|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Functionality planned at M18** | **Functionality planned at M12** | **Functionality planned at M24** |
| Brain morphological features | Further testing and code optimisation of image factorisation method (handle missing data, make more probabilistic, make computationally stable, etc). | Initial implementation of image factorisation method without distributed computing (MS126). | A functioning distributed implementation of the image factorisation method, with features obtained by the approach available for data mining. |
| Large Scale Analytics Algorithms | Additional distributed, MPI base algorithms as well as more mature implementations. | Initial prototypes of a set of algorithms run distributed on HPC via MPI. | Deployment of production grade algorithms on supercomputing infrastructure, making it available fro general use. |
| Extended multidimensional query support | The plug-in is completed and deployed as part of the Hospital Bundle. | The infrastructure to run this plug-in in the query engine will be ready. | The plug-in is completed and deployed as part of the Hospital Bundle and validated by HBP users. |
| Distributed local query engine over HPC | The federation software is integrated with the new and improved query engine API's . | The query engine extensions are implemented. | Query engine optimized using real-world datasets and queries. |
| Hospital hubs | Software extension to support provenance tracking through hospital hubs have been implemented. | A secure procedure has been defined to connect a hospital to a hub. Related data management processes are also defined. | The software required by the procedure is developed, tested and packaged, ready to be used if necessary. |
| SERVICES > Upgrade-Deploy-Release > MIP Integrated Releases Hospital Databases Bundle | Continuous process | Continuous process | Continuous process |
| Functional gene-sets relevant to human brain disorders | 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. | 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. | 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. |
| Disease signature: Distributed rule-based methods | First tests and validations on research data. | First working prototype developed. | Analysis of distributed data. |
| Longitudinal modeling: Tree-based and equation-based methods | First tests and validations on research data. | First working prototype developed. | Analysis of longitudinal data. |
| Ontologies for describing data on neurological diseases, patients | First dataset annotated according to the ontology. | Prototype of the ontology for describing data on patients with neurological diseases developed | Ontology and several datasets annotated according to the ontology. |
| GeneHeatMapper | M 8.3.2: Initial proof-of-concept algorithm for generating 3D gene expression heatmaps from gene co-expression modules. | M 8.3.2 Initial proof-of-concept algorithm for generating 3D gene expression heatmaps of single genes and SNPs | M 8.3.11 Gene Expression Maps of Disease Link to Brain Atlases |
| 3-C (Categorize, Cluster & Classify) | Methodology improvemnts - scaling up the ability to work with larger datasets. | MS 8.3.2: Initial Proof-of-concept and results of the different 3-C Algorithms | Validation of 3-C strategy on different data |
| 3-C Longitudinal Modeling | Clinical measurements data construction | Review of existing methods | MS8.3.11: Analyses of 3-C demonstrating the use of Longitudinal Data |
| Integrating multi-domain data | Increasing the variety of data sources that can be integrated into the previous analyses, in particular genomics | Developing methodologies for enriched data MS8.3.3 | Analysis of multi-domain data |
| Transformations in medical big data | Functions improvements | Improvements and use in different datasets | Prepare methods for integration into MIP, test and validate results. (preparation specification to be agreed with and provided by MIP implementation team) |
| Statistical procedures and workflows for missing values | Missing values: structure discovery | Missing values - extraction | Missing values: imputation and visualization MS8.3.10 |
| Clustering: incorporating Knowledge into the process | Exploration of new methods for incorporating Knowledge into the process | Re-evaluation of 3-C methodology as a data and knowledge combined methodology | Report on incorporating Knowledge into the process |
| Disease Signatures -concept and methodology | Summary of consultations with multi-disciplinary experts and opinion leaders | Literature Review | Suggest definition of the 'Disease signature' and proposed estimation methods. |
| The importance of transformation in medical big data | - | Paper submitted | - |
| - |  |  |  |
| Integrating multi-domain data: Methods for redescription mining | First tests and validations on research data. | First working prototype developed. | Analysis of multi-view data. |
| Integrating multi-domain data: Methods for heterogeneous networks | First tests and validations on research data. | First working prototype developed. | Analysis of text-enriched heterogeneous information networks. |
| Nifti data source for local query engine | 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 selection queries over Nifti files. | Query engine that can perform complex queries and mathematical operations over Nifti files. |
| Genetic data data source for local query engine | 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 query engine that can perform selection queries over genetic data files. | Query engine that can perform complex queries and mathematical operations over genetic data files. |
| Nifti library in local query engine | Initial proof-of-concept with a few functions defined in collaboration with Data Analysis groups. | 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. | Final Nifti library with a satisfactory number of functions. |
| Genetic data library in local query engine | Initial proof-of-concept with a few functions defined in collaboration with Data Analysis groups. | 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. | Final genetic data library with a satisfactory number of functions. |
| Secure connection between two hospitals. | Initial proof-of-concept using this component is developed. | The technical procedure describing the secure connection and which technologies to use is documented. | The component is integrated in the hospital bundle and/or data factory. |
| Anonymisation tests | Anonymization process validated in three hospitals. | Anonymization process validated in one hospital. | Anonymization process validated in all five hospitals. |
| Encrypted overlay network | Initial proof-of-concept developed. | Technology defined. | The component is integrated with the bundle in the local hospital. |
| Remote starting of services | Initial proof-of-concept developed. | dependent on delivery of: SOFTWARE > Data Factory (DF) > Workflow Engine > Encrypted overlay network Technology defined. | The component is integrated with the bundle in the local hospital. |
| Sharing processing among local hospital nodes | Initial proof-of-concept developed. | dependent on delivery of: SOFTWARE > Hospital Databases Bundle (HDB) > Local Database > Secure connection between two hospitals. Technology defined. | The component is integrated with the bundle in the local hospital. |
| Master component | Master component is integrated with 5 worker components deployed in 5 hospitals | Master component released packaged and integrated with a worker component | Master component is validated by HBP users |
| Worker/ Bridge Component | Worker/ Bridge Component is deployed in 5 hospitals | Worker/ Bridge Component released, packaged and integrated with Master component and RAW query engine | Worker/ Bridge Component is validated by HBP users |
| Web portal connector component | First version of web portal connector component deployed | Define communication protocol with web portal | Web portal connector component is validated by HBP users |
| UDFs component | Deployment of UDFs component | Implementation of UDFs component | Validation of UDFs component |
| template composer component | Deployment of template composer component | Implementation of template composer component | Validation of template composer component |
| Query template repository | Deployment of query template repository | Implementation of query template repository | Validation of query template repository |
| Management component of query template repository | Deployment of management component of query template repository | Implementation of management component of query template repository | Validation of management component of query template repository |
| Access Rights Module (ARM) | Basic access control functionality | Report on access restrictions in the MIP | Final release of ARM and integration to the platform |
| Access Rights Module (ARM) | Basic access control functionality | Report on access restrictions in the MIP | Final release of ARM and integration to the platform |
| Ontology-based data access Module (OBDA Module) | Query answering without rewriting with respect to database sources schema | Rewriting of query with respect to ontology module and preliminary integration to the platform | Final version - Optimized query answering - and full integration to the platform |
| Ontology-based data access Module (OBDA Module) | Query answering without rewriting with respect to database sources schema | Rewriting of query with respect to ontology module and preliminary integration to the platform | Final version - Optimized query answering - and full integration to the platform |
| Online Data Integration Module (ODIM) | Preliminary version of module | Report on Incremental Data Integration | Final version and integration to the Hospital Bundle |
| Online Data Integration Module (ODIM) | Preliminary version of module | Report on Incremental Data Integration | Final version and integration to the Hospital Bundle |
| Community Schema Curation Module (ComSchema) | Ability to merge and extend existing schemata and create mappings for the overall scenario | Creation of Global Mappings from privileged users and users' ability to endorse them | Visibility of user defined mappings to other users - 'friend' (or 'user I trust') feature and integration to the Web Portal |
| Community Schema Curation Module (ComSchema) | Ability to merge and extend existing schemata and create mappings for the overall scenario | Creation of Global Mappings from privileged users and users' ability to endorse them | Visibility of user defined mappings to other users - 'friend' (or 'user I trust') feature and integration to the Web Portal |
| Data Uploader | none | This component should be fully available at M12 (pending collaboration with SP7) | none |
| Data Cleaning & Formatting | none | This component should be fully available at M12 (pending collaboration with SP7) | none |
| Analytics Library | none | none | This component should be fully available at M24 (pending collaboration with SP7) |
| Data Download | none | none | This component should be fully available at M24 (pending collaboration with SP7) |
| Hospitals Information System Database | 40% of the information related to the Hospitals engaged in the SGA1 will be collected/inserted in the database and finally exposed through the KB platform | The best database solution will be identified and its structure will be finalised | 100% of the information related to the Hospitals engaged in the SGA1 will be collected/inserted in the database and exposed through the KB platform. This will facilitate the technical development allowing the propagation of the MIP solutions across new Hospitals. |
| Interactions with research initiatives | Community Engagement: the following initiatives will be contacted: EMIF-AD, ARWIBO, EuroPOND, BrainGIFT | To generate a list of research studies/consortia efforts focused on neurodegenerative diseases to be contacted | To organise at least two on-line workshops with the aim to present/promote the MIP platform |
| Knowledge Base | Knowledge base fully plugged inside the HBP relying on the GASPAR Single Sign On (SSO). | A new version of the MIP Knowledge Base will be designed in order to make it compatible with the HBP collab platform. | New user-friendly widgets and plug-in (i.e.: on-line form, forums, tutorials, variable dictionary, etc..) will be deployed in order to map every MIP tool in the Knowledge Base. The Knowledge base will allow a dynamic population of the contents allowing users to refine the messages for the user's community. It will represent the preferred MIP source of knowledge. |
| mip-microservices-infrastructure | Support for Big data applications (Spark). Hardening of servers (Linux system, SSH and HTTP), access control to allow remote administration of the servers. Distributed health monitoring to detect failures in Federation or Hospital systems. | Automated deployment of all components of the MIP, targeting Federation datacenter and Hospital data centers. Basic security in place to protect the platform against unauthorised access. | Production-level infrastructure: \* active security protection \* autiting of all user and administrative actions \* distributed performance monitoring \* distributed error reporting \* automated recovery for failing applications \* managed software upgrades, simultaneous upgrades on the Federation and hospitals. |
| Frontend and portal-backend Descriptive Stats and Stats Viewer | Creation of new composite variables as a function of existing variables. Improved versions API to be used from python Notebook with viewer capabilities. added QA testing Support for missing data. | added configuration of the variable by user profiles. filtering options. building first versions API to be used from python Notebook. added QA testing incorporate Brain Atlas viewer. | Create support for creating hierarchical models and multi-steps models. Add recommendation system for choosing the best variables based on model scoring. Allow selection of variables from the brain viewer. |
| Frontend and portal-backend Model Configuration | More advanced filtering, detection of variables with missing values. Improved versions API to be used from python Notebook with viewer capabilities. added QA testing Creation of new variables by transforming, combining or splitting existing variables. | Variable selection and filtering. building first versions API to be used from python Notebook. | Creation of a new model from an existing model by applying a pipeline of global transformations like dimentionality reduction. |
| Experiment Builder | Allow easy selection of algorithm types. | Added support for different types of algorithms. automated selection of the tools according to the variable selected. | Improved metrics for model comparisons |
| Experiment Builder Parameter setting | Automated parameters setting with recommendation add tutorials | Manual parameters setting | Included setting in the ipython notebook |
| Information and Scientific references | meta-data are stored in elastic search | meta-data are stored in static database | meta-data are stored in graph-database |
| Workflow tools for Genetic and Neuroimaging | integration of compiled matlab script in the data factory validation of the QA and QC metrics. | integration of matlab script in the data factory and bio-informatics | full validation in a large datasets. |
| Data Quality process | Test the metric in a large sample | Install and test the QA and QC metrics in a small representative sample. | validated metric available cross-cohorts |
| woken | Processing pipelines with several stages | Support for image and file processing | Processing of large datasets that do not fit in memory |
| woken-validation | Reimplementation as a Spark component for better performance, memory usage, stability. | Cross validation using random data samples. | Cross validation using hospital data |
| functions-repository | Integration of more algorithms Support for other type of repository of Docker images: CWL, Boutique... | Integration of more algorithms Creation of the document referencing all algorithms Simplification of the process to include a new algorithm | TBD |
| base-docker-images | Support for Matlab Support of files as input and output | Support for Python Integration of libraries to simplify creation of PFA documents for storing results | TDB |