| Project Number: | 284941 | Project Title: | Human Brain Project |
|---|---|---|---|

| | |
|---|---|
| Document Title: | Medical Informatics Platform v1 – set-up document |
| Document Filename: | SP08_D8.6.2 |
| Deliverable Number: | D8.6.2 |
| Deliverable Type: | Report |
| Work Package(s): | WP8.1. WP8.2, WP8.3, WP8.4, WP8.5, WP8.6 |
| Dissemination Level: | PU = Public |
| Planned Delivery Date: | M 12 / 30 September 2014 |
| Actual Delivery Date: | M 15 / 15 December 2014 |



| | |
|---|---|
| Authors: | Ferath KHERIF, P23 |
| Compiling Editors: | Tea DANELUTTI, P23 |
| Contributors: | WP8.1 team, WP8.2 team, WP8.3 team, WP8.4 team, WP8.5 team, WP8.6 team |
| STO Review: | UHEI (P45): Björn KINDLER, Sabine SCHNEIDER, Martina SCHMALHOLZ |
| Editorial Review: | EPFL (P1): Richard WALKER, Guy WILLIS, Celia LUTERBACHER |





| | |
|---|---|
| Abstract: | The aim of the Medical Informatics Platform (MIP) is to allow research queries on federated anonymised data sources, and to develop advanced data mining algorithms to identify biological signatures of brain diseases. The innovative *in situ* querying and federation technologies address the challenge of providing access to large volumes of data, while preserving privacy and data integrity with guarantees against corruption. The data mining built into the Platform inherits these properties (privacy, integrity) and gains in scalability. In addition, the data mining addresses the challenge of identifying, from the high-dimensional and heterogeneous data, the biological signatures of diseases. |
| Keywords: | Medical informatics, federated query, data mining. |

## Document Status

| Version | Date | Status | Comments |
|---|---|---|---|
| 1.0 | 01 Jul 2014 | Draft | Version sent by Writing and ST offices |
| 2.0 | 02 Sep 2014 | Draft | Compilation of SP8 Partners contributions - TD/FK |
| 3.0 | 18 Sep 2014 | Draft | Review and edit by Editorial Office (GW) |
| 4.0 | 10 Dec 2014 | Draft | Modifications by SP8 team – FK/TD |
| FINAL | 15 Dec 2014 | Final | |
| | | | |
| | | | |

# Table of Contents

# 1. Introduction

## 1.1 The Human Brain Project (HBP)

The Human Brain Project (HBP) is a major international scientific research project, involving over 100 academic and corporate entities in more than 20 countries. Funded by the European Commission (EC), the ten-year, EUR 1 billion Project was launched in 2013 with the goal "to build a completely new ICT infrastructure for neuroscience, and for brain-related research in medicine and computing, catalysing a global collaborative effort to understand the human brain and its diseases and ultimately to emulate its computational capabilities."

The fields of neuroscience, medicine and information technology each have important roles to play in addressing this challenge, but the knowledge and data that each is generating have been very fragmented. The HBP is driving integration of these different contributions.

During the Ramp-Up Phase, the HBP will collect strategic data, develop theoretical frameworks, and perform technical work necessary for the development of six Information and Communication Technology (ICT) Platforms during the Operational Phase. The ICT Platforms, offering services to neuroscientists, clinical researchers and technology developers, comprise Neuroinformatics (a data repository, including brain atlases and analysing tools); Brain Simulation (building ICT models and multi-scale simulations of brains and brain components); Medical Informatics (bringing together information on brain diseases); Neuromorphic Computing (ICT that mimics the functioning of the brain); and Neurorobotics (allowing testing of brain models and simulations in virtual environments). A High Performance Computing Platform will support these Platforms.

## 1.2 HBP Subproject 8: Medical Informatics

The goal of the Medical Informatics Platform (MIP) is to allow researchers to identify biological mechanisms that explain the complex nature of brain disease. The MIP will provide end-to-end solutions, ranging from data to advanced analytical tools. Researchers will be able to investigate questions requiring data correlations, distributions and interactions, with a view to exploring disease processes and epidemiological factors. Using analytical tools provided by the MIP (and eventually made available as personal APIs), researchers will be able to decipher the relationship between biological variables and clinical phenotypes. Simultaneously, as data accrue, and new hospitals and data generators are recruited, data mining tools will be used to explore all the data to detect recurrent patterns and to identify biological signatures of disease. The biological signatures of disease will form the basis for a new disease space that neuroscientists and clinicians can explore.

The MIP will build on data from databases belonging to public bodies and research institutes, plus hospital data federated by novel data management and query techniques. This federation software and hardware will allow researchers to query and analyse a very large volume of data without moving them from local servers and without compromising data privacy.

## 1.3 Purpose of this Document

This report will describe progress in the development of the Platform and its medical intelligence tools, plus recruitment of hospitals, in the period M1-M12.

## 1.4 Structure of this Document

The remainder of this chapter provides an SP-level overview, highlighting the SP's main accomplishments and issues encountered in the period M1-M12. Subsequent chapters look at accomplishments and in issues within individual components of the SP, as defined in D8.6.1 Medical Informatics Platform v1 – Specification Document.

- Web Portal (First Layer)
- Data Federation (Second Layer)
- Local Data (Third Layer)
- Data acquisition
- Ethics, User Support and Community Building
- Scientific Coordination

The Annexes present in tabular form what the Subproject planned to achieve in this period and what it actually achieved, including the Subproject's Scientific Key Performance Indicators (SKPIs).

# 2. Overview of Subproject 8: Achievements

## 2.1 Subproject Progress

During M1-M12, SP8 developed the different components of the Medical Informatics Platform, working towards the aim of making an internal prototype available to HBP Partners in M18. All Tasks are on track and the Milestones due in this Period have been achieved. SP8 has also worked on establishing a data integration solution, adding *in situ* querying and federation technology to pool information from different sources, as well as creating services and differentiated data mining tools. In the period M1-M6, SP8 focused on specification of the different needs and conceptual development of technological solutions as the foundation for creating the functionalities to be provided by the Platform.

The results of this work were compiled in the MIP's first Deliverable (D8.6.1 - MIP specification document). This technical document details seven Use Cases (for clinicians, neuroscientists and IT developers using the Platform); the functional and non-functional requirements; and the timeline for implementation of the Functions. It also describes key components of the MIP, namely the functional architecture (web portal layer, the federation layer and the local data layer), data collection, hospital recruitment, and data mining.

SP8 Partners have started to establish the principles for data governance in the MIP, addressing issues such as data ownership, privacy, usage, security and ethics. SP8 worked on the identification and resolution of ethical issues concerning the inclusion of data from large-scale studies, in collaboration with SP12 and Science Europe.

## 2.2 Results from Combination of Data and ICT

SP8's work has involved much collaboration between computer scientists, neuroscientists, clinicians and statisticians. New Partners admitted to SP8 via the Competitive Call have been well integrated in WP8.3. All Work Packages and Tasks collaborated closely along four main axes of development:

1) Web Portal: Forming a query is based on Ontologies, Variables and Provenance, which are stored in the Web Portal and extracted from Local Data Store Mirrors. The Web Portal also returns the provenance of Variables to users with their results.
2) Data Federation: Specification of the federation requirements, as well as data mining and distributed processing over the federation Platform.
3) Local Data collection: A Platform that through data cleaning and normalisation will populate each hospital's Local Data Store Mirror.
4) Data acquisition and data mining.

### *2.2.1 Web Portal (first layer)*

Users will access the MIP via the Unified Portal (UP), which provides a single point of access to all six HBP ICT Platforms. User authorisations and the user interface will be based on the UP specifications. The MIP services will be hosted on the UP. Users will interact with services by forming queries and launching analyses, as described in the Use Cases. Work Package 8.1, Work Package 8.2 and Work Package 8.3 (Task 8.3.2) specified the web Use Cases, plus the methods for handling metadata and enabling data integration. The MIP Web Portal specifications were developed in collaboration with SP6/SP13 (for the UP).

### 2.2.2 Data Federation (second layer)

The Data Federation layer receives queries from the Web Portal. Based on local schemas as well as summary information, the federated infrastructure forwards queries to relevant hospitals and collates their results, returning a complete result to the web interface or to relevant portal located applications. The Data Federation software hides the MIP's distributed architecture from MIP users, i.e., users won't see that the hospital data queried actually remains on server(s) of the local hospital(s) concerned. Queries can be either related to simple data retrieval or to setting up more complex data mining flows.

Work Package 8.1 (Task 8.1.3) and Work Package 8.2 developed an advance workflow for launching, monitoring and logging processing jobs. They created simple example data mining algorithms for distributed data sources. The framework similar to MapReduce will be proposed to the Tasks in Work Package 8.3 as template for their own algorithms.

Work Package 8.4 is in the process of defining the detail functional requirements for the Data Federation software. Such detailed requirements will be used for the evaluation of companies for the subcontracting. A strategy, a plan and an evaluation algorithm have almost been completed. These will be used in conjunction with the functional requirements above. In parallel, preparations for the administrative tasks (i.e., the tendering process) have started with CHUV responsible bodies.

The roadmap and first results were presented at conferences and workshops to multiple academic and non-academic audiences. Data providers (hospitals and research labs) and private companies (pharmaceutical, IT, database software companies) are showing great interest in the Medical Informatics Platform.

### 2.2.3 Local Data (third layer)

The Local Data layer is responsible for retrieving, anonymising, pre-processing and extracting features and metadata from local hospital databases. The Project will develop common protocols for data capture. Additionally, tools and workflows necessary to appropriately anonymise all data will be provided to the hospitals to ensure privacy.

The anonymised data are stored at the hospital (Local Data Store Mirror), ready to be queried. It will never leave the hospital. The Local Data layer answers queries received by the Data Federation layer. Work Package 8.3 worked on specification. Work Package 8.1 and Work Package 8.2 worked on implementation related to schema mapping, query processing, and local database schema creation.

### 2.2.4 Data acquisition and data analyses

WP8.5 worked on clinical outreach and hospital recruitment. Nineteen hospitals have expressed interest in joining the Platform at this early developmental stage. WP8.2 worked to ensure that data-capture of longitudinal data from hospitals and large-scale studies is facilitated within the MIP. Magnetic Resonance Imaging (MRI) and genetic data on 6,000 patients have been gathered from large-scale research studies, clinical cohort studies and pharmaceutical trials. The MIP database includes anonymised clinical data of 8,305 patients from CHUV (Switzerland).

WP8.2 has made the data available to WP8.1, WP8.2 and WP8.3. Work Package 8.2, Work Package 8.3 and Work Package 8.5 (Task 8.5.3) identified and gathered a large representative data set for developing and benchmarking data mining algorithms. Preliminary data mining experiments have been performed on the research data sets. A data mining strategy has been developed by WP8.3 with expert medical input. Various algorithms (supervised and unsupervised) have been combined in an iterative fashion to generate a "disease-ome" defined by clinical and biological data and to promote precision

medicine. WP8.2 is working with SP11/WP11.2 to prepare two manuscripts on the biological signatures of disease. The first is on automated diagnosis of AD based on pathologically proven subject data using a Support Vector Machine, while the second is on the use of deep learning algorithms. T8.3.3 developed and implemented a fast version of t-SNE called Barnes-Hut SNE, and made the code available to the medical informatics Platform (a manuscript has been prepared for submission). T8.3.4 developed multi-target regression with rule-based models (Model Trees and Regression Trees with Options).

## 2.3 Milestones Achieved:

- MS153 – Data federation requirements (M06)
- MS154 – Data federation concept (M12)
- MS158 – Recruitment in progress (M06)
- MS161 – Preliminary evaluation of algorithms (M06)
- MS162 - Algorithms identified (M12)
- MS166 – Specifications (M06)
- MS169 – Documentation specifications (M06)
- MS170 – Alliances identified (M12)
- MS288 – Prototype for t-SNE and t-SNE + GENESPACE in a representative large sample (M12)
- MS292 – Algorithms identified and specifications written (M12)

## 2.4 Overview of Subproject 8: Problems

There are no major problems to report. There has been some delay in hiring post-doctoral students, but they will be needed mainly when the first prototype of the Platform is in place.

## 2.5 The Next Six Months for Subproject 8

The next six months will be critical for the implementation of the first prototype of the Platform, and for putting together all the components – ranging from the web portal to the local data store mirror. In this period, SP8 will identify, evaluate and select the virtual data federation software. It will then organise and facilitate the implementation, delivery and integration of the software into the MIP and the UP. An initial version of the MIP will be made available to HBP users in M18, whose feedback will help SP8 to refine the various MIP components during the subsequent period (M18 to M30).

# 3. Data Acquisition

## 3.1 Data Acquisition: Overall Goals

The HBP will recruit at least five hospitals, research labs, industrial companies and other large-scale data gathering initiatives (e.g., large longitudinal studies) to make their data available through the Platform. Recruitment of hospitals has already begun and is already sufficient to ensure that the project reaches its recruitment targets at an early stage of the ramp-up phase. The project's clinical outreach effort is expected to attract new hospitals to the network working towards the project's longer-term targets for the operational phase.

A first framework will be developed at the Centre Hospitalier Universitaire Vaudois (CHUV) in Lausanne, Switzerland, then packaged and proposed for deployment in the other hospitals.

## 3.2 Data Acquisition: Main Achievements

Work Package 8.2 worked closely with Work Package 8.1 (Task 8.1.2) on the specification of variables for both clinical hospital data and clinical research data. Methods for extracting meta-data and ontology definitions have also been devised. SP8 representatives met several times with SP5 (Neuroinformatics) and Elsevier to discuss Ontology definitions and possible common uses.

Work Package 8.5 (Task 8.5.2) work focused on the collection of databases on patients with neurodegenerative diseases, mainly dementia and Alzheimer's disease. Three main sources have been identified:

1) Registers of drug trials conducted by major drug companies ("big pharmas") over the past 15 years (mainly in Europe).
2) Active files from the Academic "Centres de Mémoire, Recherche et Ressources" created in France by the successive "Plans Alzheimer" (15 centres).
3) Prospective cohorts of elderly people (France).

In the third quarter, Work Package 8.5 (Task 8.5.2) developed hospital recruitment strategy as well as links and agreements with large French infrastructure prospective studies in aging and neurodegeneration. Nineteen hospitals have thus far agreed in principle to share clinical data (T8.5.4).

# 4. Ethics, User Support and Community Building

## 4.1 Ethics, User Support and Community Building: Overall Goals

### *4.1.1 Ethical and regulatory issues*

In the long run, HBP research will lead to new techniques for the early diagnosis of neurological and psychiatric diseases, and for their personalised treatment (which has the potential to provide predictions of response to treatment, reductions in adverse drug reactions, etc.), which the Medical Informatics Platform will make available to the research community and, ultimately, to clinicians. These new techniques raise legal, ethical and social issues. Issues of immediate concern include questions related to the use of anonymised data, data protection and informed consent. Equally important are longer-term issues relating to therapeutic equity. Work Package 8.5 (Task 8.5.1) will work with the community and with Work Package 12.1 to analyse these issues, formulating guidelines for ethically responsible clinical decision-making that will guide the implementation and operation of the Platform.

### *4.1.2 User support*

A dedicated support centre for the MIP will be put in place. The support centre will provide users with training and technical support in the form of documentation, courses, web-based tutorials, demonstration sandboxes and a dedicated website. Work Package 8.5 (T8.5.3) will leverage the expertise, experience and, where appropriate, the technical solutions developed in the neuGRID project.

### *4.1.3 Community building and clinical outreach*

Work Package 8.5 (Task 8.5.4) will disseminate knowledge about the MIP and its clinical rationale to the clinical community (psychiatrists and neurologists, clinical researchers etc.) using lectures, presentations in conferences and workshops, summer courses teaching sessions etc.). Work Package 8.5 (Task 8.5.4) will also make a special effort towards hospital managers and CIOs with particular responsibilities for privacy and data safety. As soon as the internal release of the Platform is available, Work Package 8.5 (Task 8.5.4) will incorporate demonstrations of the Platform in its regular activities. In parallel with this effort, the project will build relationships with the pharmaceutical industry (Task 8.5.2), with the initial goal of obtaining data from clinical trials, and the longer-term objective of involving industry more closely in the work of the Project.

## 4.2 Ethics, User Support and Community Building: Main Achievements

### *4.2.1 Ethical and regulatory issues*

Task 8.5.1 focused on the identification of the main ethical issues that arise in the development of personalised medicine. Task 8.5.1 contributed to the organisation of conferences and workshops bringing together Partners from SP8 and SP12 to discuss issues about ethics.

### *4.2.2 User support*

Task 8.5.3 released a first alpha version of the MIP specific support centre Platform based on Valamis technology. Through the neuGRID Platform, Task 8.5.3 provides the so-called "research data sets" (i.e. ADNI) to HBP researchers Europe-wide. Due to the change of

affiliation of task leader Giovanni Frisoni, IRCCS-Fatebenefratelli left SP8 and was replaced by HUG. Collaboration continues however between both entities.

### *4.2.3 Community building and clinical outreach*

Task 8.5.4 work focused on the recruitment of hospitals. To date, nineteen hospitals agreed in principle. Lectures and presentations about the MIP were given at several conferences and workshops. Task 8.5.4 prepared the EU environment for Big Data based medical research and had interactions with EU structures associated with big data initiatives.

# 5. Detailed Work Progress by Work Package/Task

## 5.1 WP8.1 (Federated data management)

Milestones achieved:

- MS153 – Data federation requirements (M06)
- MS154 – Data federation concept (M12)

### *5.1.1 Task 8.1.1 (*In situ *distributed database querying)*

**Task description**

T8.1.1 will develop a set of tools and a transparent data layer allowing researchers to query heterogeneous clinical data across the distributed infrastructure provided by T8.4.2. The new data layer will allow researchers to query data *in situ*, i.e., on the servers where it was originally stored, without importing it into a DBMS or tuning the DBMS for optimal performance. The system, installed at the sites hosting the data, and incorporating plug-ins for a wide range of data formats, will import data dynamically into main memory in response to user queries. While loading the data, the system will adaptively build and maintain its own data structures (indexes, etc.), optimised to support efficient query execution.

**Main achievements (SP8_SKPI_21)**

T8.1.1 achieved MS153 by finalising the requirements for *in situ* distributed data querying, and MS154 by finalising the complete concept for *in situ* distributed data querying (M12). T8.1.1 finished requirements analysis for *in situ* distributed Query Engine (Function 8.1.1.1) in January 2014. Requirements include how the query engine should access data and execute queries, as well as the interface with the Data Federation layer. T8.1.1 finalised the specification of export mechanism for local hospitals and of data formats for Local Data Store Mirrors (Function 8.1.1.2) in March 2014. In particular, T8.1.1 defined the schema for Local Data and Provenance. Clinic data from hospitals will be exported into the Local Data Store Mirror, mapped into the Local Data schema and stored in CSV files. Image data will be stored separately and features of interest will be extracted and stored in CSV files in the Local Data Store Mirror.

T8.1.1 finished the implementation of the interface of local Query Engine with the Data Federation layer (Function 8.1.1.3) in June 2014. In particular, T8.1.1 specified the query language and results format. It also implemented data exchange (queries and results) between the Data Federation layer and the Query Engine on the side of the Query Engine. T8.1.1 is in progress of implementing the query engine for flat CSV files (Function 8.1.1.4). This Function is due in November 2014. This prototype query engine will execute the queries on flat CSV files (containing pre-processed metadata) on the local data store mirror and return the results in the format defined in Function 8.1.1.3. T8.1.1 worked out the first version of schema for Local Data and Provenance in collaboration with WP8.1 Partners. T8.1.1 discussed with SP6 possible ways of connecting the Unified Portal to the Query Engine of the MIP.

A software engineer (Cesar Matos, working full time) joined the team in M5.

**Main problems**

None.

**The next six months**

In the next six months, T8.1.1 will:

- Finish the implementation of the prototype *in situ* Query Engine
- Deploy the Query Engine in the Local Data layer and connect it with the Data Federation layer
- Finish the prototypic implementation of the *in situ* Query Engine by November 2014 (Function 8.1.1.4) and the implementation of test cases for the Query Engine by February 2015 (Function 8.1.1.5).

T8.1.1 might need to constantly modify the schema of Local Data and Provenance as more and more data sets become available. Bugs might appear during the prototypic implementation of the Query Engine. T8.1.1 will find the bugs by running numerous test cases with a broad range of data sets and query benchmarks. Bugs will be fixed before the deployment of the Query Engine.

### *5.1.2 Task 8.1.2 (Data integration)*

**Overall goals**

The purpose of T8.1.2 is to devise methods in order to integrate different sources of data either locally, in the hospital (Local Layer), or across hospitals (Federation Layer), so that the integrated data can be queried through the MIP. This involves:

- The design and implementation of an Extract/Transform/Load process that will translate each hospital's data to the Local Data Store Mirror schema employed by each hospital
- The design of the MIP ontology/schema, consisting of the definition of Variables and Ontology that will (both) provide the query language on top of which queries will be formed in the MIP
- The design and development of a visual mapping tool that will provide users the ability to declaratively define mappings among the Local Data Store Mirror of each hospital schema and the MIP ontology/schema.

**Main achievements (SP8_SKPI_22)**

T8.1.2 contributed to the achievement of MS153 and MS154 by specifying the requirements for the federation Platform. More precisely, T8.1.2 provided the architecture and workflow to integrate data both at the Local Data layer (Extract/Transform/Load process-data warehousing or data exchange), and the Data Federation layer (virtual data integration). T8.1.2 collaborated with T8.1.1 in the definition of the schema of the CHUV Local Data Store Mirror (Function 8.1.2.1). However, the Extract/Transform/Load process is still under development, as access to real data necessary for proper design has been limited.

T8.1.2 defined the Variables that will be used in the MIP schema and be made available at the Web Portal (Function 8.1.2.2). T8.1.2 extracted from CHUV data an initial set of variables and mapped them to LOINC standard. Additionally T8.1.2 extracted a set of variables, also mapped to LOINC, from medical tests obtained outside CHUV. The Ontology created so far consists of the variables identified and part of the ICD-10 ontology that refers to related diagnoses. T8.1.2 designed and initiated development of a visual mapping tool that will be used to declaratively define mappings among local hospital schemata and the MIP ontology/schema (Function 8.1.2.3). These mappings will specify how attributes in the sources (hospitals) correspond to attributes in the MIP ontology/schema, how different groupings of attributes into tables are resolved and how to resolve differences regarding the specification of data values in different sources. More precisely, the mapping tool gets as input either schemas, in xml, xsd, sql or csv format, or mapping tasks (mappings that have already been created) and provides a user friendly GUI for "drawing" mappings among elements of different schemas, allowing the use of predefined functions, e.g. string

and mathematical operations. Finally, the tool will store mappings in either xml format or as tuple generating dependencies (tgds) that can be used in query rewriting.

T8.1.2 initiated the definition of mappings across hospital schemata and the MIP ontology/schema (Function 8.1.2.4). However, this work is in preliminary stages. T8.1.2 collaborated with T8.1.1 in the design of the Local Data Store Mirror schema for CHUV. The design process was initiated by T8.1.1 (EPFL) and is proceeding in collaboration with T8.1.2. T8.1.2 team consists of one senior researcher from AUEB working on a part-time basis, and one senior researcher from AUEB working on a full time basis. The AUEB team is participating fully to workshops and SP8 videoconferences, contributing technically to the development of the Platform.

**Main problems**

The main problem that T8.1.2 faced was lack of access to data as mentioned in the section above. This led to the delay of designing the Extract/Transform/Load process (Function 8.1.2.1) that will translate data from the hospital to its Local Data Store Mirror's schema. Due to the way the Functions have been defined current delays will be easily eliminated. Nevertheless, collaboration between data management experts and data owners need to be improved in order to fill the gaps in metadata information that will make the available data useful.

**The next six months**

T8.1.2 will participate in the achievement of M155 by contributing with its data integration expertise. T8.1.2 will eliminate any gap created due to lack of data and complete the ETL process for the translation of hospital's data to their Local Data Store Mirror (Function 8.1.2.1). T8.1.2 plans to proceed with the definition of Variables in the MIP ontology/schema (Function 8.1.2.2) as well as their alignment with existing international standards, such as Loinc and ICD-10. T8.1.2 plans to proceed with the development of the visual mapping tool (Function 8.1.2.3). This tool will be used in the Web Portal in order to create mappings between the MIP ontology/schema and the schemata of the Local Data Store Mirrors (probably by users with administrative privileges). T8.1.2 plans to further create mappings across hospital schemata and the MIP ontology/schema (Function 8.1.2.4). Currently, since the MIP ontology/schema consists mostly of the schema of the CHUV Local Data Store Mirror mappings are identity mappings.

### *5.1.3 Task 8.1.3 (Workflow infrastructure)*

**Task description**

The aim of this Task is to design and develop a complex dataflow-processing engine. The engine will support dataflow processing without moving or storing data coming from local data providers. The query language of the engine will be based on SQL with user-defined functions (UDFs). UDFs will provide functionality such as validation (character encoding, date format) and statistics (Pearson).

**Main achievements (SP8_SKPI_23)**

T8.1.3 is in the process of designing and developing the complex dataflow-processing engine with native UDF support. It contains Function 8.1.3.1 and Function 8.1.3.2. Function 8.1.3.1 aims to design and develop the engine. Until now, the engine supports dataflow processing for distributed computation of average, standard deviation, covariance matrix, and principal component analysis (PCA) as well as for distributed k-means clustering over a set of variables. This Function is due in October 2015. For the execution of the aforementioned examples, the engine uses the UDFs described in Function 8.1.3.1.

Function 8.1.3.2 aims to design and develop UDFs. So far, T8.1.3 has developed the following UDFs: FILE, JDICTSPLITV, HASMD5MOD, FARITH, FSUM and GRAMIAN. The FILE function loads the input file to a table. The first column (patient's id) of the input file is stored as a normal column and all other columns (>1) are stored as a JSON dictionary. The JDICTSPLITV Function gets the JSON dictionary and splits it to two columns (key, value). HASMD5MOD Function requires the patient id as input and returns an index of range [0 – (k-1)], where k is a variable. FARITH Function requires a mathematical expression in polish notation as input and computes the result using high precision fraction computation. FSUM is an aggregate UDF, which adds the values of a column using high precision fraction computation. The reason for using high precision fraction computation is to avoid overflow and increase the computation' precision. GRAMIAN Function computes the Gramian matrix. This Function is due in May 2015.

UoA (P37) is collaborating with other WP Partners through weekly SP8 videoconferences. The team is also sharing technical details on the developed distributed data-flows and UDFs with the other WP Partners. T8.1.3 initiated collaboration with TAU (P55) (T8.3.1) for a joint work on distributed implementation of data mining algorithms. The UoA team attended the workshop meeting on EPFL (P1) campus where all the WP Partners discussed their work.

The T8.1.3 team consisted of two senior researchers working on a part-time basis, two researchers working on a part-time basis, and one junior researcher working on a full-time basis.

**Main problems**

None.

**The next six months**

T8.1.3 runs beyond the next six months. Subsequently, UoA team will continue to design and develop the complex dataflow-processing engine with native UDF support (Functions 8.1.3.1 and 8.1.3.2). Furthermore, together with TAU, T8.1.3 will start work on supporting distributed data mining algorithms.

### *5.1.4 Task 8.1.4 (Data privacy)*

**Task description**

HBP's *in situ* query technology will make it possible to query data stored on local hospital servers without moving it to a central location. When implementing the technology, the HBP will adopt a pragmatic strategy that guarantees respect for differing local regulations, the privacy policies of hospitals and any additional restrictions on use requested by individual patients (e.g. restrictions on the use of specific categories of data). To meet this need, T8.1.4 will develop novel configurable security mechanisms both at the Platform level and the general data release level. At the Platform level, T8.1.4 will develop novel access control techniques, to be deployed on local servers, giving data owners (hospitals, research institutions etc.) fine-grained control over the configuration of security permissions and ensuring that authorised personnel maintain the ability to access individual data in an emergency (e.g. when a researcher finds an undetected tumour in neuroimaging data). T8.1.4 will also develop additional methods to anonymise patient data of different modalities. At the data release level, the software will preserve researchers' ability to select data based on patients' (which will be anonymised) personal attributes (gender, age, country/region of origin, country/region of residency, ethnic origin, clinical diagnosis etc.) and execute only a set of pre-defined aggregated queries. T8.1.4 will ensure that it is impossible to identify a specific individual in a given data set.

**Main achievements (SP8_SKPI_24)**

T8.1.4 achieved MS153 – by finalising the requirements for data anonymisation tools – and MS154 – by finalising the complete concept for data anonymisation tools. T8.1.4 finished the requirements analysis for the anonymisation of patient data. The anonymisation module will have three components to strictly prevent patient data from disclosure. First, personal identifiers will be removed from patient data, image data and genetic data before they are exported from hospital systems. Secondly, queries coming from the Data Federation layer will be filtered so that only particular types of queries can be executed at the Local Data layer. Thirdly, query results will be further anonymised before they leave the local hospital. The development of the anonymisation module will be subcontracted. The procurement office of EPFL has approved the requirements written for the subcontract.

Thomas Heinis (postdoc) worked on the specification of requirements for data anonymisation. He also initialised the tendering process.

T8.1.4 has also chosen a Partner for the anonymisation subcontract. After having asked multiple providers of solutions in the anonymisation product space, Gnubila was chosen as subcontractor. While all providers submitted solutions of similar technical quality (and at similar prices), Gnubila was primarily chosen because of their extensive prior experience in developing anonymisation software for brain related data (and medical data in general) in the context of EU projects. The negotiations for the contract between HBP and Gnubila are underway and will soon be concluded.

**Main problems**

None.

**The next six months**

The subcontractor of T8.1.4 will finish the prototypic implementation of the data anonymisation tools in the next year. EU regulations, hospital privacy policies and other restrictions related to data privacy may evolve over time. T8.1.4 will develop a mechanism that can dynamically adapt the anonymisation module to such changes.

## 5.2 WP8.2 (Data acquisition and federation) (SP8_SKPI_01 – 09)

Milestones achieved:

- MS158 – Recruitment in progress (M06)

### *5.2.1 Task 8.2.1 (Analysis of historical longitudinal databases for studies in neurodegeneration)*

**Task description**

T8.2.1 will capitalise on large longitudinal study of neurodegeneration data from thousands of subjects (providing population-level data on cognitive performance, functional abilities, plasma MR imaging markers and genetics) to test HBP data mining and analysis tools for their ability to identify multi-factorial disease signatures, and prepare them for clinical evaluation.

**Main achievements**

T8.2.1 focused on the specification of variables for both clinical hospital data and clinical research data; methods for extracting meta-data and ontology definitions. The aim is to ensure that the capture of longitudinal data from hospital and large-scale studies is facilitated within the Medical Informatics Platform. In addition, T8.2.1 achieved MS158.

T8.2.1 gathered data on 6,000 patients from large-scale research studies (e.g. ADNI), clinical cohort studies (e.g. 3-C study) and pharmaceutical trials (e.g. Sanofi). T8.2.1 also gathered a first batch of data from CHUV, including 300 patients. This first batch was used to generate specifications for the MIP. T8.2.1 worked towards the integration of new data in the MIP in collaboration with WP8 (T8.1.1 for data description and T8.1.2 for Variable creation and Ontology) and T8.2.2. At the moment, the MIP database includes data of 8,305 patients from CHUV.

**Main problems**

Due to the heterogeneity of data sources within the CHUV hospital and the scale of the data, T8.2.1 experienced delays in gathering the first batch of data. This is true in particular for the extraction of images from PACS system (MRI data). To overcome this problem, a solution has been found, based on semi-manual extractions, until an automated solution will be put in place by CHUV-IT. A post-doc position was advertised but the applicants' profiles were not interesting enough. As a consequence, the position is still open. To fill the gap, a researcher has been hired for three months. His activity is shared between T8.2.1 and T8.2.2.

**The next six months**

The plan for the next six months is to pursue the integration of additional data (from CHUV, CHU-Bordeaux and CHU-Lille) in the MIP. T8.2.1 will start a new hiring process for the post-doc position.

### *5.2.2 Task 8.2.2 (Enabling the acquisition of data from prospective studies)*

**Task description**

The aim of T8.2.2 is to take advantage of existing large-scale initiatives and create data exchange standards that could later be used by hospitals contributing to the Medical Informatics Platform. T8.2.2 will also facilitate access to cognitive, imaging, biological, genetic, and molecular data sets from large on-going prospective studies.

**Main achievements**

T8.2.2 work focused on the specification of data descriptions for longitudinal and prospective studies in terms of Variables. The Variables will ensure comparability across data providers along five dimensions:

4) Scale – genetic, molecular, cellular, circuits, systems.
5) Time – data acquisition at single point/multiple points per subject.
6) Space – centred on the brain (based on brain regions).
7) Pathology – clinical labels of brain pathology.
8) Demographics – age, gender.

T8.2.2 achieved MS158 by gathering data on 8'305 patients from CHUV. Besides, T8.2.2 worked towards the integration of this data in the MIP, in collaboration with T8.2.1 and WP8 (T8.1.1 for data description and T8.1.2 for Variable creation and Ontology).

T8.2.2 initiated talks with large-scale European studies.

**Main problems**

Due to the heterogeneity of data sources within the CHUV hospital and the scale of the data, T8.2.1 experienced delays in gathering the first batch of data (see T8.2.1 above). This caused some delay. A post-doc position was advertised, but the applicants' profiles were not interesting enough. As a consequence, the position is still open. To fill the gap, a

researcher has been hired for three months. His activity is shared between T8.2.1 and T8.2.2.

**The next six months**

The plan for the next six months is to gather data from other large-scale European studies, such as IMAGEN (www.imagen-europe.com). T8.2.2 will start a new hiring process for the post-doc position.

## 5.3 WP8.3 (Medical intelligence tools for data categorisation)

Milestones achieved:

- MS161 – Preliminary evaluation of algorithms (M06)
- MS162 – Algorithms identified (M12)
- MS288 – Prototype for t-SNE and t-SNE + GENESPACE in a representative large sample (M12)
- MS292 – Algorithms identified and specifications written (M12)

### *5.3.1 Task 8.3.1 (Novel mathematical tools for the characterisation and classification of neurological and psychiatric disease)*

**Task description**

T8.3.1 develops novel tools for the systematic characterisation and classification of neurological and psychiatric disease. The tools, which will combine statistical methods with rule-based data mining and prediction algorithms, will allow researchers to perform concurrent analyses of many different classes of data, including brain images and genetics, as well as anatomical, electrical and chemical measurements. The tools will use False Discovery Rates to regularise the results from different methods. The combination of different methods within the same tool will make it easier to develop replicable classifications with meaningful clinical interpretations. The new algorithms will address the challenge of how to deal with missing data from individual patients, data that comes from very different dimensions (heterogeneous data), data generated with different diagnostic instruments and strategies (extrinsic noise), and natural individual variation (intrinsic noise).

**Main achievements (SP8_SKPI_10, 25)**

T8.3.1 developed a combined supervised and unsupervised learning method. The suggested approach includes three steps: Categorisation, Clustering and Classification. This approach is named "The 3-C strategy" and runs as follows:

- Categorisation: Using external medical knowledge three types of features were defined: the patients' assigned disease diagnosis, clinical measurements and potential biological markers, where the latter may include genomic and brain imaging information.
- Unsupervised learning: Creating new classes of biological signature of disease emerging from the clinical manifestation. These classes based on clustering of the selected clinical measurement that are further interpreted for meaningful clinical insights.
- Classification of the new clinical manifestation classes relying on the potential biological markers. This aims to determine a biological signature of disease combining clinical manifestation and biomarkers.

A paper was published in *Lecture Notes in Computer Science* (LNCS 8777, DS2014): "Categorise, Cluster, and Classify: A 3-C Strategy for Scientific Discovery in the Medical

Informatics Platform of the Human Brain Project" by Tal Galili, Alexis Mitelpunkt, Netta Shachar, Mira Marcus-Kalish, Yoav Benjamini. T8.3.1 work on data sets consisted of:

- Further identification, evaluation and estimation of the best-fitted databases that could serve as feasibility study and evaluation tools to the various analysis tools.
- Special efforts were done to register and exploit best the ADNI databases and adjust them as a joint case study to be analysed by our and the other groups analysis tools.

T8.3.1 achieved MS161 and MS162. T8.3.1 worked on distributed data analysis needs (SP8_SKPI_11 + SP8_SKPI_25 + Function 8.3.1.1). Tools, methods and algorithms, were identified and listed as potential implementation tools for the analysis of all distributed data. Two scores attached to each tool – the level of complexity in the tool implementation and its relative importance in analysing distributed data.

T8.3.1 conducted a literature review on Distributed Data Analysis (SP8_SKPI_11 + SP8_SKPI_25 + Function 8.3.1.1). A comprehensive review of available distributed implementation of algorithms is undergoing. Special emphasis is given to horizontally partitioned privacy preserving data mining algorithms. This mode of data distribution is almost unique to the healthcare system in general and addresses accurately privacy issues we face in the HBP.

T8.3.1 worked on a Semi-Automated pre-processing package (SP8_SKPI_11 + SP8_SKPI_25 + Function 8.3.1.1). A semi-automated set of Functions for data cleansing, selection and transformation is being developed. A manual prototype of which is being used on the ADNI data set. The Functions include identification of the feature's statistical type. The proper method of transformation is suggested according to each type. The Functions provide the user with analytical tools for selection of the best-fitted mode of transformation.

T8.3.1 collaborated with SP8 Partners to achieve the following results:

- TAU - UoA – AUEB (P3) collaboration: A process of collaboration was initiated on the basis of a joint work on distributed implementation of data mining algorithms. The collaboration will enable the project to benefit of each group's advantage.
- TAU – JSI collaboration: Discussion of different approaches and methodology to data analysis is occurring through mutual meetings and conferences. During these opportunities algorithms studied by each party were presented and elaborated.
- Use of NeuGrid's Platform: The NeuGrid Platform was used to access ADNI data. The Platform will also be used to share data with UoA. The Platform will enable the joint implementation of a data analysis strategy in a distributed manner.
- Review of Distributed implementation examples: TAU reviewed and commented the paper "Examples of distributed workflow_UOA_v02" written by UoA. Following the review, we organised a joint meeting. In the meeting further discussion of the implementation possibilities and limitation were raised.
- Review of local data schema: The team has reviewed and commented the paper regarding "HPB_SP8_EPFL_140528_simplified_local_schema." Comments were discussed with the team at EPFL.

Tal Kozlovski joined the TAU data mining team.

**Main problems**

Proposed data structures need to be adapted according to international standards for electronic health records.

**The next six months**

In the next six months, T8.3.1 will work to achieve MS163. The 3-C strategy resulting from MS161 will be implemented and its components studied by simulation in order to maximise its performance.

### *5.3.2 Task 8.3.2 (Advanced topographical methods for brain referencing and normalisation)*

**Task description**

The aim of this Task is to extract a feature set that may be mined from the collection of hospital scans from each patient. Fully automated computational analysis of low quality and highly variable scans is still uncharted territory, so a considerable amount of R&D is needed. The aim is to integrate the various sorts of scans of each patient in order to generate maps of different types of brain tissue. These tissue maps would then be used to determine the volumes of a number of brain structures, which would serve as features for data mining.

**Main achievements (SP8_SKPI_26)**

A label propagation prototype was delivered (Function 8.3.2.1). The prototype identifies different brain structures from tissue maps, allowing volumetric features to be computed. The remaining challenge is to compute tissue maps from poor quality scans with very thick slices, using some form of resolution recovery procedure. A basic model for this has been devised and tested for 2D toy data (Function 8.3.2.2), and was found to give promising results. Work is currently underway on developing a matrix representation system for resolution recovery of fully 3D data (Function 8.3.2.3).

On testing the 3D resolution recovery (Function 8.3.2.3), it became apparent that it is unlikely to be fast enough to keep up with processing the enormous number of scans that SP8 proposes to use. For this reason, the original resolution recovery model needs to be shelved until a time when computers are much more powerful.

While the overall objective of computing features from collections of hospital images remains intact, the strategy for achieving this goal will need to change because the resolution recovery approach would not be practical.

**Main problems**

So far, the work has proceeded according to schedule. Unfortunately, that schedule needs to change because the proposed work-plan leads to unfeasibly slow algorithms.

Solving image restoration problems with Petabyte-sized matrices requires an iterative approach, which turns out to be much slower than anticipated. An alternative strategy for dealing with collections of low-quality scans will be sought over the following six months.

**The Next six months**

The work completed so far on functions for "model construction for toy data" (Function 8.3.2.2) and "sparse matrices and intra-session alignment" (Function 8.3.2.3) will be shelved.

Over the next six months, the aim will be to explore other ways to make use of scans with thick slices. The initial approach will be to segment the multiple scans of each subject independently, and assess ways of fusing the results of this segmentation using log-odds.

There are two further options if the (as yet untried) segmentation fusion strategy proves unsatisfactory. A decision about these options would be made after discussions with MR physicists and neurologists.

1) Provide features based on the regional signal intensities extracted from "spatially normalised" versions the original images. This fall back would limit what can be achieved through data mining, as comparisons may only be made between subjects scanned with identical protocols.

2) Limit volumetric features to those hospital scans with slice thicknesses of less than about 2-3 mm. Currently, very few patient scans are performed using such protocols, but this situation may change in the coming years.

Later on, there might be problems with respect to gaining remote access to the hospital systems to properly test algorithms with real-world data.

### *5.3.3 Task 8.3.3 (Linking t-SNE maps and GENESPACE clusterings from the ABA to create, locate, score and view groups of genes harbouring a statistically significant genetic signature)*

**Task description**

t-Distributed Stochastic Neighbour Embedding (t-SNE) is a method to make a "map" of the data in which each point corresponds to a data object. Points are arranged such that nearby map points correspond to similar data objects and distant map points correspond to dissimilar data objects. t-SNE is now widely considered to be the de-facto standard for displaying complex multivariate data in maps. GENESPACE is computationally driven approach to explain how aberrations in hundreds of different genes can culminate to novel knowledge in a single neurological disorder by revealing their functional relationships throughout human brain development. Our generic method is widely applicable to other sets of neurological genes.

Both t-SNE and GENESPACE have great potential to aid identification of complex patterns in the Human Brain project data. In this task, we will to extend both techniques, and integrate them into a coherent methodology to relate molecular interaction patterns to heterogeneous sets of patient specific readouts. The proposed research has three specific objectives:

1) To develop a method that scales up t-SNE to massive data sets (Big Data) by reducing its computational complexity, and to adapt it for application on heterogeneous medical data sets in neuroscience research.

2) To develop a rule-based clustering method by interlinking t-SNE mappings with GENESPACE clustering from the Allen Brain Atlas, allowing for a robust identification of patterns within the human-brain related data that define shared, homogeneous categories with common signatures.

3) To validate the developed methods on heterogeneous data in the Human Brain Project medical informatics infrastructure

**Main achievements (SP8_SKPI_28)**

T8.3.3 achieved MS288. Since its accession to the HBP in April, LUMC (P81) has made significant progress in Task 8.3.3. A main achievement is the development of Barnes-Hut SNE, a computationally efficient t-SNE implementation that scales to millions of data points. The launch paper has recently been accepted in Journal of Machine Learning research. In addition, a paper applying this method on proteomic brain imaging data was accepted in Analytical Chemistry.

T8.3.3 worked on integration and application of BH-SNE on the Allen Brain Atlas adult and mouse transcriptome data to identify patterns in the spatial organisation of gene expression in the brain. T8.3.3 team wrote a paper on this in collaboration with the Allen

Institute, which has recently been accepted pending minor revisions in the journal Methods.

Finally, LUMC submitted a conference abstract on a distributed mining system for hospital-based, medical data collections based on light-path optical networks. The team ran a small pilot on OASIS data set, which we stored in a distributed manner. This demonstrated the feasibility of fast optical networks for secure mining of privacy sensitive data. LUMC also joined a first meeting in Lausanne, and participated in weekly teleconferences.

**Main problems**

So far, LUMC has committed part-time employees to the research, and were slightly understaffed due to the rapid accession period in April 2014. However, a post-doc was hired to begin work in October 2014.

**The Next six months**

The next six months, T8.3.3 will:

- Integrate of the BH-SNE in the distributed test framework of the HBP, and into our own distributed hospital data mining system. We will apply the developed techniques on heterogeneous medical data to strong similarities between high-dimensional patient data instances.
- Continue to apply BH-SNE on the Allen Brain atlases for discovery of networks of co-expressed genes that may be expressed in a specific disease: the so-called GENESPACE clusterings
- Finally, we will develop a user interaction paradigm for BH-SNE to enable drilling down on specific findings / clusters in the BH SNE maps.
- Deliver prototype implementations of BH-SNE for the distributed mining framework.

### *5.3.4 Task 8.3.4 (Scalable algorithms for rule-based clustering of heterogeneous data)*

**Task description**

Develop methods for rule-based predictive clustering of big data, including methods for predicting structured outputs, analysis of data streams and methods for subgroup discovery from multi-relational data, including ontologies, and demonstrate the utility of the developed approaches on HBP data.

**Main achievements (SP8_SKPI_29 - 31)**

T8.3.4 achieved MS292. Since accession of JSI (P87) to the HBP in April, T8.3.4 worked on Milestone MS293, scheduled for M18. T8.3.4 worked on the development of streaming algorithms for multi-target regression with rule-based models (Model Trees and Regression Trees with Options) (contributing to Subtask a). The team developed and evaluated an extension of the RReliefF feature-ranking algorithm for hierarchical multi-label classification (contributing to Subtask b). It developed a new methodology for analysing unlabelled data consisting of three steps: clustering with mixture models, extraction of rules from clusters and visualisation with banded matrices (contributing to Subtask c). This development contributes three new algorithm prototypes to the SP8_SKPI_25 "Medical intelligence tools for data categorisation/Supervised and unsupervised data mining algorithms."

Collaboration with TAU was established in order to specify data mining tasks and identify available data suitable for development and validation of rule-based algorithms. JSI used standard procedures for the validation of data mining algorithms (e.g. artificial and real

life validation data sets). T8.3.4 consisted of five senior researchers and five doctoral students, all of whom worked part-time on the Task.

**Main problems**

Identifying and collecting suitable data needed to develop data mining algorithms is not as easy as initially thought. One of the key problems is the very strict understanding of data privacy within the HBP, which in practice makes it impossible to have entire data set locally stored and analysed.

**The next six months**

Within the next six months the development of algorithms (for learning models from multi-target data streams, or structured output prediction, for feature ranking, for subgroup discovery from multi-resolution or heterogeneous data) should progress to the stage where a working prototype of each of the algorithms is available. This is also the requirement of Milestone MS293, which is due in Month 18. No major problems are expected in the next six months.

## 5.4 WP8.4 (Medical Informatics Platform: Integration and operations) (SP8_SKPI_16)

Milestone achieved:

- MS166 – Specifications (M06)

### *5.4.1 Task 8.4.1 (Integration, website construction, maintenance and administration)*

**Task description**

The goal of Work Package 8.4 is to design, implement and operate the MIP, federating large volumes of data in hospital and laboratory databases, making this data available to the scientific and clinical community and providing it with the tools to use the data in their research.

Task 8.4.1 will identify, evaluate and select the virtual data federation software; organise and facilitate the implementation, delivery and integration of the software into the MIP and the HBP. Task 8.4.1 will design and operate the MIP, integrating the data and tools from Work Packages 8.1, 8.2 and 8.3.

**Main achievements**

T8.4.1 achieved MS166 by contributing to the specifications document (D8.6.1). T8.4.1 is in the process of defining the detail functional requirements for the Data Federation. Such detailed requirements will be used for the evaluation of companies.

A strategy, plan and an evaluation algorithm are almost complete. These will be used in conjunction with the functional requirements above. In parallel, T8.4.1 started administrative preparations related to the tendering process with the CHUV responsible bodies. T8.4.1 identified 28 major international companies that could potentially satisfy Data Federation requirements, and made contact with 8 of them.

T8.4.1 collaborated continuously with T8.1.1 (led by EPFL) because the Data Federation infrastructure, the work on Local Data Store Mirrors (T8.1.1) and the work on data management (T8.1.4) are closely interconnected within the MIP. The EPFL team also attended meetings with potential Data Federation software providers and was asked for advice. EPFL shared with CHUV information on their work: database schemata at local nodes, progress with the anonymisation, technical details on the database configuration.

T8.4.1 identified issues and discussed them with the SP8 leader. The team kept a detailed log of all meetings, notes, outcome, actions and lessons learnt. T8.4.1 defined and maintained a clear plan of the tasks to be implemented. It ensured the stage plan was regularly updated with corrective actions if required, and remained within tolerances. T8.4.1 applied well-established project management methodologies and best practices.

Lessons learned so far:

- (+) Identifying requirements: due to the complexity of the case, the in-depth discussions and demos of the different data federation software providers, has helped us clarify further and formalise our exact requirements.
- (+) Agile plan execution: the tendering exercise initially planned for Jun-Sep 2014, was delayed by the administrative team at CHUV (due to the complex case, summer holidays, etc.). Therefore, in order to advance with the process, we have investigated tendering guidelines ourselves, started the definition of criteria and algorithms, and met with companies. Although the tendering process itself will still be allocated the same time length, we will have most of the preparations ready by then. Most importantly, we will have key knowledge about some of the software providers.
- (-) Tendering process: not as easy as initially thought to get it started (see above).

**Main problems**

The main problem encountered by T8.4.1 was the tendering process. The tendering process has not started yet due to internal administrative delays at CHUV. However, the first MIP prototype is still scheduled for M18.

**The next six months**

T8.4.1 plan for the next six months:

- Plan, design in detail, customise and test the product to meet functional requirements (with the provider).
- Define appropriate procedures, regulations and processes - i.e. the use of the product, the use of data results, etc.
- Integrate with the UP and two local hospitals databases.
- Integrate with the Local Data Store Mirror software (T8.1.1), algorithms, and data management (T8.1.4)

Thanks to this plan, T8.4.1 will be able to have the Medical Informatics Platform ready for internal release by M18 (MS167).

### *5.4.2 Task 8.4.2 (Setting up federated infrastructure)*

**Task description**

Task 8.4.2 will subcontract an industrial Partner to provide a standardised infrastructure, a set of tools and a transparent data layer allowing researchers to query heterogeneous clinical data across the distributed infrastructure developed in Task 8.1.1.

**Main achievements**

None – Tendering process not started yet.

**Main problems**

The tendering process has not started yet due to internal administrative delays at CHUV. Internal and external investigations continue in order to be able to progress with it and meet the M18 deadline.

**The next six months**

The tendering process will take place in the next six months.

## 5.5 WP8.5 (Medical Informatics Platform: user support and community building)

Milestones achieved:

- MS169 – Documentation specifications (M06)
- MS170 – Alliances identified (M12)

### *5.5.1 Task 8.5.1 (Ethical and regulatory issues)*

**Task description**

HBP work in WP8.1-WP8.3 will lead to new techniques for the early diagnosis of neurological and psychiatric disease, and for personalised treatment (predictions of response to treatment; reduction of adverse drug reactions etc.), which the MIP will make available to the research community and ultimately to clinicians. The new techniques raise immediate and longer-term legal, ethical and social issues. Issues of immediate concern include questions related to the use of anonymised data, data protection and informed consent. Equally important are longer-term issues of therapeutic equity. T8.5.1 will work with the community and with WP12.1 (HBP Foresight Lab) to analyse these issues, formulating guidelines for ethically responsible clinical decision-making that will guide the implementation and operation of the Platform.

**Main achievements (SP8-SKPI_08, 09)**

T8.5.1 work focused on the identification of the main ethical issues that arise in the development of personalised medicine. A review of the ethical balance between challenge and opportunity of personalised medicine under the aspects of adequacy, cost, and therapeutic equity is in progress.

T8.5.1 strengthened collaboration between SP12 and SP8, and initiated the third SP12 conference, to be held at the HBP Summit 2014. The conference is on the new (RDOC) approach to research funded by NIMH, which requires researchers not to use DSM/ICD diagnostic categories, but to go ‘straight to the brain’ to identify brain signatures of mental disorders.

**Main problems**

None.

**The next six months**

During the next six months, T8.5.1 will continue to work on the review of the ethical balance between challenge and opportunity of personalised medicine. Collaborations with SP12 will also be further developed.

### *5.5.2 Task 8.5.2 (Clinical trials service centre)*

**Task description**

T8.5.2 will create working relationships with the pharmaceutical industry for the provision of data from clinical trials. T8.5.2 will work with industry and industry associations to extend this kind of practice. One of T8.5.2’s goals will be to produce innovative analyses of outcomes in long-term longitudinal trials. A major goal will be to test whether the

current model of syndromic, cohort-based, RCT are more (or less) informative than studies of patients groups whose conditions share the same underlying biological cause.

**Main achievements**

T8.5.2 achieved MS170. T8.5.2 work focused on the collection of databases on patients with neurodegenerative diseases, mainly dementia and Alzheimer's disease, in three main sources of such information:

1) Registers of drug trials conducted by major drug companies ("big pharmas"), mainly in Europe, over the past 15 years. Novartis and Sanofi were contacted.
2) Active files from the Academic "Centres de Mémoire, Recherche et Ressources" created in France by the successive Plans Alzheimer. There are 15 such Centres of which 10 have already agreed, in principle, to share their data with the HBP at this stage of data integration. We now have to go and visit them individually to present to the teams the details of the project and set the practical issues of data transfer.
3) Prospective cohorts of elderly people. All of them conducted in France. The first is run by the University Hospital in Bordeaux (CRI 897) under the lead of Prof. Jean-François Dartigues and Prof. Christophe Tzourio. The second is run by the University Hospital in Lille under the lead of Prof. Florence Pasquier.

T8.5.2 also worked on developing hospital recruitment strategy as well as developing links and agreements with large French infrastructure prospective studies in aging and neurodegeneration (such as "Memento" and "3C").

T8.5.2 collaborated closely with T8.5.4 (CHUV) on developing hospital recruitment strategy.

**Main problems**

Jean-Marc Orgogozo was ill during the 3rd and 4th Quarters. He collaborated with T8.5.4 (e.g. meeting in Bordeaux with Richard Frackowiak mid-April) but activity was partially limited as a result.

**The next six months**

During the next six months, T8.5.2, in collaboration with WP8.2 and T8.5.4, will coordinate the preparation phase for the inclusion of both university hospitals (Bordeaux and Lille) in the Medical Informatics network as local data sources.

### *5.5.3 Task 8.5.3 (The Medical Informatics Platform Service Centre)*

**Task description**

The aim of T8.5.3 is to develop a support centre for the HBP Medical Informatics Platform (MIP), providing users with training and technical support. The work leverages on the expertise, experience, and the technical solutions developed in the neuGRID project (www.neugrid4you.eu).

**Main achievements (SP8_SKPI_12, 27)**

A first alpha version of the MIP specific support centre Platform based on Valamis technology has been released (http://193.204.145.212:8080) allowing the HBP end users to train their skills and access different courses, web-based tutorials, form of documentations, demonstration sand-boxes. Additionally, in the context of dissemination and promotion of the HBP initiative, the Task leader HUG (P112) contributed to the organisation of the “HBP School Alpbach, 8-14 September 2014.”

T8.5.3 collaborated with Partners inside SP8 (e.g. UoA, TAU, HUG) as well as outside SP8. T8.5.3, through the neuGRID Platform, provided the so-called "research data sets" (i.e. ADNI) to SP8 researchers.

The progression of activities is constantly monitored through the "AllWithGraph project overview" application, as well as through weekly teleconferences. T8.5.3 learned that it is mandatory to speak with end-users to plan and carefully structure the content of future open source learning contents. Due to a change of affiliation of Task leader Prof. Giovanni Frisoni, IRCCS-Fatebenefratelli (P41) was replaced by HUG (P122) in April 2014.

**Main problems**

None.

**The next six months**

In the next six months T8.5.3 will collect from Partners developing applications in the MIP Platform informative materials to be shared in order to build up wiki pages, video tutorials, manuals, etc. (Function 8.5.3.2). This will allow end-users to be informed and instructed. T8.5.3 will also work on broadening the base of users interested in using the HBP-MIP Platform collecting even new requirements. Besides, T8.5.3 will offer specific support, especially in the access and usage of data to process and discover new results.

### *5.5.4 Task 8.5.4 (Clinical outreach)*

**Task description**

T8.5.4 will disseminate knowledge about the Medical Informatics Platform and its clinical rationale to the clinical community (psychiatrists and neurologists, clinical researchers etc.) using lectures, presentations in conferences and workshops, summer courses teaching sessions etc.). T8.5.4 will also make a special effort towards hospital managers and CIOs with particular responsibilities for privacy and data safety. As soon as the internal release of the Platform is available, T8.5.4 will incorporate demonstrations of the Platform in its regular activities.

**Main achievements (SP8-SKPI_01, 02)**

T8.5.4 achieved MS170. T8.5.4 focused on the recruitment of hospitals. To date, nineteen hospitals have agreed in principle. T8.5.4 work included lectures and presentations at several conferences and workshops, mainly in Europe. For a list of dissemination activities, please refer to Section 6.3 (External meetings). Task leader Richard Frackowiak also gave several newspaper interviews. T8.5.4 worked in context of Science Europe MED SCI committee on EU regulation on data privacy in relation to derogations for medical research, and in relation to preparing the EU environment for Big Data based medical research. T8.5.4 also had interactions with EU structures associated with big data initiatives (e.g. IMI, MedDataBridges and EUDAT).

**Main problems**

None.

**The next six months**

In the next six months, T8.5.4 will continue to disseminate knowledge about the MIP to the medical community and general public.

# 6. Scientific Coordination

## 6.1 Scientific Coordination: Internal Meetings

This table lists meetings between SP staff.

| Date | Description | Location | Participants |
|---|---|---|---|
| 08-09-10-2014 | SP8 session during HBP Summit 2013 | Lausanne (EPFL) | SP8 team |
| 21-10-2013 | SP8 weekly meeting | Online | Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Bogdan Draganski, Tea Danelutti (CHUV), Thomas Heinis (EPFL), Manos Karpathiotakis, Mira Marcus-Kalish (TAU), Dimitropoulos Harry (UoA), Omiros Metaxas, John Ashburner (UCL) |
| 28-10-2013 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Bogdan Draganski, Tea Danelutti (CHUV), Thomas Heinis (EPFL), Manos Karpathiotakis, John Ashburner (UCL), Mira Marcus-Kalish (TAU), Redolfi Alberto (FBF) |
| 04-11-2013 | SP8 weekly meeting | Online | Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Bogdan Draganski, Tea Danelutti (CHUV), Anastasia Ailamaki (EPFL), Thomas Heinis (EPFL), Manos Karpathiotakis, John Ashburner (UCL), Mira Marcus-Kalish (TAU), Yoav Benjamini, Vasilis Vassalos (AUEB) |
| 11-11-2013 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Thomas Heinis (EPFL), Manos Karpathiotakis, John Ashburner (UCL) |
| 18-11-2013 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Bogdan Draganski, Tea Danelutti (CHUV), Anastasia Ailamaki (EPFL), Thomas Heinis (EPFL), Manos Karpathiotakis, Mira Marcus-Kalish (TAU), Vasilis Vassalos (AUEB), Dimitropoulos Harry (UoA), Omiros Metaxas, Bosco Paolo (FBF) |
| 25-11-2013 | SP8 weekly meeting | Online | Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Bogdan Draganski, Tea Danelutti (CHUV), Thomas Heinis (EPFL), Manos Karpathiotakis, Vasilis Vassalos (AUEB), Dimitropoulos Harry (UoA), Bosco Paolo (FBF), John Ashburner (UCL) |
| 02-12-2013 | SP8 weekly meeting | Online | Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Bogdan Draganski, Tea Danelutti (CHUV), Thomas Heinis (EPFL), Manos Karpathiotakis, Lu Xuesong, Mira Marcus-Kalish (TAU), Dimitropoulos Harry (UoA), John Ashburner (UCL) |
| 09-12-2013 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Bogdan Draganski, Tea Danelutti (CHUV), Thomas Heinis (EPFL), Lu Xuesong, Mira Marcus-Kalish (TAU), Vasilis Vassalos (AUEB), Dimitropoulos Harry (UoA), Omiros Metaxas, Bosco Paolo (FBF) |
| 16-12-2013 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Richard Frackowiak (CHUV), Bogdan Draganski, Tea Danelutti (CHUV), Thomas Heinis (EPFL), Manos Karpathiotakis, Lu Xuesong, Vasilis Vassalos (AUEB), Dimitropoulos Harry (UoA), Omiros Metaxas, Bosco Paolo (FBF), Redolfi Alberto (FBF), John Ashburner (UCL) |

| Date | Description | Location | Participants |
|---|---|---|---|
| 06-01-2014 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Bogdan Draganski (CHUV), Tea Danelutti (CHUV), Tassos Venetis (AUEB), Anastasia Ailamaki (EPFL), Thomas Heinis (EPFL), Manos Karpathiotakis, Lu Xuesong, John Ashburner (UCL), Tal Galili (TAU), Jing Cui (CHUV) |
| 13-01-2014 | SP8 weekly meeting | Online | Bogdan Draganski (CHUV), Thomas Heinis (EPFL), Manos Karpathiotakis, Xuesong Lu (EPFL), Dimitropoulos Harry (UoA), Redolfi Alberto (FBF), Tassos Venetis (AUEB), Eleni Zacharia (UoA), Mira Marcus-Kalish (TAU), Tea Danelutti (CHUV), John Ashburner (UCL), Jing Cui (CHUV) |
| 20-01-2014 | SP8 weekly meeting | Online | Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Thomas Heinis (EPFL), Manos Karpathiotakis, Xuesong Lu (EPFL), Dimitropoulos Harry (UoA), Eleni Zacharia (UoA), Tassos Venetis (AUEB), Mira Marcus-Kalish (TAU), John Ashburner (UCL), Jing Cui (CHUV), Tal Galili (TAU) |
| 27-01-2014 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Richard Frackowiak (CHUV), Bogdan Draganski (CHUV), Tea Danelutti (CHUV), Thomas Heinis (EPFL), Xuesong Lu (EPFL), Anastasia Ailamaki (EPFL), Mira Marcus-Kalish (TAU), Tassos Venetis (AUEB), John Ashburner (UCL), Dimitropoulos Harry (UoA), Eleni Zacharia (UoA), Bosco Paolo (FBF), Tal Galili (TAU) |
| 03-02-2014 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Bogdan Draganski (CHUV), Tea Danelutti (CHUV), Thomas Heinis (EPFL), Manos Karpathiotakis, Xuesong Lu (EPFL), John Ashburner (UCL), Dimitropoulos Harry (UoA), Eleni Zacharia (UoA), Tassos Venetis (AUEB), Tal Galili (TAU) |
| 10-02-2014 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Bogdan Draganski (CHUV), Tea Danelutti (CHUV), Manos Karpathiotakis, Xuesong Lu (EPFL), Eleni Zacharia (UoA), Tassos Venetis (AUEB), Richard Frackowiak (CHUV), Thomas Heinis (EPFL), John Ashburner (UCL), Tal Galili (TAU) |
| 12-02-2014 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Bogdan Draganski (CHUV), Tea Danelutti (CHUV), Manos Karpathiotakis, Xuesong Lu (EPFL), Dimitropoulos Harry (UoA), Eleni Zacharia (UoA), Tassos Venetis (AUEB), Bosco Paolo (FBF) |
| 19-02-2014 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Bogdan Draganski (CHUV), Jing Cui (CHUV), Tea Danelutti (CHUV), Tassos Venetis (AUEB), Thomas Heinis (EPFL), Manos Karpathiotakis, Xuesong Lu (EPFL), John Ashburner (UCL), Tal Galili (TAU) |
| 26-02-2014 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Richard Frackowiak (CHUV), Bogdan Draganski (CHUV), Jing Cui (CHUV), Tea Danelutti (CHUV), Tassos Venetis (AUEB), Thomas Heinis (EPFL), Manos Karpathiotakis, Xuesong Lu (EPFL), Bosco Paolo (FBF), Mira Marcus-Kalish (TAU), Eleni Zacharia (UoA), Omiros Metaxas, Lefteris Stamatogiannakis, Herald KIlapi (UoA) |

| Date | Description | Location | Participants |
|---|---|---|---|
| 05-03-2014 | SP8 weekly meeting | Online | Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Bogdan Draganski (CHUV), Jing Cui (CHUV), Mihaela Damian (CHUV), Tea Danelutti (CHUV), Tassos Venetis (AUEB), Thomas Heinis (EPFL), Xuesong Lu (EPFL), Manos Karpathiotakis, Cesar Matos (EPFL), Bosco Paolo (FBF), Mira Marcus-Kalish (TAU), Dimitropoulos Harry (UoA), Eleni Zacharia (UoA), Herald Kllapi (UoA), Lefteris Stamatogiannakis, John Ashburner (UCL), Tal Galili (TAU) |
| 12-03-2014 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Bogdan Draganski (CHUV), Mihaela Damian (CHUV), Tea Danelutti (CHUV), Thomas Heinis (EPFL), Cesar Matos (EPFL), Xuesong Lu (EPFL), John Ashburner (UCL), Tassos Venetis (AUEB), Tal Galili (TAU) |
| 19-03-2014 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Mihaela Damian (CHUV), Tea Danelutti (CHUV), Thomas Heinis (EPFL), Cesar Matos (EPFL), Xuesong Lu (EPFL), Manos Karpathiotakis, Tassos Venetis (AUEB), John Ashburner (UCL), Tal Galili (TAU) |
| 26-03-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Mihaela Damian (CHUV), Tea Danelutti (CHUV), Xuesong Lu (EPFL), Cesar Matos (EPFL), Manos Karpathiotakis, Dimitropoulos Harry (UoA), Eleni Zacharia (UoA), John Ashburner (UCL) |
| 02-04-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Mihaela Damian (CHUV), Tea Danelutti (CHUV), Xuesong Lu (EPFL), Cesar Matos (EPFL), Mira Marcus-Kalish (TAU), John Ashburner (UCL), Tal Galili (TAU) |
| 09-04-2014 | SP8 weekly meeting | Online | Dimitropoulos Harry (UoA), Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Mihaela Damian (CHUV), Bogdan Draganski (CHUV), Cesar Matos (EPFL), Thomas Heinis (EPFL), Manos Karpathiotakis, Xuesong Lu (EPFL), Eleni Zacharia (UoA), John Ashburner (UCL), Tassos Venetis (AUEB), Tal Galili (TAU), Mira Marcus-Kalish (TAU) |
| 16-04-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Mihaela Damian (CHUV), Thomas Heinis (EPFL), Xuesong Lu (EPFL), Cesar Matos (EPFL), Tal Galili (TAU), Mira Marcus-Kalish (TAU), John Ashburner (UCL) |
| 18-04-2014 | SP8 meeting | Online | Richard Frackowiak (CHUV), Jean-Marc Orgogozo (UBO) |
| 23-04-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Mihaela Damian (CHUV), Thomas Heinis (EPFL), Xuesong Lu (EPFL), Cesar Matos (EPFL), Manos Karpathiotakis, Tal Galili (TAU), Mira Marcus-Kalish (TAU), Eleni Zacharia (UoA), John Ashburner (UCL) |

| Date | Description | Location | Participants |
|---|---|---|---|
| 29-30-04-2014 | SP8 workshop (co-organised by CHUV and EPFL) | Lausanne (EPFL) | Anastasia Ailamaki (EPFL), Yoav Benjamini (TAU), Paolo Bosco (FBF), Jing Cui (CHUV), Mihaela Damian (CHUV), Tea Danelutti (CHUV), Bogdan Draganski (CHUV), Kathleen Elsig (HBP IP&TT), Richard Frackowiak (CHUV), Tal Galili (TAU), Sébastien Gouin (HBP IP&TT), Thomas Heinis (EPFL), Manos Karpathiotakis (EPFL), Ferath Kherif (CHUV), Herald KIlapi (UoA), Dragi Kocev (JSI), Boudewijn Lelieveldt (LUMC), Xuesong Lu (EPFL), Mira Marcus-Kalish (TAU), Cesar Matos (EPFLC), Jean-Marc Orgogozo (UBO), Alberto Redolfi (FBF), Marko Rognik-Sikonja (JSI), Sabine Rehberger-Schneider (HBP STO), Lefteris Stamatogiannakis (UoA), Martijn Van de Giessen (LUMC), Vasilis Vassalos (AUEB), Tassos Venetis (AUEB), John Ashburner (UCL) |
| 07-05-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Vasilis Vassalos (AUEB), Ferath Kherif (CHUV), Richard Frackowiak (CHUV), Tea Danelutti (CHUV), Bogdan Draganski (CHUV), Cesar Matos (EPFL), Manos Karpathiotakis, Bosco Paolo (FBF), Redolfi Alberto (FBF), Sašo Džeroski (JSI), Dragi Kocev (JSI), Tal Galili (TAU), Mira Marcus-Kalish (TAU), Eleni Zacharia (UoA), Lefteris Stamatogiannakis |
| 14-05-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Mihaela Damian (CHUV), Tea Danelutti (CHUV), Thomas Heinis (EPFL), Xuesong Lu (EPFL), Cesar Matos (EPFL), Manos Karpathiotakis, Sašo Džeroski (JSI), Marko Robnik-Šikonja (JSI), Dragi Kocev (JSI), Dimitropoulos Harry (UoA), Eleni Zacharia (UoA), Lefteris Stamatogiannakis |
| 21-05-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Thomas Heinis (EPFL), Xuesong Lu (EPFL), Cesar Matos (EPFL), Manos Karpathiotakis, Dragi Kocev (JSI), Bernard Zenko (JSI), Boudewijn LELIEVELDT (LUMC), Tal Galili (TAU), Alexis Mitelpunkt (TAU), Mira Marcus-Kalish (TAU), Herald KIlapi (UoA) |
| 28-05-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Mihaela Damian (CHUV), Richard Frackowiak (CHUV), Thomas Heinis (EPFL), Xuesong Lu (EPFL), Cesar Matos (EPFL), Manos Karpathiotakis, Sašo Džeroski (JSI), Boudewijn LELIEVELDT (LUMC), Alexis Mitelpunkt (TAU), Tal Galili (TAU), Dimitropoulos Harry (UoA), Eleni Zacharia (UoA) |
| 04-06-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Mihaela Damian (CHUV), Bogdan Draganski (CHUV), Thomas Heinis (EPFL), Xuesong Lu (EPFL), Cesar Matos (EPFL), Bernard Zenko (JSI), Dragi Kocev (JSI), Redolfi Alberto (FBF), Alexis Mitelpunkt (TAU), Lefteris Stamatogiannakis, Herald KIlapi (UoA) |

| Date | Description | Location | Participants |
|---|---|---|---|
| 18-06-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Mihaela Damian (CHUV), Jing Cui (CHUV), Thomas Heinis (EPFL), Xuesong Lu (EPFL), Cesar Matos (EPFL), Boudewijn LELIEVELDT (LUMC), Lefteris Stamatogiannakis, Eleni Zacharia (UoA), Dimitropoulos Harry (UoA) |
| 25-06-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Mihaela Damian (CHUV), Jing Cui (CHUV), Bogdan Draganski (CHUV), Thomas Heinis (EPFL), Xuesong Lu (EPFL), Manos Karpathiotakis, Redolfi Alberto (FBF), Bosco Paolo (FBF), Boudewijn LELIEVELDT (LUMC), Baldur van Lew (LUMC), Alexis Mitelpunkt (TAU), Lefteris Stamatogiannakis, Dimitropoulos Harry (UoA), Eleni Zacharia (UoA) |
| 09-07-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Mihaela Damian (CHUV), Jing Cui (CHUV), Xuesong Lu (EPFL), Cesar Matos (EPFL), Alexis Mitelpunkt (TAU), Mira Marcus-Kalish (TAU), Tal Galili (TAU), Lefteris Stamatogiannakis, Herald Kllapi (UoA), Dimitropoulos Harry (UoA), Eleni Zacharia (UoA) |
| 16-07-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Mihaela Damian (CHUV), Jing Cui (CHUV), Xuesong Lu (EPFL), Cesar Matos (EPFL), Thomas Heinis (EPFL), Anastasia Ailamaki (EPFL), Baldur van Lew (LUMC), Alexis Mitelpunkt (TAU), Mira Marcus-Kalish (TAU), John Ashburner (UCL), Dimitropoulos Harry (UoA), Eleni Zacharia (UoA), Lefteris Stamatogiannakis |
| 23-07-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Eleni Zacharia (UoA), Ferath Kherif (CHUV), Baldur van Lew (LUMC), Dimitropoulos Harry (UoA), Alexis Mitelpunkt (TAU), Sašo Džeroski (JSI), Cesar Matos (EPFL) |
| 19-08-2014 | SP8 weekly meeting | Online | Ferath Kherif (CHUV), Tea Danelutti (CHUV), Mihaela Damian (CHUV), Jing Cui (CHUV), Xuesong Lu (EPFL), Cesar Matos (EPFL), Thomas Heinis (EPFL), Sašo Džeroski (JSI), Bernard Zenko (JSI), Baldur van Lew (LUMC), Boudewijn LELIEVELDT (LUMC), Alexis Mitelpunkt (TAU), Tal Galili (TAU), Dimitropoulos Harry (UoA), Sabine Rehberger-Schneider (UHEI) |
| 26-08-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Jing Cui (CHUV), Anastasia Ailamaki (EPFL), Xuesong Lu (EPFL), Cesar Matos (EPFL), Alexis Mitelpunkt (TAU), John Ashburner (UCL), Eleni Zacharia (UoA) |

| Date | Description | Location | Participants |
|---|---|---|---|
| 02-09-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Jing Cui (CHUV), Mihaela Damian (CHUV), Xuesong Lu (EPFL), Cesar Matos (EPFL), Thomas Heinis (EPFL), Bosco Paolo (FBF), Sašo Džeroski (JSI), Baldur van Lew (LUMC), Alexis Mitelpunkt (TAU), Tal Galili (TAU), Tal Kozlovski, Eleni Zacharia (UoA), Dimitropoulos Harry (UoA), Lefteris Stamatogiannakis |
| 09-09-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Mihaela Damian (CHUV), Bart van Damme (CHUV), Xuesong Lu (EPFL), Darius Sidlauskas (EPFL), Thomas Heinis (EPFL), Bernard Zenko (JSI), Boudewijn LELIEVELDT (LUMC), Alexis Mitelpunkt (TAU), Tal Galili (TAU), Mira Marcus-Kalish (TAU), Eleni Zacharia (UoA), Dimitropoulos Harry (UoA) |
| 16-09-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Mihaela Damian (CHUV), Jing Cui (CHUV), Bart van Damme (CHUV), Cesar Matos (EPFL), Xuesong Lu (EPFL), Darius Sidlauskas (EPFL), Manos Karpathiotakis, Vasilis Vassalos (AUEB), Boudewijn LELIEVELDT (LUMC), Sašo Džeroski (JSI), Bernard Zenko (JSI), Baldur van Lew (LUMC), Eleni Zacharia (UoA) |
| 17-09-2014 | SP8 meeting / CHUV-TAU | Lausanne | Richard Frackowiak (CHUV), Alexis Mitelpunkt (TAU) |
| 23-09-2014 | SP8 weekly meeting | Online | Tassos Venetis (AUEB), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Bart van Damme (CHUV), Jing Cui (CHUV), Xuesong Lu (EPFL), Cesar Matos (EPFL), Thomas Heinis (EPFL), Bernard Zenko (JSI), Tal Galili (TAU), Eleni Zacharia (UoA), Dimitropoulos Harry (UoA) |
| 29-09-2014 | HBP Summit 2014 – SP8 workshop, session 1 | Heidelberg | SP8 team |

## 6.2 Scientific Coordination: HBP Meetings

This table lists meetings between this SP and other SPs.

| Date | Description | Location | Participants | Comments |
|---|---|---|---|---|
| 08-09 10 2014 | SP13 session during HBP Summit 2013 | Lausanne (EPFL) | Tea Danelutti (CHUV) | With SP13 |
| 05-02-2014 | Meeting SP8-HBP STO | Lausanne (CHUV) | Tea Danelutti (CHUV), Ferath Kherif (CHUV), Thomas Heinis (EPFL), Sabine Rehberger-Schneider (UHEI) | With SP13 |
| 07-05-2014 | Webinar on Future Medicine | Online | Ferath Kherif (CHUV) | With SP12 |
| 25-06-2014 | Meeting SP8-SP6 | Lausanne (EPFL) | Ferath Kherif (CHUV), Jeffrey Christopher Muller (EPFL) | With SP6 |
| 03-07-2014 | Meeting SP8-SP5 | Lausanne (CHUV) | Ferath Kherif (CHUV), Xavier Vasques (EPFL), Catherine Zwahlen (EPFL) | With SP5 |
| 07-07-2014 | Meeting SP8-HBP IP&TT | Lausanne (CHUV) | Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Tea Danelutti (CHUV), Kathleen Elsig (EPFL), Sebastien Gouin (EPFL) | With SP13 |
| 18-09-2014 | Meeting SP8-SP6 | Lausanne (EPFL) | Ferath Kherif (CHUV), Mihaela Damian (CHUV), Jeffrey Christopher Muller (EPFL), Cesar Matos (EPFL), Xuesong Lu (EPFL), Thomas Heinis (EPFL), Anastasia Ailamaki (EPFL), Darius Sidlauskas (EPFL), Manos Karpathiotakis | With SP6 |
| 30-09-2014 | HBP summit 2014 - cross-SP meeting: SP2, SP5, SP8 | Heidelberg | SP2 team, SP5 team, SP8 team | With SP2 and SP5 |
| 30-09-2014 | HBP summit 2014 - SP13 workshop | Heidelberg | Tea Danelutti (CHUV) + Admin Representatives from other SPs | With SP13 |

## 6.3 Scientific Coordination: External Meetings

This table lists meetings between this SP and Partners outside the HBP.

| Date | Description | Location | Participants | Comments |
|---|---|---|---|---|
| 27-11-2013 | Meeting SP8-Elsevier | Phone | Richard Frackowiak (CHUV), Bogdan Draganski (CHUV), Ferath Kherif (CHUV), Sean Hill (KI), Anastasia Ailamaki (EPFL), Tea Danelutti (CHUV) + ELSEVIER participants | |
| 03-12-2013 | SP8 Meeting with Sanofi-Aventis | Paris | Richard Frackowiak (CHUV) + Sanofi-Aventis participants | |
| 27-01-2014 | Meeting SP8-EUDAT | Lausanne (CHUV) | SP8:<br>Richard Frackowiak (CHUV), Ferath Kherif (CHUV), Bogdan Draganski (CHUV), Anastasia Ailamaki (EPFL), Thomas Heinis (EPFL), Daniel Mallmann (JUELICH), Xuesong Lu (EPFL)<br>EUDAT:<br>Daniel Mallmann (JUELICH), Claudio Cacciari (CINECA), Mark Van de Sanden (SURFsara), Peter Wittenburg (Max Planck Institute for Psycholinguistics) | |
| 01-02-2014 | SP8 meeting – Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV) | Meeting with potential software provider (exact date unknown, but before March 2014) |
| 25-02-2014 | Workshop HBP-Elsevier | Lausanne (CHUV) | Ferath Kherif (CHUV), Richard Frackowiak (CHUV), Bogdan Draganski (CHUV), Sean Hill (KI), Stephane Berghmans, Judson Dunham, Maarten Cleeren, Felisa van Hasselt, Thomas Heinis (EPFL), Jing Cui (CHUV), Mihaela Damian (CHUV) | With the participation of SP8 and SP5 |
| 10-04-2014 | Meeting with Pr Brice and Pr Agid | Paris | Richard Frackowiak (CHUV) | |
| 09-05-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Cesar Matos (EPFL), Thomas Heinis (EPFL) | With potential software provider |
| 14-05-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Mihaela Damian (CHUV) | With potential software provider |
| 28-05-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Mihaela Damian (CHUV), Richard Frackowiak (CHUV), Cesar Matos (EPFL), Thomas Heinis (EPFL) | With potential software provider |

| Date | Description | Location | Participants | Date |
|---|---|---|---|---|
| 11-06-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Mihaela Damian (CHUV) | With potential software provider |
| 13-06-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Mihaela Damian (CHUV) | With potential software provider |
| 18-06-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Mihaela Damian (CHUV) | With potential software provider |
| 23-06-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Mihaela Damian (CHUV), Thomas Heinis (EPFL) | With potential software provider |
| 09-07-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Mihaela Damian (CHUV), Cesar Matos (EPFL), Xuesong Lu (EPFL) | With potential software provider |
| 11-07-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Thomas Heinis (EPFL), Cesar Matos (EPFL), Mihaela Damian (CHUV), Richard Frackowiak (CHUV) | With potential software provider |
| 16-07-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Mihaela Damian (CHUV), Cesar Matos (EPFL), Thomas Heinis (EPFL) | With potential software provider |
| 21-07-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Mihaela Damian (CHUV), Cesar Matos (EPFL), Thomas Heinis (EPFL) | With potential software provider |
| 06-08-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Mihaela Damian (CHUV), Cesar Matos (EPFL), Jing Cui (CHUV) | With potential software provider |
| 12-08-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Bogdan Draganski (CHUV), Thomas Heinis (EPFL), Cesar Matos (EPFL), Mihaela Damian (CHUV) | With potential software provider |
| 13-08-2014 | Meeting SP8-Elsevier | Online | Richard Frackowiak (CHUV), Stéphane Berghmans (Elsevier) | |
| 22-08-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Mihaela Damian (CHUV) | With potential software provider |
| 28-08-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Mihaela Damian (CHUV) | With potential software provider |
| 29-08-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Mihaela Damian (CHUV) | With potential software provider |
| 02-09-2014 | Meeting SP8-Indian Institute of Science | Lausanne (CHUV) | Ferath Kherif (CHUV), Prof. Vijayalakshmi Ravindranath (Centre for Neuroscience) Prof. Govindan Rangarajan (Department of Mathematics) | |

| Date | Description | Location | Participants | Date |
|---|---|---|---|---|
| 04-09-2014 | Meeting SP8-Indian Institute of Science | Lausanne (CHUV) | Richard Frackowiak (CHUV), Henry Markram (EPFL), Giovanni Frisoni (HUG), Ferath Kherif (CHUV), Prof. Vijayalakshmi Ravindranath (Centre for Neuroscience)<br>Prof. Govindan Rangarajan (Department of Mathematics) | |
| 10-09-2014 | SP8 Meeting with the European Academy of Neurology | Brussels | Richard Frackowiak (CHUV) + European Academy of Neurology and the European Brain Council | |
| 15-09-2014 | Meeting SP8-Elsevier | Online | Richard Frackowiak (CHUV), Stéphane Berghmans (Elsevier) | |
| 16-09-2014 | SP8 meeting - Data Federation | Lausanne (CHUV) | Ferath Kherif (CHUV), Bogdan Draganski (CHUV) | With potential software provider |
| 19-09-2014 | SP8 meeting - Data Federation | Online | Ferath Kherif (CHUV), Bogdan Draganski (CHUV), Mihaela Damian (CHUV), Cesar Matos (EPFL), Thomas Heinis (EPFL) | With potential software provider |
| 26-09-2014 | SP8 meeting - Data Federation | Online | Ferath Kherif (CHUV), Mihaela Damian (CHUV) | With potential software provider |

## 6.4 Scientific Coordination: Monitoring and Quality Control

SP8 meetings are held online every week in order to monitor progress, and share information with Partners. A two-day workshop took place late April in Lausanne. Representatives from each SP8 Partner institution (included the two new Partners selected through the Competitive Call) joined the event and presented their work. Tasks for the third quarter were also identified and assigned. Scientific coordination was needed for the collaboration between WP8.1 and WP8.2, as well as for the collaboration with SP5 (Neuroinformatics) and SP6 (Unified Portal).

Ethics is a very important issue for the MIP. Therefore, SP8 and SP12 have to collaborate closely. Since the start of the Project, Richard Frackowiak (CHUV, P23) worked on anonymisation issues associated with proposed EU data privacy regulation (Science Europe). Ferath Kherif (CHUV, P23) has been appointed as SP8 Ethics Respondent for the HBP.

## 6.5 Scientific Coordination: Additional Comments

In the third quarter, the Scientific Coordination team has been working on the Framework Partnership Agreement (FPA) and communicating around the modifications proposed.

## 6.6 Scientific Coordination: The Next Six Months

In the next six months, SP8 will continue to have weekly meetings.

# Annex A: Milestones

| No. | Milestone Name | WP | Month Due | Month Achieved | See Page |
|---|---|---|---|---|---|
| M153 | Specification of requirements for data federation and anonymisation tools; requirements for web-based user services | 8.1 | 6 | 6 | WP8.1 |
| M158 | Recruitment of hospitals and other data sources in progress. | 8.2 | 6 | 6 | WP8.2 |
| M161 | Algorithms for deriving disease signatures, mechanisms and Milestones: identification and preliminary evaluation of clustering/data mining algorithms. | 8.3 | 6 | 6 | WP8.3 |
| M166 | Medical Informatics Platform fully specified | 8.4 | 6 | 6 | WP8.4 |
| M169 | Requirements for user documentation and support for Medical Informatics Platform; guidelines for establishing alliances. | 8.5 | 6 | 6 | WP8.5 |
| M154 | Complete concept for data federation and anonymisation tools. | 8.1 | 12 | 12 | WP8.1 |
| M162 | Identification of data mining algorithms for identification of disease signatures; final parallelisation concept for use on supercomputers/clusters. | 8.2 | 12 | 12 | WP8.2 |
| M170 | Identification of potential medical informatics alliances. | 8.5 | 12 | 12 | WP8.5 |
| MS288 | Prototype for t-SNE and t-SNE + GENESPACE in a representative large sample. (T8.3.3) | 8.3 | 12 | 12 | WP8.3 |
| MS292 | Algorithms identified and specifications written. (T8.3.4) | 8.3 | 12 | 12 | WP8.3 |

# Annex B: Subproject Functions

**Functions: WP8.1**

| No. | Function Name | Start (plan) | Start (actual) | Complete (plan) | Complete (actual) | See |
|---|---|---|---|---|---|---|
| 8.1.1.1 | Requirement analysis | Oct 2013 | Oct 2013 | Jan 2014 | Jan 2014 | T8.1.1 |
| 8.1.1.2 | Export mechanisms and data formats | Dec 2013 | Dec 2013 | Mar 2014 | Mar 2014 | T8.1.1 |
| 8.1.1.3 | Interface to Data Federation layer analysis | Mar 2014 | Mar 2014 | Jun 2014 | Jun 2014 | T8.1.1 |
| 8.1.1.4 | Implementation for flat CSV files and imaging data | Mar 2014 | Mar 2014 | Nov 2014 | - | T8.1.1 |
| 8.1.1.5 | Definition and implementation of test cases | Nov 2014 | - | Feb 2015 | - | T8.1.1 |
| 8.1.1.6 | Interface to hospital data warehouse | Jul 2015 | - | Aug 2015 | - | T8.1.1 |
| 8.1.1.7 | Exporting systems to other hospitals | Sep 2015 | - | Dec 2015 | - | T8.1.1 |
| 8.1.2.1 | Extract/Transform/Load process | Jan 2014 | Jan 2014 | Jul 2014 | Delayed | T8.1.2 |
| 8.1.2.2 | Ontology/schema mapping | Mar 2014 | Mar 2014 | Oct 2014 | Oct 2014 | T8.1.2 |
| 8.1.2.3 | Manual schema mapping | Mar 2014 | Mar 2014 | Dec 2014 | - | T8.1.2 |
| 8.1.2.4 | Schema mapping across schemata and ontologies | Aug 2014 | Aug 2014 | Jun 2015 | - | T8.1.2 |
| 8.1.2.5 | Query rewriting across hospital schemata | Apr 2015 | - | Feb 2016 | - | T8.1.2 |
| 8.1.3.1 | Complex Dataflow Processing Engine | Nov 2013 | Nov 2013 | Oct 2015 | - | T8.1.3 |
| 8.1.3.2 | User Defined Functions (UDFs) for Complex Dataflow Support | May 2014 | May 2014 | May 2015 | - | T8.1.3 |
| 8.1.4.1 | Anonymisation of patient data | Oct 2013 | Oct 2013 | Mar 2014 | Mar 2014 | T8.1.4 |
| 8.1.4.2 | Anonymisation module | Jan 2016 | - | Mar 2016 | - | T8.1.4 |

**Functions: WP8.3**

| No. | Function Name | Start (plan) | Start (actual) | Complete (plan) | Complete (actual) | See Page |
|---|---|---|---|---|---|---|
| 8.3.1.1 | Descriptive Function to provide summary statistics | Dec 2013 | Dec 2013 | Oct 2014 | Oct 2014 | T8.3.1 |
| 8.3.1.2 | Parametric tests | Oct 2014 | - | Apr 2015 | - | T8.3.1 |
| 8.3.1.3 | Supervised and unsupervised data mining algorithms | Oct 2014 | - | Apr 2016 | - | T8.3.1 |
| 8.3.2.1 | Label propagation framework | Feb 2014 | Feb 2014 | Apr 2014 | Shelved | T8.3.2 |
| 8.3.2.2 | Model construction for toy data, i.e. small representative clinical data set | May 2014 | May 2014 | Sep 2014 | Shelved | T8.3.2 |
| 8.3.2.3 | Sparse matrices and Intra-session alignment | Oct 2014 | - | Mar 2015 | - | T8.3.2 |
| 8.3.2.4 | Integration with segmentation | Apr 2015 | - | Aug 2015 | - | T8.3.2 |
| 8.3.2.5 | Testing and optimisation | Sep 2015 | - | Dec 2015 | - | T8.3.2 |
| 8.3.2.6 | Redefine features | Jan 2016 | - | Mar 2016 | - | T8.3.2 |

**Functions: WP8.5**

| No. | Function Name | Start (plan) | Start (actual) | Complete (plan) | Complete (actual) | See Page |
|---|---|---|---|---|---|---|
| 8.5.3.1 | Build the MIP Training and Support Centre | May 2014 | May 2014 | Dec 2015 | - | T8.5.3 |
| 8.5.3.2 | Preparing training materials | Jun 2014 | Jun 2014 | Dec 2015 | - | T8.5.3 |
| 8.5.3.3 | “Face 2 Face” and Basic "e-learning" | Dec 2014 | - | Dec 2015 | - | T8.5.3 |
| 8.5.3.4 | Professional e-learning | Dec 2015 | - | Mar 2016 | - | T8.5.3 |

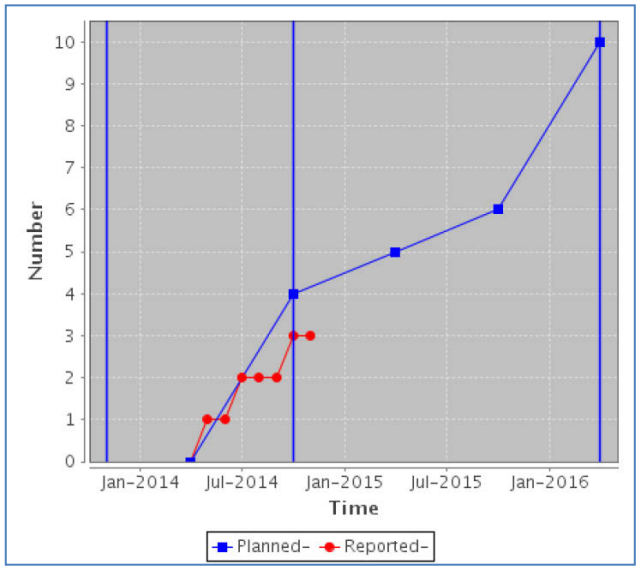
# Annex C: Scientific Key Performance Indicators (SKPIs)

Current KPI information for this SP can be found on the STO website:

https://flagship.kip.uni-heidelberg.de/jss/CollectKPI?ui=268&s=UJuR3AgTezrb&um=sPO&oSP=8

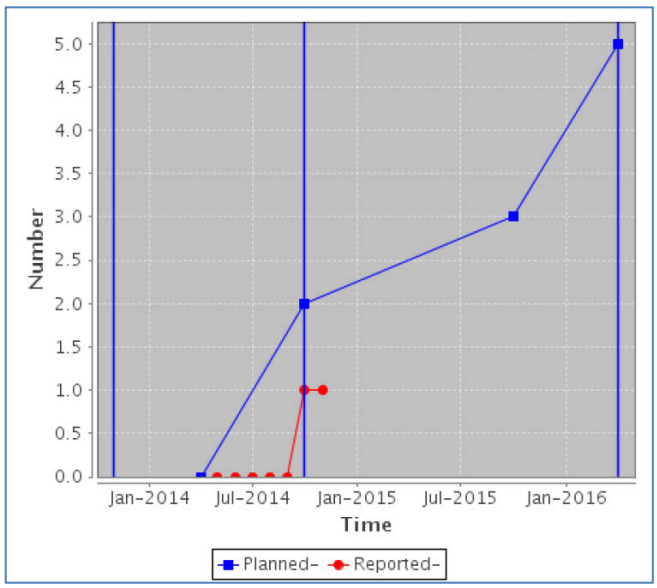
### *SP8_SKPI_13 Number of ontologies defined for the necessary vocabulary/data sets*

- Responsible: anastasia.ailamaki@epfl.ch
- 2014/10/31 (value 3): The definition of the fourth ontology has been started but not yet finished.



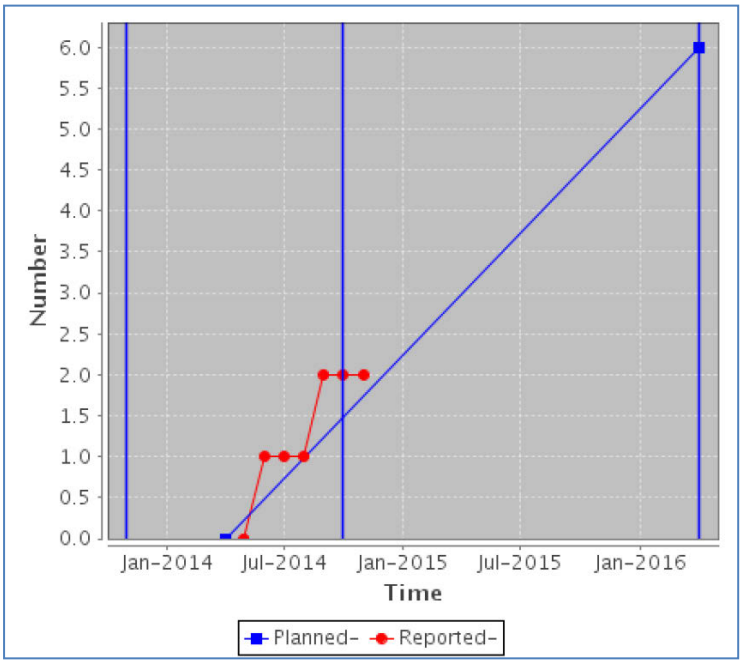

### *SP8_SKPI_14 Number of mappings between hospitals (via the Federation layer)*

- Responsible: anastasia.ailamaki@epfl.ch
- 2014/09/30 (value 1): The federation of the second hospital is delayed due to the delay in the tendering processing of the federation company.



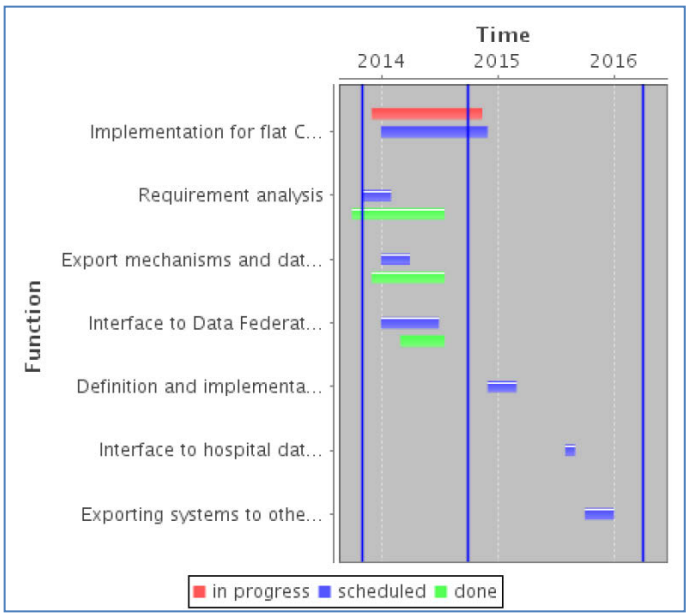
### *SP8_SKPI_15 Number of template queries*

- Responsible: anastasia.ailamaki@epfl.ch



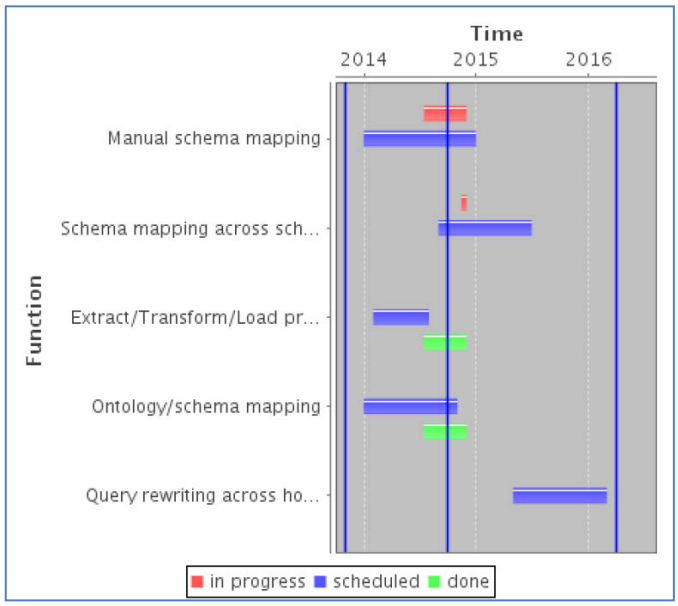

### *SP8_SKPI_21 In situ distributed database querying*

- Responsible: anastasia.ailamaki@epfl.ch
- Requirement analysis. Planned: 2013/10/31 - 2014/01/31
- Export mechanisms and data formats. Planned: 2013/12/31 - 2014/03/31
- Interface to Data Federation layer. Planned: 2013/12/31 - 2014/06/30
- Implementation for flat CSV files and imaging data. Planned: 2013/12/31 - 2014/11/30
- Definition and implementation of test cases. Planned: 2014/11/30 - 2015/02/28
- Interface to hospital data warehouse. Planned: 2015/07/31 - 2015/08/31
- Exporting systems to other hospitals. Planned: 2015/09/30 - 2015/12/31



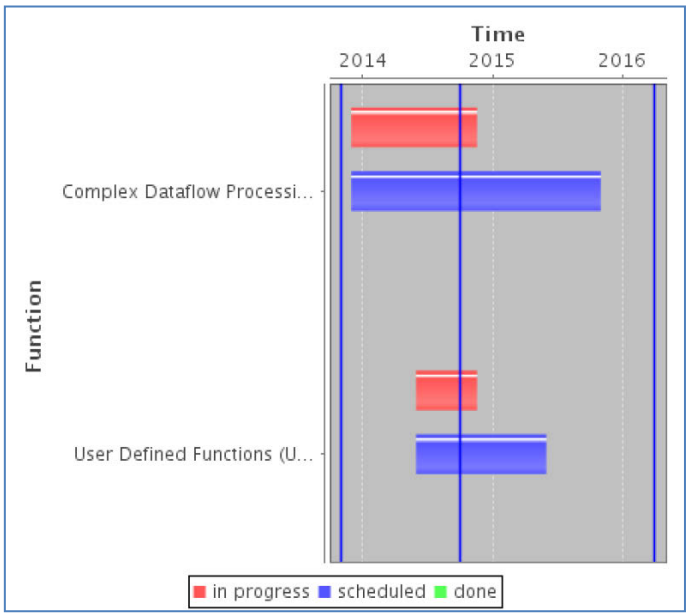
### *SP8_SKPI_22 Data integration*

- Responsible: avenet@aueb.gr
- Extract/Transform/Load process. Planned: 2014/01/31 - 2014/07/31
- Ontology/schema mapping. Planned: 2013/12/31 - 2014/10/31
- Manual schema mapping. Planned: 2013/12/31 - 2014/12/31
- Schema mapping across schemata and ontologies. Planned: 2014/08/31 - 2015/06/30
- Query rewriting across hospital schemata. Planned: 2015/04/30 - 2016/02/29



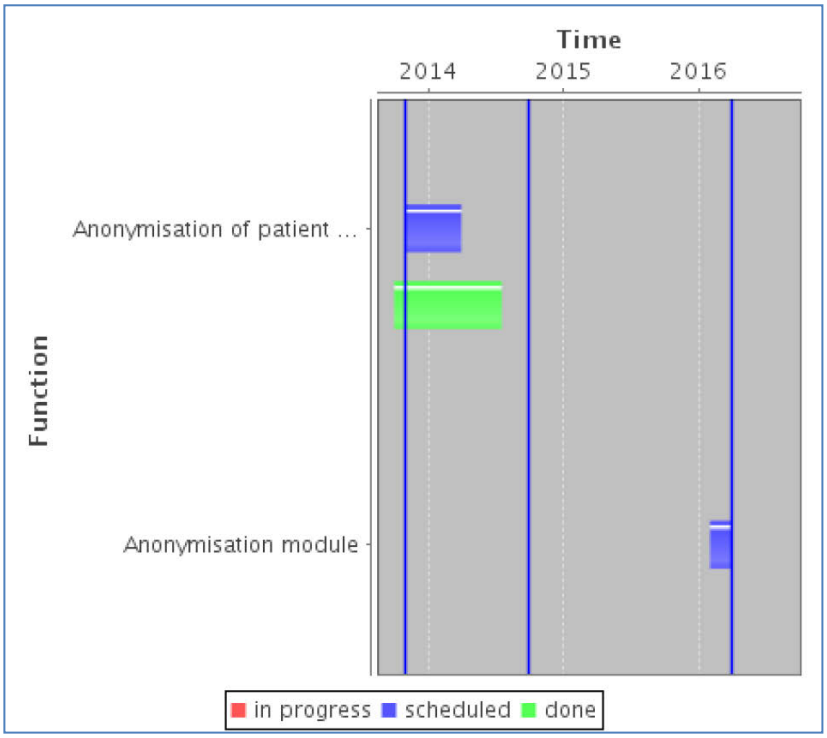

### *SP8_SKPI_23 Workflow infrastructure*

- Responsible: harryd@di.uoa.gr
- Complex Dataflow Processing Engine. Planned: 2013/11/30 - 2015/10/31
- User Defined Functions (UDFs) for Complex Dataflow Support. Planned: 2014/05/31 - 2015/05/31



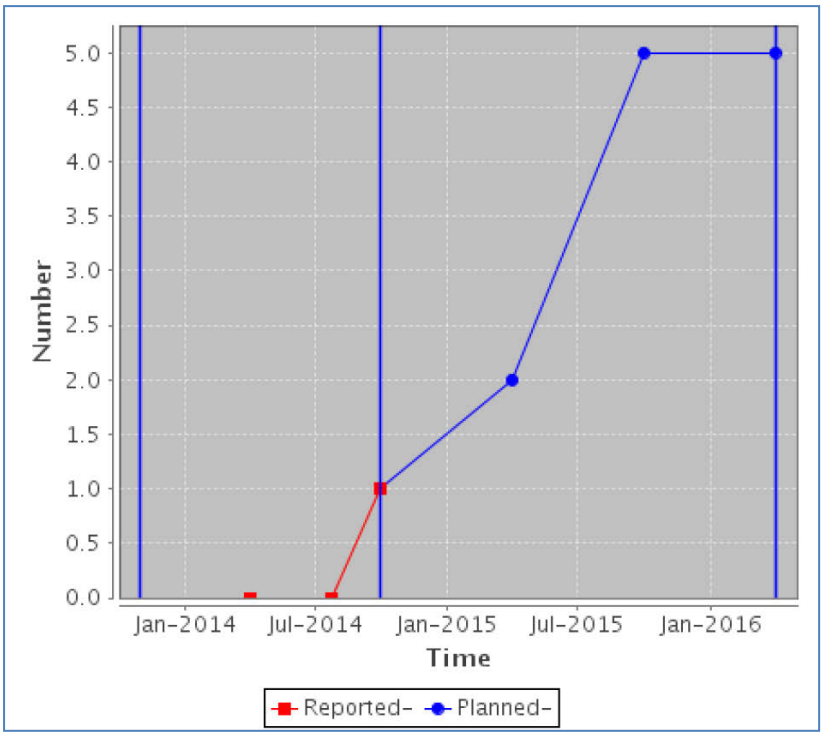
### *SP8_SKPI_24 Data privacy*

- Responsible: anastasia.ailamaki@epfl.ch
- Anonymisation of patient data. Planned: 2013/10/31 - 2014/03/31
- Anonymisation module. Planned: 2016/01/31 - 2016/03/31



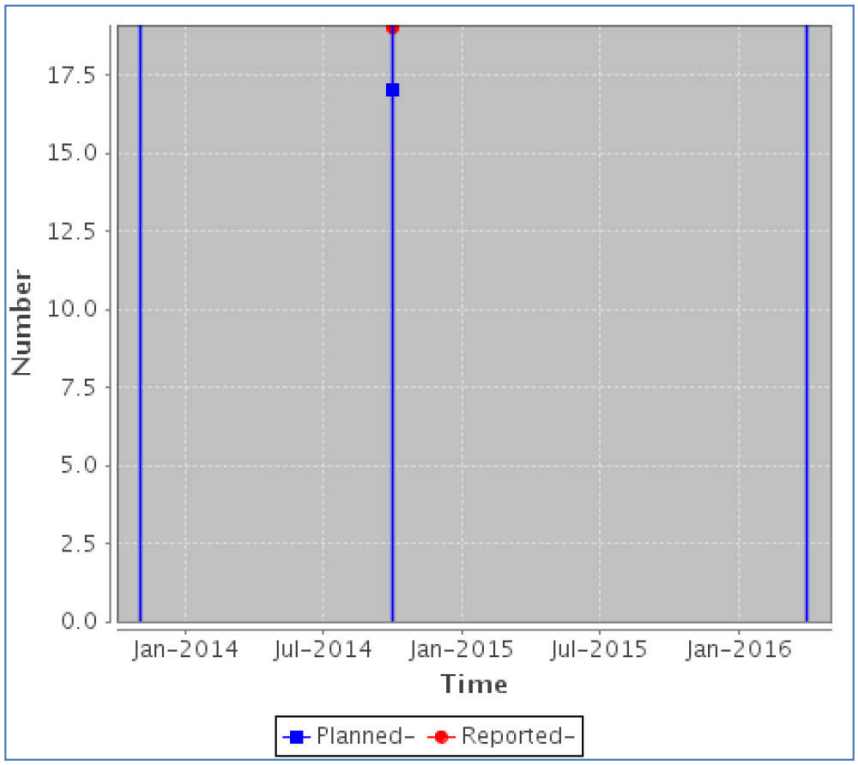

### *SP8_SKPI_01 Hospitals integrated into Federation*

- Responsible: ferath.kherif@chuv.ch



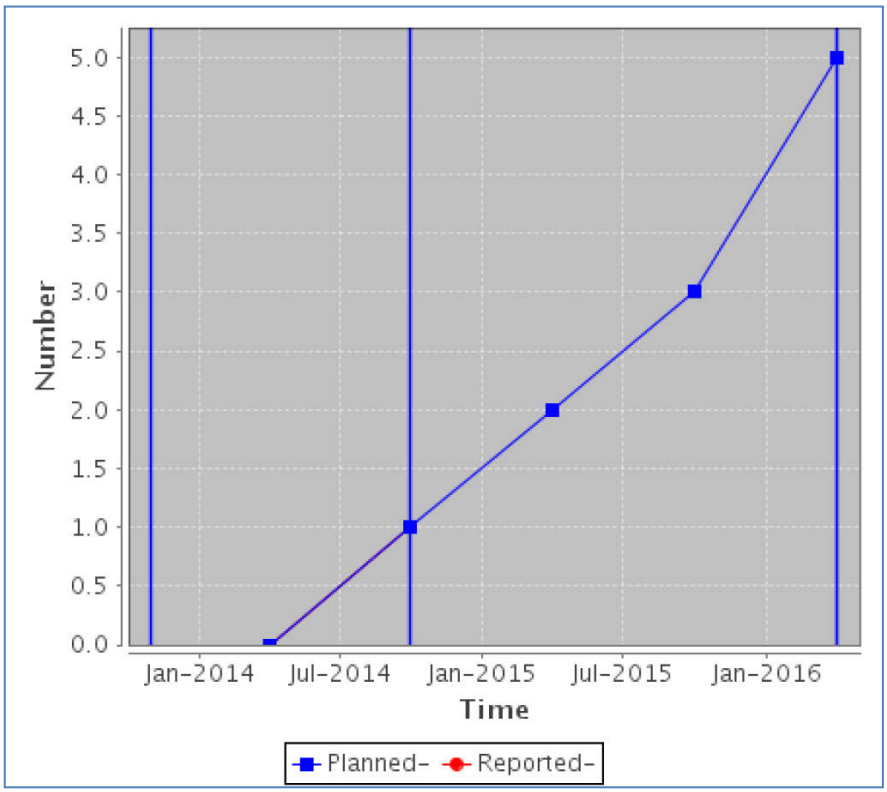
### *SP8_SKPI_02 Communication with stakeholders initiated*

- Responsible: ferath.kherif@chuv.ch



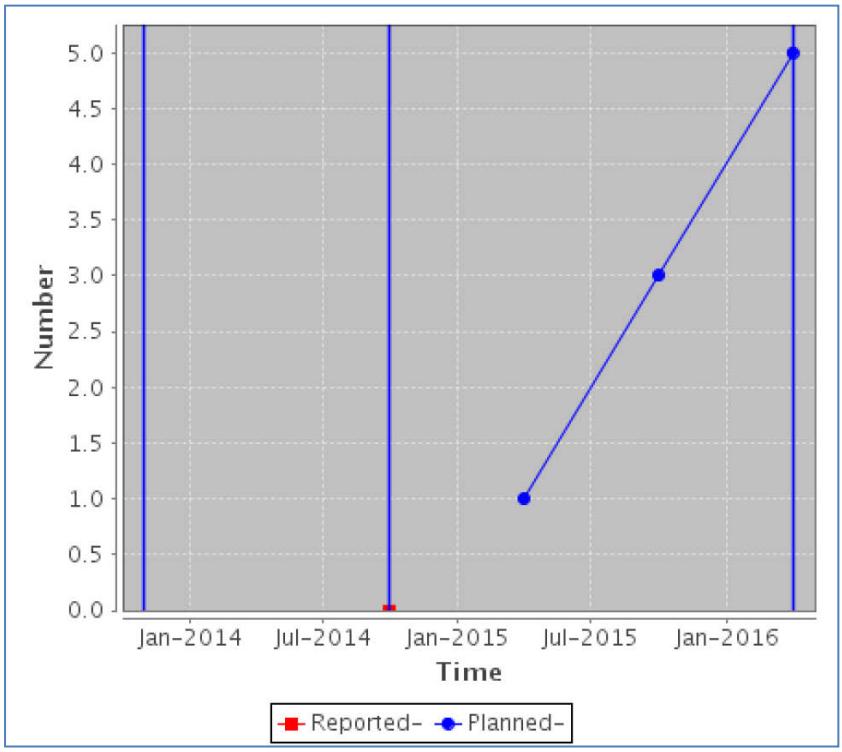

### *6.6.1 SP8_SKPI_03 Non-image data (no of hospitals)*

- Responsible: ferath.kherif@chuv.ch



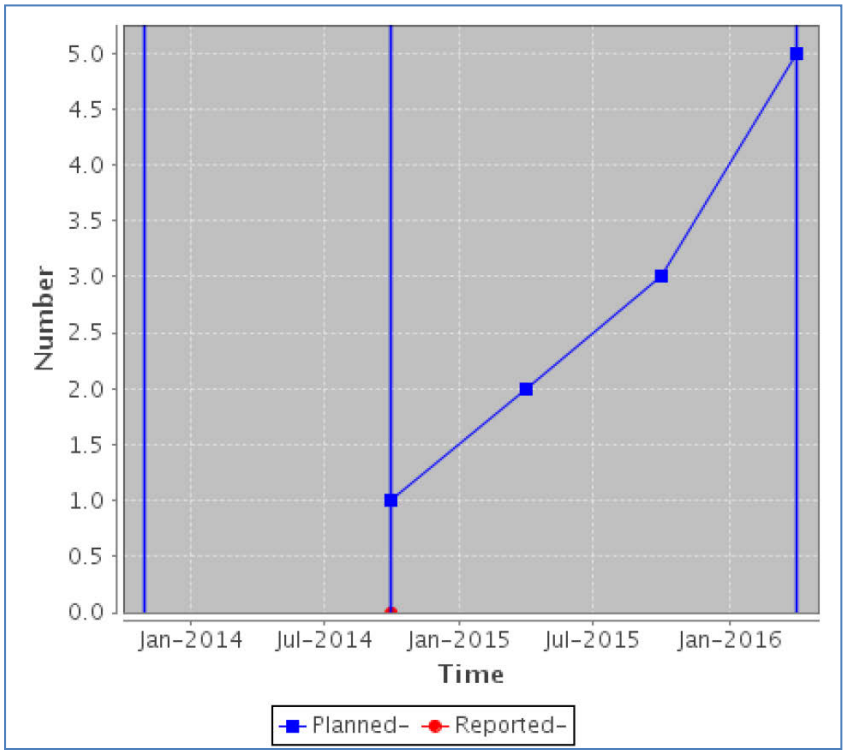
### *SP8_SKPI_04 Image data (no of hospitals)*

- Responsible: ferath.kherif@chuv.ch



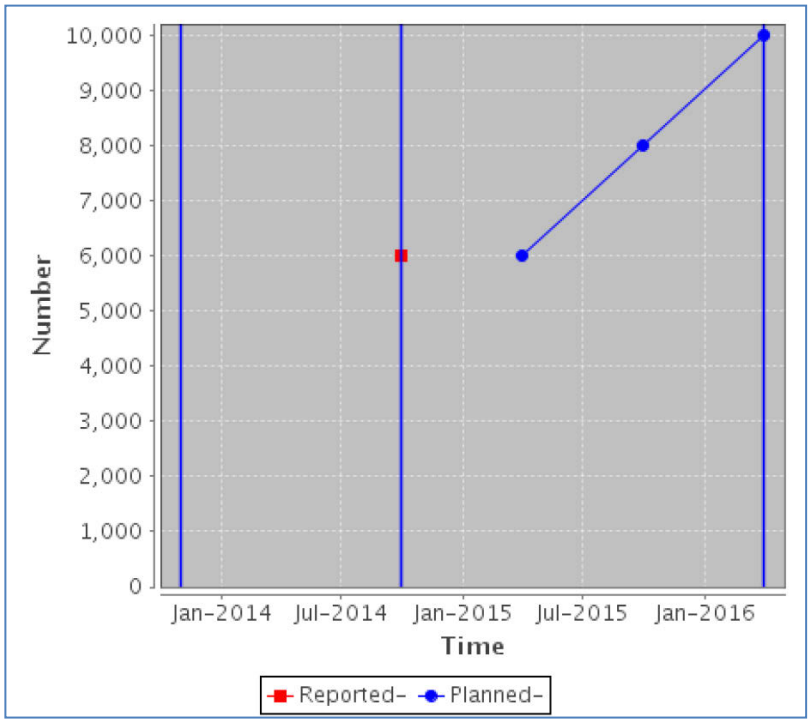

### *SP8_SKPI_05 Anonymisation of hospital data (no of hospitals)*

- Responsible: ferath.kherif@chuv.ch 2014/09/30
  anonymisation will start in M13.



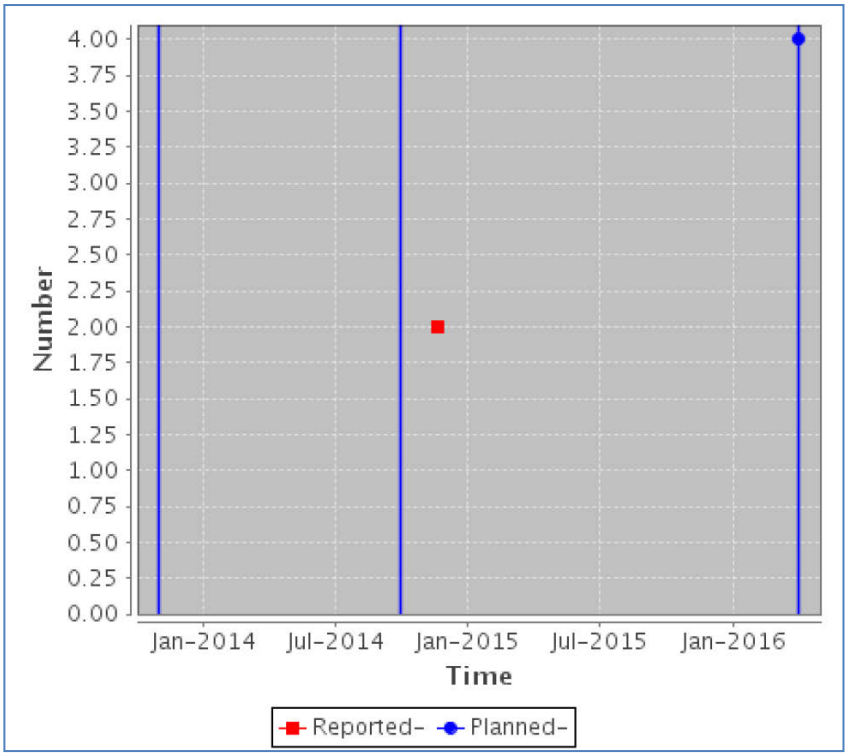
### *6.6.2 SP8_SKPI_06 Number of Images within MIP*

- Responsible: ferath.kherif@chuv.ch



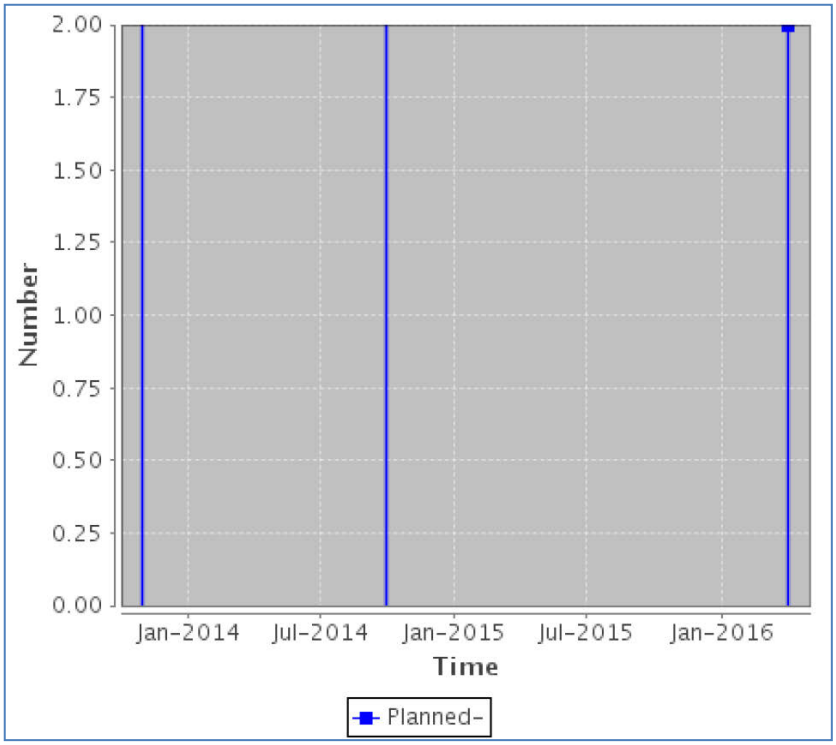

### *SP8_SKPI_07 Number of research collaboration agreements/data transmission agreements with large-scale studies*

- Responsible: ferath.kherif@chuv.ch



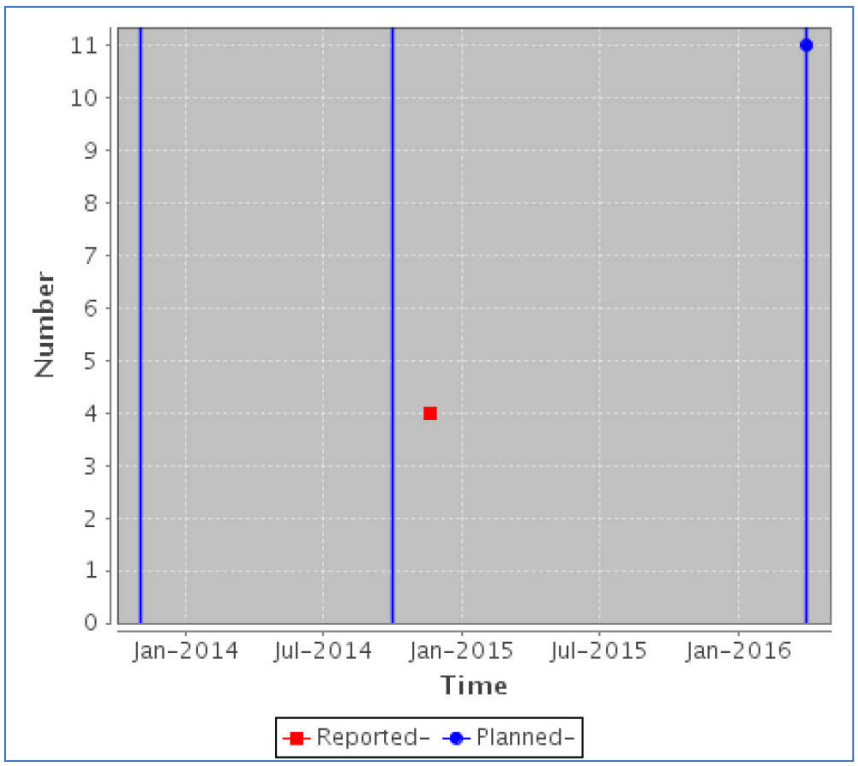
### *SP8_SKPI_08 Number of research collaboration agreements with the pharmaceutical companies*

- Responsible: ferath.kherif@chuv.ch



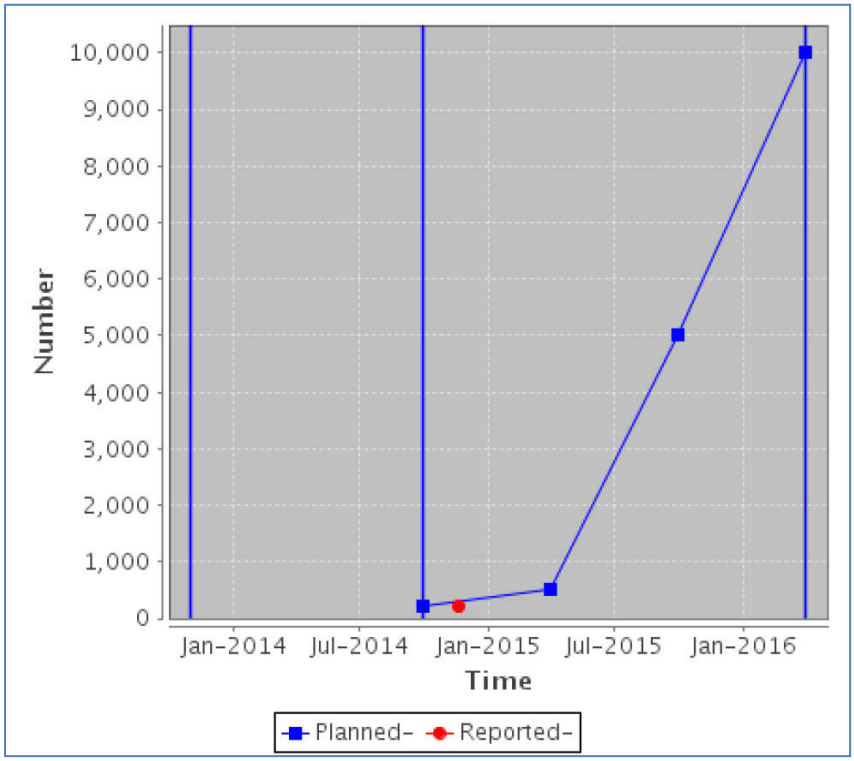

### *SP8_SKPI_09 Number of patient studies involved*

- Responsible: ferath.kherif@chuv.ch



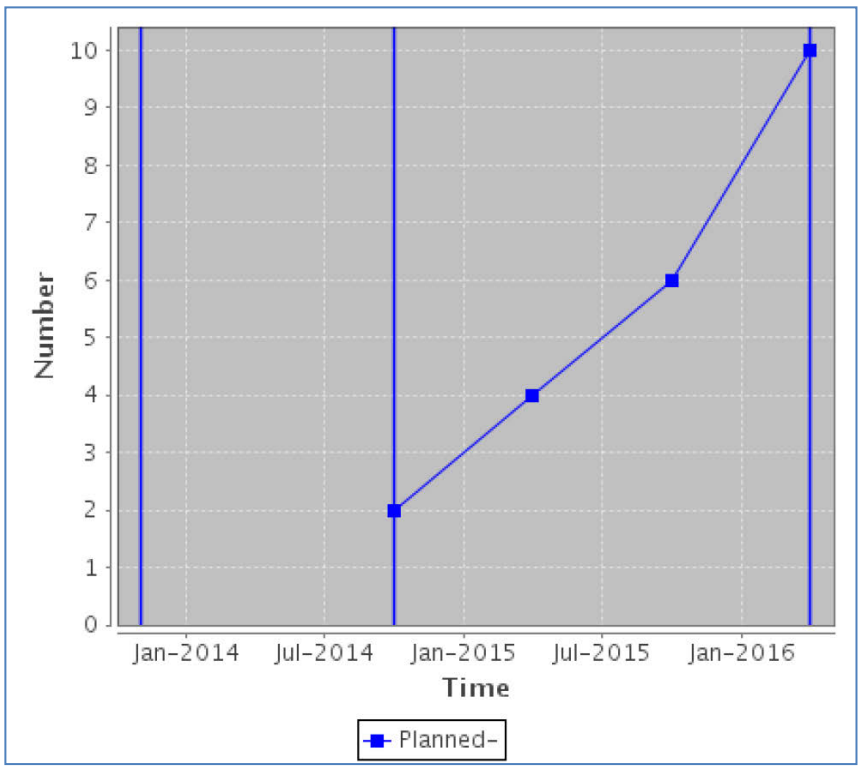
### *SP8_SKPI_10 Number of standard features available in Biological Signatures of Diseases*

- Responsible: miram@post.tau.ac.il



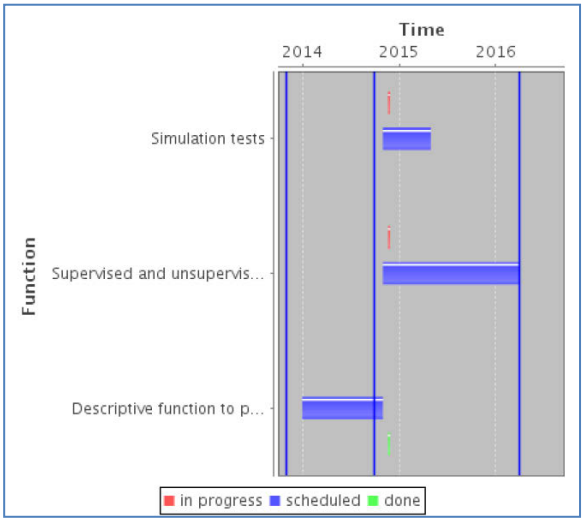

### *SP8_SKPI_11 Data mining: Number of models identified and tested*

- Responsible: miram@post.tau.ac.il



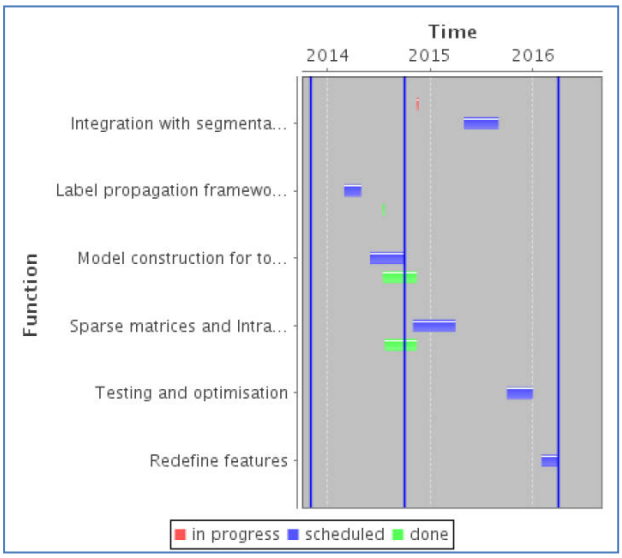
### *SP8_SKPI_25 Medical intelligence tools for data categorisation*

- Responsible: miram@post.tau.ac.il
- Descriptive Function to provide summary statistics. Planned: 2013/12/31 - 2014/10/31
- Simulation tests. Planned: 2014/10/31 - 2015/04/30
- Supervised and unsupervised data mining algorithms. Planned: 2014/10/31 - 2016/03/31



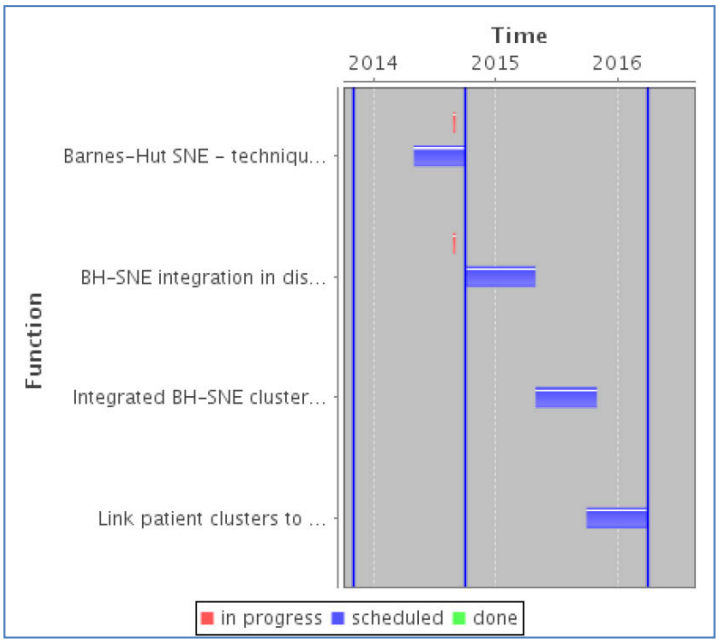

### *SP8_SKPI_26 Advanced topographical methods for brain referencing and normalisation*

- Responsible: jashburner@gmail.com
- Label propagation framework. Planned: 2014/02/28 - 2014/04/30
- Model construction for toy data, i.e. small representative clinical data set. Planned: 2014/05/31 - 2014/09/30
- Sparse matrices and Intra-session alignment. Planned: 2014/10/31 - 2015/03/31
- Integration with segmentation. Planned: 2015/04/30 - 2015/08/31
- Testing and optimisation. Planned: 2015/09/30 - 2015/12/31
- Redefine features. Planned: 2016/01/31 - 2016/03/31



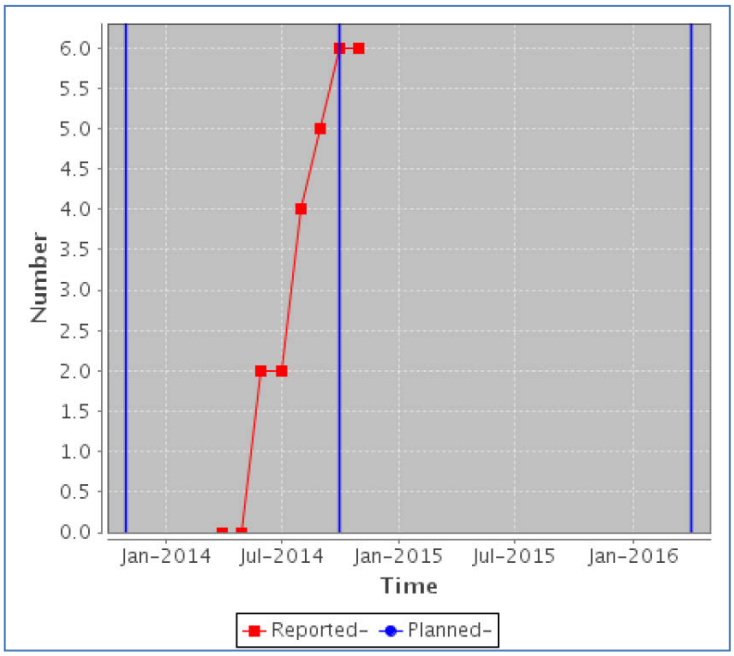
### *SP8_SKPI_28 Linking t-SNE maps and GENESPACE clusterings from the ABA*

- Responsible: b.lelieveldt@lumc.nl
  - Barnes-Hut SNE - technique validation. Planned: 2014/04/30 - 2014/09/30
  - BH-SNE integration in distributed framework. Planned: 2014/09/30 - 2015/04/30
  - Integrated BH-SNE cluster analysis to categorise patients. Planned: 2015/04/30 - 2015/10/31
  - Link patient clusters to gene expression signatures. Planned: 2015/09/30 - 2016/03/31



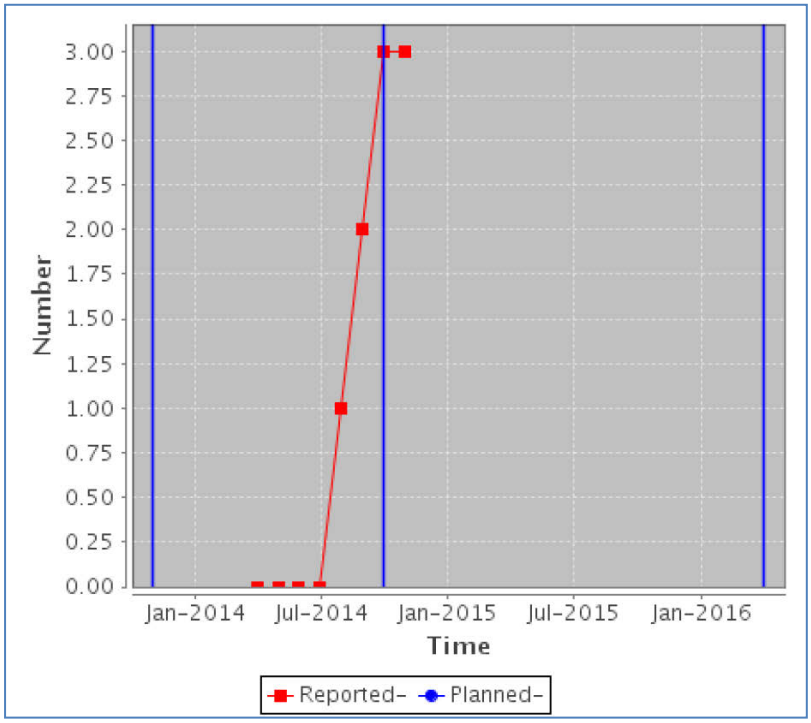

### *SP8_SKPI_29 Number of specifications of algorithms for rule-based clustering*

- Responsible: saso.dzeroski@ijs.si



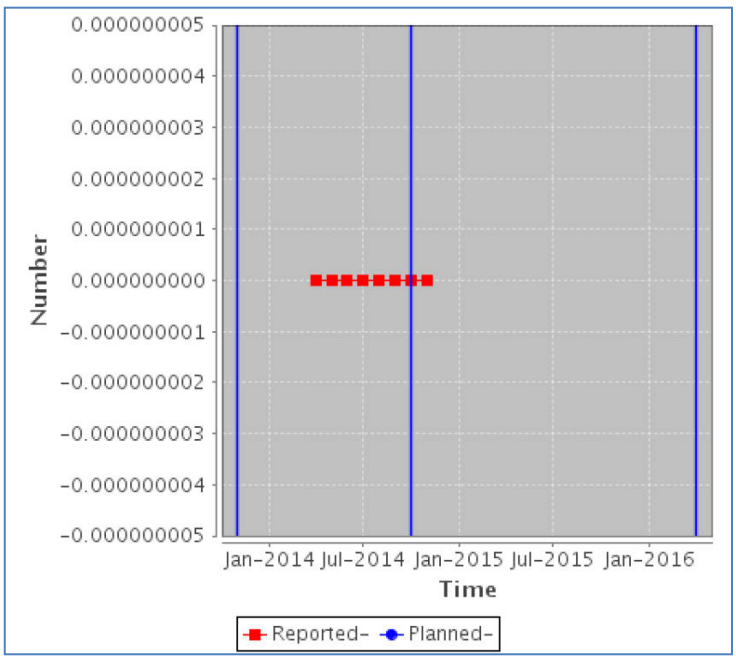
### *SP8_SKPI_30 Number of implemented prototypes of algorithms for rule-based clustering*

- Responsible: saso.dzeroski@ijs.si



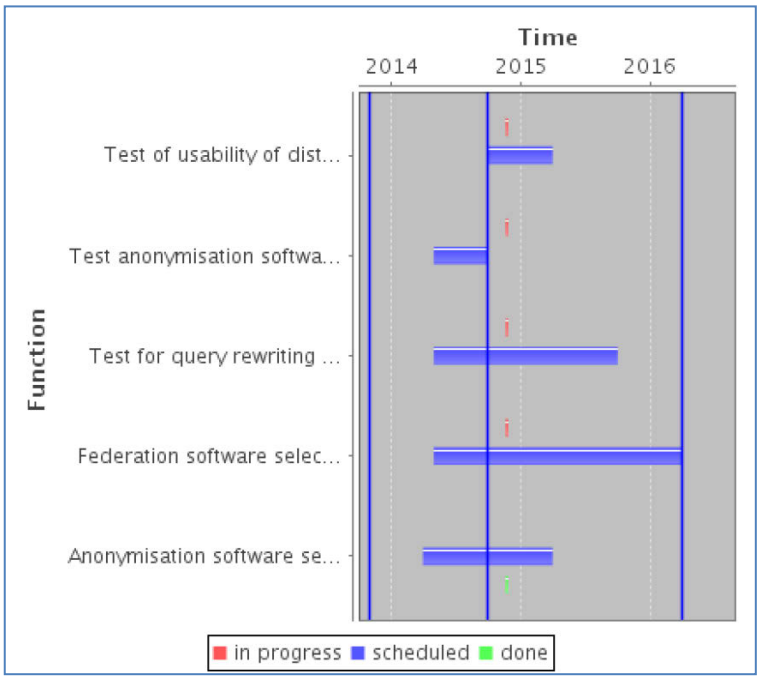

### *SP8_SKPI_31 Number of Use Cases where algorithms for RBC are applied to/ evaluated on data sets/case-studies*

- Responsible: saso.dzeroski@ijs.si



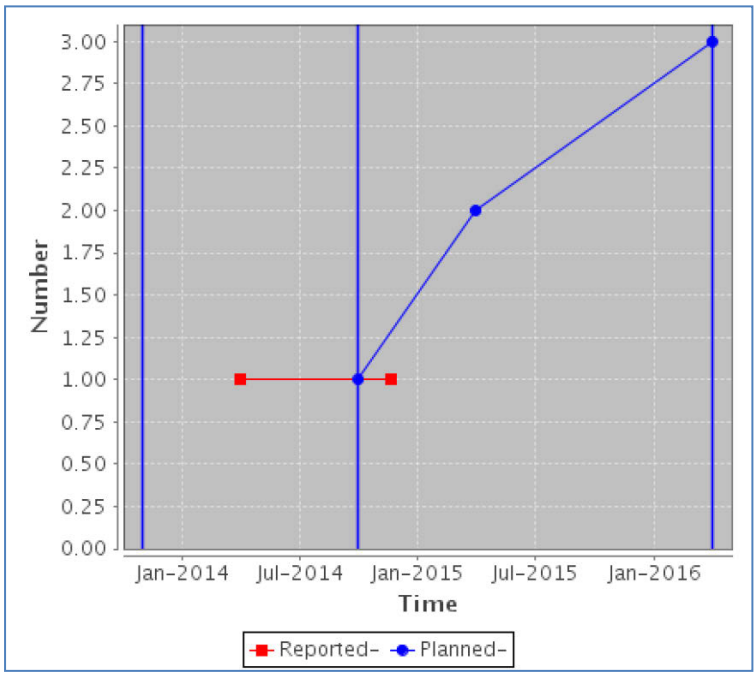
### *SP8_SKPI_16 Software implementation*

- Responsible: ferath.kherif@chuv.ch
- Test of usability of distributed processing engine. Planned: 2014/09/30 - 2015/03/31
- Test anonymisation software (subcontractor) at CHUV. Planned: 2014/04/30 - 2014/09/30
- Test for query rewriting algorithms across hospitals. Planned: 2014/04/30 - 2015/09/30
- Federation software selection. Planned: 2014/04/30 - 2016/03/31
- Anonymisation software selection. Planned: 2014/03/31 - 2015/03/31



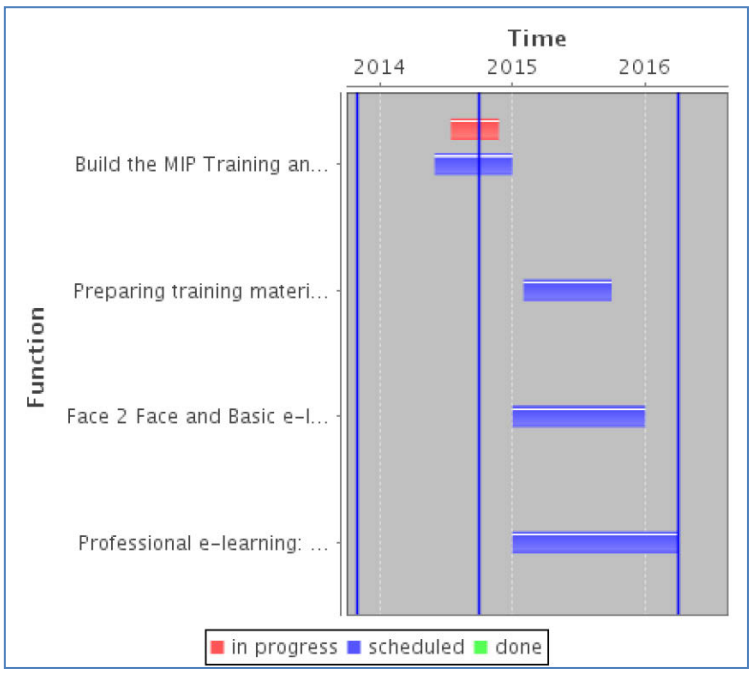

### SP8_SKPI_12 Number of workshops per year

- Responsible: aredolfi@fatebenefratelli.it



***SP8_SKPI_27 Medical Informatics Platform: user support and community building***

- Responsible: aredolfi@fatebenefratelli.it
- Build the MIP Training and Support Centre. Planned: 2014/05/31 - 2014/12/31
- Preparing training materials. Planned: 2015/01/31 - 2015/09/30
- Face 2 Face and Basic e-learning. Planned: 2014/12/31 - 2015/12/31
- Professional e-learning: online site with docs and tutorials. Planned: 2014/12/31 - 2016/03/3



# Annex D: Glossary

| Term | Definition |
|---|---|
| LOINC standard | Universal standard for identifying medical laboratory observations: http://loinc.org |
| MapReduce | A software framework that allows developers to write programs for processing and generating large data sets with a parallel, distributed algorithm on a cluster. |
| ETL process | In database usage (especially data warehousing), ETL is a process responsible for pulling data out of the source systems and placing it into a data warehouse. This process involves the following tasks: 1) Extracting data from a source, 2) Transforming them, 3) Loading them into a data warehouse or data repository |
| User-defined Function | A small program that you can write to perform an operation. User defined Functions define Functions that perform specific tasks within a larger system, such as a database or spread sheet program. |
| GUI | GUI stands for Graphical User Interface. A GUI uses icons or other visual indicators to interact with electronic devices, rather than using only text via the command line. |