# SOFTWARE

## Data Factory (DF)

### Feature Engineering

**Component(13):** Brain morphological features  
**Contributing task(s):** T8.3.11 Brain morphological features UCL  
**Description:** A privacy preserving approach for the generalised principal component analysis of large image datasets.

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| DATA | Hospital | Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Reference | TBI (Traumatic Brain Injury), PPMI (Parkinson's Progression Markers Initiative), ADNI |
| SOFTWARE | Data Factory (DF) | Data Anonymisation, Workflow Engine, Data Pipeline processes, Data Quality Processes, Data Storage |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | Further testing and code optimisation of image factorisation method (handle missing data, make more probabilistic, make computationally stable, etc). |
| Planned functionality at M12 | Initial implementation of image factorisation method without distributed computing (MS126). |
| Planned functionalities at M24 | A functioning distributed implementation of the image factorisation method, with features obtained by the approach available for data mining. |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| Neuroscientist Clinical Researcher Developer - Software Developer - Methods SP8 Platform Developer | 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. |

# SOFTWARE

## Hospital Databases Bundle (HDB)

### Local Database

**Component(14):** Access Right Module (ARM)  
**Contributing task(s):** T8.1.4 Data Integration AUEB  
**Description:** 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

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| SOFTWARE | Hospital Databases Bundle (HDB) | Local Database, Local Database, Local Database, Schema Mapping, Schema Mapping, Schema Mapping, Schema Mapping, Schema Mapping, Schema Mapping, Schema Mapping |
|  | Data Factory (DF) | Data Integration, Data Integration, Data Integration |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | Basic access control functionalityBasic access control functionalityBasic access control functionalityQuery answering without rewriting with respect to database sources schemaQuery answering without rewriting with respect to database sources schemaQuery answering without rewriting with respect to database sources schemaQuery answering without rewriting with respect to database sources schemaPreliminary version of modulePreliminary version of modulePreliminary version of module |
| Planned functionality at M12 | Report on access restrictions in the MIPReport on access restrictions in the MIPReport on access restrictions in the MIPRewriting of query with respect to ontology module and preliminary integration to the platformRewriting of query with respect to ontology module and preliminary integration to the platformRewriting of query with respect to ontology module and preliminary integration to the platformRewriting of query with respect to ontology module and preliminary integration to the platformReport on Incremental Data IntegrationReport on Incremental Data IntegrationReport on Incremental Data Integration |
| Planned functionalities at M24 | Final release of ARM and integration to the platformFinal release of ARM and integration to the platformFinal release of ARM and integration to the platformFinal version - Optimized query answering - and full integration to the platformFinal version - Optimized query answering - and full integration to the platformFinal version - Optimized query answering - and full integration to the platformFinal version - Optimized query answering - and full integration to the platformFinal version and integration to the Hospital BundleFinal version and integration to the Hospital BundleFinal version and integration to the Hospital Bundle |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
|  | 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. |

# SOFTWARE

## Algorithm Library

### Machine Learning Library

**Component(15):** Large Scale Analytics Algorithms  
**Contributing task(s):** T8.4.5 Large-scale data analytics on massively parallel architecture ICL  
**Description:** Analytics/clustering algorithms for the efficient and scalable large scale analysis of medical data.

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| SOFTWARE | Data Factory (DF) | HPC |
|  | Algorithm Library | Machine Learning Library |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | Additional distributed, MPI base algorithms as well as more mature implementations. |
| Planned functionality at M12 | Initial prototypes of a set of algorithms run distributed on HPC via MPI. |
| Planned functionalities at M24 | Deployment of production grade algorithms on supercomputing infrastructure, making it available fro general use. |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| Neuroscientist Clinical Researcher Developer - Software Developer - Methods | Large scale clustering/analysis of large amounts of (cleaned) medical data. |

# None

**Component(16):** None  
**Contributing task(s):** None  
**Description:** None

**Dependencies:**

**Releases:**



**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
|  | None |

# SOFTWARE

## Hospital Databases Bundle (HDB)

### Local Database

**Component(17):** Extended multidimensional query support  
**Contributing task(s):** T8.1.1 Infrastructure to support just-in-time analytics on raw medical data EPFL  
**Description:** Basic primitives for computation over multidimensional queries in the local hospital database. This component needs data but not necessarily data from ALL hospitals.

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| DATA | Hospital | Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital, Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital, Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
| SOFTWARE | Data Factory (DF) | Data Anonymisation, Workflow Engine, Data Pipeline processes, Data Quality Processes, Data Storage, Data Integration, Data Anonymisation, Workflow Engine, Data Pipeline processes, Data Storage, Data Integration, HPC, Data Anonymisation, Workflow Engine, Data Pipeline processes, Data Quality Processes, Data Storage, Data Integration, Data Categorization, Feature Engineering, HPC, Workflow Engine |
|  | Hospital Databases Bundle (HDB) | Schema Mapping, Schema Mapping, Hospital Bundle Package, Local Database, Federated Query, Schema Mapping, Hospital Bundle Package, Local Database, Federated Query, Schema Mapping, Hospital Bundle Package |
|  | Algorithm Factory (AF) | Workflow Engine: Woken, Package of Algorithms as Docker images, Workflow Engine: Woken |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | The plug-in is completed and deployed as part of the Hospital Bundle.The federation software is integrated with the new and improved query engine API's .Software extension to support provenance tracking through hospital hubs have been implemented.Continuous process |
| Planned functionality at M12 | The infrastructure to run this plug-in in the query engine will be ready.The query engine extensions are implemented.A secure procedure has been defined to connect a hospital to a hub. Related data management processes are also defined.Continuous process |
| Planned functionalities at M24 | The plug-in is completed and deployed as part of the Hospital Bundle and validated by HBP users.Query engine optimized using real-world datasets and queries.The software required by the procedure is developed, tested and packaged, ready to be used if necessary.Continuous process |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| SP8 Platform Developer Developer - Software Developer - Software Developer - Software | 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. |

# None

**Component(18):** Functional gene-sets relevant to human brain disorders  
**Contributing task(s):** T8.4.1 Brain scale high performance deep phenotyping CF  
**Description:** 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.

**Dependencies:**

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | 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 functionality at M12 | 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 M24 | 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. |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| Neuroscientist Clinical Researcher | 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). |

# SOFTWARE

## Algorithm Library

### Machine Learning Library

**Component(19):** Disease signature: Distributed rule-based methods  
**Contributing task(s):** T8.3.5 Methods for distributed rule-based disease signature discovery JSI  
**Description:** 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).

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| DATA | Hospital | Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital, Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Reference | PPMI (Parkinson's Progression Markers Initiative), ADNI |
| SOFTWARE | Algorithm Library | Predictive Models, Predictive Models |
|  | Algorithm Factory (AF) | Model Scoring, X-Validation module, Model Testing (PFA parsing), PFA translation, Model Training / Parameter Estimation, Package of Algorithms as Docker images, Model Scoring, X-Validation module, Model Testing (PFA parsing), PFA translation, Model Training / Parameter Estimation, Package of Algorithms as Docker images |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | First tests and validations on research data.First tests and validations on research data.First dataset annotated according to the ontology. |
| Planned functionality at M12 | First working prototype developed.First working prototype developed.Prototype of the ontology for describing data on patients with neurological diseases developed |
| Planned functionalities at M24 | Analysis of distributed data.Analysis of longitudinal data.Ontology and several datasets annotated according to the ontology. |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| Neuroscientist Clinical Researcher Neuroscientist Clinical Researcher Developer - Software Developer - Methods SP8 Platform Developer | Discovery of disease signatures on distributed data. |

# SOFTWARE

## Algorithm Library

### Brain Anatomy

**Component(20):** GeneHeatMapper  
**Contributing task(s):** T8.3.10 Methods for linkage of local SNP data (individual SNPs) to imaging data through SNP LUMC  
**Description:** Algorithm that generates a 3D expression heatmap of of an SNP name, gene name or co-expression module.

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| SOFTWARE | Web Exploration and Analytics | Image & Genetic Viewer |
|  | Algorithm Library | Feature reduction, Statistical Analytics, Brain Anatomy |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | M 8.3.2: Initial proof-of-concept algorithm for generating 3D gene expression heatmaps from gene co-expression modules. |
| Planned functionality at M12 | M 8.3.2 Initial proof-of-concept algorithm for generating 3D gene expression heatmaps of single genes and SNPs |
| Planned functionalities at M24 | M 8.3.11 Gene Expression Maps of Disease Link to Brain Atlases |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| Developer - Methods | Discovery of disease associations between gene expression data en imaging data |

# SOFTWARE

## Algorithm Library

### Statistical Analytics

**Component(21):** 3-C (Categorize, Cluster & Classify)  
**Contributing task(s):** T8.3.1 Tools to mine replicable selection and integration of hierarchical features, inter and across domains using FDR. TAU  
**Description:** Methodology for Medical big data analysis and disease sub-type identification

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| DATA | Hospital | Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital, Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital, Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital, Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital, Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Reference | PPMI (Parkinson's Progression Markers Initiative), ADNI, PPMI (Parkinson's Progression Markers Initiative), ADNI, PPMI (Parkinson's Progression Markers Initiative), ADNI, TBI (Traumatic Brain Injury), PPMI (Parkinson's Progression Markers Initiative), ADNI, PPMI (Parkinson's Progression Markers Initiative), ADNI |
| SOFTWARE | Data Factory (DF) | Data Pipeline processes, Data Quality Processes |
|  | Web Exploration and Analytics | Image & Genetic Viewer, Research & Modeling application |
|  | Algorithm Library | Statistical Analytics, Feature reduction, Statistical Analytics, Machine Learning Library, Statistical Analytics, Statistical Analytics, Statistical Analytics, Statistical Analytics |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | Methodology improvemnts - scaling up the ability to work with larger datasets.Clinical measurements data constructionIncreasing the variety of data sources that can be integrated into the previous analyses, in particular genomicsFunctions improvementsMissing values: structure discoveryExploration of new methods for incorporating Knowledge into the processSummary of consultations with multi-disciplinary experts and opinion leaders- |
| Planned functionality at M12 | MS 8.3.2: Initial Proof-of-concept and results of the different 3-C AlgorithmsReview of existing methodsDeveloping methodologies for enriched data MS8.3.3Improvements and use in different datasetsMissing values - extractionRe-evaluation of 3-C methodology as a data and knowledge combined methodologyLiterature ReviewPaper submitted |
| Planned functionalities at M24 | Validation of 3-C strategy on different dataMS8.3.11: Analyses of 3-C demonstrating the use of Longitudinal DataAnalysis of multi-domain dataPrepare methods for integration into MIP, test and validate results. (preparation specification to be agreed with and provided by MIP implementation team)Missing values: imputation and visualization MS8.3.10Report on incorporating Knowledge into the processSuggest definition of the 'Disease signature' and proposed estimation methods.- |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| Clinical Researcher Developer - Methods SP8 Platform Developer Clinical Researcher Developer - Methods SP8 Platform Developer General Public Neuroscientist Clinical Researcher Developer - Methods Neuroscientist Clinical Researcher Developer - Methods Neuroscientist Clinical Researcher Developer - Methods SP8 Platform Developer General Public Developer - Methods Clinical Researcher General Public | None |

# SOFTWARE

## Algorithm Library

### Machine Learning Library

**Component(22):** Integrating multi-domain data: Methods for redescription mining  
**Contributing task(s):** T8.3.6 Methods for redescription mining JSI  
**Description:** 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.

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| DATA | Hospital | Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Reference | PPMI (Parkinson's Progression Markers Initiative), ADNI, PPMI (Parkinson's Progression Markers Initiative), ADNI |
| SOFTWARE | Algorithm Library | Predictive Models |
|  | Algorithm Factory (AF) | Model Scoring, X-Validation module, Model Testing (PFA parsing), PFA translation, Model Training / Parameter Estimation, Package of Algorithms as Docker images, Model Testing (PFA parsing), PFA translation, Package of Algorithms as Docker images |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | First tests and validations on research data.First tests and validations on research data. |
| Planned functionality at M12 | First working prototype developed.First working prototype developed. |
| Planned functionalities at M24 | Analysis of multi-view data.Analysis of text-enriched heterogeneous information networks. |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| Neuroscientist Clinical Researcher Neuroscientist Clinical Researcher | Discovery of disease signatures. |

# None

**Component(23):** None  
**Contributing task(s):** None  
**Description:** None

**Dependencies:**

**Releases:**



**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
|  | None |

# SOFTWARE

## Hospital Databases Bundle (HDB)

### Local Database

**Component(24):** Nifti data source for local query engine  
**Contributing task(s):** T8.1.1 Infrastructure to support just-in-time analytics on raw medical data EPFL  
**Description:** Plug-in to enable local query engine to perfom queries directly on Nifti files.

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| DATA | Hospital | Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital, Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
| SOFTWARE | Data Factory (DF) | Data Anonymisation, Data Pipeline processes, Data Storage, Data Integration, Data Anonymisation, Data Pipeline processes, Data Storage, Data Integration, Data Anonymisation, Data Pipeline processes, Data Storage, Data Integration, Data Anonymisation, Data Storage, Data Anonymisation, Data Storage |
|  | Hospital Databases Bundle (HDB) | Local Database, Schema Mapping, Local Database, Federated Query, Schema Mapping, Hospital Bundle Package, Local Database, Federated Query, Schema Mapping, Hospital Bundle Package, Federated Query, Schema Mapping, Hospital Bundle Package |
|  | Algorithm Factory (AF) | Workflow Engine: Woken, Model Scoring |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | Initial proof-of-concept query engine that can perform more complex queries like joins and group by over Nifti files.Initial proof-of-concept query engine that can perform more complex queries like joins and group by over genetic data files.Initial proof-of-concept with a few functions defined in collaboration with Data Analysis groups.Initial proof-of-concept with a few functions defined in collaboration with Data Analysis groups.Initial proof-of-concept using this component is developed.Anonymization process validated in three hospitals.Initial proof-of-concept developed.Initial proof-of-concept developed.Initial proof-of-concept developed. |
| Planned functionality at M12 | Initial proof-of-concept query engine that can perform selection queries over Nifti files.Initial proof-of-concept query engine that can perform selection queries over genetic data files.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.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.The technical procedure describing the secure connection and which technologies to use is documented.Anonymization process validated in one hospital.Technology defined.dependent on delivery of: SOFTWARE > Data Factory (DF) > Workflow Engine > Encrypted overlay network Technology defined.dependent on delivery of: SOFTWARE > Hospital Databases Bundle (HDB) > Local Database > Secure connection between two hospitals. Technology defined. |
| Planned functionalities at M24 | Query engine that can perform complex queries and mathematical operations over Nifti files.Query engine that can perform complex queries and mathematical operations over genetic data files.Final Nifti library with a satisfactory number of functions.Final genetic data library with a satisfactory number of functions.The component is integrated in the hospital bundle and/or data factory.Anonymization process validated in all five hospitals.The component is integrated with the bundle in the local hospital.The component is integrated with the bundle in the local hospital.The component is integrated with the bundle in the local hospital. |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| Developer - Software Developer - Software Developer - Software Developer - Software Developer - Software Developer - Software Developer - Software Developer - Software Developer - Software | Data discovery directly on Nifti files Enable analysis over imaging data that are not provided directly by the existing pipeline. |

# SOFTWARE

## Hospital Databases Bundle (HDB)

### Federated Query

**Component(25):** Master component  
**Contributing task(s):** T8.1.5 Distributed complex workflow engine UoA  
**Description:** The master component transforms, schedules and dispatches the queries to workers

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| SOFTWARE | Hospital Databases Bundle (HDB) | Hospital Bundle Package, Hospital Bundle Package, Federated Query, Hospital Bundle Package, Hospital Bundle Package, Hospital Bundle Package, Hospital Bundle Package, Hospital Bundle Package |
|  | Algorithm Library | Predictive Models, Feature reduction, Statistical Analytics, Biological Diagnostic Tools, Machine Learning Library, Predictive Models, Feature reduction, Statistical Analytics, Biological Diagnostic Tools, Machine Learning Library |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | Master component is integrated with 5 worker components deployed in 5 hospitalsWorker/ Bridge Component is deployed in 5 hospitalsFirst version of web portal connector component deployedDeployment of UDFs componentDeployment of template composer componentDeployment of query template repositoryDeployment of management component of query template repository |
| Planned functionality at M12 | Master component released packaged and integrated with a worker componentWorker/ Bridge Component released, packaged and integrated with Master component and RAW query engineDefine communication protocol with web portalImplementation of UDFs componentImplementation of template composer componentImplementation of query template repositoryImplementation of management component of query template repository |
| Planned functionalities at M24 | Master component is validated by HBP usersWorker/ Bridge Component is validated by HBP usersWeb portal connector component is validated by HBP usersValidation of UDFs componentValidation of template composer componentValidation of query template repositoryValidation of management component of query template repository |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| Developer - Software Developer - Methods SP8 Platform Developer Developer - Software Developer - Methods SP8 Platform Developer Developer - Software Developer - Methods SP8 Platform Developer Developer - Software Developer - Methods SP8 Platform Developer Developer - Software Developer - Methods SP8 Platform Developer Developer - Software Developer - Methods SP8 Platform Developer Developer - Software Developer - Methods SP8 Platform Developer | None |

# SOFTWARE

## Hospital Databases Bundle (HDB)

### Local Database

**Component(26):** Access Rights Module (ARM)  
**Contributing task(s):** T8.1.4 Data Integration AUEB  
**Description:** 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.

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| SOFTWARE | Hospital Databases Bundle (HDB) | Local Database, Local Database, Schema Mapping, Schema Mapping, Schema Mapping, Schema Mapping, Schema Mapping, Schema Mapping |
|  | Data Factory (DF) | Data Integration, Data Integration |
|  | Web Exploration and Analytics | Data Access, Data Access |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | Basic access control functionalityBasic access control functionalityQuery answering without rewriting with respect to database sources schemaQuery answering without rewriting with respect to database sources schemaPreliminary version of modulePreliminary version of moduleAbility to merge and extend existing schemata and create mappings for the overall scenarioAbility to merge and extend existing schemata and create mappings for the overall scenario |
| Planned functionality at M12 | Report on access restrictions in the MIPReport on access restrictions in the MIPRewriting of query with respect to ontology module and preliminary integration to the platformRewriting of query with respect to ontology module and preliminary integration to the platformReport on Incremental Data IntegrationReport on Incremental Data IntegrationCreation of Global Mappings from privileged users and users' ability to endorse themCreation of Global Mappings from privileged users and users' ability to endorse them |
| Planned functionalities at M24 | Final release of ARM and integration to the platformFinal release of ARM and integration to the platformFinal version - Optimized query answering - and full integration to the platformFinal version - Optimized query answering - and full integration to the platformFinal version and integration to the Hospital BundleFinal version and integration to the Hospital BundleVisibility of user defined mappings to other users - 'friend' (or 'user I trust') feature and integration to the Web PortalVisibility of user defined mappings to other users - 'friend' (or 'user I trust') feature and integration to the Web Portal |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
|  | 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. |

# SOFTWARE

## Data Factory (DF)

### HPC

**Component(27):** Data Uploader  
**Contributing task(s):** T8.4.5 Large-scale data analytics on massively parallel architecture ICL  
**Description:** 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.

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| SOFTWARE | Algorithm Library | Statistical Analytics, Machine Learning Library, Statistical Analytics, Machine Learning Library, Statistical Analytics, Machine Learning Library, Statistical Analytics, Machine Learning Library |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | nonenonenonenone |
| Planned functionality at M12 | This component should be fully available at M12 (pending collaboration with SP7)This component should be fully available at M12 (pending collaboration with SP7)nonenone |
| Planned functionalities at M24 | nonenoneThis component should be fully available at M24 (pending collaboration with SP7)This component should be fully available at M24 (pending collaboration with SP7) |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| Neuroscientist Clinical Researcher Neuroscientist Clinical Researcher Neuroscientist Clinical Researcher Neuroscientist Clinical Researcher | Clustering of high-dimensional medical data. |

# None

**Component(28):** None  
**Contributing task(s):** T8.4.1 Brain scale high performance deep phenotyping CHUV  
**Description:** None

**Dependencies:**

**Releases:**



**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
|  | None |