3 finds tag cloud suggestive and also "augmenting" the domain knowledge a user in this domain generally have.

For participant 4 tag cloud confirms their expectation.

The task was asking to collect information in two areas: information related to "Dynamika" and "Clinical Trial". Participant 5 makes interesting observation justifying low confidence in finding by noting that there was not sufficient separation for addressing the two areas.

4.4.1 Semantic Relatedness & Usability Evaluation Task Results

This task was designed to receive feedback on the usability of Augmentor. Users were given a task to find reports that mention certain terms. 4 out of 5 participants completed the task and the comments collected indicated that they relied on the semantic relationships offered by Augmentor in completing the task. Original comments from the participants are provided as part of Appendix C. Further analysis of their comments leads to the following usability improvement suggestions:

Showing fewer details
This was pointed by two users as the information on the term (about term and related terms) becomes exhaustive and sometimes difficult to read/grasp. The problem occurs as systems such as Augmentor that utilise large collection of semantic descriptions that generally becomes even lengthier by semantic linkages making it hard for quick identification or comprehension. The highlighted problem points us to the research in the areas of entity summarisation (Cheng et al. 2011; Hoser et al. 2006; Zang et al. 2007; Erkan et al. 2004; Everett et al. 1999) where the centrality of semantic description is considered to provide summarisation.

Expanding/Collapsing Facets of information
Two of the participants commented on the expanding-collapsing of facets ("related reports" facet, "facts about" facet and "terms related" facet). These arrangements were made to address the situation where too much information is displayed on the page; hence users can have opportunity to hide some of the information. This issue also links to 1. Hence, a careful consideration will be given about the usability of the expansion-collapsing functionality. In fact, one of the potential solutions/alternative was provided by other participant who prefers mouse-over instead of mouse-click to expand/collapse facets.

Providing source of the knowledge
One of the participant suggested a feature request where the source of the knowledge (i.e. where the related terms came from – which ontology/source) is made explicit. This could lead to more confidence in the browsing.

5 Future work directions

Over all we can see that there is great potential value in the Dicode services for IMA's industrial strength platform. In this section we evaluate the readiness of services, the cost-effectiveness of including them in the platform and further services IMA would like to evaluate

5.1 Readiness

Augmentor has already been shown to have large potential utility for enhancing the workflow of clinicians using Dynamika for image analysis in randomised clinical trials. For it to incorporated into the Dynamika offering it will need to be fully quality assured by IMA according to its QMS. This will require:

- Full integration testing
- Regression testing of the Dynamika platform
- PSR (performance, scalability, reliability) testing
- Risk analysis

In order to facilitate integration IMA will need to explore with the Dicode service partners the best options with respect to:

- Web services
- API development
- Development of widgets

To be commercially ready, the Dicode services will need to integrate seamlessly with the architecture of choice for IMA. The current state of the services is experimental rather than industrial.

5.2 Cost-effectiveness

It is rather early to determine cost effectiveness of the services. At present IMA is developing an enterprise platform for its image analysis routines. The Dicode services could certainly enhance those routines. It is not yet clear if the effort required to integrate them with the workflow models of the IMA platform will make this a cost effective option.

Over the next period it will be a goal of IMA to understand fully the path to maturity for the services and how they can be incorporated.

5.3 Further services to evaluate

The work to date has focussed on Augmentor to add semantic tagging value to the core report output of Dynamika. This has been sensible because it can be evaluated in isolation of Dynamika and does not require networked connectivity of Dynamika users which was not

available. This situation is changing as IMA independently develops its enterprise platform with built in networking and collaboration. It is clear that this is the basis of incorporating further Dicode services. In particular, services of immediate interest are:

- The Dicode workbench
- The Collaboration and Decision support services
- Use case 2 related data mining services, in particular the subgroup discovery service.

References

Altman D.G. Practical Statistics for Medical Research. (1991) London England: Chapman and Hall.

Cheng, G, Tran, T & Qu, Y. (2011) RELIN: relatedness and informativeness-based centrality for entity summarization. In *Proceedings of the 10th international conference on The semantic web - Volume Part I* (ISWC'11), Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, and

Abraham Bernstein (Eds.), Vol. Part I. Springer-Verlag, Berlin, Heidelberg, 114-129.

Cohen. A coefficient of agreement for nominal scale. J. Educat Psychol Measure 1960; 20: 37-46.

Erkan, G. and Radev, D. 2004. LexRank: graph-based lexical centrality as salience in text summarization. J. Artif. Int. Res. 22, 1 (December 2004), 457-479.

Eugenio, B. D and Glass. M. The kappa statistic: a second look. Computational Linguistics, 1(30), 2004.

Everett, M.G and Borgatti, S.P "The centrality of groups and classes." Journal of Mathematical Sociology, vol. 23, no. 3, pp. 181–201, 1999.

Hoser, B., Hotho, A., Jäschke, R., Schmitz, C., and Stumme, G. Semantic Network Analysis of Ontologies. In Proceedings of LWA. 2006, 297-305.

Zhang, X., Cheng, G., and Qu, Y. 2007. Ontology summarization based on rdf sentence graph. In Proceedings of the 16th international conference on World Wide Web (WWW '07). ACM, New York, NY, USA, 707-716. DOI=10.1145/1242572.1242668 http://doi.acm.org/10.1145/1242572.1242668

Appendix A: Tasks & Evaluation guide for participants

Task 1: Semantic Summarisation Evaluation Task
Estimated Time: 10-15 minutes

You are a radiologist working for Image Medical Analysis (IMA) and regular user of Dynamika software. You are asked to use Augmentor, a tool that summarises the discussions from a clinical trial as "Tag Cloud". You are asked to look at Augmentor's tag cloud [click or See Figure 15] for one of such clinical trials and report your findings back to IMA to see if the discussions during this trial have potential to provide any useful information for IMA. You have 10 minutes to prepare one page presentation to your team on your findings.

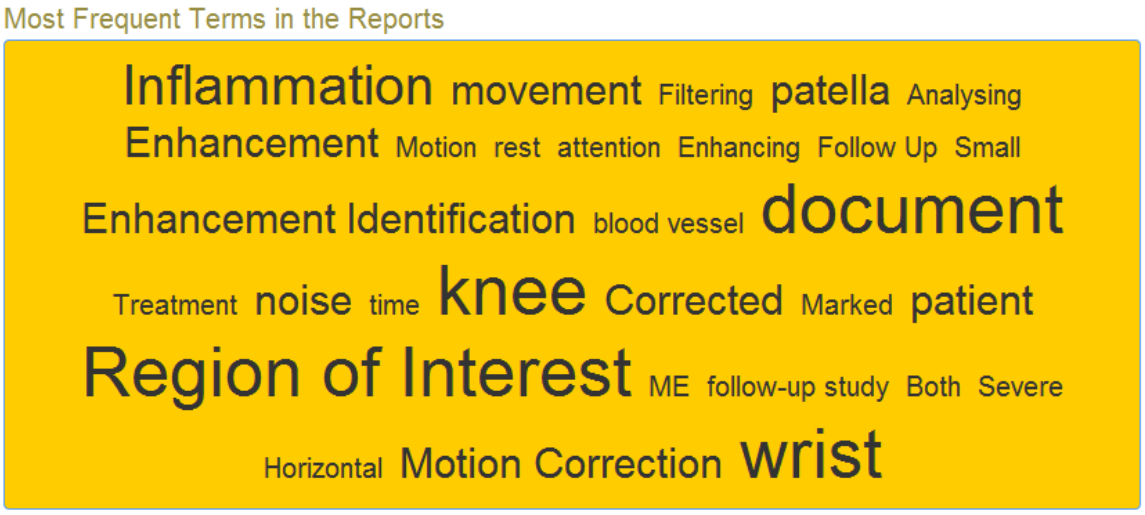

Figure 15: Most Frequent Terms in the reports as a Tag Cloud.

Overview of the Clinical Trial

Identify several most discussed issues you notice related to clinical trial

Identify several most discussed issues you notice specific to Dynamika/IMA

Please answer the following questions:

1.1 How difficult was to gather information for completing this slide?

○ Very Easy	○ Easy	○ Difficult
○ Very Difficult	○ Neutral	

Please explain why?

1. 2 How confident you are about your findings?

○ Not Confident	○ Somewhat Confident
◉ Confident	○ Very Confident

Please explain why?

Task 2: Semantic Similarity Evaluation Task
Estimated Time: 10-15 minutes

You are a radiologist using Dynamika for image processing and Augmenter for searching and browsing past diagnostic reports. You are currently working on **the report No. 5** [Click]. Read comment from the report. You want to find similar reports to the report 5. Go to this page [Click] where Augmentor has found such reports (i.e. that are similar to the report 5). Please look at these reports and answer the following questions:

2.1. Do you agree that these reports are similar to report 5? Provide your answer for each:

Report Number (similarity)	*Are they Similar?*		
Report 6 (86%)	○ Yes	○ No	○ Not Sure
Report 3 (55%)	○ Yes	○ No	○ Not Sure
Report 4 (51%)	○ Yes	○ No	○ Not Sure
Report 8 (36%)	○ Yes	○ No	○ Not Sure
Report 11 (26%)	○ Yes	○ No	○ Not Sure

2.2. Justify your answer (e.g. why a particular report is similar or why it is not?)

2.3. Is there any other report that should be listed as a similar report to 5 and Augmentor missed it? Use this URL [Click] to access all the reports. [Note: the reports are also provided in printed form].

2.4. Do you agree with the ranking order of the report (the most similar first and the least similar last)?

Rank in Augmentor	Report No	Agree with Rank?	Agree with Rank?	If No, what should be ranking order.
1	6	○ Yes	○ No	
2	3	○ Yes	○ No	
3	4	○ Yes	○ No	
4	8	○ Yes	○ No	
5	11	○ Yes	○ No	

2.5. Will the similarity service be beneficial in the task you were doing? Please tick your answer.

○ Yes	○ No	○ Not Sure

If yes, How?

Task 3: Semantic Relatedness Evaluation Task
Estimated Time: 5-10 minutes

Find **report(s)** that mention **term(s) related to** "Cardiovascular system". You can use the "semantic search" functionality from the search page [Click] on Augmentor. You will be asked to note down your findings below.

Please answer the following questions:

3.1. Did you find any reports that mention a term **related to** "Cardiovascular system"? Please tick your answer.

○ Yes	○ No

List the report numbers here (e.g. 1, 4,...) : _____

3.2. What were the terms you found that you think are related to "Cardiovascular system". List few of them.

3.3. Justify your answer(s) given in 2 (e.g how did you find out the terms are related).

3.4. Comment on your experience (possibly with examples) about the browsing in the Augmentor.

Task 4 Semantic Augmentation Evaluation Task

Estimated Time: 25-30 minutes

PART A:

Please edit the worksheet named as "**Questions**" in the provided Excel workbook. Please make sure that Macros are enabled on the document (when Macros are disabled, you will see a security warning at the top of the page, click on "options" to enable Macros). In the worksheet, you will see a table of data consisting of 4 columns:

1. **Column B - "Comment":** this is the comment from the Dynamika Report.

2. **Column C - "Important Term(s)":** this is a list of important terms that have been extracted as **relevant to the comment**.
3. **NOTE: a "term" may not necessarily appear as an exact word in the comment**

4. **Column D - "Is this correct?"**: This is a drop-down list (click on the cell and then the down arrow, and select your answer from the list). Please give your answer accordingly:
   - Please select **"Yes"** , IF you think that:

     The term relates to the comment provided. One way to think is if you search with this term and find this comment as one of the results, will it be correct result? Another way to think about it is, will you use this term to search for this comment?
   - Please select **"No"**, if the above condition is not met;
   - Please select **"I do not know"**, if you are not sure about your answer.

5. **Column E - "Missing Terms":** please write the **additional terms** that you think that are missing according to the conditions in 3, or **leave blank if not**.

NOTE: you can either enter a term that appears **either as an exact word or phrase in the comment's text**, **or should have been picked indirectly**. Again please follow the conditions in 3.

NOTE: please enter the terms separated with commas.

Appendix B: Incorrect terms originating from DON

Filtering
Follow Up
Collecting
Analysing
Enhancement Identification

Appendix C: Relatedness comments

Participant 1	Hierarchical grouping of the result would be interesting, however it would be probably challenging based on what we would like to define the hierarchy I do not feel comfortable with expanding and collapsing information... since it is in a browser just rolling is enough, and might be faster A link at the head might be appropriate that can take me directly to the related stuff
Participant 2	Provided additional "potentially" relevant information which the user may not have been aware off or taken into consideration
Participant 3	Good to have all categories (Reports, Facts, Terms) of results within a single page. - Suggestion: How about a mouse-over instead of mouse-click to expand categories? Navigation becomes easy.
Participant 4	It would have been nice to be able to see where the related terms came from. I.e. which ontology / source. The list of related terms could be ordered by the strength of the link between that and the search term. Or the 'importance' of the search terms / strength of the path they lead to. This would be particularly useful because the list is very exhaustive. The searching and navigation is intuitive enough to begin with and once you've been using it for a short time it works well. The breadcrumb trail is a nice touch but it would be nice if it worked more reliably for navigation during searching.
Participant 5	The system would be good once all problems would be solved from usability point of view as well as functionality. But the flow of searching items and the related terms look good for searching.

Table App1: Response to "Comment on your experience (possibly with examples) about the browsing in the Augmentor"