Human Brain Project

THE MEDICAL INFORMATICS PLATFORM

EXECUTIVE SUMMARY

THE MEDICAL INFORMATICS PLATFORM
MIP DEPLOYMENT EXECUTIVE SUMMARY

Purpose
The present document outlines the primary issues related to the deployment of the Medical Informatics Platform (MIP) into hospitals participating to the MIP network. It provides the overall context, short description of the MIP and its functionalities, followed by an overview of deployment activities and responsibilities.

Introduction
Brain diseases, considered as a whole, affect 165 million European citizens, a large number of whom are being at least partly managed in hospitals. The clinical data collected from these patients represent a unique source of information for better understanding and treating brain diseases but are unfortunately not usually available for research.

The MIP has been developed by the Human Brain Project (HBP), a EU Horizon 2020 Flagship project, in order to facilitate access to clinical data stored in hospitals, for research purpose, while preserving data privacy. The MIP aims at enabling breakthrough medical progress in the field of brain diseases through access to an unprecedented volume of patients' data.

MIP description
The MIP is an innovative data analysis and data collection system that provides an interface for various investigators (clinicians, neuroscientists, epidemiologists, researchers, health managers) enabling them to access and analyze anonymized medical data currently locked in hospitals, without moving the data from the hospital where they reside, and without infringing on patient privacy.

The MIP is designed to help clinicians and researchers aiming to adopt advance analytics for diagnosis and research in clinics and to promote collaborative neuroscience research using distributed hospital data.

The MIP is divided into two main components, called MIP-Local and MIP-Federated Node, which shall be installed on different servers within the participating hospitals. MIP-Local contains pseudonymised data that can only be accessed and analyzed by the Local Data Coordinator and its accredited staff from within the hospital. MIP-Federated Node contains anonymized data and can be connected to other MIP-Federated Nodes in other hospitals. Upon signed agreement between data providers from the MIP network, accredited investigators can query multiple MIP federated nodes and obtain aggregate results. Queries of the MIP-Federated Nodes do not allow to copy or upload any data, nor to see individual patient's data.
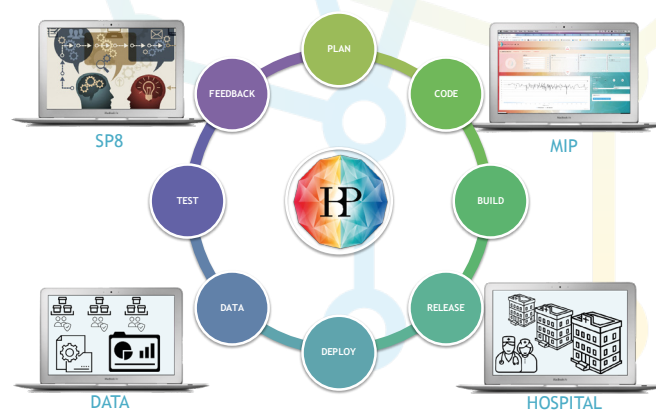
The 2-tier MIP architecture (MIP-Local, MIP-Federated Node) has been designed to address the specific challenges of:

1) local deployment adapted to each hospitals' environment,

2) capturing and processing heterogeneous type of data (e.g. socio-demographic, clinical, biological and neuroimaging data),

3) fulfilling privacy rules, policies and best practices to enable efficient and secure data sharing,

4) harmonizing data through Common Data Elements for cross-site comparisons, and

5) integrating readily available statistical and machine learning tools.

During the two completed phases of HBP, the MIP has been developed and installed in an increasing number of participating hospitals. The Lausanne university hospital (CHUV) is the HBP partner coordinating this activity

Key benefits to participate in the MIP

- Participate in the largest ever funded Europe-wide brain research initiative;
- Train and use novel state-of-the-art analytical tools, include machine learning algorithms;
- Investigate and discover novel findings from its own data using the MIP-Local;
- Participate or lead Federated analyses on big data available in the network of MIP-Federated Nodes
- Develop new scientific collaborations
- Increase the chance of future successful national or European competitive grant applications

MIP Functionalities

FUNCTIONALITY OF THE MEDICAL INFORMATICS PLATFORM	
Components	Description
Data Capture	Set of tools allowing to capture patient data from hospital's information systems, and their de-identification
Data Factory	Configurable tool chain for pre-processing neuroimaging/biomedical/demographic patients' data to extract relevant features, harmonize and map them to the MIP standard schema, as well as anonymize the resulting database.
Algorithm Library	Standard catalogue of centralised statistical methods and predictive machine learning algorithms run by the Algorithm Factory. They range from classical ones (such as linear regression, Anova, k-nearest neighbour and other naive Bayes) to more sophisticated ones (such as predictive clustering trees).
Algorithm Factory	Framework used to execute the algorithms from the Algorithm Library, as well as to train predictive models and test them using a cross-validation schema.
Web Portal	The web-based interface to offer end-user access to MIP functionalities (see list below).
Development (SDK)	Tools for integration of additional methods and algorithms, based on Docker technology.
End-user functionality	Description
Private/Public Workspace	Every user gets an account. Once authenticated he/she can access his/her private workspace. A public workspace is also provided to users to share content among the community members. The content includes data models, experiments, articles and protocols.
Variable Exploration	Interactive exploration of available metadata and data statistics.
Descriptive Statistics	Histograms of separate variables and interaction between them
Interactive Analytics and Model Creation	Creation and configuration of data models, which are sets of explanatory and predictive variables, including filters, which define filtered subsets of selected data. Analysis of variables' interactions based on this subset of cross-sectional data. Interactive visualisations for few selected variables by mean of design matrices and other plots. Data models can be saved for reuse and sharing.
Experiment Builder	Configuration and running of experiments consisting of executing algorithms on data defined in previously configured data models. Unique execution of an algorithm is possible, as well as repetitive execution of several algorithms for a cross-validation using k-random schema or using partition based on the provenance. Experiments can be saved and shared between users.
Article writing	Creation of articles where users can include results of experiments or descriptive statistics and interactive analysis plots. Articles can be saved and shared with peers.

MIP governance and management

As delineated in the next section, installation and running of the MIP in Hospitals require good collaboration between several stakeholders defined below:

1. <u>MIP-CHUV deployment team</u>: This team, based at CHUV (Lausanne), has been responsible for MIP development and now ensures its deployment, maintenance and debugging when necessary. It includes the Head of the Department of Clinical Neuroscience in charge of the MIP project, the vice-director of the CHUV IT Department, as well as project managers and IT engineers involved in MIP development and maintenance, one of whom is appointed as the MIP-CHUV Deployment Manager.

2. <u>Hospital Deployment Team</u>: This team, based in the Hospital where the MIP is to be deployed, includes at least one readily available representative from the clinical or research department aiming to use the MIP, referred to as the Hospital Deployment Manager, and one representative from the hospital IT department. Administrative and legal staff might also participate. The Hospital Deployment Manager shall be responsible for coordinating the work of the technical (IT) and medical research staff of his Hospital in close collaboration with the MIP-CHUV Deployment manager. In addition, a Local Data Coordinator shall be identified within the Hospital medical research team to supervise all activities in relation with capturing and analyzing data into MIP-Local or MIP-Federated node. The Hospital Deployment Manager and Local Data Coordinator might be the same or a different person. In case the hospital aims at using the MIP for different brain diseases, different Local Data Coordinators might be ascribed to each of these diseases.

3. <u>MIP Data Governance Steering Committee</u> (DGSC): MIP DGSC has been established in order to control all activities in relation with data analyses performed over MIP-Federated nodes and ensure that all Local Data Coordinators are duly involved in analyses that include their data. To this effect, DGSC is organized in disease-specific boards for each condition currently considered in the MIP network (i.e. dementia, epilepsy, traumatic brain injury, mental health, disorders of consciousness). Each MIP DGSC disease-specific board includes all Local Data Coordinators who have agreed to participate to federated analyses that involve their data, and have their hospital signed a MIP Data Sharing Agreement.
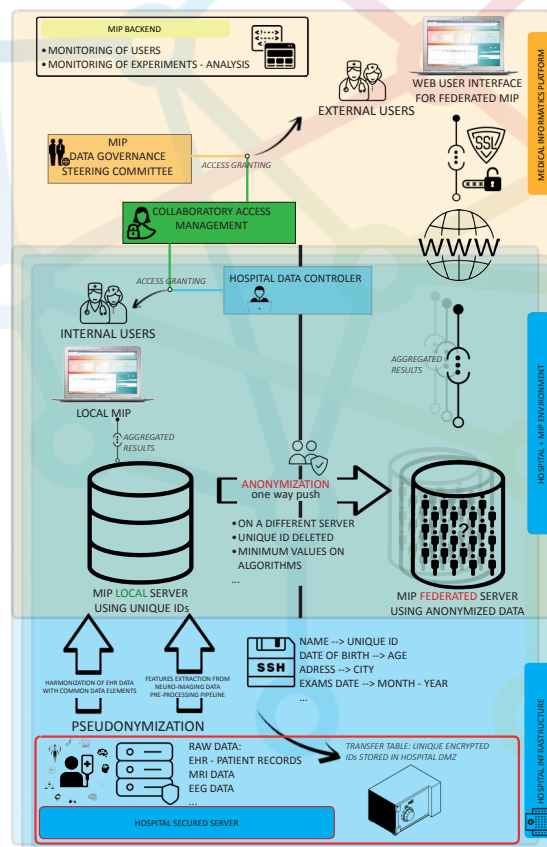
MIP deployment process

The MIP deployment process follows 12 steps, many of which are optional or can be deferred to a later stage. The rationale behind this incremental strategy is to start rapidly and easily, and progressively allow the hospital staff to gain experience along the implementation process.

The first four steps fall within the scope of the MIP installation agreement:

1. Identify all relevant hospital staff (Hospital Deployment Team) required to proceed to MIP deployment, including the Hospital Deployment manager, the Local Data Coordinator (s), as well as the MIP-CHUV Deployment Manager counterpart.

2. Secure signature of the MIP Installation Agreement by both the Hospital and CHUV legal representatives. This agreement only covers installation of the MIP software, and not data sharing.

3. Prepare the IT infrastructure (e.g. servers) needed to install the MIP-Local only or both the MIP-Local and MIP-Federated Node together, according to their specifications. This preparation is typically performed by the Hospital IT staff under the supervision of the Hospital Deployment Manager.

4. Install the MIP-Local software as a stand-alone component, or both the MIP-Local and MIP-Federated Node software together. The installation of the MIP-Federated Node component is thus optional and can be installed at a later stage. The above installations are usually being performed jointly by the Hospital IT staff and the MIP-CHUV deployment team under the co-supervision of the Hospital and MIP-CHUV Deployment Managers. Part of the installation can be performed remotely through VPN connection. On site hospital-specific tuning is however often required.

Steps 5 to 12 are optional, fall outside the scope of the MIP installation agreement, and can be deferred in time. Steps 6 to 12 require the signature of the Data Sharing Agreement:

5. Capture into MIP-Local pseudonymized patients' data from the hospital clinical or research department aiming to use the MIP, in full compliance with all local ethics and regulatory issues. This step is undertaken under the full responsibility of the above department and its Local Data Coordinator . Once data have been captured in MIP-Local, the Local Data Coordinator and his accredited local staff can use the MIP to analyse their data. No other stakeholder has access to the data.

6. Secure the signature of the MIP Data Sharing Agreement by both the Hospital and CHUV legal representatives. This agreement covers the possibility to perform federated analyses of fully anonymized data captured in the MIP-Federated node of the hospital.

7. Secure participation of the Local Data Coordinator to the relevant MIP DGSC disease-specific board.

8. Prepare the IT infrastructure (e.g. server) to install the MIP-Federated Node software if not already done during step 3 (same procedure as previously described).

9. Install the MIP-Federated Node software if not already done during step 4 (same procedure as previously described).

10. Proceed to full anonymization of the data stored in MIP-Local and then push these data into the MIP-Federated Node database. This step is undertaken under the full responsibility of the Local Data Coordinator .

11. Enable the link to the network of MIP-Federated Nodes from other hospitals. This link is enabled/disabled by the Hospital IT staff in charge of installing and controlling the MIP, under the supervision of the Hospital Deployment Manager.

12. Participate to federated analyses according to the procedures agreed upon within the relevant MIP DGSC disease-specific board. This activity is being performed by the Local Data Coordinator or its accredited staff and can only provide aggregated results with no possibility to copy or upload the database, nor that of exploring individual patient's data.

Contacts

HBP-SP8 leadership: philipperyvlinhbp@gmail.com
HBP-SP8 leadership: Sandra.Schweighauser@chuv.ch