PROJECT PERIODIC REPORT

Grant Agreement number: FP7-ICT-2011-270253

Project acronym: INTEGRATE

Project title: Driving excellence in Integrative Cancer Research through Innovative Biomedical Infrastructures

Funding Scheme: Collaborative project

Date of latest version of Annex I against which the assessment will be made:

Periodic report: **1st** ☐ **2nd** ☐ **3rd** ■ **4th** ☐

Period covered: **from 1 February 2013 to 31 January 2014**

Name, title and organisation of the scientific representative of the project's coordinator[1]:

Name: Anca Bucur

Tel: +31 40 27 49699

Fax:

E-mail: anca.bucur@philips.com

Project website[2] address: http://pothos.ics.forth.gr/integrate/

[1] Usually the contact person of the coordinator as specified in Art. 8.1. of the Grant Agreement .

[2] The home page of the website should contain the generic European flag and the FP7 logo which are available in electronic format at the Europa website (logo of the European flag: http://europa.eu/abc/symbols/emblem/index_en.htm logo of the 7th FP: http://ec.europa.eu/research/fp7/index_en.cfm?pg=logos). The area of activity of the project should also be mentioned.

Declaration by the scientific representative of the project coordinator

I, as scientific representative of the coordinator of this project and in line with the obligations as stated in Article II.2.3 of the Grant Agreement declare that:

- The attached periodic report represents an accurate description of the work carried out in this project for this reporting period;
- The project (tick as appropriate) [3]:

  ■ has fully achieved its objectives and technical goals for the period;

  ☐ has achieved most of its objectives and technical goals for the period with relatively minor deviations.

  ☐ has failed to achieve critical objectives and/or is not at all on schedule.

- The public website, if applicable

  ■ is up to date

  ☐ is not up to date

- To my best knowledge, the financial statements which are being submitted as part of this report are in line with the actual work carried out and are consistent with the report on the resources used for the project (section 3.4) and if applicable with the certificate on financial statement.
- All beneficiaries, in particular non-profit public bodies, secondary and higher education establishments, research organisations and SMEs, have declared to have verified their legal status. Any changes have been reported under section 3.2.3 (Project Management) in accordance with Article II.3.f of the Grant Agreement.

Name of scientific representative of the Coordinator: Anca Bucur

Date: 07/05/2014

For most of the projects, the signature of this declaration could be done directly via the IT reporting tool through an adapted IT mechanism.

[3] If either of these boxes below is ticked, the report should reflect these and any remedial actions taken.

TABLE OF CONTENTS

Declaration by the scientific representative of the project coordinator 2
1 Publishable summary 4
1.1 Highlighted results 4
1.2 Expected impact 5
1.3 General information 6
2 Core of the report for the period: Project objectives, work progress and achievements, project management 7
2.1 Project objectives for the period and main overall achievements 7
2.2 Work progress and achievements during the period 11
2.2.1 WP1 – IJB 11
2.2.2 WP2 – Custodix 12
2.2.3 WP3 – UPM 13
2.2.4 WP4 – FORTH 16
2.2.5 WP5 – FORTH 18
2.2.6 WP6 – Philips 20
2.2.7 WP7 – BIG 22
3 Achievements per individual partner 23
4 Project management during the period 27
4.1 Consortium Management 27
4.2 Changes in the consortium 27
4.3 Cooperation 27
4.4 Project Meetings 27
4.5 Dissemination activities 29
4.5.1 Contributions to conferences 29
4.5.2 International articles 29
4.5.3 Presentations 30
4.5.4 Project web-site 30
5 Deliverables and milestones tables 33

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-genomic 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.

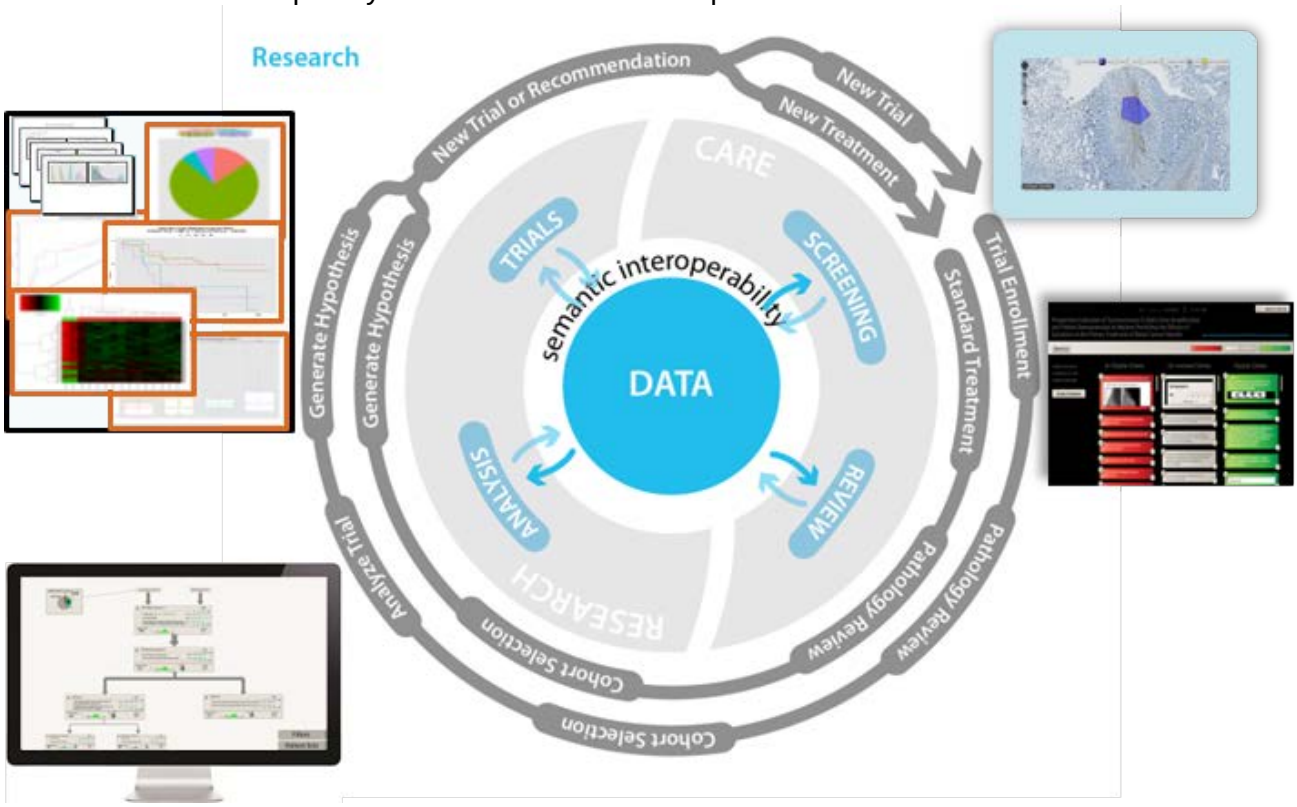
An important objective of this project is to build tools that facilitate efficient the execution of post-genomic multi-centric clinical trials in breast cancer. A range of such tools aim to support recruitment through the automatic evaluation of the eligibility of patients for trials based on matching 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.

1.1 *Highlighted results*

The INTEGRATE project provides a data environment underpinned by a standards-based semantic interoperability solution enabling the integration of both care and research data, and a range of tools for clinical research and for streamlining the screening phase of clinical trials. An

important objective of the project is to build tools that support the efficient execution of post-genomic multi-centric clinical trials in breast cancer. The Figure below depicts the main concept of the project: Closing the loop between clinical research and care in oncology. The key applications that have been developed by INTEGRATE are also depicted.



1.2 *Expected impact*

Our vision is to drive research excellence in oncology through a unique accessible biomedical infrastructure integrating diverse mega-datasets, building predictive bionetworks and offering advanced tools to guide the development of effective human therapeutics and diagnostics. These comprehensive datasets will also become available to the biomedical research community through the INTEGRATE infrastructure.

Towards personalized medicine: support for molecular screening

We are quickly moving towards an era of personalized medicine in breast cancer, with the ultimate goal of making tumour-specific “molecular fingerprints” possible. This fingerprint would consist of distinct genetic markers obtained from a simple blood draw or tumour sample, and it would allow the physician to refine a patient’s prognosis and select the best possible therapeutic options, maximizing response and minimizing toxicity. Because these distinct genetic markers are present in relatively small sub-groups of patients, the realization of this goal requires the implementation of smaller and smarter molecularly-defined clinical trials. The Breast International Group (BIG) has recognized this essential need, both for academic and pharmaceutical research, and the subsequent necessity for a molecular screening structure to support it. This platform will ultimately facilitate the efficient development of new molecules and help overcome the current hurdles of biomarker discovery.

The need for data sharing and integration

At the centre of INTEGRATE is an environment bringing together clinical, genomic, pathology and radiology imaging data, originating from multiple oncology clinical trials. Researchers will be able to select subsets of patients from the INTEGRATE repository through sophisticated queries and retrieve their data. By accessing data from multiple trials, researchers will be able to build predictive models, identify biomarkers and answer other research questions faster and with more confidence. Finally, fine-grained access control for differential access to subsets of the data by different user groups will enable flexible patterns of collaboration. But sharing of raw, unprocessed data is not sufficient. The lack of standardised medical terminology poses another challenge for the integration of data from multiple trials. Often, the same concept, such as a cancer subtype, a gene, or a medical condition, will be referenced in different ways in different studies, making meta-analyses very difficult. Thus, an important part of INTEGRATE is the identification of a core data set, i.e. a set of concepts that covers the subject domain of breast cancer clinical trials. These core concepts are then mapped by a team of information specialists and oncologists to controlled terminologies and ontologies such as SNOMEDCT for clinical terms, LOINC for laboratory and clinical observations, and MedDRA for drug safety data. INTEGRATE also extends controlled terminologies and ontologies when critical concepts in the field of breast cancer clinical trials are missing.

1.3 *General information*

General Info	
Acronym	INTEGRATE
Name	Driving excellence in Integrative Cancer Research through Innovative Biomedical Infrastructures
Web page	www.fp7-integrate.eu
Reference	FP7-ICT-2009-6-270253

2 Core of the report for the period: Project objectives, work progress and achievements, project management

2.1 *Project objectives for the period and main overall achievements*

The main project objectives for this period as described in the DoW have been to:

- Finalize the system architecture and the security framework
- Report on the predictive modeling framework and on the predictive models for therapy response
- Organize the first INTEGRATE workshop
- Prepare the evaluation and validation of the INTEGRATE solutions
- Extend the semantic solution
- Refine the project tools and integrate them with the semantic and security infrastructure
- Disseminate the project results through the production of a newsletter, publications, presentations and participation to events

During this reporting period we have achieved the following key objectives:

- 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
- Complete the implementation of the INTEGRATE tools to have them ready for evaluation and validation.
- Prepare and carry out the first INTEGRATE workshop at the European Cancer Conference.
- Define the evaluation and validation scenarios and the detailed evaluation metrics for all tools developed by the project
- Prepare the necessary clinical data to be used for the validation and evaluation of tools and to create all the necessary conditions for proper validation of the tools
- Extend the semantic interoperability solution and the Common Information Model
- Explore directions for ensuring the sustainability of the project and to facilitate future exploitation
- Disseminate the project results in the research and clinical community
- Prepare the INTEGRATE evaluation and validation workshop (scheduled for 13-14 June 2014) with expert oncologists, pathologists and bioinformaticians from the EU and beyond that are outside of the consortium and did not participate in the implementation of the project (this way having a fresh look at what we have achieved and evaluating the value and relevance for a large audience)
- Support the quick integration in the consortium of the new partner, the German Breast Group

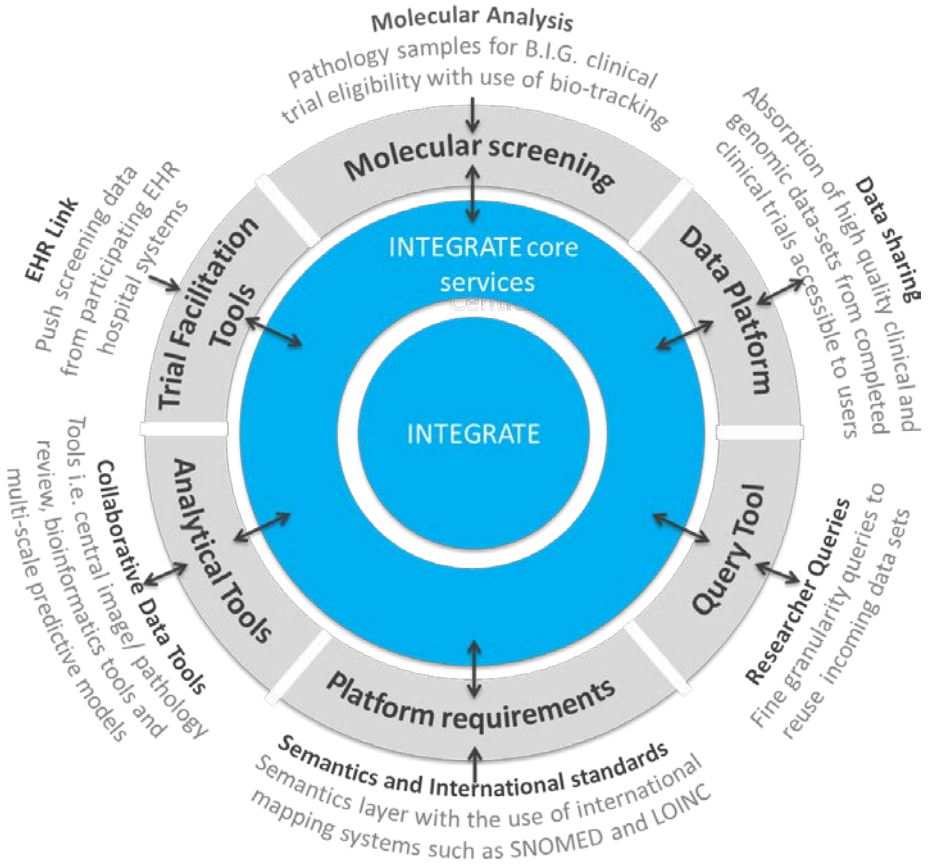
<u>Implementation effort</u>
We have focused on further implementation of the INTEGRATE tools (in the areas as described in the figure below), on the extension of semantic solution and of the INTEGRATE Common Information Model. New data was transformed through the defined ETL pipeline and uploaded to the Common Data Model. This includes large datasets from the new partner GBG and data transformation on site for the MAASTRO clinic for the validation of Patient Screening.

Evaluation and validation

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 was the topic of an additional deliverable that the consortium prepared before this review (became contractual deliverable after the amendment).

Dissemination and Exploitation

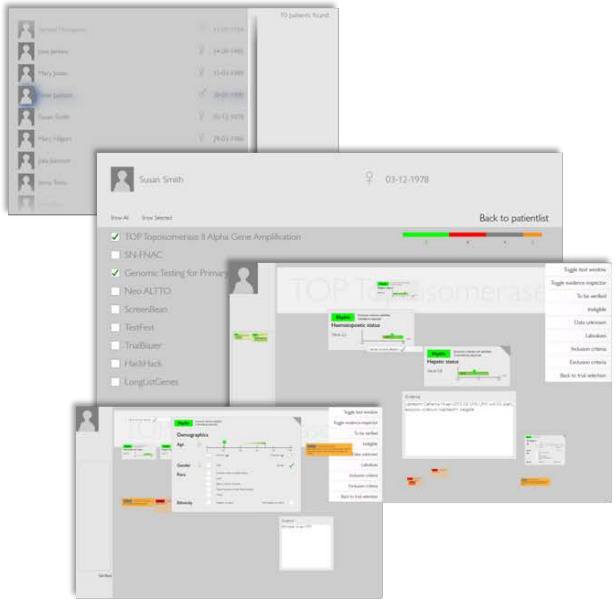
The elaboration of a sustainability plan was also an ongoing focus of this period. We strongly believe that the success of the INTEGRATE environment and semantic solution depends on our ability to collaborate with other similar initiatives, to harmonize the various solutions and to converge. With respect to our tools, their success depends on their ability to address real needs of the user community. To this end, thorough evaluation and validation is essential for the project and so is the wide dissemination of our results. To this end we have proposed an extension of the project that was approved.



Preparing the tools for validation and updates based on user feedback

Based on initial user feedback and to prepare for validation the INTEGRATE tools have undergone significant transformation that included not only the backend services but also the user interaction and the applications' UIs. An example for the patient screening application is presented in the Figure below.

Before:



After:

Recommendations from previous reviews (if applicable)

The independent experts 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 EMCC conference in September.

We implemented 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, bioinformaticians for the analysis tools). This workshop will take place on June 13-14 in Heraklion and will bring together clinical users that were not involved in the project.

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, the MAASTRO radiation oncology clinic in Maastricht, the Netherlands was selected and agreements were made for the validation of the patient screening application in that hospital. Extensions to the application were made to suit the concrete workflow of this hospital and data was transformed on site to be used for the evaluation. Currently the hospital has dedicated personnel evaluating the eligibility of each new patient for the available clinical trials; they have collected detailed information on the effort required for the manual evaluation. The scenario defined with MAASTRO is “man versus man with machine” and aims to evaluate for the same dataset the performance of the tool and the time that could be saved by the clinical experts.

To demonstrate the international cooperation dimension we continue the collaboration and alignment efforts with Sage Bionetworks, a prominent data sharing initiative in the US. We have obtained the approval to use one of their large datasets in INTEGRATE for the evaluation of the analysis tools.

The next review will take a WP-focused approach as recommended and includes 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. At the same time, technical presentations and demonstrations will be provided.

Each demonstrator will be clearly linked to the contributing WPs.

This progress reports contains more details on the clinical aspects of the project and the achieved progress. In the review the presentation discussing the clinical needs and trends will link directly to the various solutions developed in the project.

We evaluated the use of BPMN2.0 to model workflow and of the corresponding Drools framework in the INTEGRATE project. While not directly applicable to the current tools, this comment was very valuable to us as after this evaluation we selected this platform for the implementation of workflow applications.

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.

INTEGRATE Symposium on the Potential of Data Sharing

To disseminate results of the INTEGRATE project within the wider research community a couple of events have been foreseen throughout the project duration. These events are likely to take place at international (breast) cancer conferences where a large number of stakeholders are present.

A first event focusing on the relevance of data sharing concepts in oncology took place in Amsterdam, on Friday 27 September 2013 at the RAI congress centre in Amsterdam, the Netherlands, during the 2013 European Multidisciplinary Cancer Congress. Over the years, this congress has become the reference oncology meeting in Europe and beyond, providing practice-changing, high quality multidisciplinary presentations, thus representing a unique opportunity to present INTEGRATE and its solutions to a wide range of potential end-users.

This event, free of charge for the participants, took the form of a 2-hour symposium and attracted about 50 oncologists, researchers, and other potential end-users.

The symposium focused on the relevance of data sharing – and the multiple challenges it entails – in oncology. The speakers' thought-provoking talks were well received and stimulated considerable discussion.

The programme included the following presentations:

- Introduction: Clinical data - the lifeblood of clinical research – David Cameron, MD (Clinical Director and Chair of Oncology, Edinburgh Cancer Research Centre, and Professor, University of Edinburgh, UK)
- Keynote: Challenges, semantics and standards – Eric D Perakslis, PhD (Executive Director, Center for Biomedical Informatics, Harvard Medical School, USA)
- The INTEGRATE project – Anca Bucur, PhD (Philips Research, Eindhoven, The Netherlands)
- Legal and ethical aspects – Nikolaus Forgó, PhD (Professor of Law, University of Hannover, Germany)
- Security issues – Brecht Claerhout (Chief Executive Officer, Custodix N.V., St-Martens-Latem, Belgium)
- Conclusion and Q&A – David Cameron, MD and Anca Bucur, PhD

2.2 *Work progress and achievements during the period*

2.2.1 WP1 – 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.

2.2.2 WP2 – Custodix

2.2.2.1 Objectives (of the reporting period)

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

2.2.2.2 Status/progress towards objectives WP2 (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 final iteration of the architecture was finished and document in deliverable 2.7
- 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
- Work has been started for the implementation, integration and deployment of the demonstrators of year 3
- A poster about the cohort selection engine has been designed and 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
- The authorisation framework was designed and implemented. The resulting authorisation services will be integrated in the demonstrators of year 3.
- The auditing framework was specified and an initial implementation of the auditing services was done.

Task 2.5: Component integration and interfacing with external systems

- Integration guidelines were specified in Deliverable 2.5
- WP2 coordinated the integration of the different services in INTEGRATE

2.2.2.3 Deviations from the DOW and corrective actions

- Deliverable D2.6 was delayed to March 2013

2.2.2.4 Planning next period

- Final specification, integration and deployment of the security framework services, focusing on authorisation and auditing.
- Further integration of the security environment in the demonstrators of year 3.
- Integration and deployment of the services of the demonstrators of year 3.
- Provide input for the validation of the INTEGRATE tools.

2.2.3 WP3 – UPM

2.2.3.1 Objectives (of the reporting period)

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.

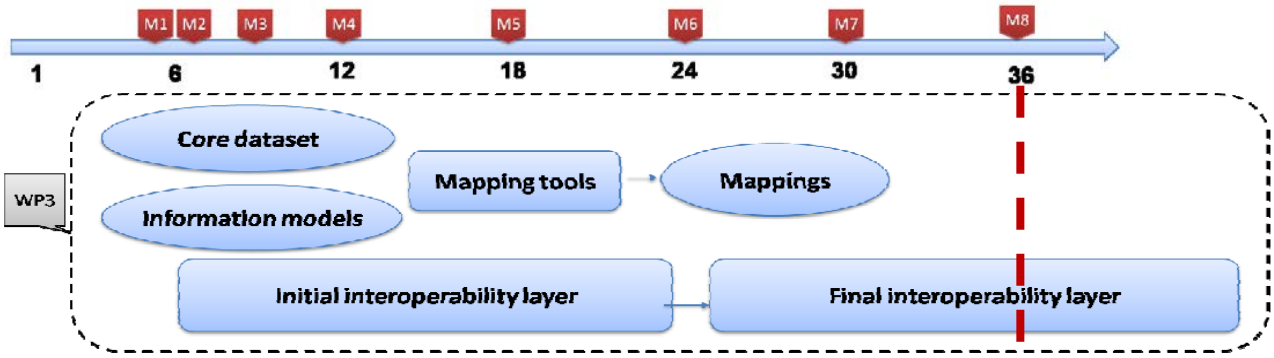


Fig A. Work Package 3 components and planning

From month 25 to month 36, 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. With the project extension of 9 months, an additional effort within WP3, and concretely Task 3.3 (mappings), is required to validate the proposed solution with additional data sources.

2.2.3.2 Status/progress towards objectives WP3 (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) since SNOMED CT did not include proper codes. 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 has been 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.

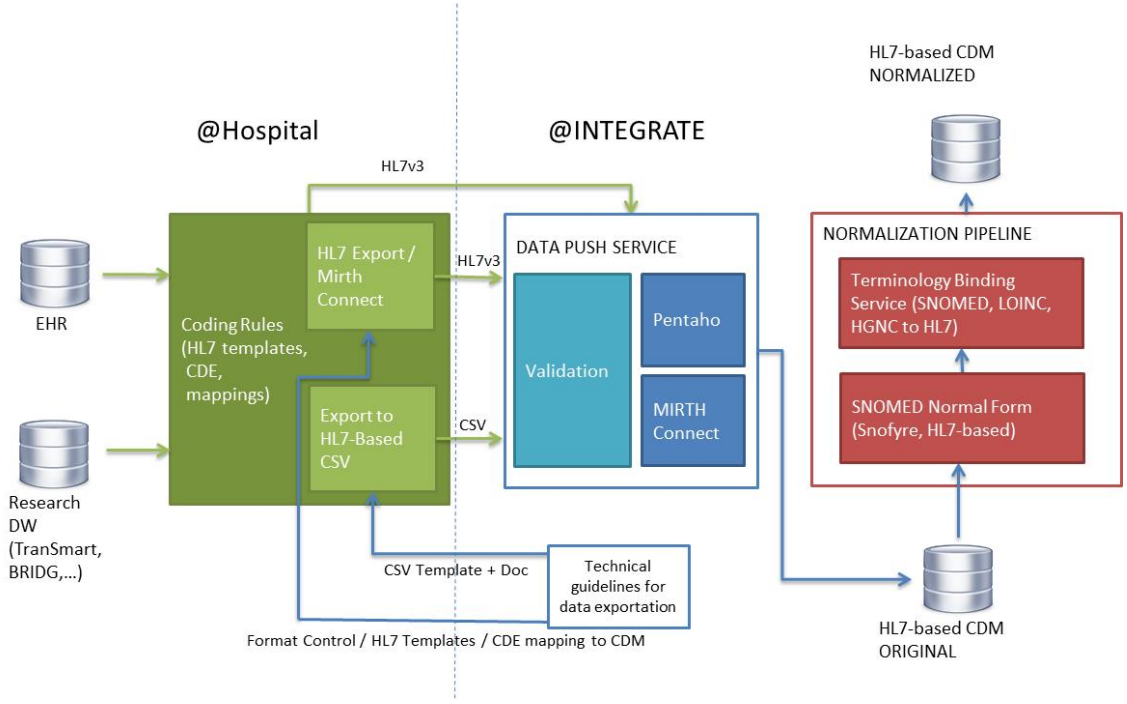


Fig. B. Deployment of the INTEGRATE platform in a new institution

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, has been also 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.

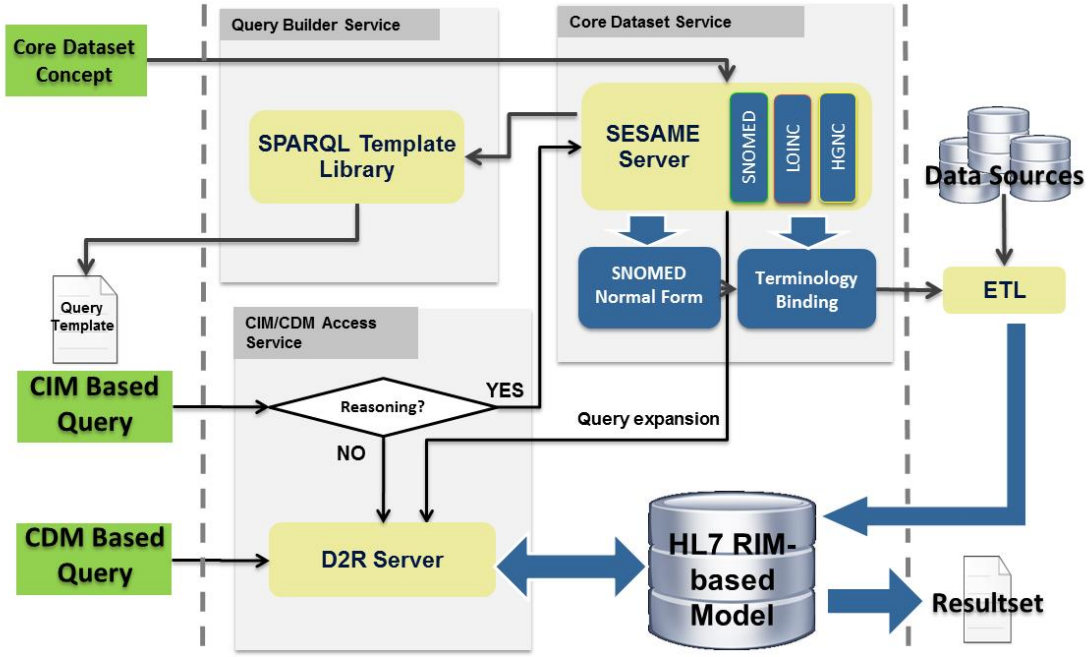


Fig. C. Query builder interaction with the rest of the components of the semantic interoperability layer

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.

2.2.3.3 Deviations from the DOW and corrective actions

There are not significant deviations from the updated version of the DoW (including a 9 month extension). WP3 have been mainly focused on Task 3.3 and 3.4 during months 25 to 36.

2.2.3.4 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.

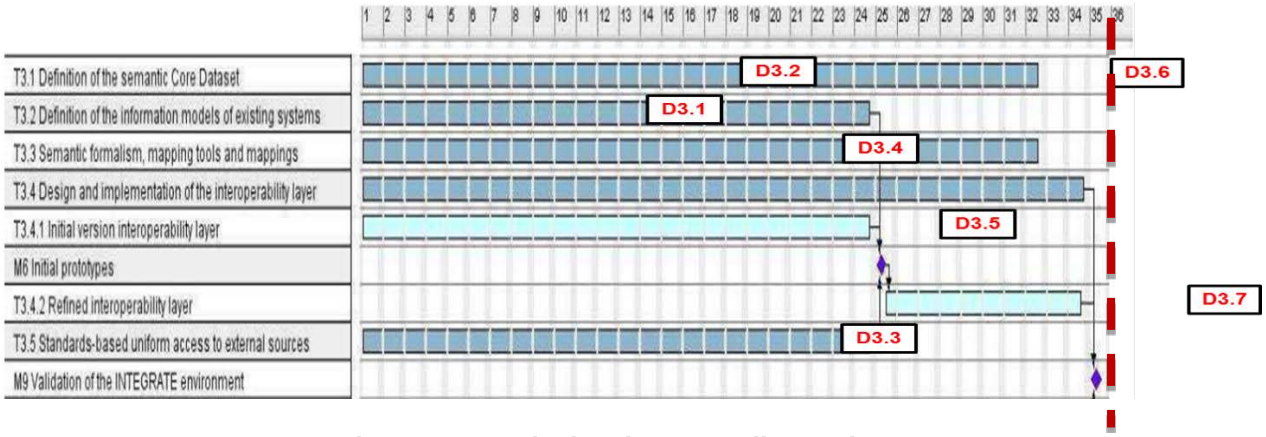


Fig. D. WP3 task planning according to the DoW

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.

2.2.4 WP4 – FORTH

For the 3rd year, the main effort in work package 4 had focused in improving the design, the functionality and the stability of the central review platform for digital pathology images. The platform allows central reviewing of digital pathology images among many reviewers, directly from the web and in the same time functions as a virtual microscope for the stored digital pathology slides.
In the section below there is a detailed description for all the activities related to the platform of central review for pathology (CRP) images.

2.2.4.1 Objectives (of the reporting period)

The following objectives had been set and fulfilled for the reporting period:

1. Updated version of the virtual collaboratory and of its services. In detail:
   a. The tiles generator of the platform has been improved, and in its current version uses a queuing stack to schedule and process automatically the uploaded digital pathology images.
   b. Connect to the Common Data Warehouse (CDW) in order to associate the pathology images which should be reviewed with the appropriate schema
   c. Connect to the Common Data Warehouse (CDW) in order to publish back the produced reports from the reviewing protocols.
   d. Added support for annotations which can include freeform shapes, predefined shapes and text for every image.
   e. Finalized CPR main workflow for the review process and implemented the corresponding functionality
   f. Added administrative utilities & tools for the management of open/closed review protocols
   g. Enhanced CPR notification center to utilize a complete inbox like solution.
2. A tool for uploading digital pathology images to the platform was developed, tightly coupled and accessible directly through the platform's web interface. The user roles of the platform have been adjusted to allow only authorized users to upload images.

3. Added further configuration options to the platform through its web interface, eliminating the need to manually edit configuration files.
4. The generation mechanism of the dynamical reports of the platform has been adjusted in order to fit exactly to the special properties of the digital pathology images, while preserving the ability to serve any kind of medical image type with almost no modifications.

2.2.4.2 Status/progress towards objectives

Task 4.1: Model, data and annotation repositories

According to this task the Central Review Platform (CRP) is using secured SOAP services for associating imaging data uploaded to the platform with the corresponding metadata in the repository.
A moderator[4] can then define a dynamic model perfectly adapted to the current needs, and then create Reviewing Protocols in which everything is controlled (images to be reviewed, reviewers, features of interest, etc.) and described (any kind of information can be used in order to complete a reviewing protocol as rich text, and URLs to documentation or other fountains of information). Additionally each time a reviewer submits any findings to the platform (in respect to a specific review protocol and patient's image) the answer is securely transmitted (secure SOAP request) and stored (in an XML format) in the CDW.

Task 4.2: Tools enabling data and knowledge sharing

According to the task the CRP provides tools enabling the clinical research community of BIG to collaboratively define research protocols and carry out all the necessary regulatory and administrative steps to carry on the success of a clinical trial.
The CRP provides wizard driven procedures and intuitive interfaces in order for the clinicians to successfully identify which patients are eligible to be included in the trial, resulting in rapid reviewing process among multiple reviewers. These procedures & interfaces were significantly altered to reflect in a more convenient way the actual patients' selection process. As a result, the process of review protocol registration is more straightforward for the end user.
The protocols are dynamically defined by the administrator/moderator of the platform, using secure mechanisms to share data among the appropriate reviewers.
The protocol definition process is enhanced with user input validation mechanisms so as to guide moderators efficiently in protocol type definition as also to prevent inconsistencies. Moreover the platform was pre-configured with a convenient set of protocol base types.

Task 4.3: Tools enabling collaboration

The concept of a centralized reviewing platform is based upon the spirit of collaboration among the reviewers who are called to review a pathology image. A review is characterized as successful only when there are no conflicts among the results of the reviewers. Thus the CRP provides means to resolve conflicts upon sharing the information available among the appropriate reviewers and by providing a seamless interface which assists the reviewers to isolate any issues appearing during a reviewing protocol. The collaboration features of the CRP include automatic notifications send by the platform, reminders and discussion like procedures until a conflict is resolved and the final status of an image under review is set.
In respect to the list submitted to previous 'Project Periodic Report' the following items has been added or altered: A brief listing of the collaboration tools provided by the CRP is:

1. Enhanced Wizard driven interface for the creation of a reviewing protocol.

[4] The CRP supports two distinct roles: moderators and reviewers.

2. Implementation of the image annotator functionality incorporated into the virtual microscope
3. Centralized, wizard like view, for protocol management which includes options for:
   a. Conflict Resolution
   b. Management of tasks per Image
   c. Management of Images' statuses per Protocol
   d. Management of Review Protocols
   e. Send reminders for overdue tasks
   f. Archive/Delete closed protocols
4. Implementation of a complete, inbox like, solution for the Notification Center
5. Implementation of the Image Uploader component which requires the definition of an additional role
6. Enhancements and finalization of the Image Tiller component
7. Additional communication services implemented (secured SOAP) for reviewers' answers storage to the CDW
8. Minor improvements in several points, such as:
   a. User input validation wherever there is such an option
   b. Improvements in user messaging
   c. Bug fixes

Task 4.4: Privacy Enhancing Processes and Services

The CRP implements SSO functionality, for the creation and the authentication of the users. All the data pushed to the platform are anonymized, while the reviewers have no access to any crucial information (regarding the identity of the patient).
A new, enhanced SSO component was deployed, provided by Custodix. Also some enhancements regarding security were implemented wherever there is communication with external components over SOAP requests.

2.2.4.3 Deviations from the DOW and corrective actions

No deviations.

2.2.4.4 Planning next period

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

- Complete the prototype of the CRP, in order to provide all the necessary functionality described in the DOW.
- Evaluate the performance and the usability of the platform in close collaboration with the clinical partners.

2.2.5 WP5 – FORTH

WP5 has focused on providing users with a collaborative, multi-functional and easy-to-use environment for exploiting, analyzing and assessing the quality of large multi-level data. Its main goal is to empower clinicians and researchers to analyze with ease clinic-genomic data in order to get simple statistics on selected parameters, perform survival analyses, compare regiments in selected trials, obtain genomic analysis results, and develop powerful multi-scale models for predicting drug response and assessing candidate biomarkers in cohorts of patients. The platform is also coupled with a security framework for enabling user authentication and authorization, a set of services that facilitate the process of loading and retrieving data from the Common Data Model (CDM) and a set of service for the visualization, storage and modification of the metadata analysis information is also allowed.

2.2.5.1 Objectives (of the reporting period)

Our effort for this reporting period has been generally directed to the following aspects:

- Change and improve the overall layout of the INTEGRATE Analysis Platform using Java Server Faces (JSF) component library (http://primefaces.org).
- Develop a web-service for enabling the interaction between the Cohort Selection tool and the Analysis platform.
- Request and gain access to SAGE data (see 2nd project periodic report for further information about the data) in order to be used for both demonstrative and evaluation purposes from the INTEGRATE Analysis Platform.
- Extend the functionality of the platform allowing users to import different datasets for analysis.
- Finalize and embed the predictive models (both Basic and Advanced) to the Analysis platform.
- Assess the performance of the predictive models using SAGE data.
- Install the latest version of the SSO extension, enabling user authentication and authorization.
- Prepare the evaluation and validation phase of the Analysis platform. Specify the evaluation/validation scenarios and create the questionnaires for the end-users and the development team according to software evaluation/validation ISO standards.

2.2.5.2 Status/progress towards objectives

Task 1.1 Definition of clinical scenario (questions) for the INTEGRATE VPH use case
No further work is needed.

Task 1.2 Definition of genetic and imaging biomarkers and of a modelling methodology
No further work is needed.

Task 1.3 Development of predictive models of response to therapy and of the modelling framework
According to the objectives outlined above, changing the layout of the platform was highly prioritized. Data tables have been replaced by widgets and data display controls, providing a more flexible, user-friendly and easily configurable deployment platform. Summarizing, the working progress, related to the development of the platform, is depicted in the following list:

- The Basic and Advanced Predictive models have been finalized and embedded to the Analysis platform.
- The table-based interface (for both analytical and predictive tools) has been replaced by a wizard-based one, assisting user in following a specific pipeline process in order to perform statistical and predictive analyses. This pipeline consists of the following steps: select a database, build a cohort, select scenarios/models, and view the results.
- The cohort selection process allows the user to constrain a request by obtaining subpopulations and build cohorts based on specific ranges of values, categories, etc.
- The scenario selection process allows user to select multiple statistical analysis scenarios or predictive models for execution in a single step.
- Controls and warnings have been added to the platform, assisting users in selecting only the eligible variables for each tool and/or model.
- The overall analysis workflow is presented in a diagram format. Every component of the diagram is functional and when clicked presents the information that corresponds to it. For example when the component that corresponds to the cohort selection process is pressed, a pop-up window appears and the generated cohort is displayed in a table format. If the

user wants to see the results of a specific analysis, then each corresponding component, related to a specific analysis, is connected to the on-the-fly generated report and the full analysis report is displayed in a pdf format.
- Optimize the software of the tools and models. The analysis (R service and Latex service) is performed remotely by a server with high specifications, contributing to the performance and enhancing the seamless user's work.
- The statistical analysis framework has been extended to allow users importing different datasets for analysis (SAGE dataset is also available for statistical analysis).
- Install the latest version of the SSO extension.

A web-service has been developed for integrating the Cohort Selection with the INTEGRATE Analysis Platform. The Cohort Selection tool sends its request via the web-service to the Analysis platform. This request includes all the necessary information for performing an analysis, i.e., selected cohort, examined variable(s) and type of the analysis. The Analysis platform performs the analysis and sends the results back to the Cohort Selection tool. Our request for accessing and using SAGE data for evaluation and demonstration purposes of the INTEGRATE Analysis Platform has been approved and we can now use SAGE for both statistical and predictive analysis.

2.2.5.3 Deviations from the DOW and corrective actions

No deviations.

2.2.5.4 Planning next period

The planning of activities for the upcoming period is mainly focused on the following fields:
- Perform the validation and evaluation of the Analysis platform.
- Refinement of the Analysis platform according to the outputs of the above procedures.
- Paper/Publications

2.2.6 WP6 – 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 are captured in the additional deliverable D6.5 Scenarios for validation and detailed metrics for the validation of tools.

Several meetings were organized with the clinical partners (including the new partner GBG) and the external clinical site interested in evaluation of our tools to agree on the validation setup and scenarios.

2.2.6.1 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.

2.2.6.2 Status/progress towards objectives

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 are described in deliverable 6.5.

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

Task 6.3 Coordinate specifications of test scenarios and of demonstrators

During this reporting period we have defined 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 defined 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.

2.2.6.3 Deviations from the DOW and corrective actions

No deviations.

2.2.6.4 Planning next period

In the next reporting period the role of this workpackage will be to coordinate, execute and report on the evaluation and validation of the INTEGRATE solutions.

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 required an extended timeline for the project.

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.

2.2.7 WP7 – BIG

2.2.7.1 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.

2.2.7.2 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 European Cancer Congress 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 participated to the mini-symposium. We have also organized all the practical details of this event.

The speakers and the titles of the talks in this mini-symposium were as follows:

- Clinical data - the lifeblood of clinical research By David Cameron, MD Prof. – University of Edinburgh, UK / BIG EB Member
- Keynote: Challenges, semantics and standards By Eric D Perakslis, PhD – Harvard Medical School, USA
- The INTEGRATE project By Anca Bucur, PhD – Philips Research, Eindhoven, The Netherlands
- Legal and ethical aspects By Nikolaus Forgó, PhD Prof. – University of Hannover, Germany
- Security and privacy issues By Brecht Claerhout – Custodix N.V., St-Martens-Latem, Belgium
- Conclusions and Q&ABy David Cameron, MD and Anca Bucur, PhD

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. Following on these discussions, the German Breast Group joined the project in February 2014.

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.

2.2.7.3 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.

2.2.7.4 Planning next period

During the next period, dissemination activities will continue (newsletter, scientific publications…). The main event is the INTEGRATE evaluation and validation workshop that will take place in June 2014. We will also participate with demonstrators and presentations in conferences and workshops, present articles in conferences and prepare additional submissions.

To support both dissemination and further sustainability we will organize a hands-on scientific workshop with related interoperability projects to evaluate the possibilities for harmonizing our solutions and to plan future collaboration.

Exploitation activities will continue through the newly identified pilot hospitals, and through the identification of additional interested stakeholders.

3 Achievements per individual partner

Partner 1 Philips
The main focus in this reporting period has been on completing the implementation of the INTEGRATE tools and on preparing the validation of these tools.

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

  Implementation work on the cohort selection prototype Nona has been completed and the tool is ready for evaluation. 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 Patient Screening tool (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. The tool has been further updated in January/February 2014 following additional feedback from users.

- 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 and to the definition of the validation scenarios and metrics.
- Presented the INTEGRATE project in the ENBC 2013 Conference and in the 2nd Summer School on Computational Oncology.
- We contributed to the preparation of the evaluation and validation of tools with the new partner GBG.
- We worked on the data transformation/ETL/extraction and annotation with SNOMED-CT of concepts out of free text (in Dutch) for the validation of the Patient screening application (Decima) at the MAASTRO Clinic in Maastricht (clinical organization that is not part of the consortium but is interested to evaluate this application).
- Together with MAASTRO we defined the evaluation metrics and scenario ("man versus man with machine") for the evaluation & validation of the Patient Screening application at MAASTRO.
- The Trial Metadata Repository was further extended with relevant metadata and populated with the descriptions of the trials from GBG and MAASTRO that will be used for the validation of the Patient Screening tool.
- Contributed to several successful paper submissions to international conferences (BIBE2013, HEALTHINF 2014, MIE 2014)

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

In the reporting period, and with respect to WP4; FORTH worked to finalize the platform for remote review of digital pathology images. The platform was updated thanks to the collaboration with our partners from BIG, and based on the feedback it was optimized in areas such as in accessibility and in its functionality as a virtual microscope. Finally enhancements have been made to the security of the platform (the SSO user authentication has been implemented as well as support for anonymized data) and also in the communication with the services among the rest of the partners.

In WP5, FORTH has focused on improving the overall layout of the Analysis platform, moving from a table-based to a wizard-based interface. Also, the SAGE dataset has been integrated for statistical analyses also. Both basic and advanced predictive models have been finalized and integrated into the platform. Functionalities such as the updated Single-sign-on (SSO) authentication mechanisms and integration between the Cohort Selection tool and Analysis platform via a web-service have been incorporated through a close collaboration with all INTEGRATE partners.

Partner 4 Custodix

- Attended telco's and technical, review and consortium meetings
- Led and contributed to the final 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, integrated 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
- Presented a poster about the DSL Query engine 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 and auditing
- Contributed to and reviewed the last iteration of the technical use cases
- Contributed to the INTEGRATE newsletter
- Contributed to deliverable 4.5: Final versions of the tools and services supporting data sharing and collaboration
- Setup environments for validation and piloting of the INTEGRATE tools

Partner 5 IJB

- Report on the preparation of the deployment environment by defining verification and validation procedures as required for testing 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 valuable uses of the platform. This work was also an opportunity to identify and present the various clinical sites involved in the validation phase, to address issues related to both security requirements and data exchange, to describe the process required to incorporate new data sources to the platform and the common data model (the Extract, Transform and Load process) and finally to define 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, oncologists, 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 of 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 will also provide functionality for processing whole-genome expression arrays, gene prognostic signatures, clinical characteristics, and imaging biomarker, perfusion and diffusion images.Study and choice of a validation methodology of computer systems. This work supported by the presence of a consultant was to learn about the various validation standards (FDA, GAMP, ISO-250x0 series) and recommendations (PIC/S, GCP) and the development of validation protocols (IQ, PQ, OQ).
- Preparing of a request to the Ethics Committee Bordet on the use of structured data for the validation phase of the Integrate project. These data from records of patients who signed a generic Informed Consent Form are used as input for the validation of demonstrators installed and tested at Bordet. This covers Cohort Selection and Patient Screening tools.
- Contributions to deliverables D6.5 (Scenarios for validation and detailed metrics for the validation of tools) and D7.10 (Project newsletter).
- Reviewed deliverables: D4.5 (Final versions of the tools and services supporting data sharing and collaboration) and D2.7 (Final system architecture and security framework).

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
- Definition of the validation environment and deployment requirements

4 Project management during the period

4.1 *Consortium Management*

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 was proposed to the EC and accepted. 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. This WP prepared all the necessary documents and carried out all the steps for the implementation of the amendment.

The consortium was extended with one partner and we have prepared together with this partner all the necessary documents.

4.2 *Changes in the consortium*

The consortium was extended with a clinical partner, the German Breast Group. This partner has already started to actively contribute to the project and will be introduced in the next review.

4.3 *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 an important focus, but even higher priority was placed on preparation of the tools for their evaluation and validation. Once the timeline was agreed and the workshop has been scheduled we entered a straight line towards the validation sessions therefore focus was essential. This was a joint effort to which all project partners were committed: All partners will contribute to the review and contributed to the prototypes that will be demonstrated.

We have also jointly organized a mini-symposium in the European Cancer 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. Several papers were accepted for publication and we were invited to present our work in several events.

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

4.4 *Project Meetings*

When	What	Organising Partner or Work Package	Where
01/02/2013	Monthly Telco	Sint-Martens-Latem/Custodix	Belgium
11/02/2013	Healthinf 2013	Barcelona/Biostec	Spain
14/02/2013	Trial metadata Telco	Sint-Martens-Latem/Custodix	Belgium
01/03/2013	Monthly Telco	Sint-Martens-Latem/Custodix	Belgium
11-12/03/2013	Consortium Meeting	Groenendael/Philips	The Netherlands
20/03/2013	Convergence Workshop	Brussels/EC	Belgium
27/03/2013	Cohort Selection GUI Telco	Sint-Martens-Latem/Custodix	Belgium
27/03/2013	Cohort Queries Telco	Sint-Martens-Latem/Custodix	Belgium
24-25/04/2013	Review Preparation Meeting	Brussels/EC	Belgium
05/04/2013	Monthly Telco	Sint-Martens-Latem/Custodix	Belgium
26/04/2013	Review Meeting	Brussels/EC	Belgium
07/06/2013	Monthly Telco	Sint-Martens-Latem/Custodix	Belgium
11-12/06/2013	Consortium Meeting	Ghent/Custodix	Belgium
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 preparation	Sint-Martens-Latem/Custodix	Belgium
23/07/2013	Seminars on the INTEGRATE project	Maastricht/MAASTRO	The Netherlands
02/08/2013	Monthly Telco	Sint-Martens-Latem/Custodix	Belgium
04-06/09/2013	Consortium Meeting	Heraklion/FORTH	Greece
17/09/2013	Cohort Dataset Telco	Sint-Martens-Latem/Custodix	Belgium
27/09/2013	ECCO Conference 2013	Amsterdam/ECCO	The Netherlands
11/10/2013	Autocomplete Service Telco	Sint-Martens-Latem/Custodix	Belgium
29/10/2013	Pilot Meeting	Maastricht/MAASTRO	The Netherlands
31/10/2013	Exploitation preparation Meeting	Eindhoven/Philips	The Netherlands
16-20/11/2013	AMIA 2013	Washington/AMIA	The United States of America
22-23/01/2014	Consortium Meeting	Groenendael/Philips	The Netherlands
26/3/2014	Validation preparation	Frankfurt/ GBG	Germany

Table 1 – Project Management Team Meetings

4.5 *Dissemination activities*

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 at the Integrate EU Project. HEALTHINF 2013: 34-41
- Brecht Claerhout, Kristof De Schepper, David Pérez del Rey and Anca Bucur: Contextualisation of ABAC Attributes through a Generic XACML Functionality Extension Mechanism. HEALTHINF 2013: 52-57
- 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.
- J. van Leeuwen, A. Bucur, B. Claerhout, K. De Schepper, D. Perez-Rey and R. Alonso-Calvo. BRIDG-based Trial Metadata Repository - Need for Standardized Machine Interpretable Trial Descriptions (HEALTHINF 2014)

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.
- Patient Screening Application to Identify Suitable Clinical Trials. A. Bucur, J. van Leeuwen, N-Z. Chen, B. Claerhout, K. de Schepper, D. Perez-Rey, R. Alonso-Calvo, K. Saini. Submitted to AMIA 2014.

4.5.1 Contributions to conferences

- (POSTER) Leveraging Dynamic Programming Languages for Efficient Implementation of a Patient Cohort Selection Engine (presented at AMIA 2013)
- 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.
- 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.
- Supporting Patient Screening to Identify Suitable Clinical Trials. Anca BUCUR, Jasper VAN LEEUWEN, Njin-Zu CHEN, Brecht CLAERHOUT, Kristof DE SCHEPPER, David PEREZ-REY, Raul ALONSO-CALVO, Lina PUGLIANO and Kamal SAINI. Accepted for presentation at MIE 2014.

4.5.2 International articles

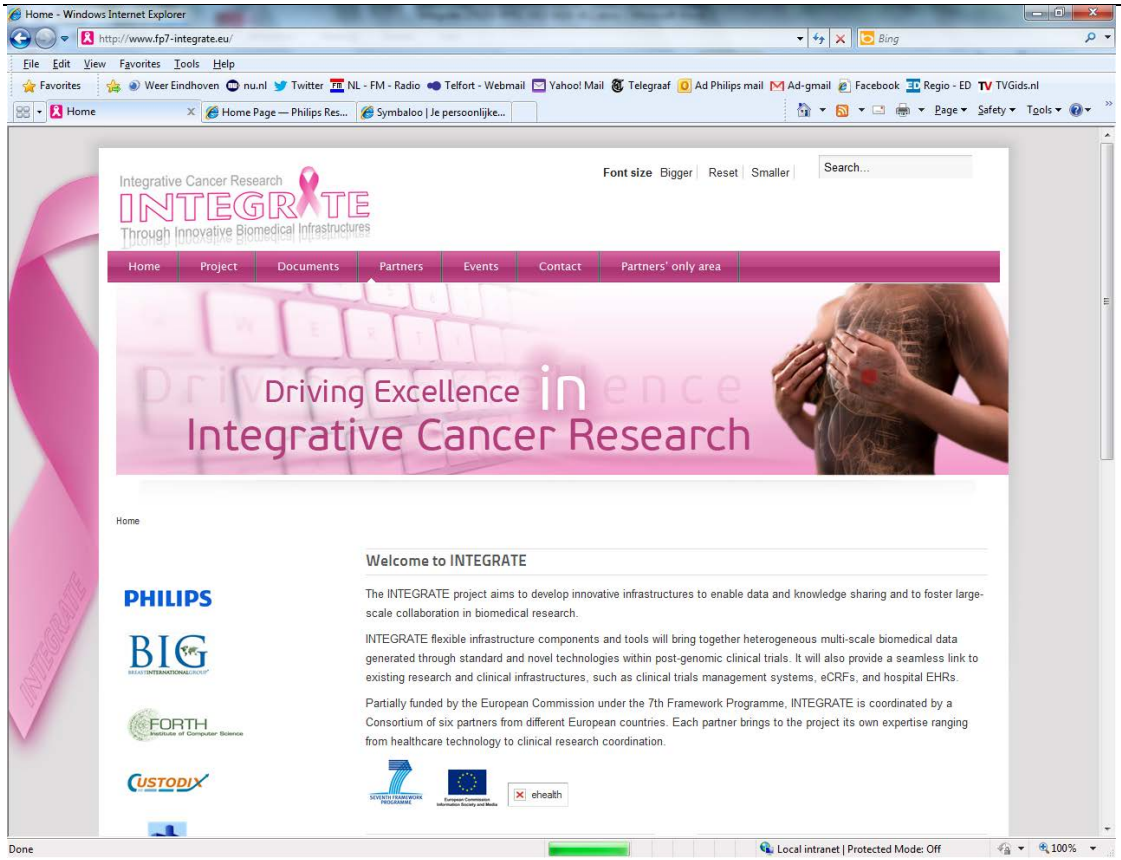
- Supporting Contextualisation of ABAC Attributes Through a Generic XACML Request Handling Mechanism (Springer-Verlag)

4.5.3 Presentations

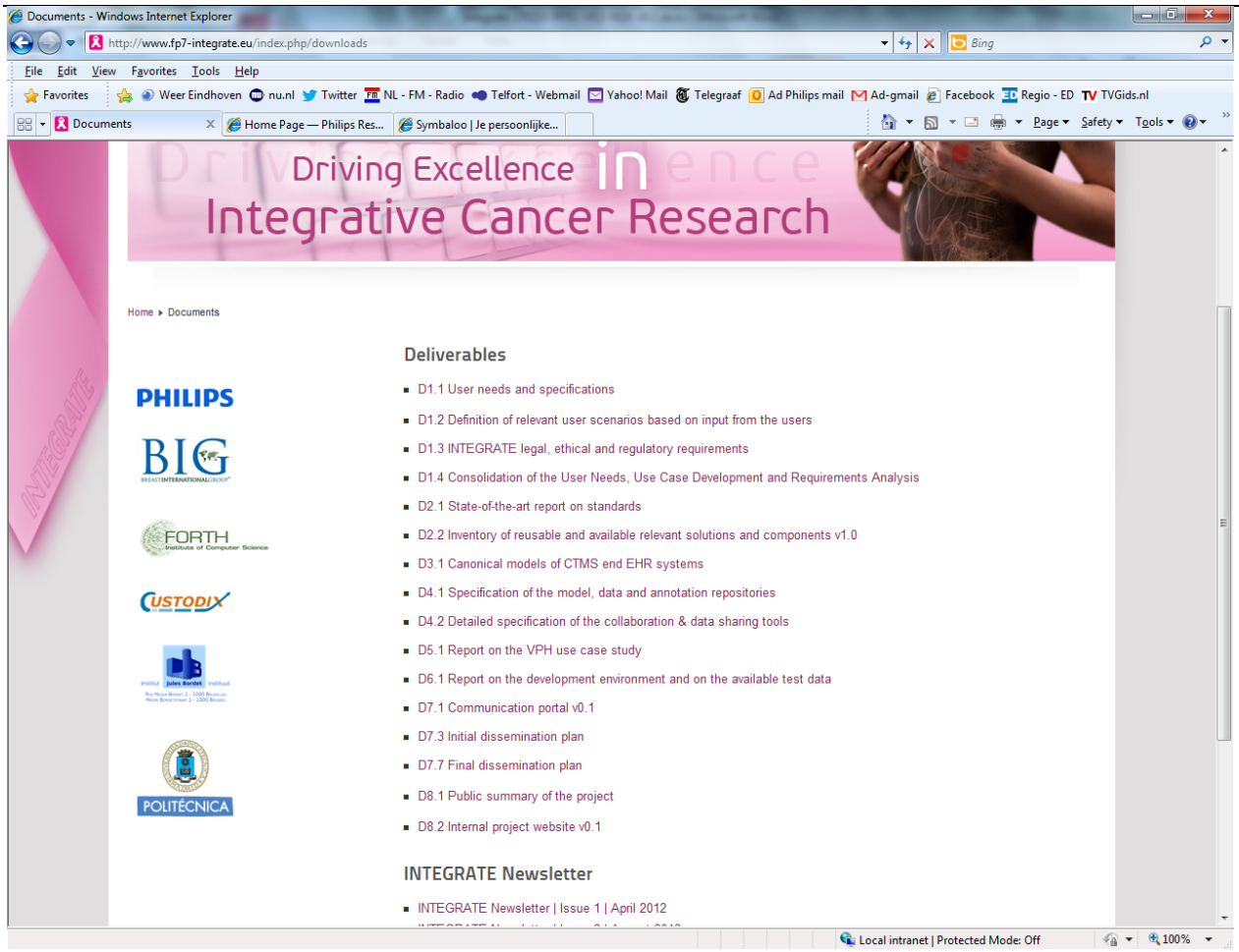
When	Where	Presentation Title	Audience	Presenting Partner(s)
11-13/02/2013	HEALTHINF 2013 Conference	Several presentations	EU	Custodix/UPM/Philips
3-7/7/2013	EMBS 2013	INTEGRATE and EURECA: Addressing semantic interoperability in healthcare	Worldwide	Philips
27/09/2013	INTEGRATE event at EMCC2013	Several presentations	EU	All
16-20/11/2013	AMIA 2013 Annual Symposium	Several presentations	Worldwide	Custodix/UPM/Philips/IJB
24/06/2013	2nd Summer School in Computational Oncology	Several presentations	EU	Custodix/UPM/Philips
18-21/12/2013	IEEE BIBE 2013	Several presentations	Worldwide	Custodix/UPM/Philips
3-6/3/2014	HEALTHINF 2014	BRIDG-based Trial Metadata Repository - Need for Standardized Machine Interpretable Trial Descriptions	Worldwide	Philips
23/7/2013	MAASTRO Clinic	Several presentations and demos	EU	All

4.5.4 Project web-site

This is a screenshot of the public website of the Integrate project, which can be found at with the following link: http://www.fp7-integrate.eu/



Referring to the webpage where all our public documents are placed:
http://www.fp7-integrate.eu/index.php/downloads



5 Deliverables and milestones tables

TABLE 1. DELIVERABLES[5]

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
7.10	Project newsletter	7	BIG	R	PU	30	Yes	July/August	
5.3	Report on the framework and on the predictive models for therapy response	5	Forth	R	PU	34	Yes	Nov/Nov	
7.9	Report on the INTEGRATE first workshop/launching event	7	BIG	R	PU	34	Yes	Nov/May	
2.7	Final system architecture and security framework	2	Custodix	R	PU	36	Yes	Jan/Feb	
6.5	Scenarios for validation and detailed metrics for the validation of tools	6	IJB	R	PU	36	Yes	Jan/Mar	

4 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).
Make sure that you are using the correct following label when your project has classified deliverables.
EU restricted = Classified with the mention of the classification level restricted "EU Restricted"
EU confidential = Classified with the mention of the classification level confidential " EU Confidential "
EU secret = Classified with the mention of the classification level secret "EU Secret "

TABLE 2. MILESTONES

Milestone no.	Milestone name	Lead participant	Due achievement date from Annex I	Achieved Yes/No	Actual / Forecast achievement date
MS7	Final Integrate architecture	Custodix	M36	Yes	February 2014
MS8	Validation of the Integrate environment	Philips	M45	Next reporting period	September 2014

6 Explanation of the use of the resources

6.1 *Manpower overview*

PMR3
Planned: Amendment (linear 12-12-12-9)
Actual is effort M25 - M36

		WP1		WP2		WP3		WP4		WP5		WP6		WP7		WP8 (MA)		Total	
		Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual
1	Philips	1,33		4,27	1,70	5,33	4,00	5,87	6,00	1,60		2,93	8,00	1,33	3,00	3,20	2,30	25,87	25,00
2	BIG	3,20	0,80	2,40	0,23	2,40	0,50	2,67	0,94	2,67	0,93	2,67	2,80	3,33	2,45	0,40	0,58	19,73	9,23
3	FORTH	1,33		4,27	15,00	1,60	6,00	6,93	25,00	7,20	26,15	2,67	11,00	1,07	3,00	0,80	1,78	25,87	87,93
4	Custodix	1,07	0,05	7,47	5,12	1,60	1,11	5,33	3,82	0,53	0,03	2,40	3,11	0,80	1,50	0,27	0,32	19,47	15,06
5	IJB	5,07		0,00		1,87		0,00		3,20	1,20	2,93	3,65	1,20	0,60	0,40	0,43	14,67	5,88
6	UPM	1,07	1,69	3,20	2,71	6,93	7,33	3,73	6,25	0,53	1,43	2,67	3,74	0,93	1,02	0,40	0,76	19,47	24,93
7	GBG (new)											0,00	0,00	0,00	0,00			0,00	0,00
	Total	13,07	2,54	21,60	24,76	19,73	18,94	24,53	42,01	15,73	29,74	16,27	32,30	8,67	11,57	5,47	6,17	125,07	168,03

The numbers in the column ‘planned’ reflect the lineair distribution of the resources over the lifetime of a work package. They do not reflect phases of high and/or low activity.

Below the planned effort over the remaining period

PMR4, effort
Planned: Amendment (linear 12-12-12-9)
Actual is effort M37 - M45

		WP1		WP2		WP3		WP4		WP5		WP6		WP7		WP8 (MA)		Total	
		Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual
1	Philips	1,00		3,20		4,00		4,40		1,20		2,20		1,00		2,40		19,40	0,00
2	BIG	2,40		1,80		1,80		2,00		2,00		2,00		2,50		0,30		14,80	0,00
3	FORTH	1,00		3,20		1,20		5,20		5,40		2,00		0,80		0,60		19,40	0,00
4	Custodix	0,80		5,60		1,20		4,00		0,40		1,80		0,60		0,20		14,60	0,00
5	IJB	3,80		0,00		1,40		0,00		2,40		2,20		0,90		0,30		11,00	0,00
6	UPM	0,80		2,40		5,20		2,80		0,40		2,00		0,70		0,30		14,60	0,00
7	GBG											6,00		3,00				9,00	0,00
	Total	9,80	0,00	16,20	0,00	14,80	0,00	18,40	0,00	11,80	0,00	18,20	0,00	9,50	0,00	4,10	0,00	102,80	0,00