redcap\_data

automatically generated from python

### Subject: 13

**Survey Timestamp**2016-09-06 18:39:39
  
**Task leader name**John Ashburner
  
**Task number attached to this component**T8.3.11 Brain morphological features UCL
  
**Name of this component**Brain morphological features
  
**Short Description of this component**A privacy preserving approach for the generalised principal component analysis of large image datasets.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > Feature Engineering
  
**Planned functionality at M12 (in line with the milestones)**Initial implementation of image factorisation method without distributed computing (MS126).
  
**Planned functionalities at M18 (in line with the milestones)**Further testing and code optimisation of image factorisation method (handle missing data, make more probabilistic, make computationally stable, etc).
  
**Planned functionalities at M24(in line with the milestones)**A functioning distributed implementation of the image factorisation method, with features obtained by the approach available for data mining.
  
**Short description of potential use case**Compute a PCA-like analysis over thousands of 3D patient scans distributed across several hospitals, in situations where some parts of the data may be missing. Principal components serve as features for data mining.
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now

### Subject: 14

**Survey Timestamp**2016-09-13 16:15:34
  
**Task leader name**Vasilis Vassalos
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Access Right Module (ARM)
  
**Short Description of this component**This is a module that enables the local database and the mediator engine to execute complex database-like queries over the hospital data while respecting complex access control schemes and schema constraints and mappings. This module will allow queries coming from the Web Portal to the LDSMs to be processed while respecting rules about who has access, and showing where and how information maps across hospitals, in a way that also takes advantage of the additional information in the multiple schemata, such as keys and foreign keys and parent-child relationships. Hence this component will take as input the access rights of the users that perform various tasks on the platform, in an appropriate format, and various known schema constraints of MIP data. This component will affect the way the Local Database and the Federation Engine work, meaning that it will only allow users to query them according to their access rights
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Local Database
  
**Planned functionality at M12 (in line with the milestones)**Report on access restrictions in the MIP
  
**Planned functionalities at M18 (in line with the milestones)**Basic access control functionality
  
**Planned functionalities at M24(in line with the milestones)**Final release of ARM and integration to the platform
  
**Short description of potential use case**Consider that a MIP user would like to run an EE query and identify the number of patients diagnosed with dementia. According to the access rules imposed by the MIP administrators only Advanced users (AU) are able to run queries that invoke aggregate data, while Simple users (SU) are only able to use the already pre-computed and stored aggregate results. Consider also an AU, who is a pathologist, and thus should not get access to brain imaging data to compute aggregates. All these restrictions can be implemented in the MIP using specific access control languages that also affect the way queries posed by users are reformulated.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Access Right Module (ARM)
  
**Short Description of this component**This is a module that enables the local database and the mediator engine to execute complex database-like queries over the hospital data while respecting complex access control schemes and schema constraints and mappings. This module will allow queries coming from the Web Portal to the LDSMs to be processed while respecting rules about who has access, and showing where and how information maps across hospitals, in a way that also takes advantage of the additional information in the multiple schemata, such as keys and foreign keys and parent-child relationships. Hence this component will take as input the access rights of the users that perform various tasks on the platform, in an appropriate format, and various known schema constraints of MIP data. This component will affect the way the Local Database and the Federation Engine work, meaning that it will only allow users to query them according to their access rights
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Web Exploration and Analytics > Data Access
  
**Planned functionality at M12 (in line with the milestones)**Report on access restrictions in the MIP
  
**Planned functionalities at M18 (in line with the milestones)**Basic access control functionality
  
**Planned functionalities at M24(in line with the milestones)**Final release of ARM and integration to the platform
  
**Short description of potential use case**Consider that a MIP user would like to run an EE query and identify the number of patients diagnosed with dementia. According to the access rules imposed by the MIP administrators only Advanced users (AU) are able to run queries that invoke aggregate data, while Simple users (SU) are only able to use the already pre-computed and stored aggregate results. Consider also an AU, who is a pathologist, and thus should not get access to brain imaging data to compute aggregates. All these restrictions can be implemented in the MIP using specific access control languages that also affect the way queries posed by users are reformulated.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Access Right Module (ARM)
  
**Short Description of this component**This is a module that enables the local database and the mediator engine to execute complex database-like queries over the hospital data while respecting complex access control schemes and schema constraints and mappings. This module will allow queries coming from the Web Portal to the LDSMs to be processed while respecting rules about who has access, and showing where and how information maps across hospitals, in a way that also takes advantage of the additional information in the multiple schemata, such as keys and foreign keys and parent-child relationships. Hence this component will take as input the access rights of the users that perform various tasks on the platform, in an appropriate format, and various known schema constraints of MIP data. This component will affect the way the Local Database and the Federation Engine work, meaning that it will only allow users to query them according to their access rights
  
**To which building block your component belongs to ? (select one)**SERVICES > Security & Monitoring > User Management
  
**Planned functionality at M12 (in line with the milestones)**Report on access restrictions in the MIP
  
**Planned functionalities at M18 (in line with the milestones)**Basic access control functionality
  
**Planned functionalities at M24(in line with the milestones)**Final release of ARM and integration to the platform
  
**Short description of potential use case**Consider that a MIP user would like to run an EE query and identify the number of patients diagnosed with dementia. According to the access rules imposed by the MIP administrators only Advanced users (AU) are able to run queries that invoke aggregate data, while Simple users (SU) are only able to use the already pre-computed and stored aggregate results. Consider also an AU, who is a pathologist, and thus should not get access to brain imaging data to compute aggregates. All these restrictions can be implemented in the MIP using specific access control languages that also affect the way queries posed by users are reformulated.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Ontology-based data access Module (OBDA Module)
  
**Short Description of this component**Ontologies play a key role in semantically defining a domain of interest. Their use in the medical domain has been extensive since they provide a standard terminology with well-defined semantics and relations among its components that allows interoperability. Bridging ontologies and data is of paramount importance for MIP. Given a query, provide answers that reflect both the data and the knowledge captured by the ontology. This component will produce such a system that will reformulate posed queries to capture the knowledge of HBP and other ontologies while also providing access to data stored on the LDSM.
  
**To which building block your component belongs to ? (select one)**DATA > MDR (Meta Data Register) > Ontology&Standards
  
**Planned functionality at M12 (in line with the milestones)**Rewriting of query with respect to ontology module and preliminary integration to the platform
  
**Planned functionalities at M18 (in line with the milestones)**Query answering without rewriting with respect to database sources schema
  
**Planned functionalities at M24(in line with the milestones)**Final version - Optimized query answering - and full integration to the platform
  
**Short description of potential use case**Consider the SNOMED ontology and the concepts Dementia, Alzheimer Dementia and Parkinson Dementia. The SNOMED taxonomy defines that the concept of Alzheimer Dementia and of Parkinson Dementia are subsumed by the concept Dementia meaning that all Alzheimer/Parkinson Dementias are also Dementias. Moreover consider that the concept Alzheimer Dementia is mapped to the variable (of the MIP schema) Diagnosis with value specific to 'AD' and Parkinson Dementia is mapped to the variable Diagnosis with value specific to 'PR'. So, when a user poses a query using the SNOMED ontology that retrieves all patients with dementia, the OBDA system would rewrite this query to also return patients with Alzheimer disease and patients with Parkinson disease and finally would translate those queries to the MIP schema so that every patient with 'AD' and 'PR' value would return.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Ontology-based data access Module (OBDA Module)
  
**Short Description of this component**Ontologies play a key role in semantically defining a domain of interest. Their use in the medical domain has been extensive since they provide a standard terminology with well-defined semantics and relations among its components that allows interoperability. Bridging ontologies and data is of paramount importance for MIP. Given a query, provide answers that reflect both the data and the knowledge captured by the ontology. This component will produce such a system that will reformulate posed queries to capture the knowledge of HBP and other ontologies while also providing access to data stored on the LDSM.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Web Exploration and Analytics > Data Access
  
**Planned functionality at M12 (in line with the milestones)**Rewriting of query with respect to ontology module and preliminary integration to the platform
  
**Planned functionalities at M18 (in line with the milestones)**Query answering without rewriting with respect to database sources schema
  
**Planned functionalities at M24(in line with the milestones)**Final version - Optimized query answering - and full integration to the platform
  
**Short description of potential use case**Consider the SNOMED ontology and the concepts Dementia, Alzheimer Dementia and Parkinson Dementia. The SNOMED taxonomy defines that the concept of Alzheimer Dementia and of Parkinson Dementia are subsumed by the concept Dementia meaning that all Alzheimer/Parkinson Dementias are also Dementias. Moreover consider that the concept Alzheimer Dementia is mapped to the variable (of the MIP schema) Diagnosis with value specific to 'AD' and Parkinson Dementia is mapped to the variable Diagnosis with value specific to 'PR'. So, when a user poses a query using the SNOMED ontology that retrieves all patients with dementia, the OBDA system would rewrite this query to also return patients with Alzheimer disease and patients with Parkinson disease and finally would translate those queries to the MIP schema so that every patient with 'AD' and 'PR' value would return.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Ontology-based data access Module (OBDA Module)
  
**Short Description of this component**Ontologies play a key role in semantically defining a domain of interest. Their use in the medical domain has been extensive since they provide a standard terminology with well-defined semantics and relations among its components that allows interoperability. Bridging ontologies and data is of paramount importance for MIP. Given a query, provide answers that reflect both the data and the knowledge captured by the ontology. This component will produce such a system that will reformulate posed queries to capture the knowledge of HBP and other ontologies while also providing access to data stored on the LDSM.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > Data Integration
  
**Planned functionality at M12 (in line with the milestones)**Rewriting of query with respect to ontology module and preliminary integration to the platform
  
**Planned functionalities at M18 (in line with the milestones)**Query answering without rewriting with respect to database sources schema
  
**Planned functionalities at M24(in line with the milestones)**Final version - Optimized query answering - and full integration to the platform
  
**Short description of potential use case**Consider the SNOMED ontology and the concepts Dementia, Alzheimer Dementia and Parkinson Dementia. The SNOMED taxonomy defines that the concept of Alzheimer Dementia and of Parkinson Dementia are subsumed by the concept Dementia meaning that all Alzheimer/Parkinson Dementias are also Dementias. Moreover consider that the concept Alzheimer Dementia is mapped to the variable (of the MIP schema) Diagnosis with value specific to 'AD' and Parkinson Dementia is mapped to the variable Diagnosis with value specific to 'PR'. So, when a user poses a query using the SNOMED ontology that retrieves all patients with dementia, the OBDA system would rewrite this query to also return patients with Alzheimer disease and patients with Parkinson disease and finally would translate those queries to the MIP schema so that every patient with 'AD' and 'PR' value would return.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Ontology-based data access Module (OBDA Module)
  
**Short Description of this component**Ontologies play a key role in semantically defining a domain of interest. Their use in the medical domain has been extensive since they provide a standard terminology with well-defined semantics and relations among its components that allows interoperability. Bridging ontologies and data is of paramount importance for MIP. Given a query, provide answers that reflect both the data and the knowledge captured by the ontology. This component will produce such a system that will reformulate posed queries to capture the knowledge of HBP and other ontologies while also providing access to data stored on the LDSM.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Schema Mapping
  
**Planned functionality at M12 (in line with the milestones)**Rewriting of query with respect to ontology module and preliminary integration to the platform
  
**Planned functionalities at M18 (in line with the milestones)**Query answering without rewriting with respect to database sources schema
  
**Planned functionalities at M24(in line with the milestones)**Final version - Optimized query answering - and full integration to the platform
  
**Short description of potential use case**Consider the SNOMED ontology and the concepts Dementia, Alzheimer Dementia and Parkinson Dementia. The SNOMED taxonomy defines that the concept of Alzheimer Dementia and of Parkinson Dementia are subsumed by the concept Dementia meaning that all Alzheimer/Parkinson Dementias are also Dementias. Moreover consider that the concept Alzheimer Dementia is mapped to the variable (of the MIP schema) Diagnosis with value specific to 'AD' and Parkinson Dementia is mapped to the variable Diagnosis with value specific to 'PR'. So, when a user poses a query using the SNOMED ontology that retrieves all patients with dementia, the OBDA system would rewrite this query to also return patients with Alzheimer disease and patients with Parkinson disease and finally would translate those queries to the MIP schema so that every patient with 'AD' and 'PR' value would return.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Online Data Integration Module (ODIM)
  
**Short Description of this component**This component will enable the move from manual to automatic cleaning, and the transformation and merging of actions whenever hospitals add new data to the Medical Informatics Platform. More precisely this component will extend MIPMap, developed during the RUP of HBP, to support incremental Data Exchange. This means that instead of re-integrating data to the hospital's LDSM, whenever new data are exported from participating hospitals (following the standard pipeline of anonymization etc), they will be integrated into the already existing data taking into account the information that has been integrated before. Hence this component will vastly affect the way Data is integrated to the platform (Data Integration & Schema Mapping/Data Exchange) and the way metadata will be enriched. The functionality provided is incremental integration of data from hospitals.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > Data Integration
  
**Planned functionality at M12 (in line with the milestones)**Report on Incremental Data Integration
  
**Planned functionalities at M18 (in line with the milestones)**Preliminary version of module
  
**Planned functionalities at M24(in line with the milestones)**Final version and integration to the Hospital Bundle
  
**Short description of potential use case**A hospital participates in the MIP by providing a first batch of data. Consider that there exists a rule stating that the target diagnostic table is populated only with patients (source patient table) who have a diagnosis (source diagnostic table). Moreover consider that in this first data batch patient 'A' did not have any examinations, hence the target diagnostics table does not contain any tuple for 'A'. Consider now that in a second batch, a month later, diagnostics for patient 'A' are entered but patient information for 'A' is not resent (as it had been sent in the first batch a month earlier). In order to populate the target diagnostic table with information for 'A' it would be essential to run again the data exchange process for all the input data (first and second batch). However, this poses a significant computational burden. Hence, it would be ideal if an incremental data exchange approach would be used that would not need to re-run the data translation process from scratch but would be able to generate the information for patient 'A' and her diagnostics as they became available.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Online Data Integration Module (ODIM)
  
**Short Description of this component**This component will enable the move from manual to automatic cleaning, and the transformation and merging of actions whenever hospitals add new data to the Medical Informatics Platform. More precisely this component will extend MIPMap, developed during the RUP of HBP, to support incremental Data Exchange. This means that instead of re-integrating data to the hospital's LDSM, whenever new data are exported from participating hospitals (following the standard pipeline of anonymization etc), they will be integrated into the already existing data taking into account the information that has been integrated before. Hence this component will vastly affect the way Data is integrated to the platform (Data Integration & Schema Mapping/Data Exchange) and the way metadata will be enriched. The functionality provided is incremental integration of data from hospitals.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Schema Mapping
  
**Planned functionality at M12 (in line with the milestones)**Report on Incremental Data Integration
  
**Planned functionalities at M18 (in line with the milestones)**Preliminary version of module
  
**Planned functionalities at M24(in line with the milestones)**Final version and integration to the Hospital Bundle
  
**Short description of potential use case**A hospital participates in the MIP by providing a first batch of data. Consider that there exists a rule stating that the target diagnostic table is populated only with patients (source patient table) who have a diagnosis (source diagnostic table). Moreover consider that in this first data batch patient 'A' did not have any examinations, hence the target diagnostics table does not contain any tuple for 'A'. Consider now that in a second batch, a month later, diagnostics for patient 'A' are entered but patient information for 'A' is not resent (as it had been sent in the first batch a month earlier). In order to populate the target diagnostic table with information for 'A' it would be essential to run again the data exchange process for all the input data (first and second batch). However, this poses a significant computational burden. Hence, it would be ideal if an incremental data exchange approach would be used that would not need to re-run the data translation process from scratch but would be able to generate the information for patient 'A' and her diagnostics as they became available.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Online Data Integration Module (ODIM)
  
**Short Description of this component**This component will enable the move from manual to automatic cleaning, and the transformation and merging of actions whenever hospitals add new data to the Medical Informatics Platform. More precisely this component will extend MIPMap, developed during the RUP of HBP, to support incremental Data Exchange. This means that instead of re-integrating data to the hospital's LDSM, whenever new data are exported from participating hospitals (following the standard pipeline of anonymization etc), they will be integrated into the already existing data taking into account the information that has been integrated before. Hence this component will vastly affect the way Data is integrated to the platform (Data Integration & Schema Mapping/Data Exchange) and the way metadata will be enriched. The functionality provided is incremental integration of data from hospitals.
  
**To which building block your component belongs to ? (select one)**DATA > MDR (Meta Data Register) > Common Variables & Metadata
  
**Planned functionality at M12 (in line with the milestones)**Report on Incremental Data Integration
  
**Planned functionalities at M18 (in line with the milestones)**Preliminary version of module
  
**Planned functionalities at M24(in line with the milestones)**Final version and integration to the Hospital Bundle
  
**Short description of potential use case**A hospital participates in the MIP by providing a first batch of data. Consider that there exists a rule stating that the target diagnostic table is populated only with patients (source patient table) who have a diagnosis (source diagnostic table). Moreover consider that in this first data batch patient 'A' did not have any examinations, hence the target diagnostics table does not contain any tuple for 'A'. Consider now that in a second batch, a month later, diagnostics for patient 'A' are entered but patient information for 'A' is not resent (as it had been sent in the first batch a month earlier). In order to populate the target diagnostic table with information for 'A' it would be essential to run again the data exchange process for all the input data (first and second batch). However, this poses a significant computational burden. Hence, it would be ideal if an incremental data exchange approach would be used that would not need to re-run the data translation process from scratch but would be able to generate the information for patient 'A' and her diagnostics as they became available.

### Subject: 15

**Survey Timestamp**2016-09-07 11:50:46
  
**Task leader name**Thomas Heinis
  
**Task number attached to this component**T8.4.5 Large-scale data analytics on massively parallel architecture ICL
  
**Name of this component**Large Scale Analytics Algorithms
  
**Short Description of this component**Analytics/clustering algorithms for the efficient and scalable large scale analysis of medical data.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Machine Learning Library
  
**Planned functionality at M12 (in line with the milestones)**Initial prototypes of a set of algorithms run distributed on HPC via MPI.
  
**Planned functionalities at M18 (in line with the milestones)**Additional distributed, MPI base algorithms as well as more mature implementations.
  
**Planned functionalities at M24(in line with the milestones)**Deployment of production grade algorithms on supercomputing infrastructure, making it available fro general use.
  
**Short description of potential use case**Large scale clustering/analysis of large amounts of (cleaned) medical data.
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now

### Subject: 16

**Survey Timestamp**[not completed]

### Subject: 17

**Survey Timestamp**2016-09-09 08:03:24
  
**Task leader name**Anastasia Ailamaki
  
**Task number attached to this component**T8.1.1 Infrastructure to support just-in-time analytics on raw medical data EPFL
  
**Name of this component**Extended multidimensional query support
  
**Short Description of this component**Basic primitives for computation over multidimensional queries in the local hospital database. This component needs data but not necessarily data from ALL hospitals.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Local Database
  
**Planned functionality at M12 (in line with the milestones)**The infrastructure to run this plug-in in the query engine will be ready.
  
**Planned functionalities at M18 (in line with the milestones)**The plug-in is completed and deployed as part of the Hospital Bundle.
  
**Planned functionalities at M24(in line with the milestones)**The plug-in is completed and deployed as part of the Hospital Bundle and validated by HBP users.
  
**Short description of potential use case**Allow ad-hoc dimensionality reduction and feature extraction, e.g. on imaging and genetic data. It can be later added to the data factory as the regular feature extraction pipelines.
  
**Task number attached to this component**T8.1.2 Installation of RAW on local computing infrastructure for hospital analytics requirements EPFL
  
**Name of this component**Distributed local query engine over HPC
  
**Short Description of this component**Extend the local query engine to enable the use of distributed computing frameworks (like Spark).
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Local Database
  
**Planned functionality at M12 (in line with the milestones)**The query engine extensions are implemented.
  
**Planned functionalities at M18 (in line with the milestones)**The federation software is integrated with the new and improved query engine API's .
  
**Planned functionalities at M24(in line with the milestones)**Query engine optimized using real-world datasets and queries.
  
**Short description of potential use case**Allow for scalability of the query engine for hospitals with big amounts of data.
  
**Task number attached to this component**T8.1.2 Installation of RAW on local computing infrastructure for hospital analytics requirements EPFL
  
**Name of this component**Hospital hubs
  
**Short Description of this component**A local data hub is a computer cluster, that is deployed within the same country as the hospital and connected together via a secure network. The objective is to allow small hospitals to share IT infrastructure.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Factory (AF) > Workflow Engine: Woken
  
**Planned functionality at M12 (in line with the milestones)**A secure procedure has been defined to connect a hospital to a hub. Related data management processes are also defined.
  
**Planned functionalities at M18 (in line with the milestones)**Software extension to support provenance tracking through hospital hubs have been implemented.
  
**Planned functionalities at M24(in line with the milestones)**The software required by the procedure is developed, tested and packaged, ready to be used if necessary.
  
**Short description of potential use case**Small hospitals with restricted resources can still contribute to the human brain project.
  
**Task number attached to this component**T8.1.3 Installation and adaptation of SP8 bundle at local hospitals with upgrade support EPFL
  
**Name of this component**SERVICES > Upgrade-Deploy-Release > MIP Integrated Releases Hospital Databases Bundle
  
**Short Description of this component**This component will prepare and install the first working version of the FCDI (Hospital Bundle) at the participating hospitals in a bottom-up fashion. This effort includes going to hospitals, installing the software, understanding the network topology,configuring the data sources and getting the right permissions. It will also integrate new features as they become available in the other tasks of WP8.1 into the FCDI.
  
**To which building block your component belongs to ? (select one)**SERVICES > Upgrade-Deploy-Release > MIP Integrated Releases Hospital Databases Bundle
  
**Planned functionality at M12 (in line with the milestones)**Continuous process
  
**Planned functionalities at M18 (in line with the milestones)**Continuous process
  
**Planned functionalities at M24(in line with the milestones)**Continuous process
  
**Short description of potential use case**Base deployment infrastructure of the hospital bundle.
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now.

### Subject: 18

**Survey Timestamp**2016-09-08 18:08:49
  
**Task leader name**Andrew Pocklington
  
**Task number attached to this component**T8.4.1 Brain scale high performance deep phenotyping CF
  
**Name of this component**Functional gene-sets relevant to human brain disorders
  
**Short Description of this component**Functional processes whose disruption contributes to the development of complex brain disorders, such as schizophrenia, will be identified through gene-set enrichment analyses of human genetic data. Genetic data will typically come from case-control studies of common and rare variation. Analytic tools will include MAGMA for common variation (SNPs) and standard logistic regression models for CNVs.
  
**Planned functionality at M12 (in line with the milestones)**Collated list of functional annotations (gene-sets) to be tested for disease association Initial list of gene-sets enriched for common (GWAS) and/or rare (CNV) variant association to schizophrenia in largest case-control datasets available to us.
  
**Planned functionalities at M18 (in line with the milestones)**Updated list of functional annotations. Additional annotations may be derived from the literature or from other HBP components. Updated list of enriched gene-sets based on emerging genetic data. Depending on access to data, updates will involve the analysis of larger datasets (increasing power to find true associations), other classes of variant (e.g. rare variants from exome/genome sequencing studies, de novo variants from trio studies) and/or a wider range of disorders.
  
**Planned functionalities at M24(in line with the milestones)**Updated list of functional annotations. Additional annotations may be derived from the literature or from other HBP components. Updated list of enriched gene-sets based on emerging genetic data. Depending on access to data, updates will involve the analysis of larger datasets (increasing power to find true associations), other classes of variant (e.g. rare variants from exome/genome sequencing studies, de novo variants from trio studies) and/or a wider range of disorders.
  
**Short description of potential use case**Use Case: Clinical Researcher Background: Gene set enrichment analyses robustly identify specific sets of synaptic genes as being enriched for schizophrenia risk variants. A clinical researcher wishes to investigate whether cognitive function is influenced by common variation in synaptic genes linked to schizophrenia. They have access to a cohort of individuals for which both cognitive measures and GWAS data are available. 1. The researcher uses the component to identify a synaptic gene-set associated with schizophrenia. 2. From schizophrenia case-control GWAS summary data they then identify common risk variants lying within genes belonging to the gene-set, recording the associated allele and effect size (log odds ratio) for each. 3. They then calculate the burden of genetic risk variants carried by each individual in their cohort, summing the number of risk alleles that they carry (weighting the count for each allele by its corresponding effect size). 4. The researcher then uses a regression model to test whether burden of genetic risk is correlated with measures of cognitive performance, using covariates to account for potential confounds (e.g. age, gender).
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now

### Subject: 19

**Survey Timestamp**2016-09-10 22:00:04
  
**Task leader name**Saso Deroski
  
**Task number attached to this component**T8.3.5 Methods for distributed rule-based disease signature discovery JSI
  
**Name of this component**Disease signature: Distributed rule-based methods
  
**Short Description of this component**Distributed versions of the tree- and rule-based methods for predictive clustering for solving different tasks of predicting structured outputs (e.g. multi-target regression).
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Machine Learning Library
  
**Planned functionality at M12 (in line with the milestones)**First working prototype developed.
  
**Planned functionalities at M18 (in line with the milestones)**First tests and validations on research data.
  
**Planned functionalities at M24(in line with the milestones)**Analysis of distributed data.
  
**Short description of potential use case**Discovery of disease signatures on distributed data.
  
**Task number attached to this component**T8.3.8 Methods for disease progression modeling JSI
  
**Name of this component**Longitudinal modeling: Tree-based and equation-based methods
  
**Short Description of this component**Machine learning methods for describing and modelling the temporal dynamics of disease and its clinical and biological markers.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Machine Learning Library
  
**Planned functionality at M12 (in line with the milestones)**First working prototype developed.
  
**Planned functionalities at M18 (in line with the milestones)**First tests and validations on research data.
  
**Planned functionalities at M24(in line with the milestones)**Analysis of longitudinal data.
  
**Short description of potential use case**Discovery of disease signatures in longitudinal data.
  
**Task number attached to this component**T8.3.9 Ontologies for describing data on neurological diseases, patients JSI
  
**Name of this component**Ontologies for describing data on neurological diseases, patients
  
**Short Description of this component**A mid-level ontology for describing various types of data on patients with neurological diseases.
  
**To which building block your component belongs to ? (select one)**DATA > MDR (Meta Data Register) > Ontology&Standards
  
**Planned functionality at M12 (in line with the milestones)**Prototype of the ontology for describing data on patients with neurological diseases developed
  
**Planned functionalities at M18 (in line with the milestones)**First dataset annotated according to the ontology.
  
**Planned functionalities at M24(in line with the milestones)**Ontology and several datasets annotated according to the ontology.
  
**Short description of potential use case**Annotation of different types of data in the MIP.
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now.

### Subject: 20

**Survey Timestamp**2016-09-09 08:01:24
  
**Task leader name**Boudewijn Lelieveldt
  
**Task number attached to this component**T8.3.10 Methods for linkage of local SNP data (individual SNPs) to imaging data through SNP LUMC
  
**Name of this component**GeneHeatMapper
  
**Short Description of this component**Algorithm that generates a 3D expression heatmap of of an SNP name, gene name or co-expression module.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Brain Anatomy
  
**Planned functionality at M12 (in line with the milestones)**M 8.3.2 Initial proof-of-concept algorithm for generating 3D gene expression heatmaps of single genes and SNPs
  
**Planned functionalities at M18 (in line with the milestones)**M 8.3.2: Initial proof-of-concept algorithm for generating 3D gene expression heatmaps from gene co-expression modules.
  
**Planned functionalities at M24(in line with the milestones)**M 8.3.11 Gene Expression Maps of Disease Link to Brain Atlases
  
**Short description of potential use case**Discovery of disease associations between gene expression data en imaging data
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now
  
**Task number attached to this component**T8.3.10 Methods for linkage of local SNP data (individual SNPs) to imaging data through SNP LUMC

### Subject: 21

**Data Access Group**sp8\_rec21
  
**Survey Timestamp**2016-09-13 21:24:46
  
**Task leader name**Mira Marcus Kalish
  
**Task number attached to this component**T8.3.1 Tools to mine replicable selection and integration of hierarchical features, inter and across domains using FDR. TAU
  
**Name of this component**3-C (Categorize, Cluster & Classify)
  
**Short Description of this component**Methodology for Medical big data analysis and disease sub-type identification
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Statistical Analytics
  
**Planned functionality at M12 (in line with the milestones)**MS 8.3.2: Initial Proof-of-concept and results of the different 3-C Algorithms
  
**Planned functionalities at M18 (in line with the milestones)**Methodology improvemnts - scaling up the ability to work with larger datasets.
  
**Planned functionalities at M24(in line with the milestones)**Validation of 3-C strategy on different data
  
**Task number attached to this component**T8.3.1 Tools to mine replicable selection and integration of hierarchical features, inter and across domains using FDR. TAU
  
**Name of this component**3-C Longitudinal Modeling
  
**Short Description of this component**Incorporating longitudinal information (day-to-day and multi-patients visits).
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Statistical Analytics
  
**Planned functionality at M12 (in line with the milestones)**Review of existing methods
  
**Planned functionalities at M18 (in line with the milestones)**Clinical measurements data construction
  
**Planned functionalities at M24(in line with the milestones)**MS8.3.11: Analyses of 3-C demonstrating the use of Longitudinal Data
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now.
  
**Task number attached to this component**T8.3.1 Tools to mine replicable selection and integration of hierarchical features, inter and across domains using FDR. TAU
  
**Name of this component**Integrating multi-domain data
  
**Short Description of this component**methodology for enriching current models
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Statistical Analytics
  
**Planned functionality at M12 (in line with the milestones)**Developing methodologies for enriched data MS8.3.3
  
**Planned functionalities at M18 (in line with the milestones)**Increasing the variety of data sources that can be integrated into the previous analyses, in particular genomics
  
**Planned functionalities at M24(in line with the milestones)**Analysis of multi-domain data
  
**Task number attached to this component**T8.3.2 Developing methods for high-dimensional data with possible informative missing values TAU
  
**Name of this component**Transformations in medical big data
  
**Short Description of this component**Symetry targeted monotone transformations, and the advantage gained in variance stability,linearity and clustering.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Statistical Analytics
  
**Planned functionality at M12 (in line with the milestones)**Improvements and use in different datasets
  
**Planned functionalities at M18 (in line with the milestones)**Functions improvements
  
**Planned functionalities at M24(in line with the milestones)**Prepare methods for integration into MIP, test and validate results. (preparation specification to be agreed with and provided by MIP implementation team)
  
**Task number attached to this component**T8.3.2 Developing methods for high-dimensional data with possible informative missing values TAU
  
**Name of this component**Statistical procedures and workflows for missing values
  
**Short Description of this component**Develop statistical procedures and workflows to help and guide the discovery of possible patterns in missing values
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Statistical Analytics
  
**Planned functionality at M12 (in line with the milestones)**Missing values - extraction
  
**Planned functionalities at M18 (in line with the milestones)**Missing values: structure discovery
  
**Planned functionalities at M24(in line with the milestones)**Missing values: imputation and visualization MS8.3.10
  
**Task number attached to this component**T8.3.3 Introducing selective inference into dimensionality reduction and clustering methods TAU
  
**Name of this component**Clustering: incorporating Knowledge into the process
  
**Short Description of this component**Evaluating disease signature clusters (by combining new approaches with tools developed for the visualization and manipulation of hierarchical clustering)
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Statistical Analytics
  
**Planned functionality at M12 (in line with the milestones)**Re-evaluation of 3-C methodology as a data and knowledge combined methodology
  
**Planned functionalities at M18 (in line with the milestones)**Exploration of new methods for incorporating Knowledge into the process
  
**Planned functionalities at M24(in line with the milestones)**Report on incorporating Knowledge into the process
  
**Task number attached to this component**T8.3.4 Statistical methods for 'Disease Signature' confidence assessment TAU
  
**Name of this component**Disease Signatures -concept and methodology
  
**Short Description of this component**Define and propose a model for disease signature
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Statistical Analytics
  
**Planned functionality at M12 (in line with the milestones)**Literature Review
  
**Planned functionalities at M18 (in line with the milestones)**Summary of consultations with multi-disciplinary experts and opinion leaders
  
**Planned functionalities at M24(in line with the milestones)**Suggest definition of the 'Disease signature' and proposed estimation methods.
  
**Task number attached to this component**T8.3.2 Developing methods for high-dimensional data with possible informative missing values TAU
  
**Name of this component**The importance of transformation in medical big data
  
**Short Description of this component**Paper. Note: Original component (Modeling Hospitals Workflow) was replaced with this, as Modeling Hospitals Workflow will be part of SERVICES > Data Governance.
  
**Planned functionality at M12 (in line with the milestones)**Paper submitted
  
**Planned functionalities at M18 (in line with the milestones)**-
  
**Planned functionalities at M24(in line with the milestones)**-
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now.
  
**Name of this component**-
  
**Short Description of this component**Component removed, as already part of SERVICES > Scientific Coordination, requirements & use cases
  
**Short Description of this component**Component removed, as already part of SERVICES > Clinical demonstrators

### Subject: 22

**Survey Timestamp**2016-09-10 22:17:06
  
**Task leader name**Nada Lavra
  
**Task number attached to this component**T8.3.6 Methods for redescription mining JSI
  
**Name of this component**Integrating multi-domain data: Methods for redescription mining
  
**Short Description of this component**Methods for redescription mining - a relatively novel data mining and knowledge discovery approach that aims to find multiple rule-based descriptions of subsets of examples (e.g. patients), where each of the descriptions is based on a different set of descriptive variables.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Machine Learning Library
  
**Planned functionality at M12 (in line with the milestones)**First working prototype developed.
  
**Planned functionalities at M18 (in line with the milestones)**First tests and validations on research data.
  
**Planned functionalities at M24(in line with the milestones)**Analysis of multi-view data.
  
**Short description of potential use case**Discovery of disease signatures.
  
**Task number attached to this component**T8.3.7 Methods for heterogeneous networks JSI
  
**Name of this component**Integrating multi-domain data: Methods for heterogeneous networks
  
**Short Description of this component**Mehods for mining text-enriched heterogeneous information networks.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Algorithm Library > Machine Learning Library
  
**Planned functionality at M12 (in line with the milestones)**First working prototype developed.
  
**Planned functionalities at M18 (in line with the milestones)**First tests and validations on research data.
  
**Planned functionalities at M24(in line with the milestones)**Analysis of text-enriched heterogeneous information networks.
  
**Short description of potential use case**Discovery of disease signatures on text-enriched data.
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now.

### Subject: 23

**Survey Timestamp**[not completed]

### Subject: 24

**Data Access Group**SP8\_rec24
  
**Survey Timestamp**2016-09-13 15:21:09
  
**Task leader name**Anastasia Ailamaki
  
**Task number attached to this component**T8.1.1 Infrastructure to support just-in-time analytics on raw medical data EPFL
  
**Name of this component**Nifti data source for local query engine
  
**Short Description of this component**Plug-in to enable local query engine to perfom queries directly on Nifti files.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Local Database
  
**Planned functionality at M12 (in line with the milestones)**Initial proof-of-concept query engine that can perform selection queries over Nifti files.
  
**Planned functionalities at M18 (in line with the milestones)**Initial proof-of-concept query engine that can perform more complex queries like joins and group by over Nifti files.
  
**Planned functionalities at M24(in line with the milestones)**Query engine that can perform complex queries and mathematical operations over Nifti files.
  
**Short description of potential use case**Data discovery directly on Nifti files Enable analysis over imaging data that are not provided directly by the existing pipeline.
  
**Task number attached to this component**T8.1.1 Infrastructure to support just-in-time analytics on raw medical data EPFL
  
**Name of this component**Genetic data data source for local query engine
  
**Short Description of this component**Plug-in to enable local query engine to perfom queries directly on genetic data files.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Local Database
  
**Planned functionality at M12 (in line with the milestones)**Initial proof-of-concept query engine that can perform selection queries over genetic data files.
  
**Planned functionalities at M18 (in line with the milestones)**Initial proof-of-concept query engine that can perform more complex queries like joins and group by over genetic data files.
  
**Planned functionalities at M24(in line with the milestones)**Query engine that can perform complex queries and mathematical operations over genetic data files.
  
**Short description of potential use case**Data discovery directly on genetic data files Enable analysis over genetic data that are not provided directly by the existing pipeline.
  
**Task number attached to this component**T8.1.1 Infrastructure to support just-in-time analytics on raw medical data EPFL
  
**Name of this component**Nifti library in local query engine
  
**Short Description of this component**Library of functions for common operations on imaging data/Nifti files in the local query engine.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Local Database
  
**Planned functionality at M12 (in line with the milestones)**dependent on delivery of: SOFTWARE > Hospital Databases Bundle (HDB) > Local Database > Nifti data source for local query engine Definition of functionality requirements. This component needs the Nifti data source to be developed first. Therefore the development effort will be focused there for the period of M01-M12.
  
**Planned functionalities at M18 (in line with the milestones)**Initial proof-of-concept with a few functions defined in collaboration with Data Analysis groups.
  
**Planned functionalities at M24(in line with the milestones)**Final Nifti library with a satisfactory number of functions.
  
**Short description of potential use case**Data discovery directly on Nifti files Enable analysis over imaging data that are not provided directly by the existing pipeline.
  
**Task number attached to this component**T8.1.1 Infrastructure to support just-in-time analytics on raw medical data EPFL
  
**Name of this component**Genetic data library in local query engine
  
**Short Description of this component**Library of functions for common operations on genetic data files in the local query engine.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Local Database
  
**Planned functionality at M12 (in line with the milestones)**dependent on delivery of: SOFTWARE > Hospital Databases Bundle (HDB) > Local Database > Genetic data data source for local query engine Definition of functionality requirements. This component needs the genetic data source to be developed first. Therefore the development effort will be focused there for the period of M01-M12.
  
**Planned functionalities at M18 (in line with the milestones)**Initial proof-of-concept with a few functions defined in collaboration with Data Analysis groups.
  
**Planned functionalities at M24(in line with the milestones)**Final genetic data library with a satisfactory number of functions.
  
**Short description of potential use case**Data discovery directly on genetic data files Enable analysis over genetic data that are not provided directly by the existing pipeline.
  
**Task number attached to this component**T8.1.2 Installation of RAW on local computing infrastructure for hospital analytics requirements EPFL
  
**Name of this component**Secure connection between two hospitals.
  
**Short Description of this component**Establish encrypted connections between two hospitals.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Local Database
  
**Planned functionality at M12 (in line with the milestones)**The technical procedure describing the secure connection and which technologies to use is documented.
  
**Planned functionalities at M18 (in line with the milestones)**Initial proof-of-concept using this component is developed.
  
**Planned functionalities at M24(in line with the milestones)**The component is integrated in the hospital bundle and/or data factory.
  
**Short description of potential use case**Secure connection to enable hospital hubs.
  
**Task number attached to this component**T8.1.3 Installation and adaptation of SP8 bundle at local hospitals with upgrade support EPFL
  
**Name of this component**Anonymisation tests
  
**Short Description of this component**Local hospital anonymization validation tests.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > Data Anonymisation
  
**Planned functionality at M12 (in line with the milestones)**Anonymization process validated in one hospital.
  
**Planned functionalities at M18 (in line with the milestones)**Anonymization process validated in three hospitals.
  
**Planned functionalities at M24(in line with the milestones)**Anonymization process validated in all five hospitals.
  
**Short description of potential use case**Show validation report to the hospitals in order to build a trusting collaboration.
  
**Task number attached to this component**T8.1.3 Installation and adaptation of SP8 bundle at local hospitals with upgrade support EPFL
  
**Name of this component**Encrypted overlay network
  
**Short Description of this component**A component that creates a secure network between hospitals for the administration of the servers.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > Workflow Engine
  
**Planned functionality at M12 (in line with the milestones)**Technology defined.
  
**Planned functionalities at M18 (in line with the milestones)**Initial proof-of-concept developed.
  
**Planned functionalities at M24(in line with the milestones)**The component is integrated with the bundle in the local hospital.
  
**Short description of potential use case**The system administrator can monitor securely the hospital bundle servers.
  
**Task number attached to this component**T8.1.3 Installation and adaptation of SP8 bundle at local hospitals with upgrade support EPFL
  
**Name of this component**Remote starting of services
  
**Short Description of this component**This component will enable the local hospital services to be remotely managed (ie start, stop)
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > Workflow Engine
  
**Planned functionality at M12 (in line with the milestones)**dependent on delivery of: SOFTWARE > Data Factory (DF) > Workflow Engine > Encrypted overlay network Technology defined.
  
**Planned functionalities at M18 (in line with the milestones)**Initial proof-of-concept developed.
  
**Planned functionalities at M24(in line with the milestones)**The component is integrated with the bundle in the local hospital.
  
**Short description of potential use case**The system administrator can manage securely the hospital bundle servers.
  
**Task number attached to this component**T8.1.2 Installation of RAW on local computing infrastructure for hospital analytics requirements EPFL
  
**Name of this component**Sharing processing among local hospital nodes
  
**Short Description of this component**This component will enable hospitals by finding a way to share processing between hospitals.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Local Database
  
**Planned functionality at M12 (in line with the milestones)**dependent on delivery of: SOFTWARE > Hospital Databases Bundle (HDB) > Local Database > Secure connection between two hospitals. Technology defined.
  
**Planned functionalities at M18 (in line with the milestones)**Initial proof-of-concept developed.
  
**Planned functionalities at M24(in line with the milestones)**The component is integrated with the bundle in the local hospital.
  
**Short description of potential use case**Hospital with less resources can off load some of their processes to another hospital/hub.
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now.

### Subject: 25

**Survey Timestamp**[not completed]
  
**Task leader name**Yannis Ioannidis
  
**Task number attached to this component**T8.1.5 Distributed complex workflow engine UoA
  
**Name of this component**Master component
  
**Short Description of this component**The master component transforms, schedules and dispatches the queries to workers
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Federated Query
  
**Planned functionality at M12 (in line with the milestones)**Master component released packaged and integrated with a worker component
  
**Planned functionalities at M18 (in line with the milestones)**Master component is integrated with 5 worker components deployed in 5 hospitals
  
**Planned functionalities at M24(in line with the milestones)**Master component is validated by HBP users
  
**Task number attached to this component**T8.1.5 Distributed complex workflow engine UoA
  
**Name of this component**Worker/ Bridge Component
  
**Short Description of this component**The workers reside on the hospital nodes and act as a bridge with the RAW query engine which executes the queries in situ.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Federated Query
  
**Planned functionality at M12 (in line with the milestones)**Worker/ Bridge Component released, packaged and integrated with Master component and RAW query engine
  
**Planned functionalities at M18 (in line with the milestones)**Worker/ Bridge Component is deployed in 5 hospitals
  
**Planned functionalities at M24(in line with the milestones)**Worker/ Bridge Component is validated by HBP users
  
**Task number attached to this component**T8.1.5 Distributed complex workflow engine UoA
  
**Name of this component**Web portal connector component
  
**Short Description of this component**It interfaces the master component with the web portal.
  
**To which building block your component belongs to ? (select one)**SERVICES > Upgrade-Deploy-Release > MIP Integrated Releases Web Exploration & Analytics
  
**Planned functionality at M12 (in line with the milestones)**Define communication protocol with web portal
  
**Planned functionalities at M18 (in line with the milestones)**First version of web portal connector component deployed
  
**Planned functionalities at M24(in line with the milestones)**Web portal connector component is validated by HBP users
  
**Short description of potential use case**The user is able to see a catalogue of algorithms, select one of them, to execute the selected algorithm by specifying particular values to its parameters and see the results of the exexution.
  
**Task number attached to this component**T8.1.6 User Defined Functions (UDFs) and query templates UoA
  
**Name of this component**UDFs component
  
**Short Description of this component**UDFs component focuses on the development of complex user-defined functions (UDFs) that are needed in SQL-based data mining workflows, adapting and supporting algorithms provided by SP8 data mining groups. UDFs that interface with external libraries/systems such as NumKit, SciKit, R will also be implemented.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Federated Query
  
**Planned functionality at M12 (in line with the milestones)**Implementation of UDFs component
  
**Planned functionalities at M18 (in line with the milestones)**Deployment of UDFs component
  
**Planned functionalities at M24(in line with the milestones)**Validation of UDFs component
  
**Task number attached to this component**T8.1.6 User Defined Functions (UDFs) and query templates UoA
  
**Name of this component**template composer component
  
**Short Description of this component**The template composer converts the template, which describes parameterized distributed workflows, into an ExaDFL query script. The template composer is responsible for the isolated execution of each algorithm template.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Federated Query
  
**Planned functionality at M12 (in line with the milestones)**Implementation of template composer component
  
**Planned functionalities at M18 (in line with the milestones)**Deployment of template composer component
  
**Planned functionalities at M24(in line with the milestones)**Validation of template composer component
  
**Short description of potential use case**The user specifies particular values to the parameters of a parameterized algorithm implemented in the query engine and then it executes the algorithm.
  
**Task number attached to this component**T8.1.7 Query templates and workflow management UoA
  
**Name of this component**Query template repository
  
**Short Description of this component**The query template repository component will provide storage, reviewing, access control (authentication and authorisation) and audit trail/logging capabilities. The repository will be hosted in a version control system (VCS).
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Federated Query
  
**Planned functionality at M12 (in line with the milestones)**Implementation of query template repository
  
**Planned functionalities at M18 (in line with the milestones)**Deployment of query template repository
  
**Planned functionalities at M24(in line with the milestones)**Validation of query template repository
  
**Short description of potential use case**Researchers who are willing to contribute a new algorithm will be able to submit to the code review system a request to add a new algorithm/template or modify an existing one. Senior researchers will then review the proposed changes and accept or deny the modification request.
  
**Task number attached to this component**T8.1.7 Query templates and workflow management UoA
  
**Name of this component**Management component of query template repository
  
**Short Description of this component**The management component of query template repository will manage user access to the query template repository. Each user will be tagged by a role in order to be able to have the corresponding access rights. The users will also have the ability to review, register, unregister, update any algorithm and monitor statistics regarding their algorithm overall execution.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Federated Query
  
**Planned functionality at M12 (in line with the milestones)**Implementation of management component of query template repository
  
**Planned functionalities at M18 (in line with the milestones)**Deployment of management component of query template repository
  
**Planned functionalities at M24(in line with the milestones)**Validation of management component of query template repository
  
**Short description of potential use case**The system administrators will be able to monitor the system usage and load and take the appropriate actions. Credentials with limited access will be used for the web portal.

### Subject: 26

**Survey Timestamp**2016-09-15 00:11:42
  
**Task leader name**Vasilis Vassalos
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Access Rights Module (ARM)
  
**Short Description of this component**This is a module that enables the local database and the mediator engine to execute complex database-like queries over the hospital data while respecting complex access control schemes and schema constraints and mappings. This module will allow queries coming from the Web Portal to the LDSMs to be processed while respecting rules about who has access, and showing where and how information maps across hospitals, in a way that also takes advantage of the additional information in the multiple schemata, such as keys and foreign keys and parent-child relationships. Hence this component will take as input the access rights of the users that perform various tasks on the platform, in an appropriate format, and various known schema constraints of MIP data. This component will affect the way the Local Database and the Federation Engine work, meaning that it will only allow users to query them according to their access rights.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Local Database
  
**Planned functionality at M12 (in line with the milestones)**Report on access restrictions in the MIP
  
**Planned functionalities at M18 (in line with the milestones)**Basic access control functionality
  
**Planned functionalities at M24(in line with the milestones)**Final release of ARM and integration to the platform
  
**Short description of potential use case**Consider that a MIP user would like to run an EE query and identify the number of patients diagnosed with dementia. According to the access rules imposed by the MIP administrators only Advanced users (AU) are able to run queries that invoke aggregate data, while Simple users (SU) are only able to use the already pre-computed and stored aggregate results. Consider also an AU, who is a pathologist, and thus should not get access to brain imaging data to compute aggregates. All these restrictions can be implemented in the MIP using specific access control languages that also affect the way queries posed by users are reformulated.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Access Rights Module (ARM)
  
**Short Description of this component**This is a module that enables the local database and the mediator engine to execute complex database-like queries over the hospital data while respecting complex access control schemes and schema constraints and mappings. This module will allow queries coming from the Web Portal to the LDSMs to be processed while respecting rules about who has access, and showing where and how information maps across hospitals, in a way that also takes advantage of the additional information in the multiple schemata, such as keys and foreign keys and parent-child relationships. Hence this component will take as input the access rights of the users that perform various tasks on the platform, in an appropriate format, and various known schema constraints of MIP data. This component will affect the way the Local Database and the Federation Engine work, meaning that it will only allow users to query them according to their access rights.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Web Exploration and Analytics > Data Access
  
**Planned functionality at M12 (in line with the milestones)**Report on access restrictions in the MIP
  
**Planned functionalities at M18 (in line with the milestones)**Basic access control functionality
  
**Planned functionalities at M24(in line with the milestones)**Final release of ARM and integration to the platform
  
**Short description of potential use case**Consider that a MIP user would like to run an EE query and identify the number of patients diagnosed with dementia. According to the access rules imposed by the MIP administrators only Advanced users (AU) are able to run queries that invoke aggregate data, while Simple users (SU) are only able to use the already pre-computed and stored aggregate results. Consider also an AU, who is a pathologist, and thus should not get access to brain imaging data to compute aggregates. All these restrictions can be implemented in the MIP using specific access control languages that also affect the way queries posed by users are reformulated.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Ontology-based data access Module (OBDA Module)
  
**Short Description of this component**Ontologies play a key role in semantically defining a domain of interest. Their use in the medical domain has been extensive since they provide a standard terminology with well-defined semantics and relations among its components that allows interoperability. Bridging ontologies and data is of paramount importance for MIP. Given a query, provide answers that reflect both the data and the knowledge captured by the ontology. This component will produce such a system that will reformulate posed queries to capture the knowledge of HBP and other ontologies while also providing access to data stored on the LDSMs.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Schema Mapping
  
**Planned functionality at M12 (in line with the milestones)**Rewriting of query with respect to ontology module and preliminary integration to the platform
  
**Planned functionalities at M18 (in line with the milestones)**Query answering without rewriting with respect to database sources schema
  
**Planned functionalities at M24(in line with the milestones)**Final version - Optimized query answering - and full integration to the platform
  
**Short description of potential use case**Consider the SNOMED ontology and the concepts Dementia, Alzheimer Dementia and Parkinson Dementia. The SNOMED taxonomy defines that the concept of Alzheimer Dementia and of Parkinson Dementia are subsumed by the concept Dementia meaning that all Alzheimer/Parkinson Dementias are also Dementias. Moreover consider that the concept Alzheimer Dementia is mapped to the variable (of the MIP schema) Diagnosis with value specific to 'AD' and Parkinson Dementia is mapped to the variable Diagnosis with value specific to 'PR'. So, when a user poses a query using the SNOMED ontology that retrieves all patients with dementia, the OBDA system would rewrite this query to also return patients with Alzheimer disease and patients with Parkinson disease and finally would translate those queries to the MIP schema so that every patient with 'AD' and 'PR' value would return.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Ontology-based data access Module (OBDA Module)
  
**Short Description of this component**Ontologies play a key role in semantically defining a domain of interest. Their use in the medical domain has been extensive since they provide a standard terminology with well-defined semantics and relations among its components that allows interoperability. Bridging ontologies and data is of paramount importance for MIP. Given a query, provide answers that reflect both the data and the knowledge captured by the ontology. This component will produce such a system that will reformulate posed queries to capture the knowledge of HBP and other ontologies while also providing access to data stored on the LDSMs.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > Data Integration
  
**Planned functionality at M12 (in line with the milestones)**Rewriting of query with respect to ontology module and preliminary integration to the platform
  
**Planned functionalities at M18 (in line with the milestones)**Query answering without rewriting with respect to database sources schema
  
**Planned functionalities at M24(in line with the milestones)**Final version - Optimized query answering - and full integration to the platform
  
**Short description of potential use case**Consider the SNOMED ontology and the concepts Dementia, Alzheimer Dementia and Parkinson Dementia. The SNOMED taxonomy defines that the concept of Alzheimer Dementia and of Parkinson Dementia are subsumed by the concept Dementia meaning that all Alzheimer/Parkinson Dementias are also Dementias. Moreover consider that the concept Alzheimer Dementia is mapped to the variable (of the MIP schema) Diagnosis with value specific to 'AD' and Parkinson Dementia is mapped to the variable Diagnosis with value specific to 'PR'. So, when a user poses a query using the SNOMED ontology that retrieves all patients with dementia, the OBDA system would rewrite this query to also return patients with Alzheimer disease and patients with Parkinson disease and finally would translate those queries to the MIP schema so that every patient with 'AD' and 'PR' value would return.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Online Data Integration Module (ODIM)
  
**Short Description of this component**This component will enable the move from manual to automatic cleaning, and the transformation and merging of actions whenever hospitals add new data to the Medical Informatics Platform. More precisely this component will extend MIPMap, developed during the RUP of HBP, to support incremental Data Exchange. This means that instead of re-integrating data to the hospital's LDSM, whenever new data are exported from participating hospitals (following the standard pipeline of anonymization etc), they will be integrated into the already existing data taking into account the information that has been integrated before. Hence this component will vastly affect the way Data is integrated to the platform (Data Integration & Schema Mapping/Data Exchange) and the way metadata will be enriched. The functionality provided is incremental integration of data from hospitals.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > Data Integration
  
**Planned functionality at M12 (in line with the milestones)**Report on Incremental Data Integration
  
**Planned functionalities at M18 (in line with the milestones)**Preliminary version of module
  
**Planned functionalities at M24(in line with the milestones)**Final version and integration to the Hospital Bundle
  
**Short description of potential use case**A hospital participates in the MIP by providing a first batch of data. Consider that there exists a rule stating that the target diagnostic table is populated only with patients (source patient table) who have a diagnosis (source diagnostic table). Moreover consider that in this first data batch patient 'A' did not have any examinations, hence the target diagnostics table does not contain any tuple for 'A'. Consider now that in a second batch, a month later, diagnostics for patient 'A' are entered but patient information for 'A' is not resent (as it had been sent in the first batch a month earlier). In order to populate the target diagnostic table with information for 'A' it would be essential to run again the data exchange process for all the input data (first and second batch). However, this poses a significant computational burden. Hence, it would be ideal if an incremental data exchange approach would be used that would not need to re-run the data translation process from scratch but would be able to generate the information for patient 'A' and her diagnostics as they became available.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Online Data Integration Module (ODIM)
  
**Short Description of this component**This component will enable the move from manual to automatic cleaning, and the transformation and merging of actions whenever hospitals add new data to the Medical Informatics Platform. More precisely this component will extend MIPMap, developed during the RUP of HBP, to support incremental Data Exchange. This means that instead of re-integrating data to the hospital's LDSM, whenever new data are exported from participating hospitals (following the standard pipeline of anonymization etc), they will be integrated into the already existing data taking into account the information that has been integrated before. Hence this component will vastly affect the way Data is integrated to the platform (Data Integration & Schema Mapping/Data Exchange) and the way metadata will be enriched. The functionality provided is incremental integration of data from hospitals.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Schema Mapping
  
**Planned functionality at M12 (in line with the milestones)**Report on Incremental Data Integration
  
**Planned functionalities at M18 (in line with the milestones)**Preliminary version of module
  
**Planned functionalities at M24(in line with the milestones)**Final version and integration to the Hospital Bundle
  
**Short description of potential use case**A hospital participates in the MIP by providing a first batch of data. Consider that there exists a rule stating that the target diagnostic table is populated only with patients (source patient table) who have a diagnosis (source diagnostic table). Moreover consider that in this first data batch patient 'A' did not have any examinations, hence the target diagnostics table does not contain any tuple for 'A'. Consider now that in a second batch, a month later, diagnostics for patient 'A' are entered but patient information for 'A' is not resent (as it had been sent in the first batch a month earlier). In order to populate the target diagnostic table with information for 'A' it would be essential to run again the data exchange process for all the input data (first and second batch). However, this poses a significant computational burden. Hence, it would be ideal if an incremental data exchange approach would be used that would not need to re-run the data translation process from scratch but would be able to generate the information for patient 'A' and her diagnostics as they became available.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Community Schema Curation Module (ComSchema)
  
**Short Description of this component**This component will allow MIP users to create, share, validate and reuse schema mappings. More precisely, MIP users (provided they have specific access rights) will be able to share their mappings, making them global. This will allow all other MIP users to view these global mappings and endorse them partially or completely to their own. Additionally users will be able to combine and extend existing mappings by adding/removing tables. Finally, users will be able to 'friend' other users allowing them access to their (non-global) mappings. This component will overall increase the scope of the MIP significantly, and allow it to deal more easily with the variety of clinical data available. This component is based on extending WebMIPMap with crowd sourcing functionalities. This component affects the ontology& standards component as it will make standardization easier. Moreover, it affects the Information and Scientific References component as it will affect the ontologies and variables used and finally it will affect schema mapping and data integration as it will affect the way mappings (that could potentially run on MIPMap) are created. The component can be used to accelerate the creation of the Knowledge Graph of SP5.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Hospital Databases Bundle (HDB) > Schema Mapping
  
**Planned functionality at M12 (in line with the milestones)**Creation of Global Mappings from privileged users and users' ability to endorse them
  
**Planned functionalities at M18 (in line with the milestones)**Ability to merge and extend existing schemata and create mappings for the overall scenario
  
**Planned functionalities at M24(in line with the milestones)**Visibility of user defined mappings to other users - 'friend' (or 'user I trust') feature and integration to the Web Portal
  
**Short description of potential use case**Use case 1: Consider an administrator that creates a mapping of some variables of a research project schema to the MIP schema, and shares this schema with the MIP users, thus making it global. Consider now that a MIP user (with no administrative privileges) would like to create a mapping for additional variables of the same schemata/ontologies. It would be beneficial for the user to be able to load the global administrator's schema and consult its correspondences (either endorse them or ignore them) to create her own mapping. Use case 2: Users can share their mappings of novel research data schemas with other users ('friends') and/or endorse their mappings, to further data sharing.
  
**Task number attached to this component**T8.1.4 Data Integration AUEB
  
**Name of this component**Community Schema Curation Module (ComSchema)
  
**Short Description of this component**This component will allow MIP users to create, share, validate and reuse schema mappings. More precisely, MIP users (provided they have specific access rights) will be able to share their mappings, making them global. This will allow all other MIP users to view these global mappings and endorse them partially or completely to their own. Additionally users will be able to combine and extend existing mappings by adding/removing tables. Finally, users will be able to 'friend' other users allowing them access to their (non-global) mappings. This component will overall increase the scope of the MIP significantly, and allow it to deal more easily with the variety of clinical data available. This component is based on extending WebMIPMap with crowd sourcing functionalities. This component affects the ontology& standards component as it will make standardization easier. Moreover, it affects the Information and Scientific References component as it will affect the ontologies and variables used and finally it will affect schema mapping and data integration as it will affect the way mappings (that could potentially run on MIPMap) are created. The component can be used to accelerate the creation of the Knowledge Graph of SP5.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > Data Integration
  
**Planned functionality at M12 (in line with the milestones)**Creation of Global Mappings from privileged users and users' ability to endorse them
  
**Planned functionalities at M18 (in line with the milestones)**Ability to merge and extend existing schemata and create mappings for the overall scenario
  
**Planned functionalities at M24(in line with the milestones)**Visibility of user defined mappings to other users - 'friend' (or 'user I trust') feature and integration to the Web Portal
  
**Short description of potential use case**Use case 1: Consider an administrator that creates a mapping of some variables of a research project schema to the MIP schema, and shares this schema with the MIP users, thus making it global. Consider now that a MIP user (with no administrative privileges) would like to create a mapping for additional variables of the same schemata/ontologies. It would be beneficial for the user to be able to load the global administrator's schema and consult its correspondences (either endorse them or ignore them) to create her own mapping. Use case 2: Users can share their mappings of novel research data schemas with other users ('friends') and/or endorse their mappings, to further data sharing.
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now.

### Subject: 27

**Survey Timestamp**2016-09-14 13:39:39
  
**Task leader name**Thomas Heinis
  
**Task number attached to this component**T8.4.5 Large-scale data analytics on massively parallel architecture ICL
  
**Name of this component**Data Uploader
  
**Short Description of this component**This component essentially allows a user to upload (medical) data to the supercomputing infrastructure. It will either be implemented and made available through command line scripts (possibly Python) or a web page.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > HPC
  
**Planned functionality at M12 (in line with the milestones)**This component should be fully available at M12 (pending collaboration with SP7)
  
**Planned functionalities at M18 (in line with the milestones)**none
  
**Planned functionalities at M24(in line with the milestones)**none
  
**Short description of potential use case**Clustering of high-dimensional medical data.
  
**Task number attached to this component**T8.4.5 Large-scale data analytics on massively parallel architecture ICL
  
**Name of this component**Data Cleaning & Formatting
  
**Short Description of this component**THis component will clean, reformat and distribute the data in the supercomputing infrastructure. It will be based on scripts and will connect to the uploading component.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > HPC
  
**Planned functionality at M12 (in line with the milestones)**This component should be fully available at M12 (pending collaboration with SP7)
  
**Planned functionalities at M18 (in line with the milestones)**none
  
**Planned functionalities at M24(in line with the milestones)**none
  
**Short description of potential use case**Clustering of high-dimensional medical data.
  
**Task number attached to this component**T8.4.5 Large-scale data analytics on massively parallel architecture ICL
  
**Name of this component**Analytics Library
  
**Short Description of this component**This component will develop and deploy a library (containing multiple clustering/classification/machine learning algorithms) for data analaysis on the distributed/supercomputing infrastructure. It will be based on MPI and most likely implemented in C++. It will contain several algotihms including approximate ones. The component connects to the data uploader and the cleaner.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > HPC
  
**Planned functionality at M12 (in line with the milestones)**none
  
**Planned functionalities at M18 (in line with the milestones)**none
  
**Planned functionalities at M24(in line with the milestones)**This component should be fully available at M24 (pending collaboration with SP7)
  
**Short description of potential use case**Clustering of high-dimensional medical data.
  
**Task number attached to this component**T8.4.5 Large-scale data analytics on massively parallel architecture ICL
  
**Name of this component**Data Download
  
**Short Description of this component**This component will enable users to download analysis results. It will either be available as scripts on the command line or as a web page. It depends on the analytics library and thus connects to it.
  
**To which building block your component belongs to ? (select one)**SOFTWARE > Data Factory (DF) > HPC
  
**Planned functionality at M12 (in line with the milestones)**none
  
**Planned functionalities at M18 (in line with the milestones)**none
  
**Planned functionalities at M24(in line with the milestones)**This component should be fully available at M24 (pending collaboration with SP7)
  
**Short description of potential use case**Clustering of high-dimensional medical data.
  
**To add a new component - click 'Next Page'**I am done. Submit the survey now.

### Subject: 28

**Survey Timestamp**[not completed]
  
**Task leader name**Bogdan Draganski
  
**Task number attached to this component**T8.4.1 Brain scale high performance deep phenotyping CHUV