FP7-ICT-2010-270253

Integrate

Driving excellence in Integrative Cancer Research through Innovative Biomedical Infrastructures

STREP
Contract Nr: 270253

6-Monthly Progress Report # 3
Feb - July 2013

Due date of report: 31 August 2013
Actual submission date: 28 November 2013

Beneficiary name	Beneficiary short name	Country	Date enter project	Date exit project
Philips Research	Philips	NL	M1	M36
Breast International Group	BIG	B	M1	M36
Foundation for Research and Technology – Hellas	FORTH	GR	M1	M36
Custodix	Custodix	B	M1	M36
Institut Jules Bordet	IJB	B	M1	M36
Universidad Politechnica de Madrid	UPM	SP	M1	M36

Start date of Project: 01 February 2011 Duration: 36 months

0 Document info

0.1 Author

Author	Company	E-mail
Anca Bucur	Philips	anca.bucur@philips.com
Kostas Marias	FORTH	kmarias@ics.forth.gr
David Perez del Rey	UPM	dperez@infomed.dia.fi.upm.es
Brecht Claerhout	Custodix	brecht.claerhout@custodix.com
Alexandre Irrthum	BIG	alexandre.irrthum@bordet.be
Cyril Krykwinski	IJB	cyril.krykwinski@bordet.be

0.2 Documents history

Document version #	Date	Change
V0.1	August 2013	Starting version, template
V0.4	October 2013	Incorporates first version of input from partners
V0.5	November 2013	Updated input from partners
V1.0	November 2013	Full version

0.3 Document data

Keywords	Interim activity report
Editor Address data	Name: Anca Bucur Partner: Philips Address: HTC 34 (5-003) 5656 AE Eindhoven The Netherlands Phone: +31 40 27 Fax: +31 40 27 E-mail: anca.bucur@philips.com
Delivery date	November 2013

0.4 Distribution list

Date	Issue	E-mailer
23.11.2013	V1.0	'CNECT-ICT-270253@ec.europa.eu'
		'Robert.BEGIER@ec.europa.eu'

Table of Contents

1 Publishable summary

Public summary
The collaborative INTEGRATE project aims to support a novel research approach in oncology through the development of innovative biomedical infrastructures enabling multidisciplinary collaboration, management and large-scale sharing of multi-level data, and the development of new methodologies and of predictive multi-scale models in cancer. The INTEGRATE infrastructure will bring together heterogeneous multi-scale biomedical data generated through standard and novel technologies within post-enomic clinical trials and seamlessly link to existing research and clinical infrastructures, such as clinical trial systems, eCRFs, and hospital EHRs, in order to enable a range of innovative applications.

INTEGRATE delivers solutions that support a large and multidisciplinary biomedical community ranging from basic, translational and clinical researchers to the pharmaceutical industry to collaborate, share data and knowledge, and build and share predictive models for response to therapies. Moving away from empirical medicine, towards evidence-based personalized care has the potential to both dramatically improve patient outcome and to reduce costs.
The project also aims to make relevant steps towards semantic interoperability. To be able to reuse previous efforts in data sharing, modeling and knowledge generation, and to access relevant external sources of data and knowledge it is beneficial to adhere whenever possible to widely accepted standards and ontologies. The use of standards will also support wide scale adoption of our solutions. A first version of our semantic interoperability layer has been implemented based on the HL7 v3 standard and on relevant medical ontologies/terminologies: SNOMED-CT, MEDDra, LOINC.
The BRIDG standard has been used to represent the clinical trial information in our environment.
An important objective of this project is to build tools that facilitate efficient the execution of postgenomic multi-centric clinical trials in breast cancer. A range of such tools aim to support recruitment through the automatic evaluation of the eligibility of patients for trials based on matching the characteristics of the patient population required by the trial to the patient data available for instance in the hospital EHR. Other range of tools focus on central review of pathology images and on the INTEGRATE Analysis Platform enabling both statistical and prediction analysis. To facilitate the use of the datasets in the INTEGRATE environment for future research, we build a flexible and intuitive cohort selection application that enables users to define, select and retrieve cohorts of patient datasets that suit their research questions. First versions of these tools have been implemented and are currently being evaluated with clinical users.

The INTEGRATE consortium focuses on sustainability beyond the scope of the research project, building a long lasting translational research infrastructure that will promote scientific collaboration among European cancer research centres, pharmaceutical companies, and biomedical research communities well beyond the FP7 funding period. While the core users of the project outcomes are members of the Breast International Group network, we will also actively promote our approach and solutions in wide user communities and in other disease domains.

2 Project objectives for the reporting period

Objectives for the reporting period

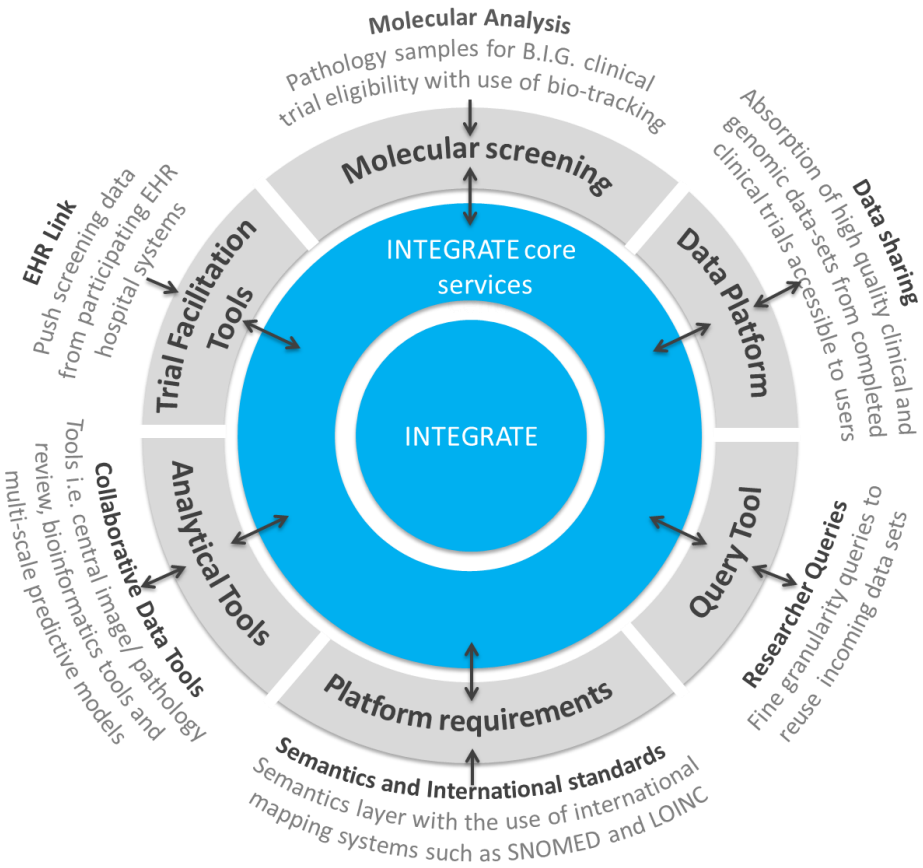
During this reporting period the main objectives as described in the DoW have been to

- Elaborate the final system architecture and security framework
- Finalize the predictive modeling framework and report on the framework and on the predictive models for therapy response
- To prepare the INTEGRATE workshop

Additionally, we have focused on further implementation of the INTEGRATE tools (in the areas as described in the figure below), on the extension of semantic solution and of the INTEGRATE Common Information Model. New data was transformed through the defined ETL pipeline and uploaded to the Common Data Model.

An important objective of this period was to prepare the evaluation and validation process of the project tools, to establish a methodology to be followed in this process and to select concrete metrics. This will be the topic of an additional deliverable that the consortium will prepare before the next review.

The elaboration of a sustainability plan was also an ongoing focus of this period.



Recommendations from previous reviews (if applicable)

The reviewers recommended that the progress from the clinical point of view will be more prominently described in the deliverables of the project and in the progress report.

The reviewers also asked that Deliverable 5.2 Report on methodology and genetic and imaging biomarkers is extended with information on the impact of the research described on the project and how the described biomarkers will be used in the project to develop predictive models.

The reviewers expressed concern with the consortium suggestion to include a non-European partner in the project.

The reviewers recommended better alignment of the technologies with the application domain, holistic integration of the dispersed activities and putting more focus on the sustainability/impact potential through exploitation activities.

The consortium was asked to consider using established standards, if relevant, from other domains in order to allow for a wider impact and deployment potential. E.g. the use of BPMN 2.0 could be a meaningful basis in order to model the information flow.

Overall exploitation and future usage of the project results was not addressed during the review; credible, partner-specific exploitation routes or at least approaches to achieve a kind of sustainability must be presented per partner and in a joint structure. Definitely the consortium should foresee more time for discussing future plans and exploitation strategy.

It is important to focus in remaining time of the project on the integration of the different tools and environments into a consolidated platform implementing realistic clinical workflows and focusing on usability and clinical validation.

<u>The response of the consortium to address the above recommendations</u>:

Deliverable D2.5 was extended to address the concerns and comments of the reviewers. It was submitted according to the defined deadline and was accepted by the reviewers.
We followed the outlined plans for organizing user evaluation and validation workshops and project events. We have prepared and organized an INTEGRATE event focused on describing the project results and emphasizing the need for data sharing. This event took place at the ECCO conference in September.

We prepare the extensions of the consortium with only one new clinical site. From the report, the GBG was accepted as a new site but the Australian Breast Cancer Group was not. Therefore we decided to only include one additional partner and reserve part of the budget for validation workshops focused on each INTEGRATE tool. In each workshop we will invite 3-5 top clinical experts in areas relevant for each of the tools (e.g. trialists for the patient screening tool, pathologists for the collaborative pathology review tool).

Additionally, to show the relevance and impact of our tools beyond INTEGRATE we are discussing with hospitals outside the consortium concerning the validation of our tool at their site. One such site will be selected before the end of 2013.

To demonstrate the international cooperation dimension we continue the collaboration and alignment efforts with Sage Bionetworks, a prominent data sharing initiative in the US.

The next review will take a WP-focused approach and include presentations for each WP next to the overall presentation of progress at project level. We will clearly specify what the progress in the reporting period was and have shorter, focused presentations as requested.

Each demonstrator will be clearly linked to the contributing WPs.

The next yearly progress reports will contain more details on the clinical aspects of the project and the achieved progress.

We will evaluate the use of BPMN2.0 to model workflow and of the corresponding Drools framework in the INTEGRATE project.

We have proposed an amendment for the extension of the duration of the INTEGRATE project to allow for extensive evaluation and validation of the developed solutions.

The final exploitation plan will address the need for sustainability and take into account the feedback and recommendations provided by the experts during the review. The amendment proposed includes effort shift to the Knowledge Management workpackage for all partners to work on the sustainability plan. We are also evaluating the opportunity of establishing an INTEGRATE foundation focused on data sharing and aiming to maintain and promote the INTEGRATE data sharing environment and tools. We will also continue the collaboration with other similar projects.

3 Workpackage progress of the period

3.1 WP 1 (IJB)

As noted in the DoW, the work to be performed in WP1 was accomplished, no effort on this work package was provided during this period.

3.1.1 Objectives (of the reporting period)

N.A.

3.1.2 Achievement/progress made in the past period (per Task)

N.A.

3.1.3 Deviations from the DOW and corrective actions

N.A.

3.1.4 Planning next period

N.A.

3.2 WP 2 (Custodix)

3.2.1 Objectives (of the reporting period)

- Implement, deploy and present the demonstrators in line with the specifications that were defined in the initial architecture document of year 2 (in cooperation with WP6 - Pilots, evaluation and validation).
- Start the design and implementation of the next iteration of the INTEGRATE demonstrators (year 3).
- Further development of the INTEGRATE security services, finishing the authentication components and start work on the authorisation (access control) and audit services.
- Finish the second iteration of the INTEGRATE architecture and security framework.
- Start work on the next and final iteration of the INTEGRATE architecture and status security framework.

3.2.2 Achievement/progress made in the past period (per Task)

Task 2.1 Identification and evaluation of relevant standards

- This task was finished in month 9

Task 2.2 Inventory of re-useable/available relevant solutions and components

- This task was finished in month 9

Task 2.3 Design and implementation of the INTEGRATE reference architecture

- The second iteration of the architecture was finished and the final iteration of the architecture was started
- A next iteration of the ‘patient screening demonstrator’, including a GUI update, was implemented, deployed and presented at the review meeting of year 2

- Two new demonstrators ('cohort selection demonstrator' and 'pathology demonstrator') were implemented, deployed and presented at the review meeting of year 2, mainly coordinated by WP2, which has dealt with task assignment, load distribution and resources allocation
- Brainstorm technical meetings were held, defining the scope of the different demonstrators for year 3
- A poster about the cohort selection engine has been written and will be presented at AMIA 2013

Task 2.4 Security for dynamic collaborative environments

- The first iteration of the INTEGRATE security framework services was finished and presented at the review meeting of year 2. The main focus was authentication which included STS/IDP components and an identity management framework
- The authentication infrastructure was integrated in the different demonstrators that were presented in year 1
- A scientific paper about the concept of contextual attributes was presented at HEALTHINF 2013 in Barcelona, which was also selected for publication in the Springer-Verlag journal
- Initial work has been done for the first iteration of the authorisation framework

Task 2.5: Component integration and interfacing with external systems

- Integration guidelines were specified in Deliverable 2.5

3.2.3 Deviations from the DOW and corrective actions

- Deliverable D2.6 was delayed to March 2013
- Deliverable D2.7 is delayed. The new deadline is M36 (January 2014) and will be aligned with the extension of the INTEGRATE project and the inclusion of an additional clinical site.

3.2.4 Planning next period

- Finishing the last iteration of the architecture and security framework (year 3), creating a final architectural document.
- Further specification, integration and deployment of the next iteration of the security framework services, focusing on authorisation.
- Further integration of the security environment in the demonstrators of year 1 and 2.
- Decide which demonstrators will be implemented (or updated) and presented in year 3 of the INTEGRATE project.

3.3 WP 3 (UPM)

The main objective in WP 3 is to facilitate a common access to clinical data for applications of the INTEGRATE platform. Common information models, vocabularies and mappings mechanisms are required to homogenize data repositories. The following objectives should be achieved: (i) identification of the initial proposals for the core dataset (common vocabulary), (ii) mapping formalisms and mappings between the core dataset and the common data model (CDM), (iii) an initial prototype of the

semantic interoperability layer to facilitate homogeneous access to data sources and (iv) a homogeneous solution to access external sources.

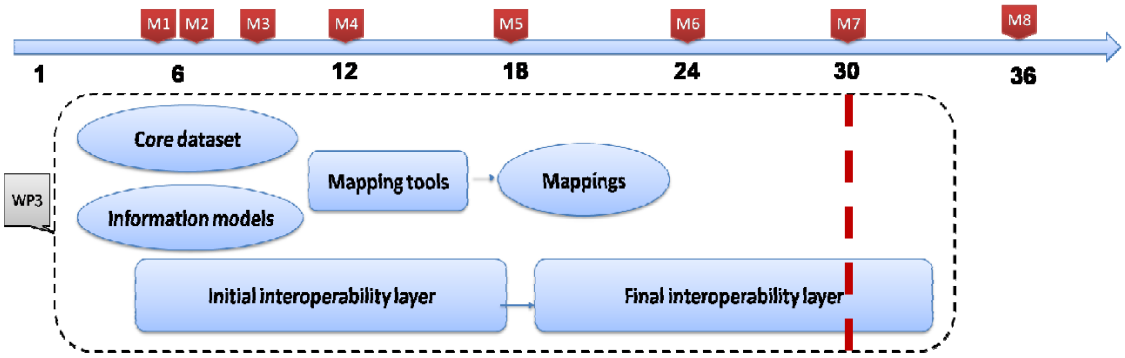


Fig A. Work Package 3 components and planning

From month 25 to month 30, WP3 have been mainly focused on the semantic interoperability layer (Task 3.4) and mappings (Task 3.3), while the core dataset (Task 3.1) and the common information models (Task 3.2) have been iteratively refined.

3.3.1 Achievement/progress made in the past period (per Task)

Task 3.1 Definition of the semantic core dataset
The set of concepts that would be included within the INTEGRATE "lingua franca" have been extended with new data sources. Concretely, a new vocabulary has been included to code gene names, HGNC (Human Genome Nomenclature Consortium). The core dataset vocabularies files have been therefore extended and stored using the OWL ontology representation language and loaded into a SESAME server to facilitate semantic reasoning.

Task 3.2 Definition of the information models of the clinical and research infrastructures
Minor issues have been updated within the CDM according to new data sources. The proposed method for data normalization following the SNOMED normal form and the terminology binding between HL7 RIM and SNOMED will be presented in the MEDINFO 2013 conference "The 14th World Congress on Medical and Health Informatics".

Task 3.3 Semantic formalism, mapping tools and mapping implementations
During the reporting period, we have explored how to map new data sources into the INTEGRATE core infrastructure. The terminology binding is mainly used to automatically build queries extracting data from the core dataset, but also to normalize such data. The following figure describe the different components and technologies used when deploying the INTEGRATE platform in a new institution.

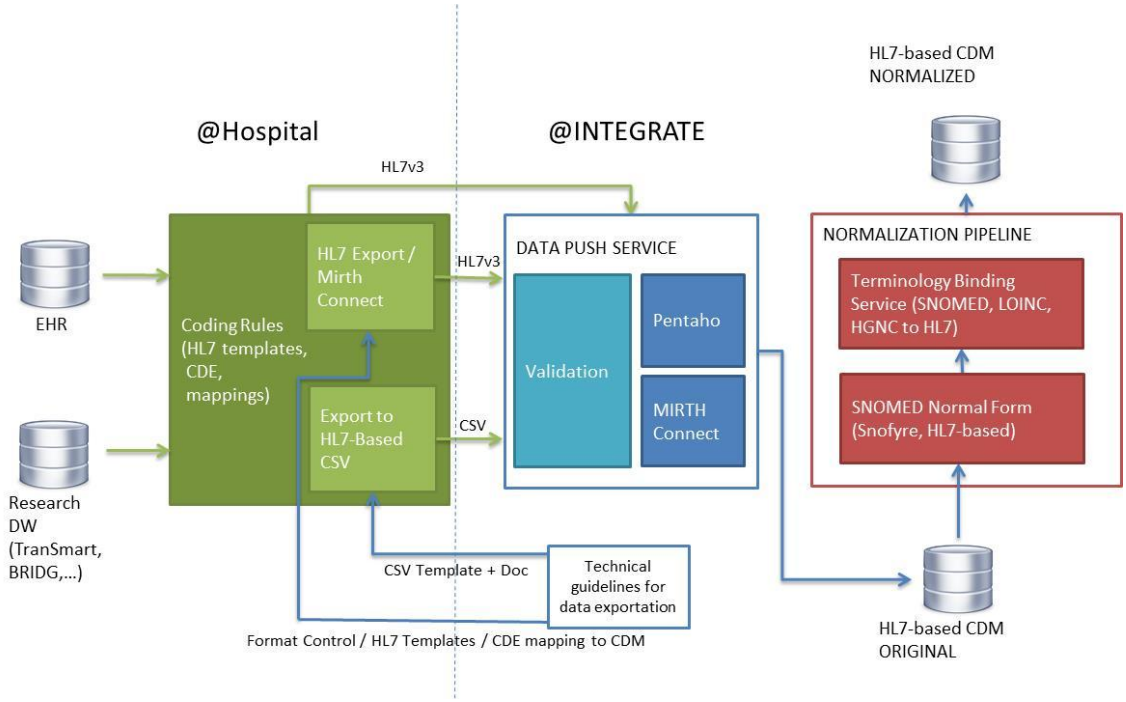


Fig. B. Deployment of the INTEGRATE platform in a new institution

Open source tools such as Mirth Connect can be used to export HL7 messages at the institution. From the INTEGRATE platform, HL7 or CSV templates are provided to follow the coding rules to be “INTEGRATE compliant”. Once such messages are generated, a data push service is available to load data into the CDM. The two tasks involved afterwards within the normalization pipeline, Terminology binding and SNOMED Normal Form, will be presented in the MEDINFO 2013 conference “The 14th World Congress on Medical and Health Informatics”.

Task 3.4 Design and implementation of the semantic interoperability layer
After a first version of the query mechanism to homogeneously retrieve data integrated through the platform, in the reporting period we have focused on: (i) normalizing queries according to data normalizations from T3.3 and (ii) encapsulating the CDM structure and the SPARQL syntax. Therefore, and additional component, the query builder, has been designed and implemented during the reporting period.

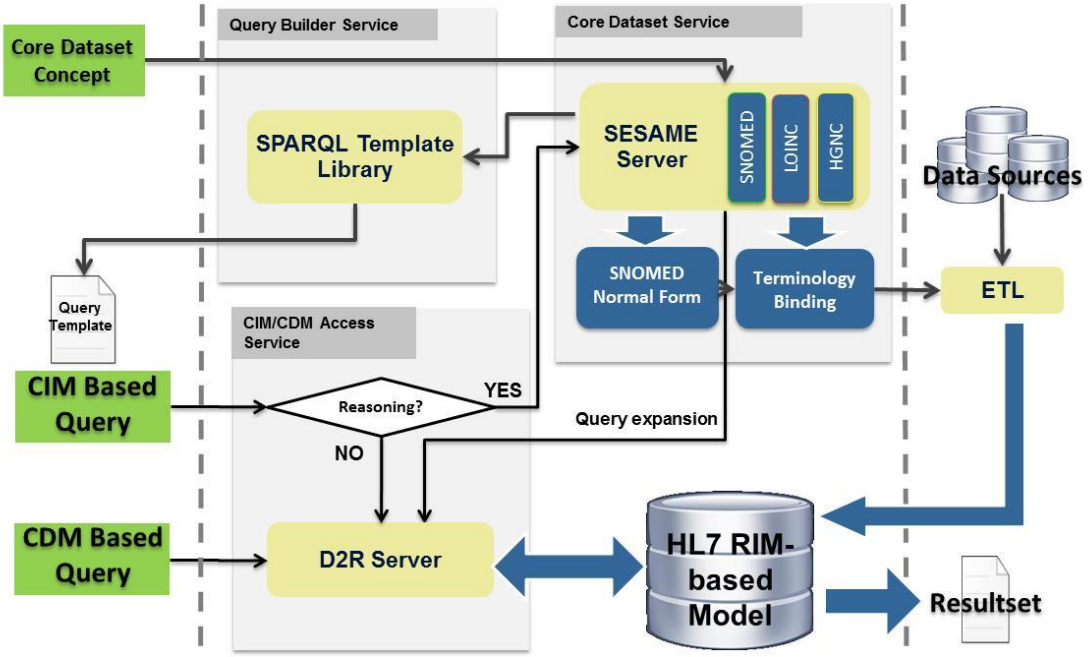


Fig. C. Query builder interaction with the rest of the components of the semantic interoperability layer

Once data is normalized and stored in the INTEGRATE CIM, the query builder receives a concept and provides the corresponding query template (SPARQL-based) according to the core dataset service. Applications using the semantic interoperability layer compose the final query and execute against the data access service to homogeneously retrieve the results.

Task 3.5 Standards-based uniform access to external sources
Besides EHRs, the main data sources until now, during the reporting period we have explored custom research databases. To avoid using HL7 messages that are not common outside the healthcare environment, we have provided CSV templates to store such external data sources into the INTEGRATE infrastructure.

3.3.2 Deviations from the DOW and corrective actions

There are not significant deviations from the DoW. WP3 have been mainly focused on Task 3.3 and 3.4 during months 25 to 30.

3.3.3 Planning next period

The next reporting period will be focused on validation of the INTEGRATE tools with new institutions. The core dataset will be extended with concepts from new data sources and new domains beyond breast cancer. The deployment of the semantic interoperability layer will be tested with new institutions involved in the validation process. And the final version of the semantic interoperability layer will be released.

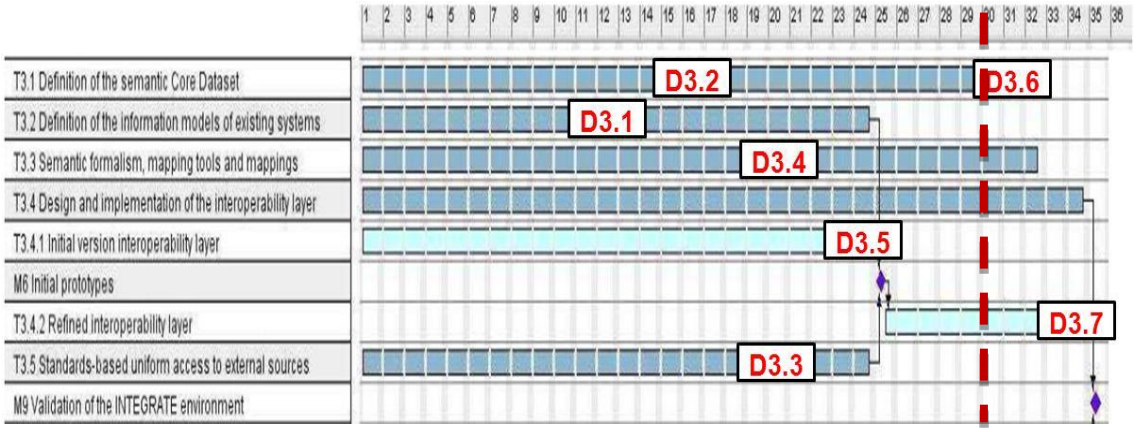


Fig. D. WP3 task planning according to the DoW

Deliverable 3.6 will explore the extension of the core dataset to new domains in oncology, while deliverable 3.7 will describe the final version of the semantic interoperability layer.

3.4 WP 4 (FORTH)

During the last 6 months WP4 focused on the quality improvement of the central reviewing platform and thus providing a more robust and user friendly web based platform for managing and reviewing pathology images.

3.4.1 Objectives (of the reporting period)

Our effort during the reporting period has been focused on:

- Resolving bugs and user interface glitches.
- Improving the user interface and the overall experience of the users.
- Finalizing the review protocol process workflow mechanisms
- Finalizing messaging services
- Finalizing the description and the design of the services for exchanging data with the INTEGRATE core platform.
- Implementing the annotations mechanism for the Central Reviewing platform.
- Updating the tilling service for the Central Reviewing platform.

3.4.2 Achievement/progress made in the past period (per Task)

An analytical report per task for the status and the progress of the WP4 -in the reporting period- follows in the sections below.

Task 4.1 Model, data and annotation repositories

In task 4.1, FORTH assisted UPM in order to define the part of the model which handles the data produced by the central reviewing platform of pathology images.

Task 4.2 Tools enabling data and knowledge sharing

In task 4.2 FORTH implemented the necessary client side listeners for the SOAP services which have been created by UPM, in order to retrieve the report data from the INTEGRATE's repositories and push updated data back to the repositories.

Task 4.3 Tools enabling collaboration

In the last 6 months, the main effort in task 4.3 was to enhance the functionality and the usability of the central reviewing platform. In more details the following areas were the ones were we focused our development.

New annotation tools have been under development, which allow the segmentation of regions of interest and their storage in the INTEGRATE repositories. The segmented areas are represented by vector graphics, which scale as needed in order to represent the correct region of interest in any zoom level in the virtual microscope module.

The resolution mechanism of the central reviewing platform, which is triggered upon a conflict among the reports submitted by the reviewers, has been upgraded in order to assist the moderator of the review to resolve the issue. Notifications and emails are send to the reviewers in order to inform them that special attention is needed, on the areas which are marked as conflicting.

The reports of the reviews can now be exported as pdf files and stored locally in the user's personal computing device.

Finally the user interface has been updated to use ajax-ified widgets whenever that was possible, in order to enhance the user experience and to provide the feeling of a native desktop application to the final user.

The developer's version of the central review for pathology images platform (aka Collaboratory) is available from FORTH's servers and along with the Analytical tools of the WP5 can be used and can be tested by any partner who has an account.

Task 4.4 Privacy Enhancing Processes and Services
[There is nothing to report regarding the task 4.4, as the implementation of the security services has been made in previous periods.]

3.4.3 Deviations from the DOW and corrective actions
No deviations in the DOW.

3.4.4 Planning next period
For the upcoming period, FORTH's development effort is focused on the following:

- Extend the tilling service in order to be easily installed and operate in the INTEGRATE's data warehouse servers (currently it is bonded to the Collaboratory server).
- Complete & test the services which communicate with the data warehouses for storing and retrieving data.
- Finalise the annotations module of the Collaboratory in order to store the regions of the interest which have been set by the reviewer in the report in the data warehouse.

3.5 WP 5 (FORTH)

WP5 has focused on providing users with a web-based multi-functional and easy-to-use platform for analyzing large multi-level datasets using a pool of statistical tools and predictive models.

3.5.1 Objectives (of the reporting period)

Our effort for this reporting period has been mainly focused in:

- Replacing the current front-end of the INTEGRATE Analysis Platform by a more user-friendly, flexible, and less complex framework.
- Establishing the integration between the Analysis Platform and the Cohort Selection Platform.
- Requesting access to the public available data from the Synapse Commons Repository (see 2nd project periodic report for further data information) to be used by the predictive analysis models.

3.5.2 Achievement/progress made in the past period (per Task)

Task 5.1 Definition of clinical scenario (questions) for the INTEGRATE VPH use case

No further work is required

Task 5.2 Definition of genetic and imaging biomarkers and of a modelling methodology

No further work is required

Task 5.3 Development of predictive models of response to therapy and of the modelling framework

The last 6 months, the main effort in task 5.3 has been moved from the development of the core platform's functionality to the re-design of the platform's layout, the validation of the existed tools through public available datasets, and the integration of the Analysis Platform with other platforms within INTEGRATE. Precisely, the overall analysis layout is being gradually replaced by Java Server Faces and PrimeFaces (http://primefaces.org) using widgets and data display controls instead of tables, providing a more flexible, user-friendly and easily configurable deployment platform.

A joint effort between FORTH and Philips partners has been made to achieve integration between the Analysis Platform and the Cohort Selection. Using this integration, the user can select a specific group of patients within the Cohort Selection and at the same time get access to statistical analysis applied to this group. This is accomplished via web services containing information about the selected cohort, the selected statistical analysis (i.e. apply descriptive statistics to tumor grading size of the selected population) and the analysis results (figures, tables, etc.). Pseudo-coding language has been generated, relating each type of the available by the Analysis Platform analysis with a unique term, and incorporated into JavaScript Object Notations for parsing data structures and associative arrays of information to both platforms.

FORTH has also prepared an application for local ethical committee approval consisting of a brief description of the rationale/methodology and an Institutional Review Board (IRB) request for local ethical approval. These two documents together

with the correspondence concerning the public available dataset from the Synapse Commons Repository were sent first to FORTH's legal department for editing/approval and then to the local committee for approval. These data will be used in analysis related to breast cancer research using the predictive modelling framework from the INTEGRATE Analysis Platform.

3.5.3 Deviations from the DOW and corrective actions

No deviations

3.5.4 Planning next period

The planning of activities for the upcoming period is mainly focused on the following fields:

- Completion of the re-design of the platform's layout using Java Server Pages and PrimeFaces.
- Once we get access to the Synapse Commons Repository dataset, an extended predictive model has been implemented to use the date in order to validate and assess its predictive efficacy in classifying different group of patients in breast cancer. This study could be further used for publications (conferences and journals).
- After the validation process, the model will be ready to be part of the predictive analysis framework within the INTEGRATE Analysis Platform.

3.6 WP 6 (Philips)

WP6 has focused in this reporting period on preparing the evaluation and validation of the INTEGRATE tools and solutions. To this end, several sessions were organized with the project partners to elaborate based on existing standards an overall methodology for evaluation and validation. For each tool we have defined concrete metrics to be measured. The results of this work will be captured in the additional deliverable D6.5 Scenarios for validation and detailed metrics for the validation of tools.

3.6.1 Objectives (of the reporting period)

The objectives of the work package for the reporting period are:

- To formulate evaluation criteria, validation procedures, and feedback report guidelines
- To coordinate the specifications of test (validation) cases and scenarios
- To coordinate evaluation and validation activities concerning all the project software components – once these are ready and delivered by the technical WPs.

Another objective of the WP is to prepare the technical and procedural infrastructure – in compliance with the defined security framework of the project – for the installation of the INTEGRATE software solutions for their extensive evaluation and validation.

3.6.2 Achievement/progress made in the past period (per Task)

Task 6.1 Building the INTEGRATE development and testing environment

No further work was required in this task

Task 6.2 Formulate evaluation criteria, validation procedures and feedback report guidelines

During this reporting period we have elaborated for each INTEGRATE tool concrete metrics and criteria that will be measured during the evaluation and validation process. We have also defined a standards-based process for carrying out the evaluation and validation steps. These will be described in deliverable 6.5.

Task 6.3 Coordinate specifications of test scenarios and of demonstrators

During this reporting period we have started the definition of the scenarios that will support the evaluation and validation of the tools. This is a joint activity of the clinical and technical partners of the project.

Task 6.4 Deployment Environment

In this reporting period we have started the definition of the deployment environment that will be set up for evaluation and validation in compliance with the legal and security framework of the project. We have agreed on the steps to be followed and the deployment context for each of the tools.

Task 6.5 Coordinate evaluation and validation activities and reporting

This activity has not started yet.

3.6.3 Deviations from the DOW and corrective actions

No deviations.

3.6.4 Planning next period

A thorough evaluation and validation of the tools requires the involvement of additional clinical experts from different sites. Therefore we have proposed the enlargement of the consortium with one clinical partner, we are planning several evaluation and validation workshops with clinical experts, and we are setting up collaborations with clinical sites outside the consortium for validation and evaluation. This also requires an extended timeline for the project.

An additional deliverable “6.5 Scenarios for validation and detailed metrics for the validation of tools” will be elaborated. This deliverable will describe for each tool developed by the project the scenarios for evaluation and validation and the detailed metrics that will be measured. This deliverable will also describe which clinical sites will validate each tool.

We aim to carry out the validation of the tools with the INTEGRATE clinical partners, in validation workshops with clinical experts from outside the consortium and at a clinical site that is not part of the consortium. Each tool will go through a thorough evaluation and validation with at least one clinical site and/or validation workshop.

3.7 WP7 knowledge management (BIG)

3.7.1 Objectives (of the reporting period)

Besides the recurring knowledge management activities (newsletter, web site…), WP7 has focused during this period on the organization of an INTEGRATE mini-symposium on "the potential of data sharing" within the context of a high profile international oncology conference. The identification of additional pilot hospitals for testing of the tools and, potentially, future exploitation, was also an objective for this period.

3.7.2 Achievement/progress made in the past period (per Task)

Task 7.1: Dissemination

The main achievement for task 7.1 for this period has been the organization of the mini-symposium "The Potential of data sharing" that will take place during the ECCO conference in Amsterdam in September 2013. This conference is one of the high-profile oncology conferences and will provide visibility of the project to the oncology community, an important part of the dissemination/exploitation target audience. Several high profile speakers have been approached and we are reaching agreements on participation to the mini-symposium. We have also organized all the practical details of this event.

Task 7.2: Exploitation

In addition to reaching to the clinical oncology community through the mini-symposium, three additional hospitals (besides IJB) have been contacted to be pilot sites for the INTEGRATE tools (in Germany, Sweden and Iceland). Besides the interest for testing and improving the tools, recruiting additional pilot hospitals is seen as a natural first step towards finding exploitation opportunities. Negotiations are on-going and, so far, contracts have been signed with one hospital in Germany.

Task 7.3 Standardisation

No activities took place for this task during the reporting period.

Task 7.4 Intellectual Property

No further work required at this stage of the project.

3.7.3 Deviations from the DOW and corrective actions

The development of common information models, vocabularies and mappings is still work in progress (nearing completion) and thus standardization at this stage is considered premature. Standardization efforts will be undertaken as soon as we consider that these products of the INTEGRATE project have reached sufficient maturity.

3.7.4 Planning next period

During the next period, dissemination activities will continue (newsletter, scientific publications…). The main event planned is the INTEGRATE mini-symposium “The potential of data sharing” (27 September 2013, ECCO conference, Amsterdam). Exploitation activities will continue through the newly identified pilot hospitals, and through the identification of additional interested stakeholders.

4 Achievements per individual partner

Partner 1 Philips

The main focus in this reporting period has been on further implementation of the INTEGRATE tools.

- Cohort selection (Nona)
  Work has focused on the concept development followed by the implementation of the prototype. This includes integration with the services:
  - Authentication service
  - Single criterion matcher service
  - Locker service

  Implementation work on the cohort selection prototype Nona has started, and is ongoing. An initial study period focused on solidifying the UI concept, so that it has a good match to both the use case and the technical boundary conditions. Nona uses the same criterion matching service as used in Decima, but here the user needs to specify the criterion in a SNAQL script himself. In addition to the functional part of looking for cohorts, Nona therefore should support in terms of (visual) templates to specify the script. This allows a wide range of users to utilize the tool: novices can use the templates, while experts are able to use the full expressive power of SNAQL. The implementation is ongoing.
- The patient screening tool prototype Decima has been extended from an initial implementation to an elaborate prototype suitable for end-user testing. The code base has been redesigned to enable a larger team to work on the code simultaneously. The interactions and visual design have been adapted according to the comments from the user study from the previous year. The criteria are now presented in sorted lists rather than supporting free placement. Clinical evidence, criterion and the computed eligibility have been strongly connected by combining them in a single visual element. Additional information on trials has been added to the trials overview, to directly aid the physician in the screening process. Decima connects to the services in the INTEGRATE ecosystem, respecting the proper authentication requirements.

We have participated in the INTEGRATE data sharing mini-symposium at the ECCO congress with a presentation of the INTEGRATE project.
We contributed to the elaboration of the evaluation and validation methodology.
Presented the INTEGRATE project in the ENBC 2013 Conference and in the 2nd Summer School on Computational Oncology.

Partner 2 BIG

During this reporting period, BIG achieved the following:

Organization of the INTEGRATE event “The potential of data sharing” (27 September 2013, ECCO conference, Amsterdam):
- Contacting and securing speakers’ participation to the event
- Practical organization

Production of INTEGRATE newsletter 4:

- Organization and gathering of contributions
- Writing of articles
- Layout and graphic design

Work towards the identification of additional pilot hospitals for the INTEGRATE tools:
- identification of three potential pilot hospitals
- Negotiations
- Securing the participation of one pilot hospital

Development of the molecular screening pilot:
- negotiations with sites and labs
- submissions to ethics committees
- development of the IT platform

Other achievements
- reviewing and participating in the writing of deliverables
- updating the INTEGRATE website
- providing on-going clinical guidance and feedback on tools
- negotiation with UNICANCER in view of obtaining additional clinic-genomic data sets for INTEGRATE

Partner 3 FORTH
FORTH has put significant effort into fast developing the collaboration platform and to this extent; a significant effort has been made to present a working version of the tool. Further work and testing is taking place to ensure that the final version will be delivered on time. The analytical tools platform has been refined and the predictive modelling components have been added. A joint effort between FORTH and Philips partners has been made to achieve integration between the Analysis Platform and the Cohort Selection.

Partner 4 Custodix
- Attended telco's and technical, review and consortium meetings
- Led and contributed to the next iteration of the architectural document
- Contribution in discussions about semantic approaches, data sources and common and local information models
- Discussed the scope of the demonstrators for the second and third review meeting
- Implemented and deployed the patient screening demonstrator and cohort selection demonstrator in collaboration with the other INTEGRATE partners
- Devised the innovative DSL Query engine core for the cohort selection application
- Started work on the final version of the privacy enhancing services
- Discussed and provided input for the scope of the INTEGRATE demonstrators in year 2 and 3
- Presented a scientific paper about contextual attributes at HEALTHINF 2013
- Written a poster about the DSL Query engine for presentation at AMIA 2013
- Integrated authentication security in the INTEGRATE demonstrators of year 1
- Finished implementation of the authentication services of the INTEGRATE security framework
- Started work on the next iteration of the security framework, focussing on authorisation

- Contributed to and reviewed the last iteration of the technical use cases
- Contributed to the INTEGRATE newsletter

Partner 5 IJB

- Report on the preparation of the deployment environment by defining verification and validation procedures as required to test the platform under legal and security requirements. This includes the definition of quality procedures for the validation of the scenarios and measurable elements to prove the valuables uses of the platform. This work was also an opportunity to identify and present the various clinical sites involved in the validation phase, address issues related to the security requirements and data exchange, describes the process required to incorporate new data sources to the platform and the common data model (the Extract, Transform and Load process) and finally defines semi-formally (through validation protocols) specific activities to validate the installation, use and performance of the tools.
- Preparation of assessment by IJB staff of developed tools (cohort selection, patient screening, central review pathology and analysis tools). The relevant actors were identified, tasks where scheduled, and contact was taken with the clinical actors (pathologists, research nurses and data centre members) who will be involved in the validation of software and prepare the technical environment required for installation tools to evaluate.
- Collaboration with FORTH for the report on the methodology and the genetic and imaging biomarkers. This work consisted into an exhaustive state of the art and a summary of the progress realized in order to provide tools necessary for doing analysis and predictive modelling within the platform. The tools will help exploiting, analysing and assessing the quality of multi-level data, and for instance estimating the correlation between them and the clinical response. Furthermore the platform will help also in selecting interesting features from the multiple level data that can be used as candidate markers, defining predictive models based on either homogeneous or heterogeneous data and validating the models within the platform. It also will provide functionality for processing whole-genome expression arrays, gene prognostic signatures, clinical characteristics, and imaging biomarker, perfusion and diffusion images.

Partner 6 UPM

- Analysis of core dataset concepts from new clinical data
- Mapping of new data sources into the CDM
- Refinement of the HL7-based CDM for INTEGRATE
- Implementation of a pipeline to normalize data sources into the CDM
- Implementation of the query builder component to encapsulate CDM structure and SPARQL syntax

5 Project management

5.1 Consortium management tasks and achievements

Several key management tasks captured the focus in this reporting period. First, we have prepared the second project review. After the review the focus shifted on coordinating the further development of the project tools and on the preparation of the validation and evaluation process. This WP also coordinated the selection of an additional clinical partner for the project (discussions with several partners were carried out to select a suitable partner with expertise in the clinical domain or the project, with expertise in running clinical trials, and willing to share data in the project and to participate in the evaluation and validation of project tools).

The WP also led the discussions concerning the elaboration of an amendment to the DoW which will be proposed to the EC. The goal of this amendment is to support effective validation of the project tools and to strengthen the focus on sustainability and exploitation of results.

5.2 Changes in the consortium

There were no changes in the consortium. However, in the next reporting period we propose the extension of the consortium with a clinical partner, the German Breast Group.

5.3 Cooperation

In this reporting period the collaboration in the consortium has been excellent, much of the work in this reporting period focusing on integration and involving all partners of the consortium.

The preparation of the demonstrators for the project review and the review itself was the main focus at the beginning of this reporting period. This was a joint effort to which all project partners were committed: All partners presented in the review and contributed to the prototypes that were demonstrated.

We have also jointly organized a mini-symposium in the ECCO congress where several partners provided presentations, a seminar at the MAASTRO clinic in Maastricht, the Netherlands (hospital interested to participate in the validation of the INTEGRATE tools), and published a newsletter.

The preparation of the evaluation and validation process has also involved all the partners in the consortium.

5.4 Meetings

Date	Event	Venue/host	Country
01-02-2013	Monthly telco	PHILIPS	Telco

05-02-2013	Trial Metadata linking to CDM	PHILIPS	Telco
11/02/2013	Healthinf 2013	Barcelona/Biostec	Spain
14/02/2013	Trial metadata Telco	Sint-Martens-Latem/Custodix	Belgium
01-03-2013	Monthly telco	PHILIPS	Telco
11-03-2013	Consortium Meeting	PHILIPS	Holland
20-03-2013	Convergence Meeting	EC	Belgium
27-03-2013	Review preparation telco	CUSTODIX	Telco
05-04-2013	Montly telco	PHILIPS	Telco
24-25/04/2013	Review Preparation Meeting	Brussels/EC	Belgium
26-04-2013	Review Meeting	EC	Belgium
03-05-2013	Monthly telco	PHILIPS	Telco
07-06-2013	Monthly telco	PHILIPS	Telco
11-06-2013	Consortium Meeting	CUSTODIX	Belgium
24-06-2013	Summer School (P-medicine project)	UdS	Germany
04/07/2013	Semantic security Telco	Sint-Martens-Latem/Custodix	Belgium
05/07/2013	Monthly Telco	Sint-Martens-Latem/Custodix	Belgium
16/07/2013	Deliverable 4.5 Telco	Sint-Martens-Latem/Custodix	Belgium
23/07/2013	Seminars on the INTEGRATE project	Maastricht/MAAS TRO	The Netherlands

6 Deliverables

Del. no.	Deliverable name	WP no.	Lead participant	Nature	Dissemination level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date	Comments
2.7	Final system architecture and security framework	2	Custodix	R	PU	30	no	36	Deliverable postponed to support the further elaboration of the environment and tools and the inclusion of an additional clinical site as it is the final version of the architecture.
5.3	Report on the framework and on the predictive models for therapy response	5	Forth	R	PU	30	yes	34	
7.9	Report on the INTEGRATE workshop/launching event	7	BIG	R	PU	30	Yes	34	
7.10	Project newsletter	7	BIG	R	PU	30	Yes	30	

6.1 List of milestones

Milestone no.	Milestone name	Lead participant	Due achievement date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
MS7	Final Integrate architecture	Custodix	30	no	36	Milestone will be reached in the next reporting period.

7 Use and dissemination

7.1 Dissemination activities

Planned/actual Dates	Type	Type of audience	Countries addressed	Size of audience	Partner responsible /involved
11-13/02/2013	HEALTHINF 2013 Conference	Health Informatics	EU		Custodix/UPM/Philips
03-07/07/2013	IEEE EMBC 2013	Biomedical Informatics	Worldwide		Philips/BIG/Custodix/FORTH/UPM/IJB
27/09/2013	INTEGRATE mini-symposium at the ECCO Congress	Oncology	Worldwide		BIG/Philips/Custodix/UPM/FORTH/IJB
16-20/11/2013	AMIA 2013 Annual Symposium	Health Informatics	Worldwide		Custodix/UPM/Philips/IJB
11-13/11/2013	IEEE BIBE 2013	Health Informatics	Worldwide		Custodix/UPM/Philips

7.2 Publications

Presented papers:
Sergio Paraiso-Medina, David Perez-Rey, Raul Alonso-Calvo, Brecht Claerhout, Kristof de Schepper, Philippe Hennebert, Jérôme Lhaut, Jasper Van Leeuwen and Anca Bucur. Semantic interoperability solution for multicentric breast cancer trials ath the INTEGRATE EU project. In proceedings of the 6th International conference on Health Informatics, HEALTHINF 2013, 11-14 Feb 2013, Barcelona.

Submitted papers:
Raul Alonso-Calvo, David Perez-Rey, Sergio Paraiso-Medina, Brecht Claerhout, Philippe Hennebert and Anca Bucur. Standard-based semantic interoperability approach for managing multi-centric clinical trials. Special issue on Managing Interoperability and compleXity in Health Systems (MIXHS) of Methods of Information in Medicine journal.

7.3 Contributions to conferences (abstracts, etc)

Juan M. Moratilla, Raul Alonso-Calvo, Gema Molina-Vaquero, Sergio Paraiso-Medina, David Perez-Rey, Victor Maojo. A data model based on semantically enhanced HL7 RIM for sharing patient data of breast cancer clinical trials. In proceedings of The 14th World Congress on Medical and Health Informatics, MEDINFO 2013, 20-23 August 2013, Copenhagen. (Accepted for publication)

Santiago Aso, David Perez-Rey, Raul Alonso-Calvo, Antonio Rico-Diez, Anca Bucur, Brecht Claerhout, Victor Maojo. Analyzing SNOMED CT and HL7 Terminology Binding for Semantic Interoperability on Post-Genomic Clinical Trials. In proceedings of The 14th World Congress on Medical and Health Informatics, MEDINFO 2013, 20-23 August 2013, Copenhagen. (Accepted for publication)

7.4 International articles

- Supporting Contextualisation of ABAC Attributes Through a Generic XACML Request Handling Mechanism (Springer-Verlag)

8 Manpower overview

Actually Spent 6-Monthly Human Resource Allocation

Partner	WP1		WP2		WP3		WP4		WP5		WP6		WP7		WP8		*Total*	
	planned	spent	Planned	spent	Planned	spent	Planned	spent	planned	spent	planned	spent	Planned	spent	planned	spent	planned	spent
Philips	0.6	1	1.6	1.2	1.7	2	1.9	2	0.6	1	0.9	1	0.3	0	1.0	1.0	8.6	*9.2*
BIG	1.5		1.1		0.1		0.9		1.0		0.8		0.9		0.12		6.42	
FORTH	0.6	1.5	2.0	6	0.5	2.5	2.3	11	2.7	15.37	0.8	2.5	0.2	1.5	0.25	1.5	9.35	*41.87*
Custodix	0.5	0.01	3.5	1.65	0.5	0.15	1.7	2.58	0.2	0	0.7	2.38	0.1	0.98	0.08	0.19	7.28	*7.86*
IJB	2.3		0		0.6		0		1.2		0.9		0.3		0.12		5.42	
UPM	0.5	0.7	1.5	1.51	2.3	4.00	1.2	3.52	0.2	0.70	0.8	3.11	0.2	0.63	0.12	0.50	6.82	*14.17*
Total WP	*12*		*19.4*		*11.4*		*16.0*		*11.8*		*9.8*		*4.0*		*3.3*		*87.8*	

(actual man months are rounded 6-monthly best estimates; final accurate man-hours are given in the cost claims)