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: 01-05-2014

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

Period covered: **from 1 February 2014 to 31 October 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://www.fp7-integrate.eu

[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: Ad de Beer (On behalf of Anca Bucur)

Date: 15-12-2014

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

1 Publishable summary

Public summary
The collaborative INTEGRATE project supports 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 brings 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.

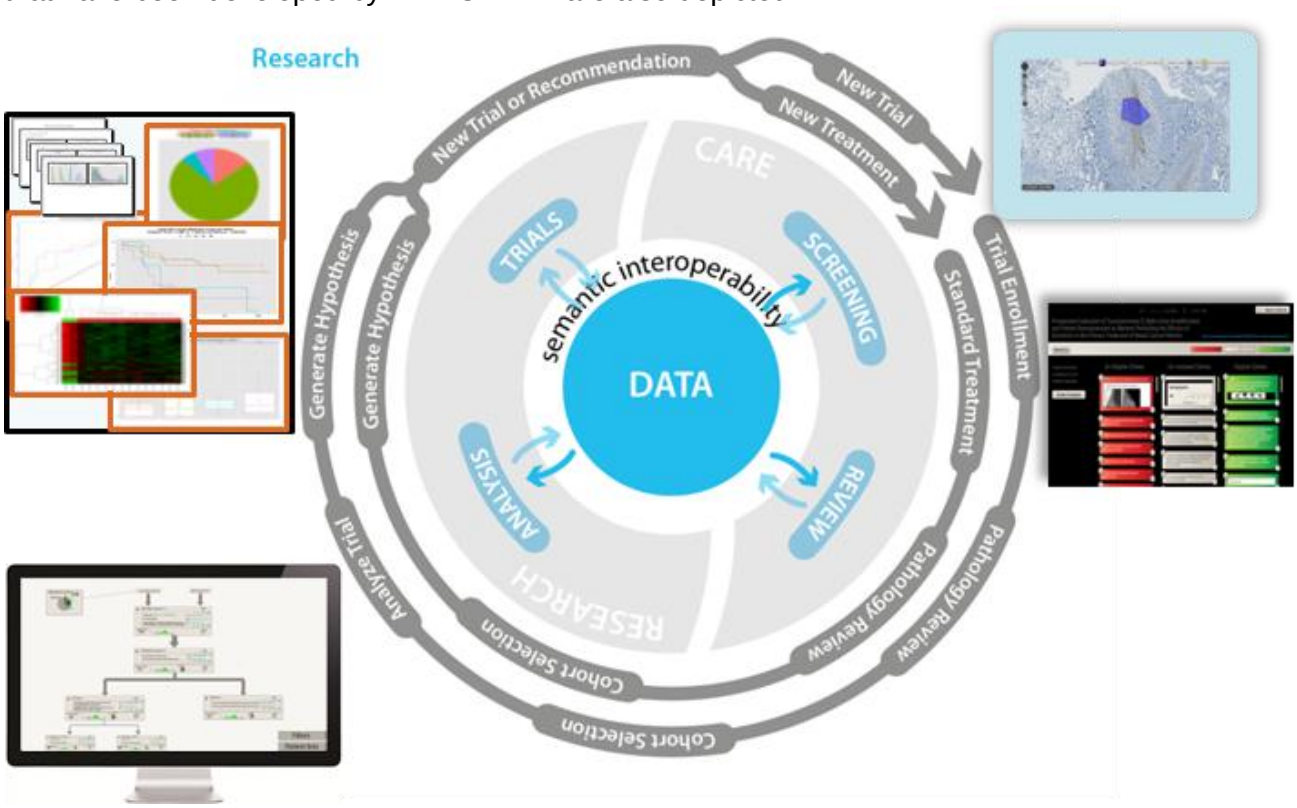
INTEGRATE delivers solutions that support a large and multidisciplinary biomedical community ranging from basic, translational and clinical researchers to the pharmaceutical industry to collaborate, share data and knowledge, and build and share predictive models for response to therapies. Moving away from empirical medicine, towards evidence-based personalized care has the potential to both dramatically improve patient outcome and to reduce costs.
The project made 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. 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 has been to build tools that facilitate efficient the execution of post-genomic multi-centric clinical trials in breast cancer. One aspect of the execution of clinical trials is the enrolment of patients into clinical trials. INTEGRATE supports efficient 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 built a flexible and intuitive cohort selection application that enables users to define, select and retrieve cohorts of patient datasets that suit their research questions.

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. For this reason a “community edition” of the semantic interoperability layer will be made available, facilitating the adoption and contribution to the INTEGRATE semantic solution. Sustainability of the core platform will be pursued via the UURECA project and “The European Institute for Innovation through Health Data”.

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.



More concretely, the outcomes of the INTEGRATE project encompass the core platform, featuring a state of the art semantic interoperability solution embedded in a Service Oriented Architecture approach.
Next to the core platform, the following applications have been developed during the runtime of the project:

- *Decima* – an application that automatically evaluates a patient's eligibility for a set of clinical trials. The application reduces the time necessary for a clinician to evaluate eligibility and potentially increases the amount of considered enrolments for patients.
- *Nona* – An application that allows a clinical researcher to intuitively and efficiently compose patient cohorts (sets of patients that share similar characteristics – as defined by the application's filters) out of all patients contained in the patient repository.
- Analytics tools. - - a web-based environment for empowering the scientists from diverse backgrounds a) to analyse with ease clinic-genomic data in order to get simple statistics on selected parameters, perform survival analyses, compare regiments in selected cohort of patient and obtain genomic analysis results, and b) to perform heterogeneous clinical data modelling for deriving and cross-validating in multiple datasets predictive clinic-genomic models of patient response, and assessing the value of candidate biomarkers.

- Pathology review – a web application that provides to the stakeholders all the necessary tools in order to have digital images reviewed by multiple experts. It provides collaboration and knowledge share capabilities and an expandable business workflow and its overall aim is to help experts conclude to the right results, among patients participating in a clinical trial.

In the last stage of the INTEGRATE project, the focus shifted from requirements analysis and development to validation and evaluation. All applications have been evaluated by clinical users during a workshop in Crete. For this workshop, the platform had been set up in a centralized configuration. Furthermore, a targeted user evaluation took place at Institut Jules Bordet. For the patient screening application, a dedicated deployment was performed and patient data from EJB was imported into the platform. Whenever a big platform is deployed, it is difficult to get a good estimate of the deployment costs. This activity was also used to keep track of the costs associated with performing a platform deployment in order to come up with good estimates which can be used in the exploitation stage. For the other tools (not residing in the clinical care domain but the clinical research domain) the central setup was used.
Finally, the patient screening application has been favourably evaluated by an external party, the MAASTRO clinic https://www.maastro.nl/.

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.

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.

The INTEGRATE project encompass the core platform, featuring a state of the art semantic interoperability solution embedded in a Service Oriented Architecture approach. This core platform provides the basis for innovative, semantic enabled clinical research and care application.

Many applications that will run in the platform will require the selection of patient cohorts - sets of patients that share similar characteristics – as defined by the clinical user by means of filters). Within INTEGRATE, an application has been developed which enables a user to construct patient cohorts in an intuitive manner.

Analytics:
Another tool has also been developed for rapidly analysing and assessing the quality of certain clinical characteristics across patient population.

Pathology review:
Central Review for Pathology images tool enable multiple users to access and review over the web, the same set of digital pathology images simultaneously.

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, and at the same time requires a bigger volume of 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.
Decima - an application developed in the INTEGRATE project – supports this trend towards a higher volume of targeted clinical trials by allowing an efficient and iterative evaluation of the eligibility of a patient for a set of clinical trials. The expected impact of the application can be one of the following:

- A more efficient evaluation of a patient's eligibility for a set of clinical trials by the clinician
- An increase of the number of clinical trials for which eligibility is checked given a patient
- An increase of the number of patients considered for enrolment in a clinical trial (impying a reduction in the recruitment time needed for a clinical trial)

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 and run validation workshops
- 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
- Prepare the final report and preparations for the final review

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

<u>Implementation effort</u>

- Security services:
    - Most effort was spend on implementing the auditing front-end and services. The front-end contains different mechanisms for querying the logs in different visualizations. The authorisation service was further refined, focusing on the special requirements applicable in the INTEGRATE project (especially policy requirements). Additional security proxy for the semantic interoperability layer were implemented and deployed. Finally the STS, IDP and authorisation security services were integrated with the auditing services.
- Patient screening tool (Decima):
    - The last iteration (within the INTEGRATE project) of the Decima tool was completed in this reporting period. The focus was on preparing the tool for the different validations sessions. The main challenges were optimizing the networking stability between the different services, customizing the GUI/workflow of Decima (depending on the requirements of each validation site) and fixing bugs encountered during the validations. Next to this further integration work was done with the semantic interoperability layer and security services.

- Nona:
    - Integration of security on all services
    - Integration of the autocomplete, analytics, and metadata services.
    - Features:
        - Examine data concepts present in a dataset, and create and modify filters directly from that.
        - Sharing of search results
        - Exporting of result sets

- Analysis platform: The implementation effort for this period was mainly focused on the following aspects:
    - The design, functionality and stability of the Analysis platform was further improved and extended based on the feedback from the review meetings and the output/comments from the two evaluation sessions.
    - Useful additional features have been incorporated within the Analysis Platform, e.g. save/restore an analysis workflow, export the selected analysis cohort in csv/pdf format, etc.
    - A new predictive model relying on a survival outcome has been developed, tested and embedded into the Analysis platform
    - The existed on-the-fly generated reports have been enriched with additional information, containing new image and table results from each analysis.
    - The analysis software code (native R language code) was updated, containing more useful information for the end-user
    - An online tutorial have been written, accessible from the Analysis platform, for quick user reference

- Pathology review:
    - Integration of security on all relevant services
    - Implementation of an extra SOAP service for pushing reviewer's answers to CDW
    - Tiling service optimization in terms of performance
    - Limitation over file size during digital image upload process removed
    - Enhance and refine protocol management process by:
        - Providing the option to the moderator not only to mark conflicting tasks but the exact parameters for which there is a conflict per task
        - Simplifying the protocol management interface (wizard) by merging certain (overlapping) views and by re-structuring the information provided to the moderator
    - Enhance and refine conflict resolution process by providing to the reviewers information from moderator about a conflict on the review form and by marking the problematic parameters in different way
    - Features:
        - Implemented a discussion board for the experts to exchange knowledge and information about a protocol image.
        - Supporting now additional image formats
        - Added the option to submit an electronic Case Report Form (eCRFs) for each review task.

<u>Evaluation and validation</u>
We have organized the INTEGRATE evaluation and validation workshop (13-14 June 2014) with expert oncologists, pathologists and bio-informaticians 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) The evaluation of this workshop is noted in the resubmission of D7.9.

In addition, evaluations were held on several pilot sites

- Patient screening (Decima) and security:
    - The patient screening tool Decima was validated at several sites, with various end-users:
        - At the Crete workshop, with a panel of invited experts

        - At the MAASTRO radio oncology clinic, a partner from outside the consortium, with an experienced trial nurse
        - At the KGU (Frankfurt University Hospital) with trial nurses of different experience levels and a trial coorndinator
        - At the IJB, with research nurses and an administrative assistant.
    - For the three evaluations on the pilot sites, anonymized local patient data was loaded into the INTEGRATE framework. Also the trials on which the patients were screened were taken from the clinical practice of the participating users.
    - Decima was found to hold great potential for improving the patient screening process. Several key issues were found that need to be resolved in order to allow this potential to be unlocked.
        - The clinical and contextual knowledge of the clinical staff must be taken into account in the formalization of the scripts. The formalization must not necessarily be logically correct in order to make clinical sense.
        - There need to be mechanisms to allow the users to prioritize and stratify the criterion evalution process. This will empower users to fit Decima smoothly in their local workflow.
        - Performance, robustness and reliability of Decima and the underlying services needs to be addressed. There need to be provisions for error handling and recovery and reporting corrections.

- Nona:
    - The cohort selection tool Nona was validated at several sites, with various end-users:
        - At the Crete workshop, with a panel of invited experts
        - At the GBG, with a bio informatician
        - At the IJB with five staff members involved in data management on clinical trials
    - For all validations, we used a large anonymized research dataset from the GBG.
    - Although Nona fits within the workflow of the skilled bioinformaticians and data handlers, it directly competes with large, general purpose statistics tools. The opportunities that Nona has are reaching out to an audience of non-programming specialists, and fitting better in specialized workflows. Especially the sharing of selection queries is mentioned: this can be an enabler for better trial feasibility studies.

- Analysis platform:
    - For the evaluation of the Analysis platform two evaluation session were carried out:
        - Quantitative evaluation workshop at Crete: two experts –outside the INTEGRATE consortium were engaged (1 oncologist, 1 bioinformatician).
        - Quantitative evaluation session at IJB: three experts from IJB participated (2 bioinformaticians, 1 epidemiologist).
    - The evaluation session at IJB was performed remotely using screen sharing and VoIP applications for the communication between FORTH and IJB.
    - Both sessions ran on the distributed INTEGRATE evaluation testbed.
    - The prototype evaluated by IJB experts had already incorporated additional features, based on observations made during the previous quantitative evaluation session.
    - For the evaluation of the Analysis platform both SAGE and TOP trial datasets were used.
    - The experts gave us a very useful feedback and inspiring from improving or adding functionalities that would lead to a more user-friendly environment for the end-users.

- The comments from the evaluators and the overall scoring indicate that the platform provides a fast and easy-to-use solution form performing specific statistical and predictive analysis, as addressed from the user needs of INTEGRATE, under a consistent and secure framework. Additional functionality will be implemented and added in order to conclude to a basic prototype which will be then used in several clinical applications.

- Pathology review:
- The CRP was validated at several sites, with experts of different working backgrounds:
  - At the Crete workshop with two experts; the first one had an academic background and more specifically is a professor in a medical school and the second was an expert pathologist in a central laboratory, responsible for collecting, scanning and evaluating large numbers of patient tissue samples, which are submitted by institutions all over the world in the context of multinational clinical trials.
  - At the IJB with one expert initially, who was a surgical pathologist and a cancer researcher. We have to mention here that the evaluation of the tool from two other experts (at IJB) is in the process of setup and will be performed until December 15th.
- For the evaluation sessions mentioned above a dataset of digital pathology images was provided by the BCTL laboratory at IJB. They were biopsies from breast cancer tumors, either haematoxylin or eosin stained (H&E) or immunohistochemistry-stained for the following markers: HER2, ER and PgR.
- All experts participated in the evaluation processes expressed they interest to stay updated about the progress of the tool, while some of them expressed interest for a future collaboration. In general the evaluation process revealed that the tool is capable of fitting in various business workflows in cases where multiple experts need to draw in common (or in cooperation) conclusions about a set of pathology images:
  - Such a case (not envisioned by the basic usage scenario as described in INTEGRATE DoW) is its use in the academic sector.
  - Evaluation sessions revealed that indeed CRP could be fit in and enhance the process of selecting patients for participating in clinical trials. At this point some key issues were identified that need to be addressed by the tool for exploiting its full potential:
    - The process of automatic conflict resolution requires further analysis and testing on various usage environments so that it can be implemented optimally
    - It was suggested to investigate the option of interconnecting the tool with Patient Record systems used by organizations internally, so that it could be incorporated as a key tool in formal internal (or joint) processes such as these of clinical trials
    - Finally, it is necessary to incorporate mechanisms of auditing / logging. This is something that is already being implemented and will probably be incorporated in the next version of the tool.

Dissemination and Exploitation

We have published the fifth and final issue of the INTEGRATE Newsletter has been published. An additionally, we have been producing a video to presenting and promoting the INTEGRATE platform and tools has been produced. This video includes interviews of end-users and project participants tools of the INTEGRATE project.

Preparing the tools for validation and updates based on user feedback All partners

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)

Although this workpackage ended in month 36 and no objectives were defined in this reporting period, there are still some supportive objectives required for:

- Implementing/integrating/deploying the demonstrators presented at the annual review meeting (year 3)
- Deploying/configuring/bugfixing the validation environments (including the security framework), part of WP6.

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

- This task was finished in month 36
- Brainstorm technical meetings were held for completing the demonstrators of year 3, presented at the annual review meeting
- Demonstrators of year 3 (patient screening, cohort selection, analysis platform, pathology) were implemented, integrated and deployed.
- Some bug fixing was done in the different tools of the INTEGRATE platform for bugs encountered during the validations.

Task 2.4 Security for dynamic collaborative environments

- This task was finished in month 36
- The final iteration of the INTEGRATE security framework services was finished and presented at the review meeting of year 3. The main focus was on authorisation and auditing.
- The authorisation part of the security framework was finished (implementation, integration and deployment) and presented at the annual review meeting.
- The audit functionality was further enhanced, including a graphical user interface, displaying the audit logs.

Task 2.5: Component integration and interfacing with external systems

- This task was finished in month 36

- Final integration between the different INTEGRATE tools (including the security framework) was coordinated by WP2.
- Tutorials were provided for security integration of the INTEGRATE tools.

2.2.3 WP3 – UPM

2.2.3.1 Objectives (of the reporting period)

The current reporting period has been focused on the validation of the INTEGRATE tools with the new partner GBG (German Breast Group) and MAASTRO as collaborator. The core dataset should be extended with concepts from the new data sources and domains beyond breast cancer provided by MAASTRO. The deployment of the semantic interoperability layer should be tested with such new datasets and external experts (including a validation workshop).

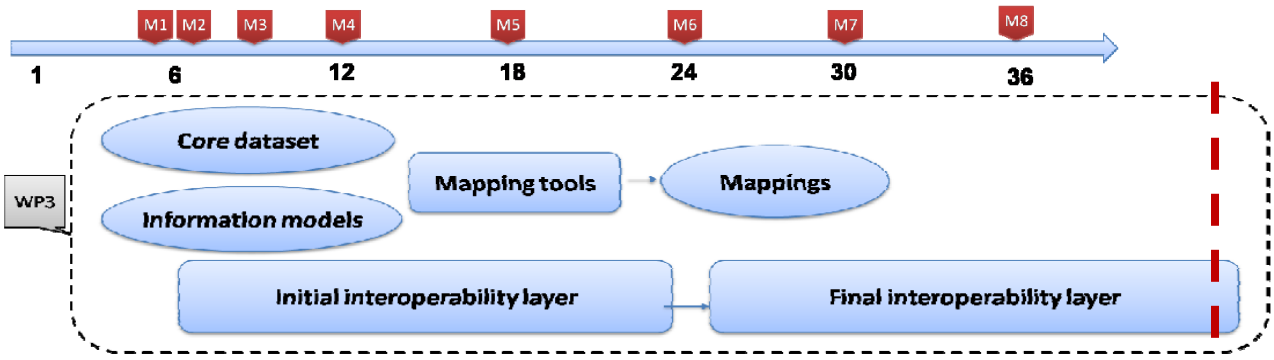


Figure 2.2.3.1 Timeline of the INTEGRATE Work Package 3

After the comprehensive validation and the extension of the INTEGRATE semantic solution, the final version should be described and reported.

2.2.3.2 Status/progress towards objectives WP3

Task 3.1 Definition of the semantic core dataset
During the reporting period, the main work within Task 3.1 was the extension of the core dataset to other domains. We identified and described two types of extensions: (i) Terminology extension (including new concepts) and (ii) Terminology binding extensions (new links from concepts to the HL7 RIM-based Common Data Model. Such mechanisms have been described in Deliverable 3.6 together with the analysis of new datasets from two new clinical partners: (i) GBG (German Breast International Group) and (ii) MAASTRO. Concepts required to code such datasets were selected from terminologies within the core dataset (more information about mappings in Task 3.3).

Task 3.2 Definition of the information models of the clinical and research infrastructures

No updates in this task for the reporting period.

Task 3.3 Semantic formalism, mapping tools and mapping implementations

Minor issues have been updated within the mapping implementations during the reporting period. Most of the work in this task, and such minor updates has been triggered by the validation process and the new datasets available in the project extension. Core dataset extensions were required and the corresponding mappings to the INTEGRATE CDM. In collaboration with data providers, we have used the Mirth Connect open source tool to export HL7 messages. In the case of MAASTRO dataset, HL7 message templates were provided from the INTEGRATE platform. The data provider then used the Mirth Connect tool to generate the corresponding HL7 messages extracted from EHR data including free text (a previous NLP process was required in this case). For GBG, the new INTEGRATE partner, the dataset was provided in CSV format without free text and including the corresponding codebook. Again the Mirth Connect tool was used in collaboration with the data provider to generate the corresponding HL7 messages. The following table include data about the data load process:

Table 3.3.1 Dataset parameters loaded into the INTEGRATE CDM

Dataset	Patients	Total Variables	Boolean variables	Simple variables	Multi-concept columns	Re-used variables
GBG TBP	156	59	32	10	17	-
GBG GAIN & Gepar Quattro	3022 + 1495	30	0	3	27	10 (TBP)
GBG Frankfurt	258	17	0	4	13	7 (TBP), 6 (Gepar Quattro), 1 (IJB)
Maastro	270	288	235	1	52	10 (TBP), 4 (Gepar Quattro)
IJB	29	99	27	29	43	5 (TBP), 8 (Gepar Quattro)

An additional tool to automate the process of normalizing SNOMED CT concepts and binding to HL7 RIM has been developed to facilitate the data load process. The SNOMED2HL7 tool is freely available (after registration) at http://kandel.dia.fi.upm.es:8078 (currently under a publication process).

Finally, and estimation of costs of the data load process has been performed:

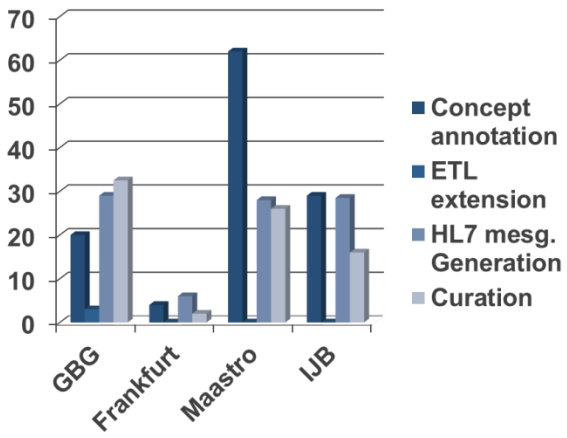


Figure 3.3.2 Estimation of cost (in hours) of the data load process to the INTEGRATE CDM.

Data load cost for a new datasets to be loaded into the CDM hardly depends on the number of concepts (annotation) and the number of patients (curation). But the total time is usually under two weeks for a trained developer. Detailed figures regarding the data load process has been also reported in deliverable 7.13.

Task 3.4 Design and implementation of the semantic interoperability layer

Besides the query builder service that was developed in the previous reporting period, the final version of the semantic interoperability layer has been modified with a new SPARQL wrapper for the CDM. The MORPH[4] engine has substituted D2R since:

- Follows specifications of the W3C[5]
- Fix performance problems of the D2R server
- The new translation mapping is implemented in the Turtle language, a syntax for RDF defined by the W3C[6]
- The new SPARQL wrapper (MORPH) and the new normalization pipeline can be observed in the following figure:

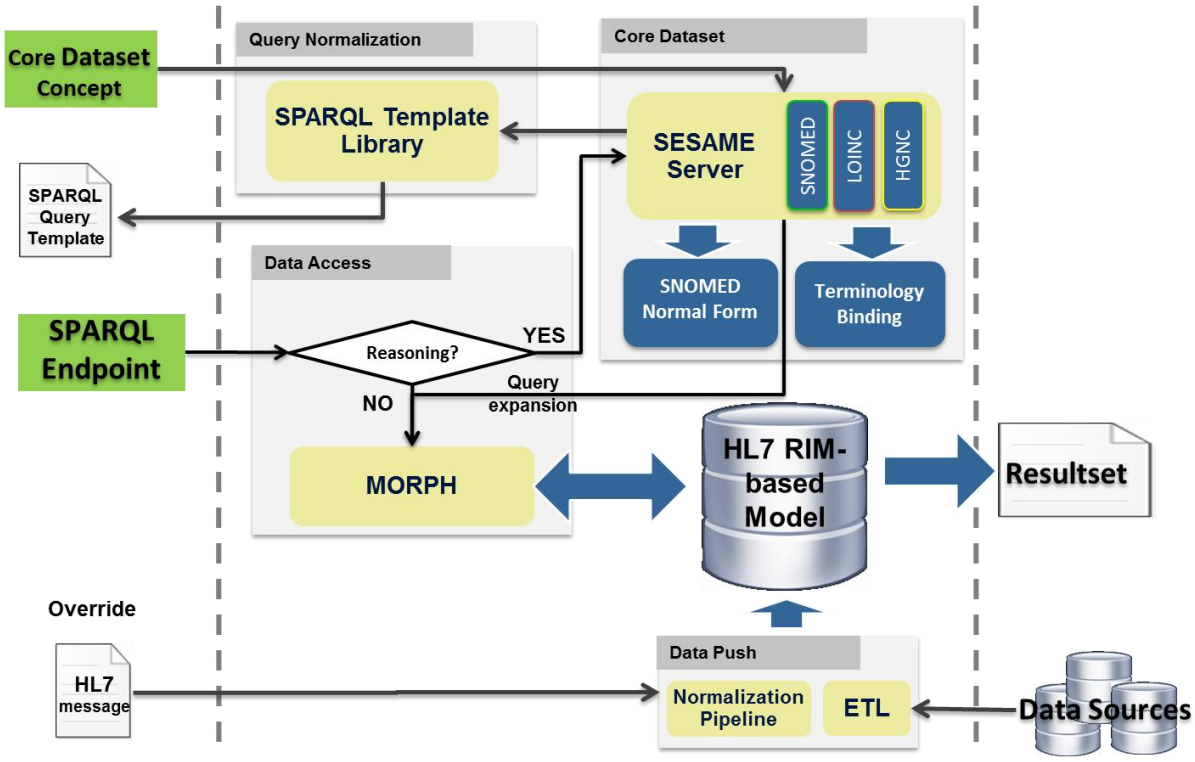


Fig. 3.4.1 Final version of the semantic interoperability layer

Deliverable 3.7 describes in detail the final version of the INTEGRATE semantic interoperability layer.

Task 3.5 Standards-based uniform access to external sources

No updates in this task for the reporting period.

[4] https://github.com/fpriyatna/morph
[5] http://www.w3.org/TR/r2rml/
[6] http://www.w3.org/TeamSubmission/turtle/

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 the semantic interoperability layer validation during months 37 to 45.

2.2.4 WP4 – FORTH

During the extension period the WP4 focused on preparing the material and configuring all the required resources, for the validation and evaluation of its tools and its services, as they have been carried out at the relative workshops of the project. After the validation of the tools, the processing of the received feedback from the evaluators provided us with valuable results for improving whatever was possible in time or for adjusting the development path of the tools towards the future. Aside of the validation and evaluation of the tools the other main focus was in the quality improvement of the tools.

2.2.4.1 Objectives (of the reporting period)

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

1. Complete the prototypes of the tools and services in the WP4, in order to provide all the necessary functionality described in the DOW.
2. Evaluate the performance and the usability of the WP4 tools, in close collaboration with the clinical partners.
3. Overall improvement of the stability and functionality of the tools and services, in order to be more user friendly and more robust in their functionalities.

2.2.4.2 Status/progress towards objectives

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

[There is nothing to report regarding the task 4.1, as the implementation of the model, data and annotation repositories has been made in previous periods.]

Task 4.2 Tools enabling data and knowledge sharing

There are three main tools (& with broader sense let's describe and a set of services as a tool) involved in the data and knowledge sharing:

1. The cohort selection tool
    a. Integration with the security framework
    b. Integration with extra services
        i. Metadata
        ii. Autocomplete
        iii. Analytics
    c. Main new features
        i. Facilities for examining the data concepts in a dataset. Filters can be created and modified straight from these concepts.
        ii. Sharing of search results
        iii. Exporting of result sets (patient IDs only for now)
2. The central pathology review platform
    a. Integration with the security framework
    b. Main new features
        i. Discussion Board with unique subject per protocol image
        ii. Additional digital pathology formats supported
3. The integration of the semantic interoperability services with the collaborative tools.

a. During this period an additional HL7 message template (following the IHE standard) has been developed to store Pathology Image metadata to the Common Data Model

Task 4.3 Tools enabling collaboration

In the last 9 months, the main effort in task 4.3 was to evaluate the performance and the usability of the platform as also complete the prototype of the CRP, in order to provide all the necessary functionality described in the DOW. In more details the following areas were the ones were we focused our development.

The evaluation of the platform in terms of performance and usability has been performed in close collaboration with the clinical partners. For this purpose two evaluation sessions have been scheduled and conducted, the first one at FORTH and the second one at IJB premises. The outcome of these sessions gave us valuable feedback about Central Review for Pathology platform's current state. During these sessions several suggestions about changes or additions to already implemented functionality have been made and some important issues have been identified.

In the following sections we briefly describe the most important additions & changes that have been implemented during this last period:

- In case of Conflict in Tasks the instructions provided for its resolution are depicted on the review forms for helping reviewers during conflict resolution process. Also conflicting tasks are marked on the forms with different colours for the same purpose.
- For each protocol image there is a dedicated discussion board where the reviewers can exchange knowledge or opinions about a specific image scoring/review process. This board can also be used during conflict resolution process.
- During protocol management process the moderator can not only mark conflicting tasks, but the exact scoring variables (per tasks) for which there is a conflict.
- Protocol management process has been simplified by removing and merging specific steps in the wizard which is used for it. Additionally the first step of protocol management wizard has been re-designed and now depicts the overall protocol status in a more convenient way.
- Tiling service has been improved in terms of performance and the 2 GB file size limitation during image upload has been removed.
- Additional digital pathology image formats are now supported
- An additional SOAP service for storing expert answers to the DCW.
- During protocol review process we have added the option to include (per protocol task) a second, standardized electronic Case Report Form (eCRFs) which the reviewers have to fill in and submit to the system.
- Finally minor enhancements alongside the platform have been made. Indicatively: terminology in several parts have been enhanced in order to be more conceivable for the average user; image annotator have been enhanced by adding/removing specific annotation tools; review forms have been updated by including several pieces of useful information etc.

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
The final iteration of specifying the privacy enhancing processes was done and documented in deliverable 4.6 (Integrated Privacy Enhancing Services and Processes)

Further implementation work was done on the privacy enhancing tools (CATS and PIMS), focussing on deployability and configurability. New features required by the INTEGRATE processes were added to the CAT tool.

The privacy enhancing tools were integrated with the authentication framework (STS service), part of WP2.

2.2.4.3 Deviations from the DOW and corrective actions

(None)

2.2.5 WP5 – FORTH

WP5 has focused on providing users with a collaborative, multi-functional and easy to use environment for exploiting, analysing and assessing the quality of large multi-level data. Its main goal is to empower clinicians and researchers to analyse 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:

- Evaluate the performance and the usability of the Analysis platform in close collaboration with the clinical partners.
- Further improvement and refinement of the Analysis platform according to the feedback from the latest review meeting and the output/comments gathered from the two evaluation sessions carried out at FORTH and IJB.
- Inclusion of a regression model for predicting survival outcome.
- Update the existed analysis software code including information regarding the selected variables for analysis, the filtering process during the selection of a cohort for the analysis, etc
- Include pre-processing steps for data exploration and assessment of significant degree of the cohort variables in the predictive modelling framework.
- Include proportional hazard regression models at the survival analysis.
- Assess the performance of the updated predictive models using SAGE dataset.
- Inclusion of the latest version of the PrimeFaces library (http://primefaces.org/), which is a Java Server Faces library for building the user interface (UI) of the platform.
- Submit and produce a paper/publication for illustrating the functionalities of the Analysis platform and as part of the dissemination activities.

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, the final iteration of the design, functionality and stability of the Analysis platform was highly prioritized. Summarizing, the working progress, related to the development of the platform, is depicted in the following list:

- The framework used for the development of the user interface of the Analysis platform has been updated to the latest version (Primefaces 5.1)
- A refinement of the **History portlet** has been performed, which includes:
  - A more intuitive procedure has been adopted for comparing the results of two or more analysis results. The user has to simply select the corresponding rows at the History table and then press a button.
  - The History table is automatically updated in specific time intervals. Thus the user does not need to manually refresh the page, in order to be notified if an analysis has been completed.
  - The analysis report can be viewed by just double clicking on the corresponding row, (instead of using the pop-up menu).
- A refinement of the **Data Sources** portlet has been performed, which includes:
  - The difference between the scheduled and non-scheduled data retrieval has been better clarified using instruction messages.
  - A description and the size of the retrieved datasets has been included, in order the user to be more aware of the available datasets.
- A general refinement of the analysis wizard has been performed, which includes:
  - The tabs are now clickable and in combination with the “Previous/Next” buttons, a more user-friendly navigation has been achieved.
  - A better navigation has also been achieved, by keeping the selections at each step (dataset-cohort-scenarios/models-results) during navigation.
  - The entire workflow can now be saved in an xml format and be loaded whenever is needed. Namely, the user’s selections at each specific step are stored and can be retrieved either partially (e.g. only the cohort selections) or in whole. The workflow can be completed (with produced results) or non-completed.
  - The functionality of the mindmap has been improved. The mindmap is the PrimeFaces component used at the final step of analytical and predictive modelling tools wizards for presenting the results in a diagram format The overall workflow is now presented in one-level style, so the user no longer has to be specific with the clicks.
  - Several changes regarding the overall layout of the wizard has been made, i.e.,
    - instruction messages have been added wherever considered necessary,
    - text descriptions of the datasets/cohorts/scenarios have been added at the corresponding wizard steps,
    - reset functionality of cohort and scenario/model selections has been implemented,

        - exporting feature of the selected cohort in xls/pdf format has also been incorporated etc
- The analysis software code was updated in order to include additional information regarding the selected input variables, the selected cohort, a summary of the filtering options.
    - A short description of the selected variables for analysis is given by a table at the top of the analysis report.
    - The pre-processing step of the cohort selection is also described in the analysis report.
- A predictive model for survival outcome has been added to the platform. This model is performed in both homogeneous and heterogeneous analysis cohorts.
    - Lassoed principal components (LPC) technique for testing significance of the Gene Expressions and/or the SNPs in case of been selected for the analysis is applied as a pre-processing step. LPC is followed by False Discovery Rate (FDR) estimations.
    - Exploratory tools (e.g. heatmaps and volcano plots) are also generated at the pre-processing phase giving a clear indicator about the significant level of the potential biomarkers in the analysis.
    - A backward variable selection using a Cox regression model has been implemented
    - A Random Forest regression model has also been added to the predictive model
    - Both models are tested during training using bootstrapping techniques and several statistical measures of performance.
    - Both models are tested and evaluated using new cohorts and return predicted survival outcomes.
    - The predictive modelling framework is separated into 3 different sub-models (TRAIN, TEST and COMPLETE study) allowing users a) selecting training a model for future predictive model analysis, b) testing a trained model that is stored to the platform's database using new testing cohorts c) performing a complete with defined train and test sets.
- The survival analysis software code has been updated using proportional hazard regression models.
- The analysis report of the descriptive statistics has been updated including new graphs and tables.
- The existed predictive models for a binary outcome have been refined and enriched with pre-processing steps for variable selection and data visualization.
    - Lassoed principal components (LPC) technique for testing significance of the Gene Expressions and/or the SNPs in case of been selected for the analysis is applied as a pre-processing step. LPC is followed by False Discovery Rate (FDR) estimations.
    - Exploratory tools (e.g. heatmaps and volcano plots) are also generated at the pre-processing phase giving a clear indicator about the significant level of the potential biomarkers in the analysis.
    - The generated reports have been updated making them more user-friendly and understandable.
- The GBG dataset has been incorporated into the platform for performing statistical analysis.
- Several bugs were detected during evaluation sessions and corrected afterwards.
- The misleading term “Simple Statistics” was replaced by “Descriptive Statistics”.
- The paper entitled “Multi-Modal Medical Data Analysis Platform (3MDAP) for analysis and predictive modelling of cancer trial data” has been submitted and presented at the 6th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation - The CHIC Project Workshop (IARWISOCI) that was held on November 3-4, in Athens.

2.2.5.3 Deviations from the DOW and corrective actions

No deviations

2.2.6 WP6 – Philips

2.2.6.1 Objectives (of the reporting period)

The current reporting period has focussed on the validations at the workshop in Crete, at GBG, KGU and IJB. Additionally, we have cooperated with the MAASTRO radio oncology clinic for a dedicated study of the patient screening tool Decima. For each site, this included setting up the INTEGRATE service framework, filling in the clinical content where required, deploying the end-user tools, and conducting the actual validation.

2.2.6.2 Status/progress towards objectives

Task 6.1 Building the INTEGRATE development and testing environment
In order to conduct the evaluations, two test environments have been constructed.
The first is the Distributed Testbed, with servers located at the developing partner's site. The testbed is accessible from any normal internet connection, provided that the user is authenticated through the INTEGRATE security layer.
A second test environment was set up within the firewall of the IJB. This implied an installation of all INTEGRATE components on local servers, with the exception of the trial metadata and the STS. The local installation ensured that all patient related data was held within the IJB's secure environment.

Task 6.2 Formulate evaluation criteria, validation procedures and feedback report guidelines
No updates in this task for the reporting period.

Task 6.3 Coordinate specifications of test scenarios and of demonstrators
The focus in this period has been on filling in the data and tailoring the test scenarios to the clinical validation sites.

The patient screening tool Decima has several input datasets: a set of formalizations of trial criteria, and the patient list, with underlying patient data. The latter was taken from the output of WP3, as indicated in Table 3.3.1.

For the Crete workshop, and for the evaluation at KGU, the GAIN, Gepar Quattro and TPB trials have been used. The MAASTRO evaluation used the formalization of 13 trials on lung and breastcancer. At IJB, the set comprised four clinical trials.

The cohort selection tool Nona has been evaluated with the complete GBG patient data set. All supporting services, such as the analytics, the autocomplete and the metadata were configured to work with the same underlying data.

For the validation at IJB, use of the *Breast structured dataset* was submitted to and approved by the IJB Ethics Committee.

Task 6.4 Deployment Environment
For the evaluation and validation in Crete invited BIG a panel of international experts to act as test users. For the pilot sites, the test users were selected in collaboration with the particular site. These were clinical and research staff from the respective institutes who deal in their work with the task at hand, either because of it being part of their daily work, or because they are a stakeholder in the task being evaluated.

The coordination of the execution of the evaluations has been done mainly by mail and telco's. In a few cases, it was necessary to visit the site prior to the evaluation to coordinate and test the technical setup and data load.

Task 6.5 Coordinate evaluation and validation activities and reporting
We have conducted evaluations of the four end user tools with a diverse set of users.

- During the INTEGRATE event in Crete, invited experts from outside the INTEGRATE consortium evaluated and commented on the tools in a workshop setting.
- A dedicated evaluation for the patient screening tool Decima was held with clinical staff of the MAASTRO radio oncology clinic. We performed a simulation of the daily screening work of the trials nurses.
- Through our new partner GBG we had access to clinical staff at the KGU for testing the patient screening tool Decima. GBG staff assisted in the evaluation of the cohort selection tool Nona.
- A series of evaluation for all tools was finally done with clinical and research staff at IJB.

The evaluations confirmed the envisioned workflows and uncovered many local variations. Our tools proved to be relevant and adaptable to a wide range of workflows. For all tools, we have identified improvements and additional functionality to bridge the gap between the current prototypes, and actual clinical deployment. Our main goal of having usable, relevant tools has been met. The results of the various evaluations have been recorded in detail in deliverables 6.6 and 7.14.

2.2.6.3 Deviations from the DOW and corrective actions

No major deviations.

2.2.7 WP7 – BIG

2.2.7.1 Objectives (of the reporting period)

Besides the recurring knowledge management activities (e.g., maintenance of the web site…), WP7 published the INTEGRATE Newsletter – Issue 5. And in particular, WP7 has focused during this period on the production of a video to present and promote the tools of the INTEGRATE project.

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

Task 7.1: Dissemination
The first major achievement for task 7.1 for this period has been the production of the INTEGRATE Newsletter (issue 5 – visible on the INTEGRATE website http://fp7-integrate.eu/).

The second major achievement has been the preparation, filming and final production of a video to present and promote the tools of the INTEGRATE project. The video gives a general

introduction on the INTEGRATE project, then focuses on presenting the four tools developed. Through a series of interviews, INTEGRATE developers and end-users explain what these tools are and how they can facilitate the work of oncologists and breast cancer researchers. The video will be made available on the INTEGRATE website (www.fp7-integrate.eu), as well as on the BIG website (www.BIGagainstbreastcancer.org) and others (e.g., eCancer).

Task 7.2: Exploitation
BIG has co-organised the validation workshops. By reaching towards users outside of the consortium, these workshops serve an exploitation purpose besides the more technical validation aspects. The first Evaluation Workshop was held at the premises of FORTH, in Heraklion, Greece, on Friday, 13 June 2014. The IJB Evaluation Workshop took place in Brussels from Wednesday, 29 October 2014 to Friday, 31 October 2014 at the premises of the *Institut Jules Bordet*.

Task 7.3 Standardisation
Besides continuing work on the semantic interoperability layer (see WP3, above), an effort has been made towards standardisation of machine-interpretable clinical trial descriptions, with the publication of an article on our standards-based trial metadata repository (van Leeuwen et al, 2014).

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
During the reporting period, D7.9 was modified and re-submitted. The original version of D7.9 was not approved, as the event organized was not considered a proper launching event. According to the recommendations of the reviewers, D7.9 was modified to become a summary of all events organized for end-users.

WP7 has been producing a video to present the INTEGRATE project and its achievements.

2.2.7.4 Planning
WP7 plans to make the video available, and, more generally, to keep information about the INTEGRATE project on the BIG website (http://www.bigagainstbreastcancer.org/scientific-projects1/research-pro).

3 Achievements per individual partner

Partner 1 Philips

We have constructed and deployed the Locker service as a general component in the INTEGRATE service network. The Locker is a personal data storage facility, integrated in the security framework, and can be used by services and end user applications for general data storage.

We have continued with the development of two end user tools: Decima and Nona. Both tools have been prepared for deployment the various evaluations, with full security integration. For Decima, installers are available. Field support was and is available during evaluations and tests.

We have assisted in the data loads for the Maastro and GBG data. We have also created the formalizations of trial criteria for the evaluations at Maastro, KGU and IJB. The formalizations covered trials relevant to the patient data and trials for that particular hospital. The formalizations allow the automatic matching of criteria against patient data. The formalizations and data loads were direct inputs for performing the evaluations.

The trial metadata service and the Locker service have been deployed on staging servers to allow evaluation and demonstrations while work continues on the development servers.

We have conducted evaluations of the four end user tools with a diverse set of user. During the INTEGRATE event in Crete, invited experts from outside the INTEGRATE consortium evaluated and commented on the tools. A dedicated evaluation for the patient screening tool was held with the clinical staff of the MAASTRO radio oncology clinic. Through our new partner GBG we had access to clinical staff at the KGU for testing the patient screening and cohort selection tool. A series of evaluation for all tools was finally done with clinical and research staff at IJB. The evaluations confirmed the envisioned workflows and uncovered many local variations. Our tools proved to be adaptable to a wide range of workflows.
We have presented the INTEGRATE framework at ECRIN.

The results of the validations in Crete have been recorded in deliverable 7.14. The validations on the pilot sites are described in deliverable 6.6. Additionally, we have overviewed and contributed to deliverable 7.13 on exploitation, and 4.5 on the tools and services supporting data sharing and collaboration.

Partner 2 BIG

During this reporting period, BIG achieved the following:

- Production of the INTEGRATE video
- Production of INTEGRATE newsletter, issue 5:
- Organization and gathering of contributions
- Writing articles
- Layout and graphic design

Other achievements

- reviewing and participating in the writing of deliverables
- updating the INTEGRATE website
- providing on-going clinical guidance and feedback on tools

Partner 3 FORTH

FORTH hosted and successfully organized the evaluation workshop that was held at its facilities on June 13, 2014. FORTH was responsible for the entire logistic part (e.g. coffee/lunch/dinner

arrangements, accommodation booking, travel reimbursement – in collaboration with BIG) and for several organization issues (e.g. agenda, provision of local monitors and laptops. meeting rooms booking, etc).

FORTH also contributed to the following deliverables:

- Contributed to Deliverable 4.5: Final versions of the tools and services supporting data sharing and collaboration
- Contributed to Deliverable 4.6: Integrated privacy enhancing services and processes
- Overviewed and contributed to Deliverable 4.7: Final version of the virtual collaboratory and its services
- Contributed to Deliverable 7.11: The results from the evaluation session at Crete have been recorded in this deliverable
- Contributed to Deliverable 6.6: The results from the evaluation session at IJB have been recorded in this deliverable
- Contributed to Deliverable 7.13 regarding the final exploitation plan

FORTH has participated at the Clinical Research Informatics' Solution Days: Advanced IT-support for challenges in Clinical research workshop on May 26-27th, in Düsseldorf, Germany, presenting the Central Review for Pathology images platform.

During this period FORTH has also published two papers:

A. Iliopoulos, I. Karatzanis, M. Tsiknakis, V. Sakkalis and K. Marias. (2014). "A collaborative central reviewing platform for cancer detection in digital microscopy images", 6th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation (IARWISOCI) - The CHIC Project Workshop, Athens, Greece, 3-4 November, 2014. {To be published}
B. Manikis, G.C., Maniadi, E., Tsiknakis, M., & Marias, K. (2014). Multi-Modal Medical Data Analysis Platform (3MDAP) for analysis and predictive modelling of cancer trial data. 6th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation (IARWISOCI) - The CHIC Project Workshop, Athens, Greece, 3-4 November, 2014. {To be published}

Regarding the Analysis platform in specific, the tool had also prepared for participating to the two evaluations sessions at FORTH and IJB. Based on the outputs from these evaluation sessions and on the reviewers' feedback the further improvement and development of the tool was continued. Moreover, some of the proposed functionalities have already been incorporated into the platform.

In respect to the CRP platform, the improvement and the development of the tool has been continued during the last period based on the feedback we have got from the 3rd annual review as also based on the recommendations and the outcome of the two evaluation sessions in which we participated. Additionally, the tool was prepared and participated successfully in the two evaluation sessions performed at FORTH and IJB. The results of the sessions gave us valuable feedback about its current state; several of the suggestions made during these sessions have been already incorporated into the platform whether several else are currently being implemented. It should also be noted that during the evaluation sessions we got feedback about its future expansion. More details about the features incorporated into the platform during the previous period can be found at paragraph 4.3.

Partner 4 Custodix

- Attended telco's and technical, review and consortium meetings
- Discussed the scope of the demonstrators for third review meeting

- Implemented, integrated and deployed the patient screening demonstrator and cohort selection demonstrator (year 3) in collaboration with the other INTEGRATE partners
- Finished work on the final iteration of the security framework, focussing on authorisation and auditing
- Overviewed the integration of the de different tools as WP2 lead
- Contribution in discussions about the ongoing work in WP3 (semantic approaches, data sources and common and local information models)
- Assisted setting up the security proxies of the semantic interoperability framework of WP3
- Contributed to deliverable 3.7: Final version prototype of the semantic interoperability layer
- Privacy enhancing processes and services (CATS tool, PIMS tool) were further developed and designed.
- Assisted setting up the final security environment for the sharing and collaborative tools and services
- Overviewed and contributed to deliverable 4.6: Integrated privacy enhancing services and processes
- Contributed to deliverable 4.5: Final versions of the tools and services supporting data sharing and collaboration
- Assisted with final security integration in the modelling framework of WP5.
- Deployed and configured the validation environments for IJB, Frankfurt, Maastro and Crete validations/pilots.
- Attended and recorded the validations/pilots at IJB, Frankfurt, Maastro and Crete.
- Contributed to deliverable 6.6: Report on the validation of the INTEGRATE technologies on pilot sites
- Contribution in discussions about exploitation (including sustainability plan).
- Contributed to deliverable D7.13 Final exploitation plan including sustainability plan
- Contributed to reporting and planning of the INTEGRATE project

Partner 5 IJB

- Submitting a request to our internal Ethics Committee to provide additional structured dataset for the evaluation sessions of NONA and DECIMA tools. Helping the mapping of our data to the terminology core dataset chosen within the project. Providing local servers and installing on it services used for evaluation/validation purposes.
- Evaluation/validating all INTEGRATE tools:
    - NONA: evaluation with 5 users (1 epidemiologist, 1 biostatistician, 1 project manager for clinical trials, 1 translational research coordinator/clinical data manager, and 1 clinical fellow).
    - DECIMA: evaluation with 3 research nurses.
    - Analysis: evaluation with 3 users (1 epidemiologist, 1 biostatistician, 1 project manager for clinical trials).
    - Central Pathology Review with 1 pathologist.

  Contributing to D6.6 for reporting evaluation activities related with all NONA, DECIMA, Analysis platform and Central Pathology Review tools evaluation/validation at IJB. Contributing to the newsletter in D7.5 to generally describe the evaluation process of INTEGRATE tools onsite.
- Contributing to D7.13 on final exploitation and sustainability plan for the tools and the INTEGRATE platform (security and semantic frameworks).
- Reviewing D3.6, D3.7, D4.5, D4.7, D6.5 and D6.6
- Participating to the INTEGRATE Dissemination video filmed onsite with potential users of the tools.
- Organising the Consortium meetings in June 2014 in Brussels, and attending the following one in October 2014 in Eindhoven. Participating to the INTEGRATE-EHR4CR convergence

meeting. Attending the evaluation/validation workshop in June 2014 in Crete. Finally welcoming tool developers and users to a 3 days evaluation/validation workshop in October 2014 in Brussels.

Partner 6 UPM

- Data load process of new clinical datasets into the INTEGRATE CDM: GBG and MAASTRO
- New mappings and core dataset extensions from new data sources
- Estimation of costs of the data load process
- Validation of the INTEGRATE semantic solution by external experts
- Publication of the normalization process in a high impact international journal

Partner 7 GBG

The contribution of GBG comprised two parts.

Part 1: GBG

- Preparing the data module for export of data for INTEGRATE from MedCODES (the GBG proprietary electronic data collection system)
- Preparing Data for transfer to Custodix
- Code Book generated for GBG Trial data for GAIN, TBP, GeparQuattro
- Review and finalized SNOMED mapping of the data
- Prepared Inclusion and Exclusion Criteria for GBG Trials GAIN, TBP, GeparQuattro
- Took part in various TCs in INTEGRATE
- Took part in technical, review and consortium meetings
- Organization of the Meeting at the GBG to present the tools of INTEGRATE and discuss the possibility of potential pilot sites

Part 2: Pilot Site Frankfurt KGU Womens Hospital data

- Organizing the first Meeting in Frankfurt KGU Hospital with Philips and Custodix to introduce the INTEGRATE Project and discuss potential collaboration
- Selection and export of trial data for INTEGRATE from Hospital HIS ORBIS system
- Anonymization and curation of the data
- Codebook generation
- Selection of concept annotation SNOMED-CT for the Frankfurt data
- Prepared Inclusion exclusion criteria for the Frankfurt trials
- Organization of the Evaluation/Validation workshop of the Decima tool at Frankfurt hospital with potential end users, namely study nurses, physicians and statisticians.
- Contributed to discussions about potential exploitation of the tools

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 third project review.
After the review the focus shifted on coordinating the remaining tasks and deliverables. We put a lot of effort in the development of the project tools and on the preparation of the validation and evaluation process.
This WP also prepared and coordinated implementation of the new clinical partner, 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).
This WP prepared all the necessary documents and carried out all the steps for the implementation of the last amendment.

4.2 *Changes in the consortium*

No other changes in the consortium

4.3 *Cooperation*

In this reporting period the collaboration in the consortium has been excellent, all partners knew that all the deliverables and other reports and requirements towards the consortium and to the EU should be omitted in due date.
The preparation of the demonstrators for the final 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 evaluation workshop in June (Crete) where several partners provided presentations, and published a last newsletter. Several papers were accepted for publication and we were invited to present our work in several events.
The finalisation 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
Febr 25	Service performance	Custodix	Telco
April 22	Technical preparation for evaluation	Philips	Maastro radiooncology clinic, Netherlands
May 11	Review Meeting	Brussels Jules Bordet	Belgium
Mar 13	Technical preparation for evaluation	Philips	Maastro radiooncology clinic, Netherlands

When	What	Organising Partner or Work Package	Where
May 14	Performance & deployment	UPM	Telco
May 20	Technical meeting cohort selection BIG/Philips	BIG	Brussels, Belgium
May 27	Maastro dataset processing	PHILIPS	Telco
June 11-14	Evaluation Workshop	Crete	Greece
June 17	Consortium Meeting	Brussels	Belgium
June 27	Maastro dataset processing	PHILIPS	Telco
July 4	Meeting Frankfurt Hospital	Frankfurt (GBG)	Germany
July 4	Evaluation planning	GBG	Frankfurt, Germany
July 8	Application integration	UPM	Telco
July 19	SNOMED CT Data Mapping	Neu Isenburg	Germany
July 24	Values vs. Categories data types	UPM	Telco
October 3	Evaluation Maastro, Session 1	Philips	Maastro radiooncology clinic, Netherlands
October 7-9	Evaluation KGU	GBG	Frankfurt, Germany
October 10	Evaluation Maastro, Session 2	Philips	Maastro radiooncology clinic, Netherlands
October 16,17	Consortium Meeting	Philips	Eindhoven, Netherlands
October 29,30	Evaluation IJB	IJB	Brussels, Belgium

Table 1 – Project Management Team Meetings

4.5 *Dissemination activities*

Presented papers:
J. van Leeuwen, A. Bucur, B. Claerhout, K. De Schepper, D. Perez-Rey, R. Alonso-Calvo.
BRIDG-based Trial Metadata Repository - Need for Standardized Machine Interpretable Trial Descriptions
Presented at the Health Inform (2014)

4.5.1 International articles

Paraiso, S., D. Perez del Rey, Anca Bucur, Brecht Claerhout, and Raul Alonso-Calvo. "Semantic normalization and query abstraction based on SNOMED-CT and HL7: Supporting multi-centric clinical trials." IEEE J Biomed Health Inform (2014) (In press) - http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6901196 - JCR Impact Factor = 2.072

A. Iliopoulos, I. Karatzanis, M. Tsiknakis, V. Sakkalis and K. Marias. “A collaborative central reviewing platform for cancer detection in digital microscopy images”. 6th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation - The CHIC Project Workshop (IARWISOCI). November, 3-4, 2014, Athens. {To be published}

Manikis, G., Maniadi, E., Tsiknakis, M., and Marias, K. “Multi-Modal Medical Data Analysis Platform (3MDAP) for analysis and predictive modelling of cancer trial data”. 6th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation - The CHIC Project Workshop (IARWISOCI). November, 3-4, 2014, Athens. {To be published}

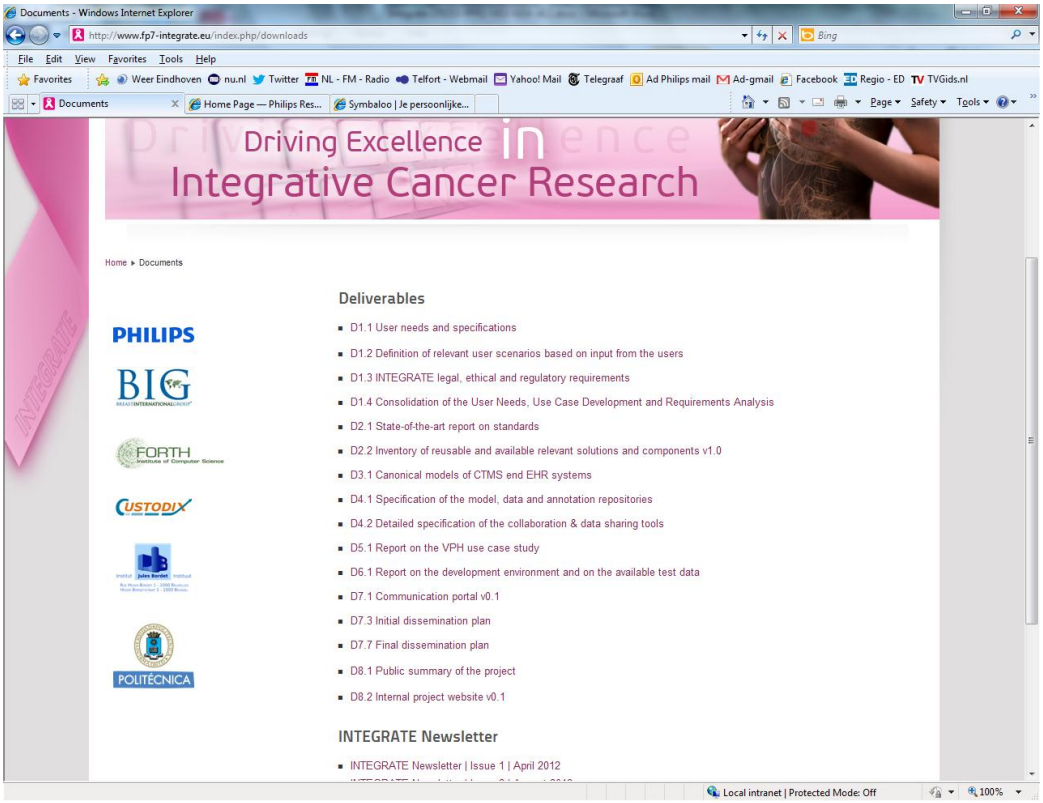
4.5.2 Presentations

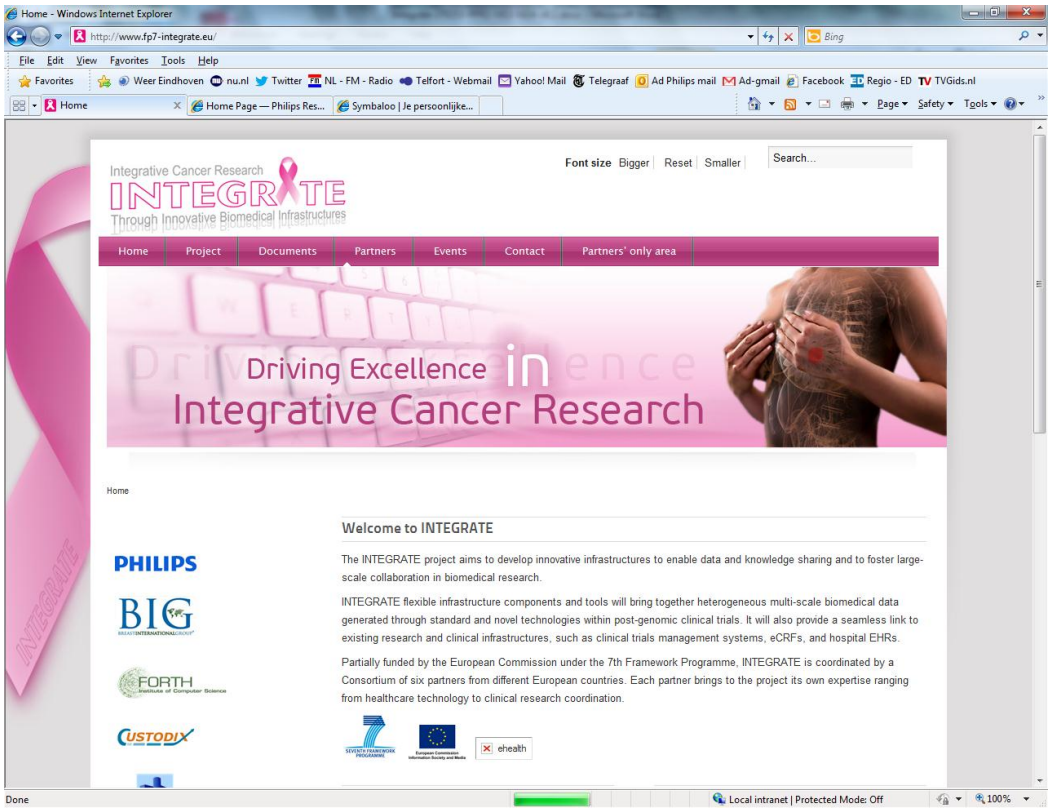
When	Where	Presentation Title	Audience	Presenting Partner(s)
26 & 27 May 2014	Clinical Research Informatics' Solution Days: Advanced IT-support for challenges in Clinical research, Düsseldorf, Germany	Central Review for Pathology Images		FORTH
3 & 4 Nov 2014	6th International Advanced Research Workshop on In Silico Oncology & Cancer Investigation, Athens, Greece	A collaborative central reviewing platform for cancer detection in digital microscopy images		FORTH
3 & 4 Nov 2014	6th International Advanced Research Workshop on In Silico Oncology & Cancer Investigation, Athens, Greece	Multi-Modal Medical Data Analysis Platform (3MDAP) for analysis and predictive modelling of cancer trial data		FORTH

4.5.3 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


Documents - Windows Internet Explorer
http://www.fp7-integrate.eu/index.php/downloads
Driving Excellence in Integrative Cancer Research
Home ▸ Documents
Deliverables
D1.1 User needs and specifications
D1.2 Definition of relevant user scenarios based on input from the users
D1.3 INTEGRATE legal, ethical and regulatory requirements
D1.4 Consolidation of the User Needs, Use Case Development and Requirements Analysis
D2.1 State-of-the-art report on standards
D2.2 Inventory of reusable and available relevant solutions and components v1.0
D3.1 Canonical models of CTMS end EHR systems
D4.1 Specification of the model, data and annotation repositories
D4.2 Detailed specification of the collaboration & data sharing tools
D5.1 Report on the VPH use case study
D6.1 Report on the development environment and on the available test data
D7.1 Communication portal v0.1
D7.3 Initial dissemination plan
D7.7 Final dissemination plan
D8.1 Public summary of the project
D8.2 Internal project website v0.1
INTEGRATE Newsletter
INTEGRATE Newsletter | Issue 1 | April 2012
PHILIPS
BIG
FORTH
CUSTODIX
POLITÉCNICA


Home - Windows Internet Explorer
http://www.fp7-integrate.eu/
Integrative Cancer Research
INTEGRATE
Through Innovative Biomedical Infrastructures
Font size Bigger Reset Smaller
Search...
Home
Project
Documents
Partners
Events
Contact
Partners' only area
Driving Excellence in Integrative Cancer Research
Home
Welcome to INTEGRATE
The INTEGRATE project aims to develop innovative infrastructures to enable data and knowledge sharing and to foster large-scale collaboration in biomedical research.
INTEGRATE flexible infrastructure components and tools will bring together heterogeneous multi-scale biomedical data generated through standard and novel technologies within post-genomic clinical trials. It will also provide a seamless link to existing research and clinical infrastructures, such as clinical trials management systems, eCRFs, and hospital EHRs.
Partially funded by the European Commission under the 7th Framework Programme, INTEGRATE is coordinated by a Consortium of six partners from different European countries. Each partner brings to the project its own expertise ranging from healthcare technology to clinical research coordination.
PHILIPS
BIG
FORTH
CUSTODIX
ehealth

5 Deliverables and milestones tables

TABLE 1. DELIVERABLES[7]

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
4.5	Final versions of the tools and services supporting sharing and collaboration	4	UPM	P	RE	M37	Yes	M37	
3.6	Study on the extension of the Core Dataset	3	Philips	R	PU	M38	Yes	M38	
4.6	Integrated privacy enhancing services and processes	4	Custodix	P	PU	M38	Yes	M38	
7.11	Project newsletter	7	BIG	R	PU	M38	Yes	M38	
3.7	Final version prototype of the semantic interoperability layer	3	UPM	P	PU	M42	Yes	M42	
4.7	Final version of the virtual collaboratory and its services	4	Forth	P	PU	M42	Yes	M42	
7.13	Final exploitation plan including sustainability plan	7	Philips	R	PU	M43	Yes	M43	
6.6	Validation of the INTEGRATE technologies on pilot sites	6	IJB	R	RE	M45	Yes	M45	

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 1. DELIVERABLES[7]

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
7.9	Report on the Integrate first workshop / launching event	7	Philips	R	PU	M45	Yes	M45	Resubmission of M34

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	M37	Yes	M37
MS8	Validation of the Integrate environment	Philips	M45	Yes	M45