# SOFTWARE

## Data Factory (DF)

### Feature Engineering

|  |  |
| --- | --- |
| **Component** | Brain morphological features |
| **Description** | T8.3.11 Brain morphological features UCL |
| **Contribution task** | A privacy preserving approach for the generalised principal component analysis of large image datasets. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Reference data: TBI (Traumatic Brain Injury), PPMI (Parkinson's Progression Markers Initiative), ADNI |
| **SOFTWARE** | Data Factory (DF): Data Storage |

# SOFTWARE

## Data Factory (DF)

### Data Anonymisation

|  |  |
| --- | --- |
| **Component** | Anonymisation tests |
| **Description** | T8.1.3 Installation and adaptation of SP8 bundle at local hospitals with upgrade support EPFL |
| **Contribution task** | Local hospital anonymization validation tests. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
| **SOFTWARE** | Data Factory (DF): Data Storage |

# SOFTWARE

## Data Factory (DF)

### Workflow Engine

|  |  |
| --- | --- |
| **Component** | Encrypted overlay network |
| **Description** | T8.1.3 Installation and adaptation of SP8 bundle at local hospitals with upgrade support EPFL |
| **Contribution task** | A component that creates a secure network between hospitals for the administration of the servers. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Data Factory (DF): Data Storage |
|  | Hospital Databases Bundle (HDB): Hospital Bundle Package |
| **SERVICES** | Security & Monitoring: Data governance |

|  |  |
| --- | --- |
| **Component** | Remote starting of services |
| **Description** | T8.1.3 Installation and adaptation of SP8 bundle at local hospitals with upgrade support EPFL |
| **Contribution task** | This component will enable the local hospital services to be remotely managed (ie start, stop) |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Hospital Bundle Package |
| **SERVICES** | Upgrade-Deploy-Release: Micro-services |
|  | Security & Monitoring: Data governance |

# SOFTWARE

## Data Factory (DF)

### Data Integration

|  |  |
| --- | --- |
| **Component** | Ontology-based data access Module (OBDA Module) |
| **Description** | T8.1.4 Data Integration AUEB |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | MDR (Meta Data Register): Ontology&Standards |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Schema Mapping |

|  |  |
| --- | --- |
| **Component** | Online Data Integration Module (ODIM) |
| **Description** | T8.1.4 Data Integration AUEB |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Data Factory (DF): Data Integration |
|  | Hospital Databases Bundle (HDB): Schema Mapping |

|  |  |
| --- | --- |
| **Component** | Community Schema Curation Module (ComSchema) |
| **Description** | T8.1.4 Data Integration AUEB |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | MDR (Meta Data Register): Ontology&Standards |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Schema Mapping |

# SOFTWARE

## Data Factory (DF)

### HPC

|  |  |
| --- | --- |
| **Component** | Data Uploader |
| **Description** | T8.4.5 Large-scale data analytics on massively parallel architecture ICL |
| **Contribution task** | 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: Machine Learning Library |

|  |  |
| --- | --- |
| **Component** | Data Cleaning & Formatting |
| **Description** | T8.4.5 Large-scale data analytics on massively parallel architecture ICL |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Algorithm Library: Machine Learning Library |

|  |  |
| --- | --- |
| **Component** | Analytics Library |
| **Description** | T8.4.5 Large-scale data analytics on massively parallel architecture ICL |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Algorithm Library: Machine Learning Library |

|  |  |
| --- | --- |
| **Component** | Data Download |
| **Description** | T8.4.5 Large-scale data analytics on massively parallel architecture ICL |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Algorithm Library: Machine Learning Library |

# SOFTWARE

## Data Factory (DF)

### Data Pipeline processes

|  |  |
| --- | --- |
| **Component** | Workflow tools for Genetic and Neuroimaging |
| **Description** | T8.5.2 Web API and microservices architecture for community driven data analyses and workflows CHUV |
| **Contribution task** | add community tools from for the analyzing Genetic and CT and MRI data (e.g SPM) or compilation of matlab script |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Data Factory (DF): HPC |
| **SERVICES** | Upgrade-Deploy-Release: MIP Integrated Releases Hospital Databases Bundle |

# SOFTWARE

## Data Factory (DF)

### Data Quality Processes

|  |  |
| --- | --- |
| **Component** | Data Quality process |
| **Description** | T8.5.2 Web API and microservices architecture for community driven data analyses and workflows CHUV |
| **Contribution task** | Metrics for data Quality Process using existing tools |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | MDR (Meta Data Register): Common Variables & Metadata, Ontology&Standards |
|  | Normative and aggregated clinical data: No description |
|  | Reference data: Atlas for Brain Disease, TBI (Traumatic Brain Injury), PPMI (Parkinson's Progression Markers Initiative), ADNI |
| **SOFTWARE** | Data Factory (DF): HPC |
| **SERVICES** | Upgrade-Deploy-Release: MIP Integrated Releases Algorithm Library |

# SOFTWARE

## Algorithm Library

### Machine Learning Library

|  |  |
| --- | --- |
| **Component** | Large Scale Analytics Algorithms |
| **Description** | T8.4.5 Large-scale data analytics on massively parallel architecture ICL |
| **Contribution task** | 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 |

|  |  |
| --- | --- |
| **Component** | Disease signature: Distributed rule-based methods |
| **Description** | T8.3.5 Methods for distributed rule-based disease signature discovery JSI |
| **Contribution task** | 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 Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
| **SOFTWARE** | Algorithm Library: Predictive Models |
|  | Algorithm Factory (AF): Package of Algorithms as Docker images |
| **MODELS** | Biological Signature of Diseases: Healthy Aging |

|  |  |
| --- | --- |
| **Component** | Longitudinal modeling: Tree-based and equation-based methods |
| **Description** | T8.3.8 Methods for disease progression modeling JSI |
| **Contribution task** | Machine learning methods for describing and modelling the temporal dynamics of disease and its clinical and biological markers. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Normative and aggregated clinical data: No description |
|  | Reference data: PPMI (Parkinson's Progression Markers Initiative), ADNI |
| **SOFTWARE** | Algorithm Library: Predictive Models |
|  | Algorithm Factory (AF): Package of Algorithms as Docker images |
| **MODELS** | Biological Signature of Diseases: Healthy Aging |

|  |  |
| --- | --- |
| **Component** | Integrating multi-domain data: Methods for redescription mining |
| **Description** | T8.3.6 Methods for redescription mining JSI |
| **Contribution task** | 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 Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Normative and aggregated clinical data: No description |
|  | Reference data: PPMI (Parkinson's Progression Markers Initiative), ADNI |
| **SOFTWARE** | Algorithm Factory (AF): Package of Algorithms as Docker images |
| **MODELS** | Biological Signature of Diseases: Healthy Aging |

|  |  |
| --- | --- |
| **Component** | Integrating multi-domain data: Methods for heterogeneous networks |
| **Description** | T8.3.7 Methods for heterogeneous networks JSI |
| **Contribution task** | Mehods for mining text-enriched heterogeneous information networks. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Reference data: PPMI (Parkinson's Progression Markers Initiative), ADNI |
|  | Normative and aggregated clinical data: No description |
| **SOFTWARE** | Algorithm Library: Predictive Models |
|  | Algorithm Factory (AF): Package of Algorithms as Docker images |
| **MODELS** | Biological Signature of Diseases: Parkinson Disease |

|  |  |
| --- | --- |
| **Component** | base-docker-images |
| **Description** | T8.5.2 Web API and microservices architecture for community driven data analyses and workflows CHUV |
| **Contribution task** | Collection of Docker images used as building blocks to support the integration of an algorithm and its runtime environment in the platform |

**Dependencies:**

|  |  |
| --- | --- |
| **SERVICES** | Upgrade-Deploy-Release: QA (Quality Assurance) Tools |

# SOFTWARE

## Algorithm Library

### Brain Anatomy

|  |  |
| --- | --- |
| **Component** | GeneHeatMapper |
| **Description** | T8.3.10 Methods for linkage of local SNP data (individual SNPs) to imaging data through SNP LUMC |
| **Contribution task** | 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: Brain Anatomy |

|  |  |
| --- | --- |
| **Component** | Brain Scale High Performance Deep Phenotyping |
| **Description** | T8.4.1 Brain scale high performance deep phenotyping CHUV |
| **Contribution task** | Understand the organisation of the neural circuits underlying normal and adaptive behaviours in a genetically homogeneous population with high risk for brain disorders to enable differentiation from normal behaviour. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Reference data: Atlas for Brain Disease |
| **SOFTWARE** | Data Factory (DF): Data Storage |

|  |  |
| --- | --- |
| **Component** | Tools for Macro- to Micro-Scale Data Analysis and Atlasing |
| **Description** | T8.4.3 Tools for macro- to micro-scale data analysis and atlasing CHUV |
| **Contribution task** | Models of MRI data will be developed for specific characterisation of tissue microstructure in vivo. Particular emphasis will be placed on implementation, allowing direct use of the models on in vivo data. The relationship of the newly developed biomarkers with histological and existing qMRI biomarkers will be investigated using the state-of-the-art facilities. To this end, a direct comparison will be conducted between in vivo and ex vivo three-dimensional MRI and histological data in brain surgery specimens from focal epilepsy. This Task will demonstrate the causal relationship between expected neuronal cell loss, fibrillary gliosis, granule cell dispersion, focal dysplasia and estimates of grey matter volume and cortical thickness and will analyse the degree of overlap between MRI- and histology-based definition of cortical thickness. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Reference data: Atlas for Brain Disease |
| **SOFTWARE** | Data Factory (DF): Workflow Engine |
|  | Algorithm Factory (AF): X-Validation module |

# SOFTWARE

## Algorithm Library

### Statistical Analytics

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

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Normative and aggregated clinical data: No description |
|  | Reference data: PPMI (Parkinson's Progression Markers Initiative), ADNI |
| **MODELS** | Biological Signature of Diseases: Alzheimer's Disease |

|  |  |
| --- | --- |
| **Component** | 3-C Longitudinal Modeling |
| **Description** | T8.3.1 Tools to mine replicable selection and integration of hierarchical features, inter and across domains using FDR. TAU |
| **Contribution task** | Incorporating longitudinal information (day-to-day and multi-patients visits). |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Normative and aggregated clinical data: No description |
|  | Reference data: PPMI (Parkinson's Progression Markers Initiative), ADNI |
| **SOFTWARE** | Algorithm Library: Statistical Analytics |

|  |  |
| --- | --- |
| **Component** | Integrating multi-domain data |
| **Description** | T8.3.1 Tools to mine replicable selection and integration of hierarchical features, inter and across domains using FDR. TAU |
| **Contribution task** | methodology for enriching current models |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Normative and aggregated clinical data: No description |
|  | Reference data: PPMI (Parkinson's Progression Markers Initiative), ADNI |
| **SOFTWARE** | Algorithm Library: Machine Learning Library |
| **MODELS** | Biological Signature of Diseases: Healthy Aging |

|  |  |
| --- | --- |
| **Component** | Transformations in medical big data |
| **Description** | T8.3.2 Developing methods for high-dimensional data with possible informative missing values TAU |
| **Contribution task** | Symetry targeted monotone transformations, and the advantage gained in variance stability,linearity and clustering. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Data Factory (DF): Data Quality Processes |
|  | Algorithm Library: Statistical Analytics |

|  |  |
| --- | --- |
| **Component** | Statistical procedures and workflows for missing values |
| **Description** | T8.3.2 Developing methods for high-dimensional data with possible informative missing values TAU |
| **Contribution task** | Develop statistical procedures and workflows to help and guide the discovery of possible patterns in missing values |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Reference data: TBI (Traumatic Brain Injury), PPMI (Parkinson's Progression Markers Initiative), ADNI |
| **SOFTWARE** | Web Exploration and Analytics: Research & Modeling application |
|  | Algorithm Library: Statistical Analytics |

|  |  |
| --- | --- |
| **Component** | Clustering: incorporating Knowledge into the process |
| **Description** | T8.3.3 Introducing selective inference into dimensionality reduction and clustering methods TAU |
| **Contribution task** | Evaluating disease signature clusters (by combining new approaches with tools developed for the visualization and manipulation of hierarchical clustering) |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Normative and aggregated clinical data: No description |
|  | Reference data: PPMI (Parkinson's Progression Markers Initiative), ADNI |
| **SOFTWARE** | Algorithm Library: Statistical Analytics |
| **MODELS** | Biological Signature of Diseases: Parkinson Disease |

|  |  |
| --- | --- |
| **Component** | Disease Signatures -concept and methodology |
| **Description** | T8.3.4 Statistical methods for 'Disease Signature' confidence assessment TAU |
| **Contribution task** | Define and propose a model for disease signature |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Algorithm Library: Statistical Analytics |
| **MODELS** | Biological Signature of Diseases: Healthy Aging |

# SOFTWARE

## Algorithm Library

### Biological Diagnostic Tools

|  |  |
| --- | --- |
| **Component** | Brain Scale Disease Bayes Modelling |
| **Description** | T8.4.2 Brain scale disease bayes modelling CHUV |
| **Contribution task** | Provide models that capture the dynamic changes that occur in the brain due to the disease (e.g. overproduction and aggregation of the amyloid peptide or cognitive decline). The Bayesian approach has theadvantage of explicitly incorporating the uncertainty on the parameter values of each hypothesis and results in posterior probabilities of the parameters of the model and the competing hypotheses. Bayesian Models can therefore be used to learn about the data by exploring different hypotheses based on a prior knowledge of parameter variations and fluctuations. The data type we deal with is heterogeneous and biased since it gathers clinical, genetic and neuroimaging data. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | MDR (Meta Data Register): Common Variables & Metadata, Ontology&Standards |
| **SOFTWARE** | Algorithm Factory (AF): Package of Algorithms as Docker images |

# SOFTWARE

## Hospital Databases Bundle (HDB)

### Local Database

|  |  |
| --- | --- |
| **Component** | Extended multidimensional query support |
| **Description** | T8.1.1 Infrastructure to support just-in-time analytics on raw medical data EPFL |
| **Contribution task** | 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 Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | MDR (Meta Data Register): Ontology&Standards |
| **SOFTWARE** | Data Factory (DF): Data Integration |
|  | Hospital Databases Bundle (HDB): Schema Mapping |
| **SERVICES** | Upgrade-Deploy-Release: MIP Integrated Releases Hospital Databases Bundle |
|  | Data governance: No description |
|  | Security & Monitoring: Data governance |

|  |  |
| --- | --- |
| **Component** | Distributed local query engine over HPC |
| **Description** | T8.1.2 Installation of RAW on local computing infrastructure for hospital analytics requirements EPFL |
| **Contribution task** | Extend the local query engine to enable the use of distributed computing frameworks (like Spark). |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
| **SOFTWARE** | Data Factory (DF): HPC |
|  | Hospital Databases Bundle (HDB): Hospital Bundle Package |
| **SERVICES** | Upgrade-Deploy-Release: Micro-services |
|  | Data governance: No description |
|  | Security & Monitoring: Data governance |

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

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Data Factory (DF): Data Integration |

|  |  |
| --- | --- |
| **Component** | Genetic data data source for local query engine |
| **Description** | T8.1.1 Infrastructure to support just-in-time analytics on raw medical data EPFL |
| **Contribution task** | Plug-in to enable local query engine to perfom queries directly on genetic data files. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Schema Mapping |
|  | Algorithm Factory (AF): Model Scoring |

|  |  |
| --- | --- |
| **Component** | Nifti library in local query engine |
| **Description** | T8.1.1 Infrastructure to support just-in-time analytics on raw medical data EPFL |
| **Contribution task** | Library of functions for common operations on imaging data/Nifti files in the local query engine. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Data Factory (DF): Data Integration |

|  |  |
| --- | --- |
| **Component** | Genetic data library in local query engine |
| **Description** | T8.1.1 Infrastructure to support just-in-time analytics on raw medical data EPFL |
| **Contribution task** | Library of functions for common operations on genetic data files in the local query engine. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Data Factory (DF): Data Integration |

|  |  |
| --- | --- |
| **Component** | Secure connection between two hospitals. |
| **Description** | T8.1.2 Installation of RAW on local computing infrastructure for hospital analytics requirements EPFL |
| **Contribution task** | Establish encrypted connections between two hospitals. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
| **SOFTWARE** | Data Factory (DF): Data Anonymisation |
| **SERVICES** | Data governance: No description |
|  | Security & Monitoring: Data governance |

|  |  |
| --- | --- |
| **Component** | Sharing processing among local hospital nodes |
| **Description** | T8.1.2 Installation of RAW on local computing infrastructure for hospital analytics requirements EPFL |
| **Contribution task** | This component will enable hospitals by finding a way to share processing between hospitals. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Hospital Bundle Package |
| **SERVICES** | Upgrade-Deploy-Release: Micro-services |
|  | Security & Monitoring: Data governance |

|  |  |
| --- | --- |
| **Component** | Access Rights Module (ARM) |
| **Description** | T8.1.4 Data Integration AUEB |
| **Contribution task** | 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): Federated Query |
|  | Web Exploration and Analytics: Data Access |
| **SERVICES** | Security & Monitoring: User Management |

# SOFTWARE

## Hospital Databases Bundle (HDB)

### Federated Query

|  |  |
| --- | --- |
| **Component** | Master component |
| **Description** | T8.1.5 Distributed complex workflow engine UoA |
| **Contribution task** | The master component transforms, schedules and dispatches the queries to workers |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Hospital Bundle Package |

|  |  |
| --- | --- |
| **Component** | Worker/ Bridge Component |
| **Description** | T8.1.5 Distributed complex workflow engine UoA |
| **Contribution task** | The workers reside on the hospital nodes and act as a bridge with the RAW query engine which executes the queries in situ. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Hospital Bundle Package |

|  |  |
| --- | --- |
| **Component** | UDFs component |
| **Description** | T8.1.6 User Defined Functions (UDFs) and query templates UoA |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Hospital Bundle Package |
|  | Algorithm Library: Machine Learning Library |

|  |  |
| --- | --- |
| **Component** | template composer component |
| **Description** | T8.1.6 User Defined Functions (UDFs) and query templates UoA |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Hospital Bundle Package |
|  | Algorithm Library: Machine Learning Library |

|  |  |
| --- | --- |
| **Component** | Query template repository |
| **Description** | T8.1.7 Query templates and workflow management UoA |
| **Contribution task** | 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). |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Hospital Bundle Package |

|  |  |
| --- | --- |
| **Component** | Management component of query template repository |
| **Description** | T8.1.7 Query templates and workflow management UoA |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Hospital Bundle Package |

# SOFTWARE

## Hospital Databases Bundle (HDB)

### Schema Mapping

|  |  |
| --- | --- |
| **Component** | Ontology-based data access Module (OBDA Module) |
| **Description** | T8.1.4 Data Integration AUEB |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | MDR (Meta Data Register): Ontology&Standards |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Schema Mapping |

|  |  |
| --- | --- |
| **Component** | Online Data Integration Module (ODIM) |
| **Description** | T8.1.4 Data Integration AUEB |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Data Factory (DF): Data Integration |
|  | Hospital Databases Bundle (HDB): Schema Mapping |

|  |  |
| --- | --- |
| **Component** | Community Schema Curation Module (ComSchema) |
| **Description** | T8.1.4 Data Integration AUEB |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | MDR (Meta Data Register): Ontology&Standards |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Schema Mapping |

# SOFTWARE

## Algorithm Factory (AF)

### Workflow Engine: Woken

|  |  |
| --- | --- |
| **Component** | Hospital hubs |
| **Description** | T8.1.2 Installation of RAW on local computing infrastructure for hospital analytics requirements EPFL |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
| **SOFTWARE** | Data Factory (DF): HPC |
|  | Hospital Databases Bundle (HDB): Hospital Bundle Package |
|  | Algorithm Factory (AF): Package of Algorithms as Docker images |
| **SERVICES** | Upgrade-Deploy-Release: MIP Integrated Releases Algorithm Factory |
|  | Data governance: No description |
|  | Security & Monitoring: Data governance |

|  |  |
| --- | --- |
| **Component** | woken |
| **Description** | T8.5.2 Web API and microservices architecture for community driven data analyses and workflows CHUV |
| **Contribution task** | Workflow for Analytics An orchestration platform for Docker containers running data mining algorithms. This project exposes a web interface to execute on demand data mining algorithms defined in Docker containers and implemented using any tool or language (R, Python, Java and more are supported). It relies on a runtime environment containing Mesos and Chronos to control and execute the Docker containers over a cluster |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | MDR (Meta Data Register): Common Variables & Metadata |
| **SOFTWARE** | Algorithm Library: Machine Learning Library |
|  | Algorithm Factory (AF): Package of Algorithms as Docker images |
| **SERVICES** | Upgrade-Deploy-Release: MIP Integrated Releases Algorithm Library |

|  |  |
| --- | --- |
| **Component** | Integration of Spark |
| **Description** | T8.5.2 Web API and microservices architecture for community driven data analyses and workflows CHUV |
| **Contribution task** | Spark is a fast and general engine for large-scale data processing. Spark provides many different data analytics and machine learning algorithms and is a popular platform for data science. Some tasks in SP5 leverage Spark and will provide image processing libraries for Spark. Woken can leverage many features of Spark, starting from its data model and existing algorithms to build advanced features such as Cross Validation. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Algorithm Factory (AF): PFA translation |
| **SERVICES** | Upgrade-Deploy-Release: QA (Quality Assurance) Tools |

# SOFTWARE

## Algorithm Factory (AF)

### X-Validation module

|  |  |
| --- | --- |
| **Component** | woken-validation |
| **Description** | T8.5.2 Web API and microservices architecture for community driven data analyses and workflows CHUV |
| **Contribution task** | Cross validation of the result of a machine learning algorithm from its PFA model and a set of test data not used during training |

**Dependencies:**

# SOFTWARE

## Algorithm Factory (AF)

### Package of Algorithms as Docker images

|  |  |
| --- | --- |
| **Component** | functions-repository |
| **Description** | T8.5.2 Web API and microservices architecture for community driven data analyses and workflows CHUV |
| **Contribution task** | Scripts for building the Docker images representing the functions and a document describing the set of functions available in the MIP |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Algorithm Library: Machine Learning Library |

# SOFTWARE

## Web Exploration and Analytics

### Data Access

|  |  |
| --- | --- |
| **Component** | Access Rights Module (ARM) |
| **Description** | T8.1.4 Data Integration AUEB |
| **Contribution task** | 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): Federated Query |
|  | Web Exploration and Analytics: Data Access |
| **SERVICES** | Security & Monitoring: User Management |

# SOFTWARE

## Web Exploration and Analytics

### Knowledge Base Application

|  |  |
| --- | --- |
| **Component** | Hospitals Information System Database |
| **Description** | T8.2.2 Hospitals and information system departments relationship management UNIGE |
| **Contribution task** | A relational database collecting all the ITC standards currently in use in the engaged Hospitals will be created and populated. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | Hospital Data: Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | MDR (Meta Data Register): Ontology&Standards |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Hospital Bundle Package |

|  |  |
| --- | --- |
| **Component** | Knowledge Base |
| **Description** | T8.5.1 Web-based medical data analyses foundation UNIGE |
| **Contribution task** | Wiki-spaces, video tutorials, on-line questionnaires, courses management, I-python space, explanations of all the variables inside the MIP will be connected for the benefit of the end-users to the HBP Collab platform (https://collab.humanbrainproject.eu/ ). |

**Dependencies:**

|  |  |
| --- | --- |
| **SERVICES** | User Training & Communication: No description |

# SOFTWARE

## Web Exploration and Analytics

### Epidemiological Exploration (EE)

|  |  |
| --- | --- |
| **Component** | Frontend and portal-backend Descriptive Stats and Stats Viewer |
| **Description** | T8.5.1 Web-based medical data analyses foundation CHUV |
| **Contribution task** | software applications supporting the MIP portal providing access to the variables-ontology and provenance databases. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | MDR (Meta Data Register): Common Variables & Metadata, Ontology&Standards |
| **SOFTWARE** | Algorithm Library: Statistical Analytics |
|  | Algorithm Factory (AF): Workflow Engine: Woken |
| **SERVICES** | Upgrade-Deploy-Release: MIP Integrated Releases Web Exploration & Analytics |
|  | Community: Clinical demonstrators |
|  | User Training & Communication: No description |

# SOFTWARE

## Web Exploration and Analytics

### Model Configuration

|  |  |
| --- | --- |
| **Component** | Frontend and portal-backend Model Configuration |
| **Description** | T8.5.1 Web-based medical data analyses foundation CHUV |
| **Contribution task** | Software applications supporting the MIP portal that let the user build a model from a selection of variables and filters |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | MDR (Meta Data Register): Common Variables & Metadata, Ontology&Standards |
| **SOFTWARE** | Algorithm Library: Machine Learning Library |

# SOFTWARE

## Web Exploration and Analytics

### Experiment Builder

|  |  |
| --- | --- |
| **Component** | Experiment Builder |
| **Description** | T8.5.1 Web-based medical data analyses foundation CHUV |
| **Contribution task** | design and explore the model space. use Data mining and Bayesian metrics for comparing the models based on cross-validation. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Web Exploration and Analytics: Data Access |
| **MODELS** | Biological Signature of Diseases: Healthy Aging |

|  |  |
| --- | --- |
| **Component** | Experiment Builder Parameter setting |
| **Description** | T8.5.1 Web-based medical data analyses foundation CHUV |
| **Contribution task** | User interface for the parameter setting of the experiment builder. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | MDR (Meta Data Register): Common Variables & Metadata, Ontology&Standards |
| **SERVICES** | Upgrade-Deploy-Release: QA (Quality Assurance) Tools |
|  | User Training & Communication: No description |
| **MODELS** | Biological Signature of Diseases: Healthy Aging |

# SOFTWARE

## Web Exploration and Analytics

### Information and Scientific references

|  |  |
| --- | --- |
| **Component** | Information and Scientific references |
| **Description** | T8.5.1 Web-based medical data analyses foundation CHUV |
| **Contribution task** | Access to reference databases: Variables, ontologies, research object and catalog (Data, workflow, provenance and Models). |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | MDR (Meta Data Register): Common Variables & Metadata, Ontology&Standards |
| **SERVICES** | Upgrade-Deploy-Release: MIP Integrated Releases Web Exploration & Analytics |
|  | Data governance: No description |
|  | Community: Clinical demonstrators |
|  | User Training & Communication: No description |
|  | Security & Monitoring: Data governance |
| **MODELS** | Biological Signature of Diseases: Healthy Aging |

# SERVICES

## Upgrade-Deploy-Release

### MIP Integrated Releases Hospital Databases Bundle

|  |  |
| --- | --- |
| **Component** | SERVICES > Upgrade-Deploy-Release > MIP Integrated Releases Hospital Databases Bundle |
| **Description** | T8.1.3 Installation and adaptation of SP8 bundle at local hospitals with upgrade support EPFL |
| **Contribution task** | 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. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Data Factory (DF): Workflow Engine |
|  | Hospital Databases Bundle (HDB): Hospital Bundle Package |
|  | Algorithm Factory (AF): Workflow Engine: Woken |
| **SERVICES** | Upgrade-Deploy-Release: Micro-services |

# SERVICES

## Upgrade-Deploy-Release

### MIP Integrated Releases Web Exploration & Analytics

|  |  |
| --- | --- |
| **Component** | Web portal connector component |
| **Description** | T8.1.5 Distributed complex workflow engine UoA |
| **Contribution task** | It interfaces the master component with the web portal. |

**Dependencies:**

|  |  |
| --- | --- |
| **SOFTWARE** | Hospital Databases Bundle (HDB): Hospital Bundle Package |

# SERVICES

## Upgrade-Deploy-Release

### Micro-services

|  |  |
| --- | --- |
| **Component** | mip-microservices-infrastructure |
| **Description** | T8.5.2 Web API and microservices architecture for community driven data analyses and workflows CHUV |
| **Contribution task** | Provides the foundation to deploy and manage the MIP components on the Federation and on the Hospital nodes. The infrastructure provides basic services: clustering, high availability, configuration, security, service catalog, monitoring, deployment that are essential to manage a small to medium scale datacenter. It supports the execution of Big data services on a cluster and the execution of Docker containers in production. It promotes the use of micro-services,i.e. small applications dedicated to perform one task and composed together to achieve richer functionality. It provides the following services: SERVICES > Upgrade-Deploy-Release > Micro-services SERVICES > Upgrade-Deploy-Release > MIP Integrated Releases Data SERVICES > Upgrade-Deploy-Release > MIP Integrated Releases Data Factory SERVICES > Upgrade-Deploy-Release > MIP Integrated Releases Hospital Databases Bundle SERVICES > Upgrade-Deploy-Release > MIP Integrated Releases Algorithm Factory SERVICES > Upgrade-Deploy-Release > MIP Integrated Releases Algorithm Library SERVICES > Upgrade-Deploy-Release > MIP Integrated Releases Web Exploration & Analytics SERVICES > Security & Monitoring > Security |

**Dependencies:**

|  |  |
| --- | --- |
| **SERVICES** | Upgrade-Deploy-Release: QA (Quality Assurance) Tools |

# SERVICES

## Community

### Clinical demonstrators

|  |  |
| --- | --- |
| **Component** | Interactions with research initiatives |
| **Description** | T8.2.3 Research initiatives UNIGE |
| **Contribution task** | To contact European and International research initiatives interested in connecting to or use the MIP and its tools |

**Dependencies:**

|  |  |
| --- | --- |
| **SERVICES** | Upgrade-Deploy-Release: MIP Integrated Releases Web Exploration & Analytics |
|  | User Training & Communication: No description |
|  | Security & Monitoring: User Management |

# DATA

## MDR (Meta Data Register)

### Ontology&Standards

|  |  |
| --- | --- |
| **Component** | Ontologies for describing data on neurological diseases, patients |
| **Description** | T8.3.9 Ontologies for describing data on neurological diseases, patients JSI |
| **Contribution task** | A mid-level ontology for describing various types of data on patients with neurological diseases. |

**Dependencies:**

|  |  |
| --- | --- |
| **DATA** | MDR (Meta Data Register): Common Variables & Metadata |