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 2011 to 31 January 2012**

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: 02 April 2012

[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

1.1 *Introduction to the INTEGRATE project*

The INTEGRATE project aims to deliver 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.

Our 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. A unique quality of the INTEGRATE approach is the full commitment of the Breast International Group, as a partner in the project, to contribute patient data and the extensive basic, translational and clinical research expertise of their network to build solutions based on challenging but realistic use cases.

To achieve adoption of such scale we need to make use whenever possible of existing standards and terminologies. The INTEGRATE research infrastructure will store and manage a wide range of datasets, such as clinical, bio-molecular, imaging, models, annotations and other metadata, and put a strong focus on data privacy and security. The semantics of the clinical terms should be captured by standard terminology systems such as SNOMED CT, ICD, LOINC. The scalability of the solution needs to be achieved by modularization, e.g. instead of aiming at inclusion of the complete SNOMED terminology (more than 300 000 concepts) we will identify a core subset that covers the chosen clinical domain and the datasets in our repositories. Such core data set shall be validated both by clinical and knowledge engineering experts to assure proper coverage and soundness.

The INTEGRATE project aims to support 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. We will actively promote our approach and solutions in wide user communities and in other disease domains. The main objectives of the project are as follows.

BUILD INFRASTRUCTURE COMPONENTS AND TOOLS FOR THE STORAGE, SHARING AND MANAGEMENT OF DATA, INFORMATION, KNOWLEDGE AND MODELS

INTEGRATE will build reusable components based on which we will set up a dynamic infrastructure supporting our user community to store, manage and share biomedical data, models, tools, methodologies, and knowledge.

The current heterogeneity in healthcare-related research, manifest at the level of methodologies, workflows, data processing, and ICT infrastructures, tools and services, has significant negative impact on medical knowledge discovery, on the validation of clinical research results and on the adoption of the new results in clinical care for more predictive, individualized, effective and safer healthcare. On the ICT side, a main barrier is the lack of interoperability among relevant infrastructures, services and tools, due to the low adoption of common standards and terminologies. INTEGRATE will build solutions based on established standards for storing,

annotating and exchanging biomedical data, metadata, models and knowledge. The use of established standards and terminologies also supports the integration with existing infrastructures and the access to external relevant repositories adhering to those standards and terminologies.

BUILD TOOLS TO ENABLE COLLABORATION

We will support the description and execution of shared, multi-disciplinary and multi-site workflows. An important requirement for emergent collaborations is a shared workspace that is accessible to all collaborators. Ideally, this workspace should include all the important transactions that have taken place among scientific workers. In addition to a shared and open workspace, emergent collaboration requires meta-level information that highlights the significance of the transactions that are occurring or have occurred within a group of collaborators. Meta-level information that indicates the significance of transactions within a group can help a group determine which issues it should tackle first and which contributions should be given more weight than others.

BUILD PREDICTIVE MODELS AND A MODELLING METHODOLOGY AND FRAMEWORK

The INTEGRATE project proposes an approach and a methodology and builds a framework enabling the development of multi-scale predictive models of response to therapy in breast cancer, making use of multi-level heterogeneous data provided by clinical trials in the neo-adjuvant setting.

By proposing a methodology and building a framework for predictive models development within clinical trials we support more efficient development and validation of such models and faster adoption into clinical practice through the process of clinical trial validation. The aim of our biomedical modelling and simulation research will be focused on predicting the responsiveness of patients to specific drugs. This could lead to a more targeted and personalized treatment of the patient, avoiding at the same time a great deal of suffering due to unnecessary or ineffective treatment.

LINK TO EXTERNAL SOURCES OF INFORMATION

Biomedical research often relies on access to the many external repositories of data, information and knowledge. In the INTEGRATE project we will provide uniform standardized interfaces to external resources relevant to our user community. The external data will be used in the development of predictive models and to provide input to various analysis and communication tools.

ENABLE SEMANTIC INTEROPERABILITY TO EXISTING RESEARCH AND CLINICAL INFRASTRUCTURES

The ability to interface to existing medical research infrastructures is an important objective of INTEGRATE, as it is the basis for reaching a large community of users. Furthermore, to promote the fast adoption of the clinical research results into clinical care, we need to also target standards-based interoperability to existing clinical infrastructures.

From the technology viewpoint, in order to provide an efficient, robust and semantically interoperable solution, one needs to move from plain keyword matching to a combined approach where keywords are mapped to higher level concepts with clearly defined semantics. Such concepts are usually organized in concept hierarchies and include domain specific attributes and relations. Reasoning at this level, rather than at keyword level, is expected to enable us to move from error-prone lexical matching to more robust semantic-aware solutions.

PROTOTYPING AND VALIDATION OF THE INTEGRATE INFRASTRUCTURE COMPONENTS AND TOOLS

The capability of the INTEGRATE environment to achieve the above mentioned objectives will be demonstrated through prototypes implementing realistic clinical scenarios. These scenarios will also allow us to demonstrate the interoperability of our solutions with existing infrastructures.

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.

1.2.1 More predictive, individualized, effective and safer healthcare

The development in the INTEGRATE project of a modelling framework and of predictive models of response in the context of post-genomic clinical trials has the potential to contribute towards a 'modelling aided' optimal treatment design for cancer patients that will positively influence the treatment outcome. We also aim to initiate a paradigm shift in breast cancer treatment selection supported by cancer treatment planning, treatment monitoring and outcome prediction in silico. More effective healthcare is achieved by selecting the optimal therapy for an individual patient. Using the INTEGRATE platform it will be possible to efficiently run trials that could shed light in the optimal prediction for the candidate therapeutic schemes/schedules based on the patient's specific data.

1.2.2 Improved interoperability of biomedical information and knowledge

Taking into account the need for reuse, efficiency and wide-scale integration, the INTEGRATE project will have a strong focus on standards-based interoperability. We will build a flexible infrastructure consisting of interoperable components interconnected by standard interfaces, we will develop uniform access to relevant external resources and services, and we will insure interoperability with relevant existing infrastructures in clinical research and care. The INTEGRATE project will assess relevant existing infrastructures from an adopt, adapt and interoperate perspective.

1.2.3 Social Impact

Allowing for discoveries in the laboratory to be quickly transferred to the clinical management and treatment of patients can bring important societal benefits by significantly improving patient outcomes. Additionally, there is a strong need to enable the rational and personalized use of treatments that suit individual patients, and to move away from the current predominantly empirical approach.

Providing the necessary infrastructure, tools and services to the clinical research community will enable them to reduce costs by more efficiently setting up and carrying out clinical trials; better reuse of data, knowledge and tools; reduced duplication of efforts; easier access to all relevant information out of external sources; and more insightful generation of new research hypotheses. In the end this means quicker validation of new discoveries in clinical trials and transfer of new results into clinical care to become part of new treatments and improve patient outcomes. Providing standards-based infrastructure and services enabling biomedical researchers to build comprehensive molecular and clinical datasets will also support the definition of validated disease models that can improve the speed and efficiency of therapeutic drug development.

1.3 *Objectives during the first reporting period and progress*

In the first year of the project a crucial objective was to create the right conditions to enable the consortium to achieve the desired objectives. In this context, an essential activity was the definition by our clinical users of realistic and detailed clinical scenarios that are of top priority. That enabled us to ground our research in the realities of the clinical community and to shape our technical developments on their needs and priorities. Based on these scenarios we elaborated technical use cases and extracted requirements that will drive the technical work in the project. Next to the user needs, we have also focused on the legal, ethics and regulatory requirements which are of key importance when building solutions to be applied in the clinical field. This work guarantees that research and outcome of INTEGRATE are in line with current legal and ethical European legislation and with relevant regulatory frameworks.

Another important task was to build the development environment. The IT experts of our clinical partners provided schema-level data describing the structure and content of both the clinical and research infrastructures. Additionally, sufficient (and well-matched) patient data from both care and research were provided to enable the development of our first prototypes. This data was prepared according to the legal, privacy and security requirements. The data was subsequently stored in a "surrogate" data warehouse in a data model that adheres to the HL7 RIM standard.

The initial architecture of the INTEGRATE environment has been defined, together with all relevant use cases. This also includes the initial security architecture which is an integral part of the overall architecture of the system.

Of core importance for our environment is the development of a suitable semantic solution. In this reporting period we have defined the core dataset that sufficiently covers our application domain based on existing terminologies: SNOMED-CT, LOINC and MedDRA. The first solution to the semantic layer has also been defined and was implemented within the first INTEGRATE demonstrator.

We developed prototypes that cover several relevant clinical scenarios: molecular screening (patient recruitment), underlying semantic interoperability environment, collaborative environment for pathology review and data analysis tools. The implementation of the demonstrators has been a collective effort that involved all project partners and required close collaboration.

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

2.1 *Project objectives for the period*

The main tasks during the first year of the INTEGRATE project (February 1st, 2011 – January 31st, 2012) have focused on the following objectives:

- Define the clinical scenarios, the user needs, the legal, ethics and regulatory requirements.
- Carryout extensive state of the art reviews concerning relevant technologies, tools and services.
- Select relevant standards for the development of the INTEGRATE solutions.
- Define the technical use cases and the first integrate architecture.
- Build the development environment, i.e. create the set-up allowing the INTEGRATE infrastructure components and tools to be designed and built.
- Build the initial semantic interoperability layer based on the definition of the core dataset covering the clinical domain of interest and of the information models of the sources.
- Set up a coherent and efficient way of working in the consortium, including through the use of the project wiki and of the internal document repository.
- Build the INTEGRATE website.
- Set up collaborations with relevant initiatives and disseminate our results.

All the objectives for the first year have been achieved in this reporting period. An overview of the progress is presented in the next section. Further, the document details the achievements per each work package of the project.

2.2 *Work progress and achievements during the period*

In the first year of the project a crucial objective was to create the right conditions to enable the consortium to achieve the desired objectives. In this context, an essential activity was the definition by our clinical users of realistic and detailed clinical scenarios that are of top priority for the users. That enabled us to ground our research in the realities of the clinical community and to shape our technical developments on their needs and priorities. Based on these scenarios we elaborated technical use cases and extracted requirements that will drive the technical work in the project. Next to the user needs, we have also focused on the legal, ethics and regulatory requirements which are of key importance when building solutions to be applied in the clinical field. This work guarantees that research and outcome of INTEGRATE are in line with current legal and ethical European legislation and with relevant regulatory frameworks.

Both the scenario definition and the requirements analysis follow an iterative process during the project. The iteration completed in year one enables the development of the first versions of the prototypes which will be further refined based on evaluation with the users and on the definition of new scenarios and requirements. This approach has the goal to develop solutions that closely satisfy the needs of the users.

Another important task was to build the development environment, i.e. to create the set-up allowing the INTEGRATE infrastructure components and tools to be designed and built. The IT experts of

our clinical partners provided schema-level data describing the structure and content of both the clinical and research infrastructures. Additionally, sufficient (and well-matched) patient data from both care and research were provided to enable the development of our first prototypes. This data was prepared according to the legal, privacy and security requirements. The data was subsequently stored in a "surrogate" data warehouse in a data model that adheres to the HL7 RIM standard.

The initial architecture of the INTEGRATE environment has been defined and described in deliverable 2.4, together with all relevant use cases. It is a comprehensive document describing all aspects relevant for the development of the INTEGRATE environment and tools from an architectural perspective. This document also includes the initial security architecture which is an integral part of the overall architecture of the system.

Since we target wide-scale adoption, we make use whenever possible of existing standards. During the first year of the project we have identified and evaluated all VPH, ICT, healthcare IT, modelling and data standards, and ontologies/terminologies relevant for the INTEGRATE environment.

To support reuse and large scale integration we evaluated existing components and tools that could be integrated within the INTEGRATE framework and could provide useful functionality within the defined clinical scenarios. During the project, to make efficient use of our resources, we will evaluate whenever applicable the possibility of extending the functionality of existing standards-based and widely adopted components to suit the needs of our user community, if the effort of building upon existing solutions is not prohibitively higher than developing targeted new tools.

Of core importance for our environment is the development of a suitable semantic solution. In this reporting period we have defined the core dataset that sufficiently covers our application domain based on existing terminologies: SNOMED-CT, LOINC and MedDRA. The first solution to the semantic layer has also been defined and was implemented within the first INTEGRATE demonstrator.

Prototypes were developed that cover several relevant clinical scenarios: molecular screening (patient recruitment), underlying semantic interoperability environment, collaborative environment for pathology review and data analysis tools. The development of the demonstrators has been a collective effort that involved all project partners and required close collaboration.

During the first 12 months of the project 21 deliverables (including the first project newsletter) were planned. All these deliverables were completed and submitted for review. Out of those, two deliverables of WP2 Architecture and integration (D2.3 and D2.4) were merged and submitted as a single document in order to provide a coherent and complete picture and to avoid content overlap.

2.2.1 WP1 User needs and requirements (IJB)

2.2.1.1 Objectives (of the reporting period)

The main objectives of the WP for this period were to identify the users and their needs, to define and prioritize comprehensive user scenarios, to contribute to the development of use cases, and to define legal and regulatory requirements.

Work for this WP has been done in close collaboration with BIG, with many meetings held in common.

2.2.1.2 Status/progress towards objectives WP1 (per Task)

Task 1.1 Identification of the users and their needs

User needs for the INTEGRATE environment were initially gathered through interviews and discussions with leading oncologists and researchers from the NeoBIG research program that promotes data sharing in the context of neoadjuvant breast cancer therapy.

More detailed user requirements for the INTEGRATE environment were elicited from a larger panel of potential end-users and advisors from BIG and IJB, including oncologists, translational researchers, clinical trial administrators, legal advisors, health IT specialists and data analysts.

Working reunions were held weekly within BIG and IJB, and regularly through teleconferences and face-to-face meetings between all members of the consortium. The opinions of external advisers were also solicited and gathered during some of these meetings and during other events such as conferences.

The different categories or roles of end users of the INTEGRATE platform have been identified during this reporting period. These roles include clinicians, core laboratory staff, administrators and researchers from academia and pharmaceutical companies. Sub-categories of users have also been identified. For example, clinician has been sub-divided into investigator, radiologist, pathologist, clinical research nurse, etc.

Access requirements and access rights associated with these user roles have also been identified and a document describing these requirements and rights has been drafted.

The main product of these activities of IJB toward completion of this task is deliverable D1.1 “User needs and specifications for the INTEGRATE environment”.

Task 1.2 Definition of user scenarios

A large part of the effort during this reporting period concerned definition of the user scenarios. Seven important scenarios have been identified. These are:

Scenario 1 - Molecular testing (including biotracking, eligibility criteria testing, and link to the eHR)
Scenario 2 - Uploading of completed trial data to the data-sharing platform
Scenario 3 - Research queries on completed trial trial data
Scenario 4 – Retrospective central review of pathology
Scenario 5 - Data administration
Scenario 6 - Sharing of predictive models
Scenario 7 - Integrated analytical tools

The molecular testing scenario corresponds to a user need that was not identified during the drafting of the project. It acknowledges the need, in addition to data sharing, for a tool to support identification of eligible patients for clinical trials with molecularly-targeted anti-cancer agents. The molecular screening service of INTEGRATE will seamlessly accommodate multiple hospitals and testing laboratories and ensure a consistent reporting and tracking of biological samples (encompassing tumor tissues, blood and derivatives) across Europe, with real time access of molecular screening tests results to investigators, clinicians and trial administrators. This will offer a unique opportunity to enhance patient selection and increase the numbers of patients enrolling in clinical trials in a cost and time efficient manner. Additionally, the molecular screening data (clinical, pathological and genomic) could also be linked to the long-term clinical data within a specific clinical trial.

An important sub-task to ensure that user scenarios can be exploited correctly is agreement on a shared, unequivocal vocabulary. To this end, a glossary of terms from the user scenarios has also been constructed.

These scenarios and the glossary of terms are presented in deliverable D1.2 “Definition of relevant user scenarios based on input from the users”.

Task 1.3 Legal and regulatory compliance requirements
Legal and regulatory compliance requirements were given considerable consideration throughout the reporting period. Legal advisors and clinical trial specialists from BIG participated in the working meetings related to user requirements and user scenarios, allowing early identification of potential legal and regulatory issues.

Several meetings specifically devoted to the discussion of these matters were held within IJB and BIG. External advisors with a relevant experience in data sharing for clinico-genomic trials (Sage bionetworks, I-SPY/TRANSCEND) were also identified and contacted and their recommendations were gathered during face-to-face and teleconference meetings.

Deliverable D1.3 “INTEGRATE legal, ethical and regulatory requirements” presents all the relevant topics that have been identified, including topics related to data protection, informed consent, intellectual property rights and contractual matters.

Task 1.4 Definition of the relevant use cases and requirements analysis
IJB participated in the elaboration of use cases, for deliverable D1.4 “Consolidation of the user needs, use-case development and requirements analysis (draft)”. We revised and commented on some use cases especially, those dealing with the screening scenario and cohort selection. We also defined use cases related to the problem of synchronization between, on one hand, the E.H.R. replica stored on the intermediate Integrate platform and, on the other hand, data coming from the clinical environment. We focused on some use case issues about the management of access rights (especially, the granularity level) and proposed some suggested security model.

2.2.1.3 Deviations from the DOW and corrective actions

During the user requirement gathering phase, a new user need was identified which is not described in the initial description of work and is reflected by scenario 1 “Molecular testing”. This new requirement was discussed with the other partners and deemed in scope with the rest of the project, as it will leverage the same semantic interoperability layer and many of the software components developed for data sharing. It will also provide a perfect “test bed” for many of these software components.

As this new user requirement has been embraced by INTEGRATE partners and is in line with the rest of the project, no corrective actions are required.

2.2.1.4 Planning next period

With respect to WP1, the next period will be devoted to the consolidation of user needs, use cases and requirement analysis. During this period, IJB and BIG will continue to provide to the other partners clinical information relevant to the interpretation of the user scenarios and use cases.

A refinement of access requirements and access rights for data sharing is also needed and will be pursued by BIG and IJB, in collaboration with the other INTEGRATE partners, during the next period.

Finally, legal and regulatory compliance requirements will also be consolidated during the next period, leading to deliverable D1.3 “INTEGRATE legal, ethical and regulatory requirements”.

2.2.2 WP2 Architecture and integration (Lead: Custodix)

2.2.2.1 Objectives (of the reporting period)

Finalizing the first document of the initial architecture which integrates the different components and tools (modules and services) provided by the INTEGRATE project.
This contains a first iteration of the security/component/data/metadata/information/... models and semantic solutions, based on the provided stakeholders scenarios and requirements defined in WP1. Also the identification and evaluation of the relevant standards/technologies and re-useable/available relevant solutions need to be investigated and written down in the corresponding deliverables. Finally a status update on the INTEGRATE implementation work needs to be given.

2.2.2.2 Status/progress towards objectives WP2 (per Task)

Task 2.1 Identification and evaluation of relevant standards

- The state-of-the-art document for deliverable D2.1 (month 9) was finalized and sent to the EU for review. It contains research in following topics:
    - Relevant ontologies and vocabularies
    - Semantic repositories
    - Automated reasoning in the semantic web
    - Ontology mediation, alignment and merging
    - Ontologies for the life sciences
    - Data and ontology sources
    - Query languages for semi-structured data
    - Security and privacy standards
- This task is now completed

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

- D2.2 (Inventory of reusable/available relevant solutions and components) was written, finalized, internally reviewed and send to the EU. Each partner contributed to one or more of the following topics:
    - Data management
    - Reasoners
    - Ontology mediation
    - Security and privacy standards
    - Biotracking
- This task is now completed

Task 2.3 Design and implementation of the INTEGRATE reference architecture

- Several brainstorm meetings were held, defining the core influential aspects of the INTEGRATE architecture (i.e. semantic approach, information models, possible interface technology)
- A table of content was created for the architectural document that assigns the responsible contributor to each section.
- Different sections were written in the architectural document:
  - Abstract
  - Introduction
  - 10000 feet view
  - Different architectural views (functional view, information view, ...)
  - Link between the architectural views
- D2.3 (Initial report on the INTEGRATE security framework) that was originally defined as a separate deliverable in task 2.4, was merged in D2.4 (Initial system architecture and implementation status).
- The initial implementation status was written and added as part of D2.4

Task 2.4 Security for dynamic collaborative environments

- The initial security framework was researched and written down as part of D2.4

2.2.2.3 Deviations from the DOW and corrective actions

- D2.3 (Initial report on the INTEGRATE security framework) was merged in D2.4 (Initial system architecture and implementation status).

2.2.2.4 Planning next period

- Start of the next iteration of the architecture, detailing the high-level view of the current specified architecture.
- Define the interfaces and links between the components of the initial implementation
- Build the first implementation of the framework
- Start writing the integration guidelines

2.2.3 WP3 – Data Models and Interoperability (UPM)

2.2.3.1 Objectives (of the reporting period)

The main objective of this WP is to enable interoperability within the INTEGRATE environment, providing services such as data extraction, transformation and mapping among data models of clinical infrastructures. The objectives for the first part of the reporting period were mainly focused on the core dataset (Task 3.1) and information models (also called common data models) (Task 3.2). In the second part of the period, the mapping formalism (Task 3.3) and semantic interoperability layer (Task 3.4) has been also tackled to provide the first version of the platform.

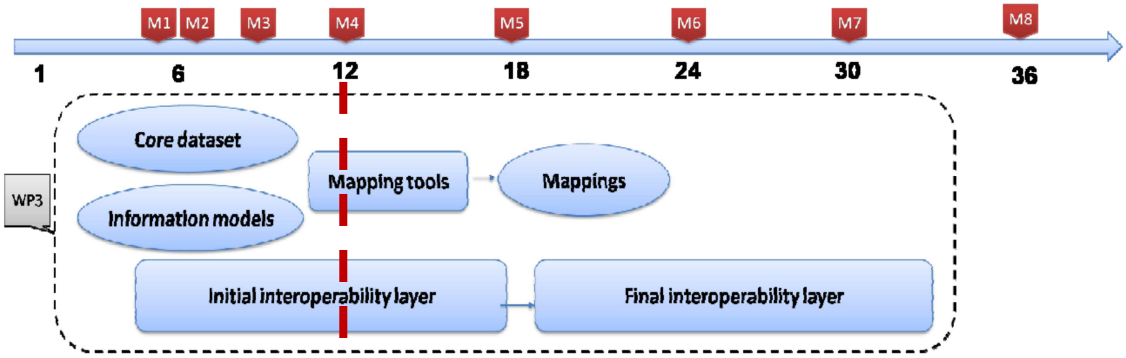

Fig 1. Work Package 3 components and planning

The core dataset is the shared vocabulary, including the corresponding relationships (also known as ontologies), required to accomplish the semantic interoperability among heterogeneous systems. To integrate heterogeneous data models from different sources, mappings linking the core dataset and common data model concepts are required. The semantic interoperability layer will use these mappings to provide a uniform and semantically interoperable platform for Electronic Health Records and Clinical Trial data (including external sources).

2.2.3.2 Status/progress towards objectives WP 3 (per Task)

Task 3.1 Definition of the semantic core dataset

Within the core dataset task, a set of relevant domain concepts, describing the semantics of the clinical domain, was identified. Since INTEGRATE aims to reuse available terminologies, different standardized terminologies have been analyzed as core dataset candidates, i.e. SNOMED CT, ICD-10, MedDra, LOINC and MeSH. The analysis suggests that the INTEGRATE platform may require various terminologies to cover the scenarios described by the user requirements, in areas such as adverse events or laboratory test results. However, SNOMED CT may provide the majority of the core dataset concepts, since there exist mappings from MedDra (i.e. adverse events) and LOINC (i.e. laboratory test results) terminologies to SNOMED. When new vocabularies are required, these terminologies will be handled by selecting a “default” vocabulary for each area, or by including new terms as SNOMED CT extensions (e.g. Clinical trial eligibility criteria unique codes, where there is a lack of standardized classification).

To provide an environment allowing reasoning, required to perform semantically-aware queries, different ontology languages, such as RDF, SKOS or OWL, have been analyzed to store the core dataset. Existing tools provide method to automatically transform certain terminologies to such languages, i.e. SNOMED CT to OWL. Database models, that have been extended to store the core dataset, have been also analyzed. They offer an improved performance but more complex or limited reasoning.

A preliminary set of terms to develop the core dataset have been manually identified by clinical partners and has being compared with available terminologies. Results and comparison with other core dataset from similar projects suggest that we should minimize the use of post-coordination or extensions to describe new concepts. Automatic methods to extract core dataset terms from case report forms and eligibility criteria have been also performed with promising results that will be used in subsequent versions of the core dataset. For the first implementations, SNOMED CT, with a limited set of extensions has been selected.

Task 3.2 Definition of the information models of the clinical and research infrastructures
The aim of a Common Data Model is to provide a canonical view, reflecting the content and the structure of each data source. The following features, among others, are being taken into account to design the common data model of INTEGRATE: modeling capabilities to store data from requirements, performance, potential for reasoning and minimizing structural modifications when new sources are integrated. There is a trade-off between minimizing future modifications, requiring a simple design with dynamic updates for new sources, and performance. Simple solutions will have lower performance, while complex schemas will require more changes in the future. A set of data models candidates were analysed: i2b2 (informatics for integrating biology & the bedside), OMOP (Observational Medical Outcomes Partnership) and HL7 RIM (Health Level 7 Reference Information Model). The outcome suggested that i2b2 or OMOP models would require extensions, while just a portion of HL7 RIM would not be used. For the first implementation, a simplified version of HL7 RIM with a SPARQL endpoint has been developed.

Tools to Extract Transform and Load (ETL) original data sources into the common data model have been required. A state of the art has been carried out, regarding ETL tools available. Pentaho Kettle and Talend are open source projects that can be used for this task within INTEGRATE. Both have a large community of users and have been previously used within biomedical integration projects. For the first implementation, Pentaho Kettle have been applied to populate the INTEGRATE common data model with anonymized EHR data.

Task 3.3 Semantic formalism, mapping tools and mapping implementations
This task aims to identify the requirements to link core data set concepts, EHR sources and the clinical trial management system. According to the INTEGRATE architecture, there is an instance of each source stored at each institution, following the common data model from previous task. Mapping requirements were considered to design the core data set and common data model. We have also analyzed user scenarios and the corresponding use cases to identify requirements for reasoning.

The mapping implementation of the INTEGRATE platform have required the supervision of domain experts to link concepts and data models. Schema transformations, driven for such mappings, are performed by ETL tools and the semantic interoperability layer, to provide a uniform view of the data sources within INTEGRATE.

Task 3.4 Design and implementation of the semantic interoperability layer
The semantic interoperability layer execute mappings during the data retrieval phase, instantiating the semantic concepts with patient data and/or clinical trial data. It enable the linkage between patient data at INTEGRATE repositories and patient data at existing clinical and research systems. This capability allows tools and services of INTEGRATE to access all the necessary data from INTEGRATE repositories and relevant existing research and clinical systems in a semantically-aware and uniform way.

The core dataset, the common data model and the mapping approach are included within the semantic interoperability layer. During the reporting period, the suitability of the core dataset and common data models to perform the required integration have been the focus of this task. In addition we have carried out an analysis of current query languages that can be used to retrieve data from the platform: RDQL, RQL, SerQL, SPARQL. For the first implementation, SPARQL is

used to query the semantic interoperability layer and SESAME repository was applied to provide SNOMED CT reasoning.

Task 3.5 Standards-based uniform access to external sources
A solution based on uniform interfaces and existing standards is the objective of this task, to enable that INTEGRATE tools and services can access data and knowledge from external repositories. Structured data sources can be queried by using the adopted standards, while unstructured datasets should be transformed into a structured format. Even with structured data such as EHRs, some information is still stored as free text. Simple transformations can be performed using ETLs capabilities. EHRs from the clinical partners are at the moment, the external sources of the platform. Further versions of the platform will include molecular information from samples.

2.2.3.3 Deviations from the DOW and corrective actions

There are not significant deviations from the DoW apart from expected delays to obtain surrogate data from the clinical side. The WP3 have been therefore mainly focused on Task 3.1 and 3.2 during the first part of the period. During the second part of the reporting period, the high complexity of HL7 data model adopted produced and increment of WP3 effort. Some of the WP4 work load has been moved to the WP3, since previous work on the core dataset and common data models is required to provide tools enabling data and knowledge sharing (Task 4.2).

2.2.3.4 Planning next period

After the development of a first implementation of the semantic interoperability layer, including the eligibility criteria matcher, feed-back from annual review (4th May) will be used to refine the current approach (figure below).

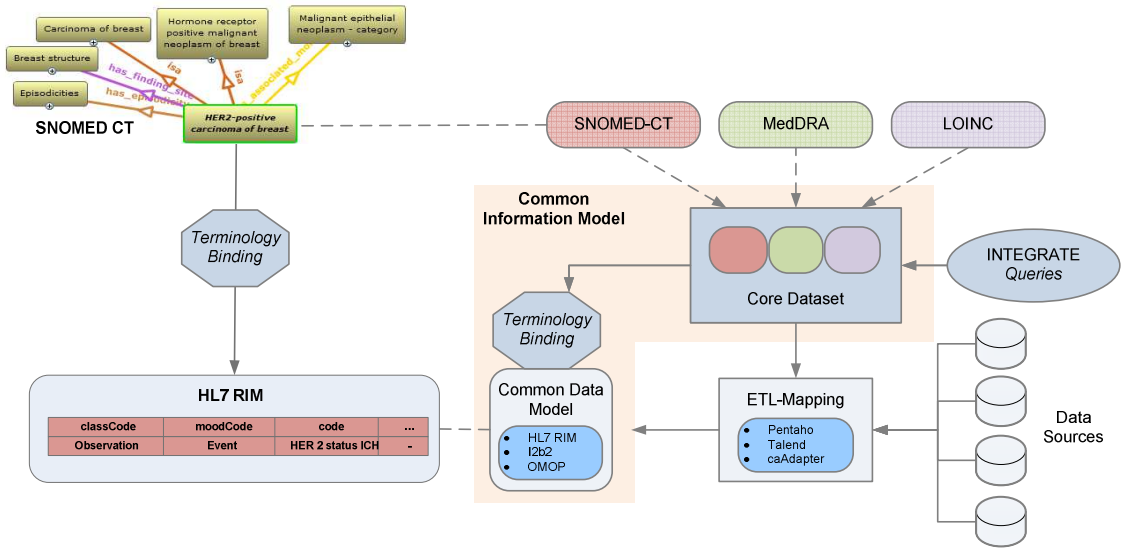

Fig. 2. Current approach of the semantic solution within the INTEGRATE project

Within the next period, semantic solution implementations will also cover the cohort selection scenario in addition to the criteria matcher. Additional reasoning scenarios will be analyzed to

provide a semantically sustainable platform. Currently, the core dataset has been manually identified by clinical users and an initial exploration has been applied to public eligibility criteria. Results will be analyzed by clinical experts and compared to manual results to improve the automatic identification process. A common data model based on HL7 RIM, with a SPARQL endpoint, has been designed and implemented. It has been initially populated with EHR surrogate data from the IJB partner. In the next period, additional elements will be included within the common data model to cover EHR data required to solve further eligibility criteria. As shown in the next figure, data models of current systems have been already provided (D3.1 month 12).

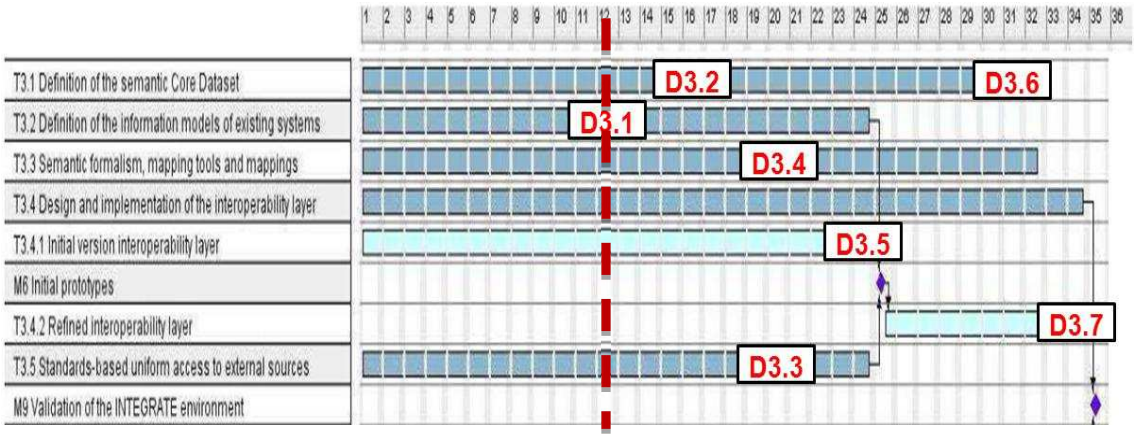

Fig 3. WP3 task planning according to the DoW

A first implementation of the common data model to cover current systems has been developed during the reporting period. In the next period, the first release of the core dataset will be delivered (D3.2 month 16), together with the semantic formalism and uniform access to external sources (D3.4, D3.3 month 20) and the initial version of the interoperability layer (D3.5 month 24).

2.2.4 WP4 Sharing and Collaborative Tools and Services (Lead: FORTH)

2.2.4.1 Objectives (of the reporting period)

The main objective of this work package is to design and develop a virtual “collaboratory” to be initially deployed and demonstrated for the BIG scientific community. Our definition and vision of scientific collaboratories is “a network-based facility and organizational entity that spans distance, supports rich and recurring human interaction oriented to a common research area, fosters contact between researchers who can be either known or unknown to each other, and provides access to data sources, artifacts and tools required to accomplish research tasks.” For this specific period a number of possible scenarios were examined regarding pathology remote collaboration concepts within BIG. Several technical discussions took place regarding the different possibilities to establish a robust collaboration framework amongst BIG participating pathologists.

2.2.4.2 Status/progress towards objectives WP4 (per Task)

Task 4.1 Model, data and annotation repositories
This task will develop the model library infrastructure using a common, XML-based format for the model with associated metadata description (relevant information from 3rd party data resources or literature, annotation with controlled vocabularies, results
of reference analysis etc.).
Initial work on this task included the preparation of D4.1 Specification of the model, data and annotation repositories by partner Philips dealing with wide variety of data available within INTEGRATE, including clinical trial data, imaging studies, molecular (genetic) data and clinical care data, providing access to high volumes of heterogeneous biomedical data at a wide variety of spatial scales. Predictive models and simulations – stored in the model repository – will exploit this wealth of (multiscale) biomedical data to – for instance – predict therapy sensitivity for patients, and unprecedented meta-analyses can be performed across trials. In order to efficiently access data and models, metadata and annotations are stored in metadata repositories. The work in this task first deals with the INTEGRATE scenarios (involving Molecular screening, Trial meta analysis, Predictive modelling, Central Review, etc.) in combination with relevant formats, standards and guidelines to arrive at the requirements for the data repositories, (predictive) model repositories and annotation repositories.

Task 4.2 Tools enabling data and knowledge sharing
This task will be focused on delivering a set of services and tools of the virtual collaboratory of the BIG community exploiting innovative community annotation, crowd-sourcing and scientific accreditation tools as well as semantic approaches to interoperability and automated reasoning. In addition, in order to support clinical research, the task will develop tools enabling the clinical research community to collaboratively define research protocols and carry out all the necessary regulatory and administrative steps to set up a clinical trial.
FORTH provided guidance in the technical discussion concerning the slide scanner that will be acquired by BIG, taking into account the collaboration parameters that the workflow between pathologists must have.
The list of the slide scanners of interest is the following:

- Roche (Scanner: iScanCoreo, Software: Virtuoso Digital Pathology Application Software)
- Aperio (Scanner was not specified)+Definiens Software
- Hamamatsu (Hamamatsu HPF-Nanozoomer RS2.0 PACK)
- Leica (scanner SCN400)
- Olympus (scanner & software were not specified)

In evaluating the Aperio platform, a webinar from Aperio was carefully studied. The first evaluation is that the software seems capable and expandable but the most suitable to determine that would be the Clinicians themselves.
BIG finally proceeded to the purchase of a Hamamatsu device, and has provided the first 3 test images at FORTH, in Hamamatsu format, on March 2, 2012.

Task 4.3 Tools enabling collaboration
An important requirement for emergent collaborations is a shared workspace that is accessible to all collaborators. Ideally, this workspace should include all the important transactions that have taken place among scientific workers. In addition to helping a group of collaborators learn from past transactions and take the best step forward, the workspace will facilitate stigmergy, i.e., it will

enable a worker's contribution to stimulate others to build on that contribution without any direct communication between the workers.
FORTH initially suggested to create a central imaging review tool for BIG trials. Eventually BIG suggested to drive the development of a collaboration environment for pathologists instead.

In the webinar that was organized on the 18th of July 2011, from Pixcelldata an interesting software platform was presented, Collibio. Subsequently, a project-internal discussion between FORTH and BIG took place. Based on the details of the presentation and the discussion that took place, we have concluded that the pros and cons of the specific platform are:

Pros:

- The main concept of the platform is USERS and IMAGE SERVERS, which are brought together in PROJECTS or WORKSPACES. This appears to be an interesting approach to support collaboration.
- Multiplatform: A web based environment, accessible from any type of desktop operating system (because it is flash based, mobile operating systems are excluded, except android).
- A Remote pathology viewer: Images are not downloaded locally, but are accessed directly from the database of the Image Server of the slide scanner (it has to be supplied from the manufacturer itself). Because the images are not downloaded locally the user does not have to wait for them (which might be very time consuming).
- The images are available as soon as they are scanned.
- Support for users with configurable roles and permissions that can share projects contacting multiple images.
- Collaboration capabilities for asynchronous reporting.
- It has a very configurable mechanism to create custom forms.
- Supports navigation modes (zoom in and out) and Annotate mode.
- In general it seems to be a modular platform, providing APIs for image access and upload.

Cons:

- The database which stores the links between the images and the metadata information (annotations, forms, etc) is handled by Pixcelldata, in their farm.
- It does not support a centralized image repository.
- It does not support all the major Image scanners (e.g Roche bio-imaging platform is NOT supported).
- It does not have any predefined form or pathology. The forms are generic, and cannot be imported or distribute the template of a form between the users.
- Does not support SYNCHRONOUS COLLABORATION, i.e. multiple users collaborating and interacting on the same image simultaneously (although many users can open the image at the same time, their number and the efficiency of operation is set by the total amount of users accessing the image server and the hardware capabilities of the image server. In case that the collaborators try to do a synchronous operation, database hierarchy and logic is applied and the first user gains the lock, while the changes from the rest users are rejected until the lock is released).
- The segmentation and annotation functionality of the image viewer is limited.
- The export formats are limited to only one (EXCEL). Lack of XML export is crucial for collaboration with other toolkits. Some XML export functionality is planned, but we do not know when it will be available.
- Image analysis functionality is not supported.

Based on the experience FORTH suggested a number of alternatives.

- Alternative A:
  A commercial off-the-shelf solution is selected to fulfill the collaboration needs in INTEGRATE – to whatever degree available platforms allow.
- Alternative B:
  FORTH undertook the responsibility to coordinate the effort (in the context of WP4) to develop a solution that encapsulates as much functionality of the investigated products as is considered necessary, in order to create a customized, INTEGRATE specific, collaboration environment that is able to handle BIG's requirement for a centralized data (including pathology image) warehouse. Such an approach would enable to integrate additional functionalities including image analysis (e.g. for cross-image intensity normalization, estimation etc.) and support for synchronous collaboration.

FORTH has the resources required to provide such a solution that will be tailored to the specific needs of BIG and the INTEGRATE project. FORTH can also commit to provide the necessary technical support of this dedicated collaboration and analysis platform, even after the end of the INTEGRATE project.
In order to accelerate the decision process and to clarify the situation regarding the central review, a face to face meeting was organised on December 20 2011, in Brussels, where FORTH presented many available open source technologies & tools that could be the basis for a state of the art central review platform, along with a draft plan for implementation. BIG on the other hand announced that it prefers to use an under development project of the Univercity of Liege, called Cytomine, which upon its release will bare a commercial / proprietary license. During the face to face meeting, representatives from the university of Liege demonstrated their software at FORTH. BIG has decided to use the Cytomine software as an image viewer for the pathology images and also for annotating the pathology images. All the rest functionality of the collaborative platform will be developed by FORTH. Cytomine is currently under investigation with what type of license it can be provided to the BIG. During the time where this report was compiled Cytomine was not available for integrating in the platform, API and/or other technical details have not been provided, and details like these create doubts on which degree the Cytomine software could be seamlessly integrated with the rest of the platform.
FORTH has provided the detailed use cases and has specified the architecture of the collaboration environment in the deliverables D1.4 and D 2.2, and based on these development has already started.

Task 4.4 Privacy Enhancing Processes and Services
Initial work in this task is carried out by partner CUSTODIX aiming to ensure privacy and security within the specific architecture of INTEGRATE and will in essence input from the security framework discussed in the architectural deliverables of INTEGRATE as a whole.

2.2.4.3 Deviations from the DOW and corrective actions

No major deviations time wise. On the other hand the collaboration concepts where somehow different between the conceptual level presented in the DoW and the actual needs of BIG. However, FORTH in collaboration with BIG partners tried to balance both in the discussions regarding the collaboration environment that needs to be established within BIG pathology central review and now it has become clear what exactly needs to be done in the collaboration tool development front.

2.2.4.4 Planning next period

With a basis of user requirements and needs, and relying on the specified use cases and the selected tools, the development process for the collaboration tools and services has started.

2.2.5 WP5 Support for predictive modeling and simulators (FORTH)

2.2.5.1 Objectives (of the reporting period)

The main objectives of this work package are to propose an approach and a methodology and to build a framework enabling the development multi-scale predictive models of response to therapy in breast cancer, making use of multi-level heterogeneous data provided by clinical trials in the neo-adjuvant setting. The models developed in this WP will be based on realistic clinical research scenarios, as outlined in the neoBIG research program, and on comprehensive data sets from rigorously conducted clinical trials. The models will also be used to validate the INTEGRATE approach and the appropriateness of the INTEGRATE infrastructure.

Our effort for this reporting period has been mainly focused on the definition of the clinical scenarios (questions) for the INTEGRATE VPH use case, and the implementation of the predictive analysis framework that addresses a part of these clinical scenarios using multi-modal data. Moreover, during the reporting period, a complementary analysis has been pursued enabling the statistical analysis of the data, providing an integrated analysis platform where clinical scenarios related to the statistical analysis and the prediction modelling can be tackled directly within the platform.

Moreover, a demonstrative simplified version of our analysis framework has been also decided for the needs of the first review.

2.2.5.2 Status/progress towards objectives WP5 (per Task)

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

This task uses as input the clinical scenarios elaborated in WP1, based on which will develop VPH-focused scenarios. This Task takes input of Task1.2, Task3.4-5 in order to exploit the possibilities of sharing data both provided by our clinical partners within the INTEGRATE environment but also from public databases. After several discussions data from the TOP trial are gradually becoming available including clinical and genomic data generated during the clinical trial. The report of the VPH use case has been delivered in month 8.
The scenarios that have been defined and reported in D.5.1 address the research questions as described in D.1.2, including statistical analysis in clinical, genomic and imaging biomarking data, and sophisticated pattern recognition techniques for integrating and selecting the most relevant heterogeneous data that contribute to the prediction of the tumor response to a specific regimen.

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

The consortium decided to share the BIG data from the TOP trial in order to develop novel predictive models and investigate new biomarkers. The genomic information provided by BIG includes gene expression, DNA methylation and SNP data, essential for the integrated translational research investigations that the platform aims to facilitate. In addition to the genomic data, imaging data comprising pathology, radiology and nuclear medicine data are expected to be shared by the consortium.

The main objective for this reporting period, regarding Task 5.2, has been to collect the several types of the genomic data for the statistical analysis and the design of the prediction models. An analytical report has been also conducted in which several techniques for extracting relevant imaging markers (i.e. volume metrics, image texture, shape analysis metrics, etc.) are described. Based on the report, tools and UI if needed will be provided to the clinicians for extracting the imaging biomarkers and shared with the VPH community. This task will be highly active however in the second year of the project.

Task 5.3 Development of predictive models of response to therapy and of the modelling framework
A number of actions have been taken in order to define the more efficient methods for building a prediction model from different data sources.

Among several existing techniques, a pattern recognition approach using multiple kernel learning (MKL) architecture for integrating the heterogeneous data was selected to be applied to our prediction model. . A detailed description is given in D5.1. The prediction modeling approach is in its initial phase of implementation.

Besides the prediction modelling approach, several statistical analysis techniques was decided to be adopted to our analysis platform, related to the clinical scenarios that refer to statistics. Therefore, during the reporting period, we have started implementing our routines under the R platform, using open source libraries for the basic functionality of our analysis. This analysis is consisted of:

- in case of analyzing clinical data:
  - simple statistics for the population of the clinical data,
  - survival analysis using Kaplan-Meier and non-parametric tests for comparison between the survival curves,
  - correlation techniques for assessing the correlation degree between different regimens and parameters (Odds Ratios, forestplots, etc.)
- in case of analyzing the genomic data:
  - techniques for assessing the quality of the genomic data (boxplots, MA, density plots of the probe-level data).
  - volcano plots, heatmaps and quantile plots to demonstrate the use of fold changes in conjunction with statistical tests.
  - statistical gene selection techniques for determining interesting genes between two groups (i.e. discrimination based on the pathological complete response).
- In case of analyzing imaging biomarkers:
  - an approach similar to the statistical analysis approach is adopted.

All the above statistical analysis is followed by dynamically created reports which can be updated automatically if data or analysis change, using Latex documentation.

A simplified version of the above analysis will be included to the demonstration version of the first review.

2.2.5.3 Deviations from the DOW and corrective actions

No deviations for this reporting period.

2.2.5.4 Planning next period

D5.1 was submitted on time. This set out precisely the basis for all the future work in WP5. The next step is to optimize the predictive/analysis tools and incorporate them in the platform in an interoperable fashion with the other technologies that the project tis developing (e.g. data management services).

2.2.6 WP6 Pilots, evaluation and validation (Philips)

2.2.6.1 Objectives (of the reporting period)

Considering the user needs as described in WP2 and the corresponding intended pilots, this work package will identify specific application objectives to be tested, define the evaluation criteria and devise monitoring procedures to be executed by the involved stakeholder groups. Special care will be taken to involve the biomedical and clinical end-user community as early as possible in the evaluation and validation effort.
Technical validation will be conducted in tight collaboration with WP3, WP4 and WP5, and the procedures for the assessment of the adequacy of treatment of personal data will be established jointly with WP5. Adequate personal-data treatment is of special importance in the project as foreseen pilots will involve real clinical data.
Specifically the objectives of the work package with respect to evaluation and validation are:

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

A crucial objective of WP6 is to coordinate the efforts with the technical staff and the IT departments of the pilot sites, so that the Consortium receives all information required for developing the information models of the existing infrastructures, and all the data necessary for the testing and validation of the INTEGRATE infrastructure components and tools. It is a critical activity for the project and corresponds to a project milestone.

2.2.6.2 Status/progress towards objectives WP6 (per Task)

Task 6.1: Building the INTEGRATE development and testing environment
The objective of this task is to coordinate all efforts that need to take place locally at each and every pilot site – early enough in the project implementation period – to build the development environment (e.g. “surrogate” databases), and provide access to suitable schema- and instance-level datasets to be used by the project prototypes.

In this context, deliverable 6.1 has been produced, reporting on the development environment and on the available test data. An EHR surrogate DB has been setup.
Relevant clinical trial data (for the TOP trial) and the relevant anonymized patient data has been made available in the EHR surrogate.

Task 6.3: Coordinate specifications of test scenarios and of demonstrators
In this task, minimal scenarios based on the user needs expressed in WP2 will be elaborated to test targeted software components. This includes notably the identification and preparation of

relevant test data, as well the preparation of standalone validation results to which the outcome of the execution of the components can be compared.
In addition, complete demonstrators of clinical relevance will be designed, jointly with the clinical partners of the project, to illustrate the progress of the project during reviews and to serve as evaluation material. As much as possible, this material will be reused as external demonstration and training material in the context of WP7.

The coordination and work for the demonstrators or the review meeting takes place in this context.

2.2.6.3 Deviations from the DOW and corrective actions

No major deviations

2.2.6.4 Planning next period

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

In this task the procedures for the evaluation and validation (E+V) activities will be established. In general evaluation criteria will be continuously adapted to the current state of development of the environment, considering the end-user scenarios and clinical pilots as general guideline. Usability, user-friendliness, clarity of on-line documentation, speed and robustness will be key criteria in the evaluation process. Quantitative measures of the benefits of the project as a whole will also be developed in Task 6.2. The validation of the platform will essentially be conducted by the design of and execution of test cases with known results, those will adapted to the specificities of the software issued in each work package. The outcome of this task will be a set of procedures to guarantee a proper monitoring of the adequacy of the software developed in INTEGRATE to its intended goal, as well as guidelines describing the feedback procedure to developers.

Task 6.4: Deployment Environment

The objective of this task is to prepare the technical and procedural environment – in compliance with the defined legal and security framework of the project – for the installation of INTEGRATE technologies and tools for their extensive evaluation and validation. It is also its responsibility to design, oversee and execute all activities, including training, for preparing the clinical pilots for their validation activities.

Task 6.5: Coordinate evaluation and validation activities and reporting

Evaluation workshops involving software developers and users will be held periodically during the development of the INTEGRATE environment. Evaluation groups will be provided with the criteria and validation procedures defined in Task T6.2 and will execute the scenarios designed in Task T6.3 to provide feedback to the developers.
Feedback reports will suggest possible improvements, modifications, and additional functionalities.
In this task, WP6 will report to management board and steering groups on progresses achieved and remaining problems to be addressed, on the extent to which the project environment meets the identified evaluation criteria. The objective of this task is also to carry out together with WP2 the validation of the INTEGRATE environment at the clinical sites, according to the validation methodology and scenarios defined.

3 Project management during the period

During this period the main tasks in WP8 were to set up a coherent way of working in the consortium and to support collaboration in order to achieve the proper implementation of the work plan and to reach the desired progress. The specific objectives for year one of the project and the progress towards these objectives are described in detail in the previous sections.

3.1 *Consortium Management*

Three consortium workshops and two meetings with members of the External Advisory Board were organized. Additionally, the setting up of collaborations with relevant external initiatives was supported. The project external website, a document repository/ internal website and a project wiki were implemented and are frequently used in the consortium. For example, the project wiki is actively used in the co-development of the first INTEGRATE demonstrator to reflect status, progress, issues, solutions, etc.

In this first period the cooperation in the consortium has been excellent, the regular meeting and t-cons enabling a clear definition of tasks and close collaboration on the development of the technical solutions. The scenarios, use cases, the architecture and the approach towards the semantic solution are all the result of collective effort. Additionally, the partners have closely collaborated on the design and development the first INTEGRATE prototype that implements the molecular screening scenario and which will be demonstrated during the first project review.

All the deliverables planned for the first year were completed through a tight collaboration among the project partners.

Additionally, we have set up relevant external collaborations with prominent initiatives in our area of research:

- SAGE Bionetworks
- TRANSCEND /UCSF

There are no changes in the consortium or in the planning.

3.2 *Project Meetings*

3.2.1 Internal workshops & Management Team Meetings

When	Where	Meeting	Participants
02-04/02/2011	Brussels - Philips	Kick off meeting	All partners
21-22/06/2011	Brussels – IJB	1st Internal workshop	All partners
11-12/10/2011	Greece - Forth	2nd Internal workshop	All partners
10-11/01/2012	Madrid – UPM	3rd Internal workshop	All partners
06-08/02/2011	Ghent – Custodix	4th Internal workshop	All partners

Table 1 – Internal workshops & Project Management Team Meetings

3.2.2 Technical meetings

When	Event	Venue	Participants
02/03/2011	IJB-BIG-Philips meeting	IJB, Brussels	IJB, BIG, Philips
10/04/2011	Meeting with Dr Flamen (Bordet radiologist)	IJB, Brussels	IJB, Bordet
14/04/2011	Meeting with Sarah Davis (TRANSCEND)	University of California	IJB, BIG
15-16/04/2011	SAGE Meeting	University of California	IJB, Bordet
0/05/2011	Meeting with Stephen Friend (SAGE)	IMPAKT conference, The Square, Brussels	IJB, Bordet
10/05/2011	Meeting with Dr Lemort (Bordet radiologist)		
12/05/2011	1st Technical Meeting	Eindhoven / Philips	All partners
07/06/2011	Use Cases Telco		Philips-UPM-Custodix
15/07/2011	Meeting with Dr Salgado (Bordet pathologist)	IJB, Brussels	IJB
18/07/2011	Webinar about Pixcelldata digital pathology solution		
19/07/July 2011	Common Data Model Telco		Philips-CUSTODIX-UPM
21-22/9/2011	INTEGRATE consortium technical meeting	SOST (Spanish Ministry of Science), Brussels	All partners
21-22/09/2011	4th Technical Meeting	Brussels / UPM	Belgium
01/12/2011	Common Data Model Telco	Philips-UPM	Philips-UPM
20/12/2011	BIG/IJB/FORTH meeting about WP4	BIG, Brussels	IJB, BIG
22/12/2011	Common Data Model Telco	Philips-CUSTODIX-UPM	Philips-CUSTODIX-UPM
10-11/01/2012	5th Technical Meeting	Madrid / UPM	Madrid / UPM
19/01/2012	Data modelling Telco	Spain	Philips-IJB-UPM
01/02/2012	Demonstrator Telco	Philips-IJB-CUSTODIX-UPM	Philips-IJB-CUSTODIX-UPM
16/03/2012	Demonstrator Telco	Philips-IJB-CUSTODIX-UPM	Philips-IJB-CUSTODIX-UPM

Table 2 – Technical Team Meetings

3.2.3 Work Packages Meetings

When	Where	Subject	Organising Partner or WP	Participating Partners
Every week	IJB/BIG	IJB/BIG Joint meeting		IJB, BIG
First Friday of the month		WP2 Telco	Custodix	WP2 members
21/12/2011		WP7 Telco (exploitation)	Custodix	WP7 members
12/05/2011	Eindhoven	Technical Meeting	Philips	All partners
07/06/2011		WP1 Telco (Use cases)	Philips	Philips - UPM
19/07/2011		WP4 Telco (Common Data Model)	Philips	Philips-Custodix - UPM
01/12/2011		WP4 Telco (Common Data Model)	Philips	Philips - UPM
22/12/2011		WP4 Telco (Common Data Model)	Philips	Philips-Custodix - UPM
19/01/2012		WP4 Telco (Data modelling)	Philips	Philips-IJB-UPM
01/02/2012		WP6 Telco (Demonstrators)	Philips	Philips – IJB – Custodix - UPM
16/03/2012		WP6 Telco (Demonstrators)	Philips	Philips – IJB – Custodix - UPM

Table 2 – Work Packages and Working Group Meetings

3.3 *Exploitation*

3.3.1 Adoption of the INTEGRATE solutions

The objective of this task is to build on the existing BIG network to guaranty the implementation of the INTEGRATE solutions beyond the consortium borders. The starting point of this task will be the use of the molecular screening platform in the context of a BIG clinical trial which is currently in preparation. Project priorities and timelines have been redefined to meet this specific objective. More information on these developments is provided in the WP1 report.

3.3.2 Ensuring project sustainability

An important goal of the exploitation task is the sustainability of the INTEGRATE solutions beyond the duration of the project and their use. This is the object of the INTEGRATE (Initial) Exploitation Plan which is currently being developed by BIG in collaboration with all the partners participating in this task. The elaboration of such a plan requires a comprehensive analysis of the possible use scenarios, of the user groups, of the possible sources of funds as well as an estimation of the project overall costs (see also Deviations from the DOW and corrective actions). For this reason, BIG has engaged in one-on-one consultations with representatives of similar initiatives such as SAGE Bionetworks and I-SPY/TRANSCEND, as well as with senior representatives of the Pharmaceutical Industry (in the context of the BIG Executive Board retreat held at Stockholm in September 2011). BIG also had several contacts with Dr David Cameron, professor of oncology in Edinburgh and member of BIG's Executive Board. Prof. Cameron has experience in data-sharing initiatives in the context of cancer research and has accepted to be consulted by BIG on an ad-hoc

basis. Information gathered through these consultations will support the business models to be developed in the INTEGRATE exploitation plans.

3.4 *Use of Foreground*

The information reported in this section is delineated in the Initial Dissemination Plan which was produced in Month 9 of the project and which underlies the INTEGRATE approach to dissemination.

The first priority of the dissemination task was to optimise the sharing of knowledge at the project-level from the start of the project. To achieve this objective, the project coordinator, Philips, and the participant in charge of the knowledge management work package, BIG, have put in place a variety of tools:

3.5 *Dissemination activities*

3.5.1 Publications and abstracts

During this reporting period BIG and IJB have presented 2 poster about the INTEGRATE platform at international oncology conferences. These posters presented INTEGRATE to the community of oncologists and translational researchers. Secondly, two articles describing INTEGRATE were published by BIG in the widely distributed BIG newsletter, increasing the visibility of the project. Finally BIG has coordinated and contributed to preparation of the first INTEGRATE newsletter, which will be issued on April 1st (Month 14). See the dissemination section for more details.

When	Where	Publication Title	Author(s)
	Press media	A browser extension to retrieve EHR-based biomedical literature. BMC Medical Informatics and Decision Making	Perez-Rey, D., Jimenez-Castellanos, A., Garcia-Remesal, M., Crespo, J., Maojo, V. CDAPubMed:

3.5.2 Conferences – other activities

BIG, as clinical partner of INTEGRATE, contributes to dissemination by sharing information on the project within its collaborative network[4]. This occurs throughout the year through the bi-annual BIG scientific meetings or through BIG's own channels of communication such as the BIG newsletter or the BIG website.

[4] The BIG Headquarters is responsible for the coordination of a network of 50 groups based in Europe, Canada, Latin America, Asia and Australasia. These research entities are tied to several thousand specialised hospitals and research centres worldwide.

Planned /actual Dates	Type	Type of audience	Countries addressed	Size of audience	Partner responsible /involved
Sep 23-27, 2011	Poster in international oncology conference	Medical, radiation, and surgical oncologists, translational researchers, basic scientists, healthcare workers, patient advocates,	All European countries, and also strong representation from most other countries	Estimated 15000	BIG
Nov 3-5, 2011	Poster in international oncology conference	Medical, radiation, and surgical oncologists, translational researchers, basic scientists, healthcare workers, patient advocates,	European as well as other continents	Estimated 2500	BIG
Dec, 2011	Article in BIG newsletter	Clinical trialists, Medical, radiation, and surgical oncologists, translational researchers, basic scientists, healthcare workers, patient advocates,	European as well as other continents	50 clinical trial groups worldwide	BIG
Q4, 2011	Article in magazine	General public	Mainly Belgium		BIG

3.5.3 Project web-sites

3.5.3.1 Project Wiki

INTEGRATE Consortium members use the INTEGRATE Wiki[5] (D7.1) as a database and knowledge management tool. This user-friendly website is structured around the different work packages and can be easily developed and modified by any identified/authorized INTEGRATE project participant. Work package activities, reports, specific questions about the project as well as any other relevant information are immediately available to all participants. Access is restricted to participants of the Consortium and it is protected by user authentication.

[5] http://atlas.ics.forth.gr/INTEGRATE/wiki/index.php/Main_Page

3.5.3.2 BSCW Document Server

The project BSCW Shared Document Server[6] was conceived as a communication platform for the project administration (e.g., templates, minutes of the conferences). Both the Wiki and BSCW platform constitute the project intranet and are accessible from the INTEGRATE website for project participants.

The second objective of this task is to disseminate information about the project and its objectives to the user communities i.e., the scientific community, academic institutions, other research organizations, pharmaceutical companies, or the lay public. During the first year of project this objective has been achieved through various activities such as:

3.5.3.3 Project public website

The INTEGRATE public website (D7.2) presents general project information, participant information, downloadable publications and public deliverables[7]. Furthermore, it informs viewers about previous and forthcoming public events and activities of the project as needed. While designed and hosted by FORTH the INTEGRATE website is being updated by BIG, the partner responsible for the website content.

The Integrate project website is running and up-to-date. The URL of the public pages is: http://pothos.ics.forth.gr/integrate/

Below a screenshot of the homepage

[6] http://atlas.ics.forth.gr/bscw

4 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
8.1	Public summary of project	8	Philips	O	PU	1	Yes	Feb	
8.2	Internal project website	8	Philips	O	RE	3	Yes	Apr	
7.1	Communication portal Twiki	7	Forth	O	PP	3	Yes	Apr	
1.1	User needs and specifications for the INTEGRATE environment	1	IJB	R	PU	6	Yes	Jul	
7.2	External project website	7	BIG	O	PU	6	Yes	Jul	
8.3	Interim Progress report #1	8	Philips	R	CO	7	Yes	Aug	
5.1	Report on the VPH use case study	5	Forth	R	PU	8	Yes	Jan	
1.2	Definition of relevant user scenarios based on input from the users	1	IJB	R	PU	9	Yes	Nov	
1.3	INTEGRATE legal, ethical and regulatory requirements	1	BIG	R	PU	9	Yes	Dec	
2.1	State of the art report on standards	2	Philips	R	PU	9	Yes	Nov	
4.1	Specification of the model, data and annotation repositories	4	Philips	R	PU	9	Yes	Nov	
6.1	Report on the development environment and on the available test data	6	Philips	R	PU	9	Yes	Jan	

7.3	Initial dissemination plan	7	BIG	R	PU	9	Yes	Jan	

TABLE 2. MILESTONES

Milestone no.	Milestone name	Due achievement date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
MS1	Formation of boards and committees	Month 6	Yes	Jul 2011	
MS2	Initial requirements wrt the Integrate environment	Month 6	Yes	Jul 2011	
MS3	User scenarios based on the user requirements	Month 9	Yes	Oct 2011	
MS4	Initial Integrate architecture	Month 12	yes	Jan 2012	

5 Explanation of the use of the resources

5.1 *Manpower overview*

Actually Spent Human Resource Allocation Year 1

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

Partner	WP1		WP2		WP3		WP4		WP5		WP6		WP7		WP8		*Total*	
	planned	spent	Planned	spent	Planned	spent	Planned	spent	planned	spent	planned	spent	Planned	spent	planned	spent	planned	spent
Philips	2.5	2	5	3	7	1.14	7	7	2	8	4	0.52	1	0.5	4	0.55	32.5	*19.71*
BIG	6	13.3	3	4.15	3	1.80	3	3.00	3	3.7	3	1.60	4	5.50	0.5	3.00	25.5	*36.06*
FORTH	**2.5**	**2**	**5**	**3**	**2**	**1.14**	**9**	**7**	**9**	**8**	**3**	**0.52**	**1**	**0.5**	**1**	**0.55**	**32.5**	***19.71***
Custodix	2	3.51	9	13.6	2	1.12	7	2.77	0.9	0	3	0	0.9	0.13	0.3	0.15	25.1	*21.34*
IJB	9.5	10.8	0	0	2.5	8.1	0	0	4	1.3	4	1.2	1.2	0.4	0.5	0	21.7	*21.8*
UPM	1.0	1,38	3.0	4,99	4.6	9,92	2.4	2,79	0.4	0,12	1.6	2,4	0.4	0,95	0.24	0,55	13.64	*23,10*
Total WP	*24.5*	*32.99*	*26*	*28.74*	*25.5*	*23.22*	*31*	*22.56*	*19.8*	*21.12*	*20*	*6.24*	*8.9*	*7.98*	*6.9*	*4.8*	*165.5*	*147.65*