SP8 MIP - Results for SGA2 Year 1 (D8.1.1 – SGA2)

Figure 1: MIP user interface including graphical elements modified in the latest release of the MIP front end

The user interface belongs to component no. 3056 - Web Exploration & Analytics, and specifically to the Portal Front End. It is described in Chapter 4.1

Project Number:	785907	Project Title:	Human Brain Project SGA2

Document Title:	SP8 MIP - Results for SGA2 Year 1
Document Filename:	D8.1.1 (D48.1 D23) M12 SGA2 ACCEPTED 190722
Deliverable Number:	SGA2 D8.1.1 (D48.1, D23)
Deliverable Type:	Report
Work Package(s):	WP8.1, WP8.2, WP8.3, WP8.4, WP8.5, WP8.7, WP8.8, WP8.9, WP8.10 WP8.6 does not contribute to this deliverable, but elaborated its own document as CDP6
Dissemination Level:	PU = Public
Planned Delivery Date:	SGA2 M12 / 31 Mar 2019
Actual Delivery Date:	SGA2 M14 / 14 May 2019; ACCEPTED 22 Jul 2019

Authors:	Philippe RYVLIN (P27)
Compiling Editors:	Jacek MANTHEY (P27)
Contributors:	Philippe RYVLIN, CHUV (P27), all chapters. David ATIENZA, EPFL (P1), ch.2,5,8. Florent GAILLARD, CHUV (P27), ch.3,4. Jacek MANTHEY (P27), ch.4,5,7. Erika BORCEL, CHUV (P27), ch.5. David STEINBERG, TAU (P57) , ch.8. Evidokia MAILI, UoA (P43), ch.2,3,5,8. Petra RITTER, CHARITE (P122), ch.2,7,8. Olivier DAVID, UGA (P125), ch.2,5,8. Gunter SCHUMANN, KCL (P38), ch.2,8. Sumit MADAN, FG (P22), ch.2,8.
SciTechCoord Review:	
Editorial Review:	Annemieke MICHELS, EPFL (P1)

Description in GA:	For consistent presentation of HBP results, SGA2 M12 Deliverables describing the accomplishments of an entire SP or CDP have been prepared according to a standard template, which focuses on Key Results and the outputs that contribute to them. Project management elements such as Milestones and Risks will be covered, as per normal practice, in the SGA2 Year 1 Report.
Abstract:	This Deliverable is the annual compound of SP8 'outputs' organised by SP8's Key Results. The main outputs are the continuation of the consolidation and deployment in hospitals of the Medical Informatics Platform, according to SP8 work plan, as well as that of the dementia use-case. Preliminary outputs from new Work Packages selected through Calls of Expression of Interest are available.

Keywords:	SP8, Medical Informatics Platform, SGA2, ethics, MIP deployment, MIP maintenance, MIP algorithms, privacy-aware, iEEG, the Virtual Brain, ontologies, quantitative neurobehavioural, IMAGEN, dementia, use cases.
Target Users/Readers:	Neuroscientists, brain researchers, data scientists, neuro-clinicians, companies and other potential users of HBP results

Table of Contents

Table of Tables

Table of Figures

1. Overview

The Medical Informatics Platform (MIP) aims at leveraging the capacities of medical researchers to share patients' data in a way enabling access to very large volume of information (Big Data) while fully preserving data privacy. To this end, the MIP is offering a unique IT solution whereby patients' data always stay in the hospital where they were originally collected (i.e. never copied, sent or uploaded on a server outside the hospital or in the cloud) and where they can be queried through so-called federated analyses. Federated analyses consist in sending the same query to different patients' datasets located in different hospitals, simultaneously, in order to obtain a result reflecting all these datasets. To this purpose, the MIP is installed in each participating hospital where it manages the curation and analyses of the local dataset while being connected to MIP from other hospitals in order to deliver federated analyses.

In the first year of SGA2, the MIP has been consolidated by the different teams in charge of its development. In parallel, an ambitious deployment plan was undertaken with a MIP currently installed in 16 hospitals, while another 11 hospitals have signed a MIP installed agreement and 30 more have declared being interested. In particular, a partnership was signed with the European Reference Network EpiCARE, a network of 28 major hospitals specialised in rare and complex epilepsies.

The first use case developed using the MIP has focused on Alzheimer's disease and dementia. Preliminary results on neuroimaging-based biomarkers (MRI) have been recently obtained and support the value of the MIP for testing novel biomarkers in hospitals. Other use cases have been recently launched in the fields of Epilepsy, Traumatic Brain Injury and Mental Health, including novel projects selected in 2018 through open-calls.

2. Introduction

According to the DoA, SP8 has undergone major changes in SGA2, primarily characterised by:

- a hand-over of MIP development from CHUV to UoA (WP8.5)
- a shift of CHUV activities towards an ambitious MIP deployment plan (WP8.2/WP8.3)
- integration of new activities: CDP6 (WP8.6) and 4 open calls (WPs 8.7 to 8.10)

1) **WP8.1** (coordination and communication, CHUV)

The reorganisation of SP8 activities between partners as well as the integration of many new partners has required significant coordination efforts. Accordingly, CHUV has gathered partners of the SP8 consortium at four two-day workshops in Geneva and Lausanne, as well as at two single-day meetings in Maastricht and Lyon. Staff from the CHUV IT team also visited UoA and AUEB in Athens twice to facilitate MIP hand-over, and they have organised weekly TCs.

As part of this WP, UoA has ensured technical coordination of SP8 (T8.1.2), while CHUV thoroughly tackled all ethics and data privacy issues (T8.1.4, KR8.1).

2) **WP8.2** (MIP deployment, operation and maintenance, CHUV)

The CHUV IT team has debugged and upgraded the seven previously installed MIP instances and installed nine more instances. Several MIP upgrades were delivered that contributed to KR8.1, KR8.2 and KR8.3.

3) **WP8.3 (**Coordination of clinical use cases and ontologies, CHUV, EPFL, UBO)

A very active community building was undertaken, resulting in the addition the MIP to 9 new hospitals. 10 more hospitals have already signed an installation agreement, which contributed to KR8.2 and KR8.3. The use case on dementia has continued to develop across several of these hospitals. A pilot study aiming at introducing data from wearable devices into the MIP was performed. Many contacts were laid with various stakeholders to develop a business model towards a public-private partnership (PPP) ensuring a sustainable MIP.

4) **WP8.5** (MIP Development, Enhancement, and Robustisation, UoA, AUEB)

An automated anonymisation tool was developed that contributed to KR8.1. A Data catalogue component and a new QC tool were developed. Scientific Workflow Engine (Galaxy) was integrated with Federated Data Processing Engine, allowing via a visual interface the construction, management, and execution of complex scientific workflows. New machine learning algorithms were added into the Federated Data Processing Engine.

5) **WP8.6** (Modelling for drug discovery, IP)

This WP corresponds to CDP6 which outputs are reported in D8.6.1, the CDP6 compound Deliverable.

6) **WP8.7 to WP8.10**

These WPs correspond to the projects selected through the 2018 SP8 Calls of Expression of Interest (CEoIs). Due to administrative reasons unrelated to SP8 activities, these WPs could only start in November 2018. Nevertheless, they contributed significantly to KR8.2, KR8.3 and KR8.5. WP8.7 (Charité) developed personalised "The Virtual Brain" models of neurodegenerative and ageing brains which contributed to KR8.5. WP8.7 describes its results in the CDP8 M12 Periodic Project Report. WP8.8 (UGA) has developed a new sharable data format for Human intracerebral EEG (iEEG) within the framework of the international BIDS specification (Brain Imaging Data Structure). BIDS-iEEG will facilitate integration with other modalities such as fMRI, MEG, and EEG and iEEG data sharing using the MIP. WP8.9 (Fraunhofer) has been working on extending MIP's infrastructure with a comprehensive framework of high quality brain disease ontologies with application services and two components to be released in May 2019. WP8.10 (KCL, UKAachen) has developed a new method to find relations between behavioural symptoms, and neuroimaging measures of brain structure and function in the longitudinal IMAGEN-study. Their original findings provide a framework for developing a classification of psychiatric illness based on quantitative neurobehavioural measures.

3. Key Result KR8.1 MIP infrastructure and operational activities comply with EU ethics and data privacy/security regulatory requirements

3.1 Outputs

3.1.1 Overview of Outputs

1) Output 1: Ethics and Data Governance (CHUV, P27)
2) Output 2: Anonymisation module (AUEB, P4) – C3088
3) Output 3: New privacy aware algorithms for federated analysis (UoA, P43) – C3000

Table 1: Overview of releases and major updates related to Key Result KR8.1

ID	Component Name	Type	Contact	Info on releases and major updates
No C ID	Guidelines and requirements for Anonymisation in the MIP	Other	F. GAILLARD (CHUV, P27)	Included in D12.4.8.
No C ID	Installation and License agreement	Other	F. GAILLARD CHUV, P27)	Produced in collaboration with a specialised lawyer
No C ID	Data Sharing Agreement	Other	F. GAILLARD (CHUV, P27)	Produced in collaboration with a specialised lawyer
No C ID	Ethics and Legal Requirements - documents for hospitals	Other	F. GAILLARD (CHUV, P27)	https://hbpmedical.github.io/deployment-pack/4-MIP%20Ethics%20and%20Legal%20Requirements.pdf
No C ID	Results of the first DPIA	Other	F. GAILLARD (CHUV, P27)	Already available to the EC, provided to the EC by email in answer to the questions from the reviewers
No C ID	Organisation of the MIP Data Governance Steering Committee and the Pathology-specific sub-committees	Other	F. GAILLARD (CHUV, P27)	Terms of Reference of the MIP Data Governance Steering - document already available to its members and currently being reviewed and edited by them. Charter of the MIP Data Governance Steering – is available as milestone MS8.3.3.
C3088	Anonymisation for the federation	Software	Vasileios VASSALOS (AUEB, P4)	Release Name: Data anonymisation Release date: 31 Mar 2019 URL: https://github.com/aueb-wim/anonymization-4-federation
C3000	Federated / distributed	Software	Yannis IOANNIDIS (UoA, P43)	Release Name: v.18 Release date: 15 Feb 2019

	data processing engine			URL: https://github.com/madgik/exareme/releases/tag/v18 https://github.com/madgik/mip-algorithms

3.1.2 Output 1 – Ethics and Data Governance (CHUV, P27)

The following work relative to this output has been done:

- Guidelines and requirements for Anonymisation in the MIP were written
- Installation and License agreement was produced in collaboration with a specialised lawyer
- Data Sharing Agreement was produced in collaboration with a specialised lawyer
- The document on Ethics and Legal Requirements for hospitals was produced
- First DPIA was completed
- Organisation of the MIP Data Governance Steering Committee and the Pathology-specific sub-committees was defined and work was launched.

3.1.3 Output 2 - Anonymisation module (AUEB)

This new software module (C3088), allowing that analyses performed through the MIP federate network will only have access to fully anonymised datasets, has been developed and provided for integration. It enables hospitals that want to share their data through their MIP federation node to anonymise these data in order to comply with the privacy standards set by the HBP Ethics committee in coordination with SP8. Data in hospitals will reside on two databases, a pseudonymised one for local hospital use, and an anonymised one containing a fully anonymised version of the pseudonymised datasets. On updates of datasets, a fully anonymised new database will be produced to replace the current one. The different versions of the anonymised database will be kept using a version control system to track provenance.

3.1.4 Output 3 - New privacy aware algorithms for federated analysis (UoA)

The Federated Data Processing Engine Exareme (C3000) was further enhanced with new user-defined functions in order to support the implementation of novel federated algorithms/workflows, as well as the integration of algorithms developed by other SP8 Partners. New algorithms are Pearson correlation, Anova (testing phase), id3 (testing phase), K-means accurate (testing phase).

3.2 Validation and Impact

All ethics-related outputs described in 3.1.1 were validated with appropriate ethics bodies (including HBP ethics board) and specialised lawyers. The produced documents have enabled developing an appropriate framework for the management of ethics and data privacy issues with the MIP (e.g. development of the anonymisation module) and have facilitated the deployment of the MIP in new hospitals.

The anonymisation module now enables to perform MIP federated analyses of anonymised patients' data in full compliance with EU regulations, given the fact that such anonymised data are not covered by GDPR.

Analytical tools for federated analysis are implemented. Testing is conducted based on specifications provided by CHUV and TAU, so that they are compliant with the privacy requirements described in Deliverables D8.5.1 and D12.4.8.

3.2.1 Actual Use of Output(s) / Exploitation

Output 1: The MIP Installation and License agreement, together with the Ethics and Legal Requirements document, is currently being used successfully to get hospitals effectively installing the MIP. The DPIA was sent for opinion to CNIL (Commission Nationale d'Informatique et des Libertés) and was evaluated as very relevant. The guidelines and requirements for Anonymisation in the MIP was used by AUEB to develop the anonymisation module.

Output 2 and 3: These outputs are currently being tested and integrated.

3.2.2 Potential Use of Output(s)

Output 1: The current MIP deployment phase is focusing on the installation of MIP local as a first step for engaging hospitals. The next phase will make use of the data sharing agreement and the anonymisation module to install the MIP federated node and proceed to federated analyses. This in turn will allow for the implementation of important Use Cases, such as advanced phenotyping of the ageing brain cognitive diseases at early stage, and discovery of novel disease models and their application to Parkinson's Disease.

Output 2 and 3: These outputs will be used in the future releases of the MIP.

3.2.3 Publications

None.

3.2.4 Measures to Increase Impact of Output(s): Dissemination

The documents from Output 1, in particular Ethics requirements and agreements, are constantly being disseminated among hospitals in the form of the Deployment Pack. A presentation on these requirements was done during the HBP summit, for the Data Governance session. Another presentation is planned during the Symposium in April 2019 at CHUV, and a lecture in September 2019 in Graz, as part of a student workshop.

We always plan to actively communicate on the progresses made on MIP compliance with ethics and data privacy regulations, to expand adoption of the MIP by potential end-users.

4. Key Result KR8.2 MIP is operated over a large network of European Hospitals (≥ 30)

4.1 Outputs

4.1.1 Overview of Outputs

1) Output 1: Data Governance Steering Committee (CHUV, P27)
2) Output 2: MIP operations and maintenance (CHUV, P27)
3) Output 3: MIP software upgrades (CHUV, P27)

Table 2: Overview of releases and major updates related to Key Result KR8.2

C ID	Component Name	Type	Contact	Info on releases and major updates
No C ID	QA environment maintenance	service	M. NASUTI (CHUV, P27)	Cluster for the central node and two worker nodes Release name: 2019.02.25 Release: 2.8.5 Release date: 25 Feb 2019
No C ID	Improved security on the servers	other	M. NASUTI (CHUV, P27)	Firewall rules, fail2ban, restricted SSH access, etc Release name: 2019.03.04 Release: 2019.03.04 Release date : 4 Mar 2019
No C ID	MIP instances upgrades	service	M. NASUTI (CHUV, P27)	Latest stable features of the MIP Release name: continuous work Release: 2.8.5 Release date: continuous work URL: N/A
No C ID	API test scripts	software	M. SPUHLER (CHUV, P27)	Input and output model and algorithm API test. Preversion. To be integrated fully in Front end test suite Release name: 2018.11.27 Release: 2018.11.27 Release date: 27 Nov 2018 URL: https://github.com/HBPMedical/mip-api-tests
3056 - Web Exploration & Analytics	New version of Portal Front end	software	M. SPUHLER (CHUV, P27)	Robustisation, error handling, normalised data structure. Migration performed in anticipation of complex visualisation and third-party integrations. Release name: 2019.02.07 Release: 2.15.0 Release date: 7 Feb 2019 URL: https://github.com/HBPMedical/portal-frontend/archive/2.15.0.zip
3056 - Web Exploration & Analytics	New release of Portal Back end	software	L. CLAUDE (CHUV, P27)	Release highlights: stabilisation, error recovery and internal health checks Release name: 2019.02.20 Release: 2.9.3 Release date: 20 Feb 2019 URL:

				https://github.com/LREN-CHUV/portal-backend/releases/tag/2.9.3
2938 Algorithm Orchestrator	New release of Algorithm Factory (Woken)	software	L. CLAUDE (CHUV, P27)	Release highlights: new distributed algorithms (PCA, k-means), updates to existing algorithms. Fixes on k-fold validation, numerous bug fixes and stabilisation, error recovery and internal health checks, support for multiple tables Detailed release notes: https://github.com/LREN-CHUV/woken/blob/master/CHANGELOG.md Release name: 2019.02.22 Release: 2.8.5 Release date: 22 Feb 2019 URL: https://github.com/HBPMedical/woken/releases/tag/2.9.3 https://github.com/LREN-CHUV/woken-messages/releases/tag/2.9.5
645 Cross-validation module	New release of Algorithm Factory (Woken validation part)	software	L. CLAUDE (CHUV, P27)	Release highlights: stabilisation, error recovery and internal health checks Detailed release notes: https://github.com/HBPMedical/woken-validation/blob/master/CHANGELOG.md Release name: 2019.02.22 Release: 2.6.4 Release date: 22 Feb 2019 URL: https://github.com/HBPMedical/woken-validation/releases/tag/2.6.4
669 Common Data Elements database	Updates to MIP Common Data Elements and their taxonomy.	other	L. CLAUDE, M. NASUTI (CHUV, P27)	Updates to MIP Common Data Elements and their taxonomy URL: https://github.com/LREN-CHUV/meta-db-setup/releases/tag/2.4.2 https://github.com/LREN-CHUV/mip-cde-meta-db-setup/releases/tag/1.3.3
647 Algorithm Repository	Integration of t-SNE algo visualisation on front end	software	M. SPUHLER (CHUV, P27)	Packaged t-SNE. Algorithm rules in Woken. Front end integration, input form, visualisation output Release name: 2018.12.13 Release: 0.4.3 Release date: 13 Dec 2018 URL: https://github.com/LREN-CHUV/algorithm-repository/tree/master/python-tsne

4.1.2 Output 1 – Data Governance Steering Committee (CHUV, P27)

The Data Governance Steering Committee (DGSC) (T8.3.1) and its charter were defined and launched. They are mandatory to get the different medical and research centres which compose the MIP network accepting to contribute to data sharing through the MIP.

4.1.3 Output 2 – MIP operations and maintenance (CHUV, P27)

The following WP8.2 outputs were delivered for MIP operations and maintenance:

- QA environment maintenance
- Improved security on the servers
- MIP instances upgrades
- API test scripts

More details are provided in Table 2.

4.1.4 Output 3 – MIP software upgrades (CHUV, P27)

The following upgrades of the MIP software modules were delivered by WP8.2 to strengthen MIP robustness:

- Portal
- Algorithm Orchestrator
- Cross-validation module
- Common Data Elements database
- Algorithm Repository

More details are provided in Table 2.

4.2 Validation and Impact

The MIP DGSC charter was validated by its members. Its impact is a reinforcement of the transparency and trust for data sharing with the MIP.

The MIP operations and software upgrades outputs were validated according to the usual test/operational procedures. The impact of these upgrade is an improvement in the quality of MIP operations.

4.2.1 Actual Use of Output(s)

Output 1: The MIP DGSC and its pathology-specific sub-committees are operational and hold meetings.

Output 2, 3: The MIP operations and software upgrades outputs are being used in the MIP instances.

4.2.2 Potential Use of Output(s)

Output 1: The MIP DGSC will continue organising regular meetings in order to contribute to MIP operations and extensions.

Output 2, 3: The MIP operations and software upgrades outputs will continue to be used in the future MIP instances.

4.2.3 Publications

None.

4.2.4 Measures to Increase Impact of Output(s): disseminations

Numerous meetings were held with clinicians from hospitals to elaborate the charter of the DGSC. Clinicians from additional hospitals are informed about DGSC and its sub-committees and invited to join.–The MIP DGSC will continue organising regular meetings in order to contribute to MIP operations and extensions.

Hospitals' IT contacts are informed about MIP maintenance and upgrade activities.

The MIP DGSC will continue organising regular meetings in order to contribute to MIP operations and extensions.

5. Key Result KR8.3 Established large-scale network of MIP-data providers, including clinical departments and research consortia, collating data from more than 30,000 patients with various brain diseases

5.1 Outputs

5.1.1 Overview of Outputs

The outputs of KR9.3 are:

1) Output 1: Development and dissemination of MIP deployment package (CHUV, P27)
2) Output 2: Partnership with the European Reference Network (ERN) EpiCARE (CHUV, P27)
3) Output 3: Partnership with the EU-Funded TBI cohorts (CHUV, P27)
4) Output 4: Additional hospitals in the MIP network (CHUV, P27)
5) Output 5: Improvements in the deployment procedures of the MIP (CHUV, P27)
6) Output 6: Shareable data format for intracranial EEG (UGA, P125)
7) Output 7: Data Transfer Agreement for iEEG data from F-TRACT project (UGA, P125)
8) Output 8: Clinical Data Catalogue (AUEB, P4) - C2990
9) Output 9: Data Cleansing components (AUEB, P4) – C2996
10) Output 10: Android mobile app for acquisition of multimodal data in outpatient settings (EPFL, P1)
11) Output 11: SVM resource-constraining methods (EPFL, P1)
12) Output 12 – Analysis results of the IMAGEN cohort (KCL, P38)

Output 13 – Protocol for the Systematic Assessment of Novel Types of Data Integration into the MIP (EPFL, P1)

Table 3: Overview of releases and major updates related to Key Result KR8.3

C ID	Component Name	Type	Contact	Info on releases and major updates
No C ID	MIP Deployment Pack - various documents	Other	Erika BORCEL (CHUV - P27)	Release name: continuous updates URL: https://hbpmedical.github.io/deployment-pack/ Release date: continuous releases
No C ID	BIDS-iEEG	Other (data format)	Olivier David (UGA - P125)	Release name: 0.2 URL: http://bids.neuroimaging.io https://psyarxiv.com/r7vc2/ Release date: 14 Dec 2018
No C ID	Data Transfer Agreement (Convention de mise à	Other	Olivier David (UGA - P125)	Release Name: V1 URL: document already provided to Inserm and UGA; not publicly accessible

	disposition des données)			
C2990	Clinical Data Catalogue	Software	Vasileios VASSALOS (AUEB - P4)	Release Name: Data Catalogue Release date: 22 Feb 2019 URL: https://github.com/HBPMedical/DataCatalogue
C2996	Data cleansing component	Software	Vasileios VASSALOS (AUEB - P4)	Release Name: QC tools Release date: 22 Feb 2019 URL: https://github.com/HBPMedical/DataQualityControlTool
No C ID	Android app for the acquisition of multimodal data in outpatient settings	Software	David ATIENZA (EPFL, P1)	Release Name: lab release 2019.02. URL: https://c4science.ch/diffusion/7361/grandjean-stressmonitoring.git Release Date: Feb 2019
No C ID	New methods for tailoring the resource consumption of a machine learning algorithm to wearable devices (SVM resource-constraining methods).	Other	David ATIENZA (EPFL, P1)	URL: https://infoscience.epfl.ch/record/261315/files/DATE19-SVM.pdf Release Date: Feb 2019
C102 MIP Microservice Infrastructure	New release of Deployment scripts for MIP platform	software	Ludovic CLAUDE (CHUV - P27)	Release name: 2018.05.23 Release: 2.8.5 Release date: 23 May 2018 URL: https://github.com/LREN-CHUV/mip-microservices-infrastructure/releases/tag/2.8.5
C671 Omics pipeline for feature engineering	New release of the hierarchizer	Software	Mirco NASUTI (CHUV - P27)	Support new data formats Release name: 07.08.2018 Release: 1.3.8 Release date: 07 Aug 2018 URL : https://github.com/HBPMedical/hierarchizer

5.1.2 Output 1 - Development and dissemination of MIP deployment package (CHUV, P27)

We have developed and disseminated to all MIP stakeholders a comprehensive MIP Deployment Pack which contains all relevant documents including MIP installation agreement, data sharing agreement, Ethics and data privacy, IT requirements and instructions on how to install the MIP, MIP user manual, and the detailed IT description of the MIP components.

The executive summary of the MIP Deployment Pack provides a precise roadmap describing the different administrative and technical steps that need to be undertaken for installing a new MIP

instance. Indicators from this roadmap are being used to track the level of advancement of MIP installation in every new hospital on a weekly basis.

5.1.3 Output 2 – Partnership with the European Reference Network (ERN) EpiCARE (CHUV, P27)

A partnership between HBP and DG Sanco funded ERN EpiCARE for rare and complex epilepsies has been approved by the two parties. EpiCARE being a consortium of 28 major EU University hospitals, we have requested all 28 centres to further endorse the partnership. 24 of these hospitals have now signed the corresponding memorandum of understanding (MoU), while the 4 pending shall sign it in the near future. The implementation of the MIP in EpiCARE hospitals is supported by a specific voucher obtained through the 2018 HBP voucher call.

5.1.4 Output 3 – Partnership with the EU-Funded TBI cohorts (CHUV, P27)

The 2018 HBP voucher call has provided support for deploying the MIP in the two large TBI (traumatic brain injury) cohorts funded by DG research (CREACTIVE and Center-TBI). A MIP instance has been installed at the CREACTIVE center (Bergamo), while that at Center-TBI (Karolinska Institute, Stockholm) shall follow in year 2.

5.1.5 Output 4 – Additional hospitals in the MIP network (CHUV, P27)

According to the three above outputs and MIP dissemination strategy, a total of 55 hospitals were contacted across Europe and 1 in South America to install the MIP. 47 of these hospitals have expressed their interest in collaborating on the MIP with HBP SP8. 26 have already signed a MIP Installation Agreement, 9 of which have installed a MIP.

Novel MIP instances have been installed in the following institutions:

- Centre for Health Technologies, University of Pavia, Italy
- IRCSS Neurological Institute Carlo Besta, Milan, Italy
- St. Anne's Hospital, Brno, Czech Republic
- Motol university hospital, Prague, Czech Republic
- Danish epilepsy Filadelfia, Dianalund, Denmark
- Hospital del Mar, Barcelona, Spain
- Institute Mario Negri, Bergamo, Italy
- Sahlgrenska University Hospital, Gothenburg, Sweden
- University of Grenoble, Grenoble, France.

5.1.6 Output 5 – Improvements in the deployment procedures of the MIP (CHUV, P27)

The following outputs which improve the MIP deployment IT procedures were produced:

- New release of Deployment scripts for MIP Platform (C102)
- New release of the hierarchizer (C671)

5.1.7 Output 6 - Shareable data format for intracranial EEG (UGA, P125)

A sharable data format for Human intracranial EEG (iEEG), which facilitates data sharing between hospitals was defined and published.

This format is based on the BIDS (Brain Imaging Data Sharing, http://bids.neuroimaging.io) initiative already developed for MRI and MEG. An international group of experts of intracranial data, to which UGA and UCBL belong, defined BIDS for iEEG with full compatibility for WP8.8 objectives.

A BIDS-iEEG extractor of the F-TRACT database has been developed by AMU and UGA. An initial data set of 16 patients included in F-TRACT has been converted into BIDS-iEEG and transferred by UGA to UKLFR and AMU. UKLFR and AMU are validating the structure of the dataset by UGA and will test their software to detect interictal events and compare their respective metrics for epileptic circuits characterisation.

5.1.8 Output 7 – Data Transfer Agreement for iEEG data from F-TRACT project (UGA, P125)

The Data Transfer Agreement (DTA, in French: “Convention de mise à disposition de données”) has been signed by Inserm and UGA on 18 March 18 2019.

It concerns the access to the intracranial recording and stimulation data of 300 patients of the F-TRACT cohort, coming from the 5 University hospitals (Grenoble, Lyon, Marseille, Milan, Freiburg) involved in WP8.8. The data have been gathered by the F-TRACT project in Grenoble under an Inserm IRB (14-140). The DTA specifies that UGA has access to the data for the duration of the HBP, and can give access to the data to WP8.8 partners, only for achieving the goals of the project. To allow gathering iEEG data during the whole duration of HBP, Inserm has amended, on 13 February 2019, the end date of the F-TRACT IRB to the end date of HBP.

Following the DTA signature, UGA has prepared a first dataset at the end of March 2019 and distributed it to UKLFR and AMU.

5.1.9 Output 8 - Clinical Data Catalogue (AUEB, P4)

The Clinical Data Catalogue (C2990) has been defined and developed. It is now in the test phase. It provides descriptive information (metadata) to MIP users. MIP users will be able to search for variables, read variables metadata from the GUI, download hospital variables metadata, explore visually the local variables-to-CDEs transformations, and choose between versions. The MIP Data Factory team is responsible for inserting variables for a new hospital, inserting/editing/deleting variables in an existing hospital, and managing metadata versions.

5.1.10 Output 9 - Data Cleansing components (AUEB, P4)

Quality Control tools (C2996) for outliers and error detection have been developed to perform quality checks of hospital data. The profiling tools generate statistical reports for the input datasets, both for tabular data and for images metadata (DICOM).

5.1.11 Output 10 – Android mobile app for acquisition of multimodal data in outpatient settings (EPFL, P1)

In the scope of the systematic assessment for the integration of new types of data into the MIP, we defined a pilot study on the acquisition, transmission, storage, feature extraction and incorporation

of data from different wearable devices into the MIP. In this way, the resulting reports C2975 (*In-depth assessment of potential new data integration into the MIP*) and C2976 (*Recommendations for the MIP technical development during SGA3*) will benefit from a quantitative evaluation of the specific technologies that are used in the pilot study.

This pilot study involves the development of a mobile app that will capture data from different commercial wearable devices (currently the Microsoft Band and the Empatica E4) and store them securely in an Oracle cloud infrastructure at CHUV. This application has been developed and is available. Then, a feature extraction pipeline will be developed to include the target biomarkers into the MIP. These biomarkers will be focused on epilepsy monitoring, since it is a well-known use case for wearable devices, and we recently obtained some promising results in epilepsy detection with wearable-EEG data.

5.1.12 Output 11 - SVM resource-constraining methods (EPFL, P1)

Research was done on the number of optimisations that can be made to the SVM (Support Vercot Machine) training algorithm to allow a direct execution on wearable devices, with the aim of distributing the workload of training personalised models for epilepsy detection. A paper is published on this work.

5.1.13 Output 12 – Analysis results of the IMAGEN cohort (KCL, P38)

The results of the IMAGEN cohort have been analysed. They provide a framework for the identification of symptom clusters based on shared biology, thus generating the basis for a new psychiatric nosology that is based upon quantifiable neurobiological measures, as opposed to the biologically heterogeneous categorical classifications used today. A related publication is under revision in Nature Neuroscience.

5.1.14 Output 13 – Protocol for the Systematic Assessment of Novel Types of Data Integration into the MIP (EPFL, P1)

The proposed protocol identifies a set of tasks, the reference partners and resources required to accomplish the integration of three types of data in MIP (omics, brain connectivity, sensory), as well specific challenges together with the expected outcomes of each task. This output contributes to components C2975 and C2976.

5.2 Validation and Impact

Regarding the MIP deployment pack, all contracts (MIP installation agreement and data sharing agreement) were validated by lawyers with expertise in the field (from CHUV and a private law firm). The Ethics and data privacy section was reviewed and validated by SP12 ethics support.

The upgrade of MIP deployment script and hierarchizer were validated by testing.

The impact of the reshuffled MIP deployment procedures and strategy (from selection of and communication with new hospitals, to the release of novel documentation and upgraded installation script) have enabled an effective fast track deployment in line with the DoA.

The SVM resource constraining methods are validated by peers, further validation is planned by journal reviewers.

5.2.1 Actual Use of Output(s)

All outputs relating to MIP deployment are used for the installation of all new MIP instances.

The model of iEEG data format has been published on Bioarchives (https://psyarxiv.com/r7vc2/) on December 14, 2018. It has been downloaded 265 times. This preprint is also accessible on ResearchGate (https://www.researchgate.net/publication/329637653_BIDS-iEEG_an_extension_to_the_brain_imaging_data_structure_BIDS_specification_for_human_intracranial_electrophysiology). It counts 352 reads and 5 recommendations, which corresponds to a Research Interest score of 14.6 (higher than 86% of research items on ResearchGate). This report in its final form is under minor revision at Nature Scientific Data (initial submission on January 29, 2019; revision scheduled to be submitted on April 8, 2019). Potential Use of Output(s)

The Android app is being used in tests.

5.2.2 Potential Use of Output(s)

All outputs relating to MIP deployment will continue to be used for the installation of future MIP instances. MIP deployment will enable to increase the number of patients' records available in the MIP to peform the use-cases planned in SP8-SGA2.

The BIDS-iEEG format will be used for all future Human intracranial EEG use-cases developed in HBP. It will be first used by WP8.8 partners for upgrading research software to make them compatible with the BIDS-iEEG format. A BIDS-manager program will be developed by AMU and UGA to make the data analysis software solutions interoperable on the same data structure (results saved as BIDS-iEEG derivatives), which will constitute the core of SEEG MIP. This initial version of SEEG MIP is under deployment in a beta test mode on a local hospital network at CHUGA.

The Android app will serve a double purpose: on the one hand, it will help to define the interfaces that will be needed to integrate the data from wearable devices into the MIP, thus helping to analyse the architecture changes that will be necessary to include new types of data into the MIP. On the other hand, it will be used for the obtaining data from actual patients and provide a valuable tool for the acquisition of multiple physiological parameters in a continuous manner and in outpatient settings, with the potential to increase to a large extent both the number of patients and the amount of data available for federated analysis. Quantitative evaluation of the specific technologies that are used in the pilot study will benefit the resulting reports C2975 (*In-depth assessment of potential new data integration into the MIP*) and C2976 (*Recommendations for the MIP technical development during SGA3*).

Regarding the SVM resource-constraining methods, they may influence the architecture design for the incorporation of data from wearable devices. Some computing tasks might be revealed as more efficient if they are executed in the wearables, so adopting an edge computing model for some tasks can be a good option to consider.

5.2.3 Publications

This is a selection of the publications for this KR. The complete list will be attached to the report.

Noteworthy, 2 papers are under revision at Nature Scientific data and Nature Neuroscience, but they will be published after Y1 (outside the scope of this Deliverable) if accepted.

1) P1669 Christopher Holdgraf, Stefan Appelhoff, Stephan Bickel, Kristofer Bouchard, Sasha D'Ambrosio, **Olivier David,** Orrin Devinsky, Ben Dichter, Adeen Flinker, Brett Foster, Krzysztof Gorgolewski, Iris Groen, David Groppe, Aysegul Gunduz, Liberty Hamilton, Christopher Honey, Mainak Jas, Robert Knight, **Jean-Philippe Lachaux**, Jonathan Lau, Brian N. Lundstrom, Christopher Lee-Messer, Kai Miller, Jeffrey G. Ojemann, Robert Oostenveld, Giovanni Piantoni, Natalia Petridou, Andrea Pigorini, Nader Pouratian, Nick Ramsey, Arjen Stolk, Nicole C. Swann, **Francois Tadel,** Bradley Voytek, Brian Wandell, Jonathan Winawer, Lyuba Zehl, Dora Hermes.

BIDS-iEEG: an extension to the brain imaging data structure (BIDS) specification for human intracranial electrophysiology. PsyArXiv. December 13. doi:10.31234/osf.io/r7vc2.

This is a preprint that is under revision at Nature Scientific Data.

Significance: very high as it defines an international format for iEEG data exchange, as defined by a large international (USA-Europe) consortium of ICT and iEEG experts.

Output: 6

2) P1778 Pascual, D., Aminifar, A., and Atienza, D. *"A Self-Learning Methodology for Epileptic Seizure Detection with Minimally-Supervised Edge Labeling"*. Design, Automation and Test in Europe (DATE) conference. 2019. Already available at:https://infoscience.epfl.ch/record/261181/files/Self_learning_paper_CameraReady.pdf ; (DOI will be available later)

   Significance: This paper shows the potential of a wearable EEG platform to improve the personalised detection of epileptic seizures, using a self-learning approach and requiring minimum input from the patient.

   Output: 11

3) P1780 Ferretti, L., Ansaloni, G., Pozzi, L., Aminifar, A., Atienza, D., Cammoun, L., and Ryvlin, P "Tailoring SVM Inference for Resource-Efficient ECG-Based Epilepsy Monitors". Design, Automation and Test in Europe (DATE) conference. 2019.

   Already available at (DOI will be delivered later):

   https://infoscience.epfl.ch/record/261315/files/DATE19-SVM.pdf

   Significance: This paper explores a number of optimisations that can be made to the SVM training algorithm to allow a direct execution on wearable devices, with the aim of distributing the workload of training personalised models for epilepsy detection.

   Output: 11

5.2.4 Measures to Increase Impact of Output(s): disseminations

The current MIP deployment strategy is very effective and does require further measures for an increased impact.

The Android app could be applied to multiple medical studies and with different purposes beyond the HBP scope. This will be pursued with the further development of the app.

6. Key Result KR8.4 MIP analytical tools enable federated analyses of multidimensional longitudinal data for advanced biological disease signature

6.1 Outputs

6.1.1 Overview of Outputs

1) Output 1: Description of models and pseudo-code for the 3C strategy (TAU, P57)

C ID	Component Name	Type	Contact	Info on releases and major updates
No C ID	Description of models and pseudo-code for the 3C strategy	software	S. ROSSET (TAU, P57)	The objective "pseudo code for incorporation.." is rendered unnecessary by having the actual code implemented in R packages https://github.com/HBPMedical/CCC and an interface to this code in MIP: https://github.com/HBPMedical/algorithm-repository/tree/master/r-3c More details are available in Milestone MS8.4.3.

6.1.1 Output 1 - Description of models and pseudo-code for the 3C strategy (TAU, P57)

This output provides a description of the 3C multi-steps algorithm, in which the transitions between the steps are based on the user feedback.

The algorithm usage has the goal to connect potential biomarkers with classes of clinical measurements. The steps are reducing the number of clinical features by selecting the best ones, clustering these measures into groups, and finally predicting the clusters from the biological measures.

6.2 Validation and Impact

The algorithm was validated by the peers. The impact

6.2.1 Actual Use of Output(s)

6.2.2 Potential Use of Output(s)

Output 1: The algorithm will continue to be used to examine whether the measures can provide clustering, differentiating phenotypes and whether these can contribute to prediction of disease trajectories. It will also be used to validate models by using them to make predictions on a new set of data provided from a different facility. In addition to validating the models, this will allow us to explore their sensitivity and specificity to allow future MIP users a novel classification of their own clinical samples based on the "biological signatures" of disease.

6.2.3 Publications

A publication is under review in Mov Dis.

6.2.4 Measures to Increase Impact of Output(s): disseminations

Conference Presentations:

Talma HENDLER: Early classification of neurobehavioral indices unveils the clinical trajectory of recent trauma survivors. 10th Conference of the Eastern Mediterranean Region of the International Biometric Society, December 2018, Jerusalem.

Mira MARCUS-KALISH: Reliable, replicable knowledge & data analysis towards precise "disease signatures". 10th Conference of the Eastern Mediterranean Region of the International Biometric Society, December 2018, Jerusalem.

Significance: In addition to presenting the results, the presentations serve to increase the general awareness about the analysis methods.

Outputs: They both concern Output 1.

7. Key Result KR8.5 MIP performs advanced automatised extraction of clinical relevant data from hospitals electronic health records (EHR)

7.1 Outputs

7.1.1 Overview of Outputs

1) Output 1 – Optimised data integration into MIP (CHUV, P27)
2) Output 2 – Automated pipeline for TVB (CHARITE, P122)

Table 4: Overview of elements contributing to automatised extraction of clinical relevant data

Component ID	Component Name	Type	Contact	Info on releases and major updates
2927 Data Catalogue	Specification for easier integration of EHR data	Other	Ludovic. CLAUDE, Mirco NASUTI (CHUV, P27)	Specification and implementation of a simpler method of integrating data into MIP, based on Frictionless Data packages; Release name: 2018.09.27; Release: Release date: 27 Sep 2018 URL: https://github.com/LREN-CHUV/data-db-setup/releases/tag/2.5.5
No C ID	Virtual Brain usage for-behavioural-and-neural-research	Other	Petra RITTER (CHARITE, P122)	URL: : https://www.elsevier.com/books/molecular-genetic-and-statistical-techniques-for-behavioral-and-neural-research/gerlai/978-0-12-804078-2 Release date : 26 April 2018

7.1.2 Output 1 - Optimised data integration into MIP (CHUV, P27)

This output offers a simpler method for integrating data into MIP, using Frictionless Data packages. It shall also contribute to automated data extraction from EHR.

7.1.3 Output 2 - Automated pipeline for TVB (CHARITE)

Au automated pipeline has been developed to leverage the processing of The Virtual Brain (TVB) from empirical MRI data. We updated the pipeline to account for atrophy, and also have quality control checks to minimise artefacts.

7.2 Validation and Impact

The optimised method of data integration into MIP was validated by usual testing of software and is ready to use.

The pipeline for personalised modelling of neurodegenerative and aged brains has resulted in several publications.

7.2.1 Actual Use of Output(s)

The pipeline for personalised modelling of neurodegenerative and aged brains has been used to process large public imaging data sets of healthy individuals and patients.

7.2.2 Potential Use of Output(s)

Output 1: The optimised method of data integration into MIP will be used to facilitate future data ingestions, to enable automatic EHR data extraction and to perform automatic processing of the incoming data for quality control.

Output 2: The pipeline for personalised modelling of neurodegenerative and aged brains shall be used to process large public imaging data sets of healthy individuals and patients.

7.2.3 Publications

This is a selection of the publications for this KR. The complete list will be attached to the report.

1) 1) P1781 Zimmermann, Perry, Breakspear, Schirner, Sachdev, Wen, Kochan, Mapstone, Ritter, McIntosh, Solodkin. Differentiation of Alzheimer's disease based on local and global parameters in personalized Virtual Brain models. Neuroimage Clinical https://doi.org/10.1016/j.nicl.2018.04.017

   Significance: First report on personalised large-scale brain network modelling of Alzheimer's disease, mild cognitive impairment and healthy controls.

   Output: 2.

7.2.4 Measures to Increase Impact of Output(s): disseminations

This is a selection of the disseminations for this KR. The complete list will be attached to the report.

1) The Virtual Brain processing was presented by Petra RITTER (CHARITE, P122) in the following workshop BrainModes Workshop „Computational Modelling of Neurodegeneration and the aging brain, 2-4 December 2018 in Havana, Cuba.

   https://www.humanbrainproject.eu/en/follow-hbp/events/computational-modelling-neurodegeneration-and-aging-brain/

   Significance: This workshop brought together leading scientists developing brain simulation methods for clinical translation in the field of neurodegenerative diseases.

   Output: 2. (modelling of VB).

8. Conclusion and Outlook

During the first year of SGA2, SP8 has focused on three main activities:

- Consolidate the MIP (WP8.1, WP8.2, WP8.4, WP8.5) which contributed to **KR8.1**, **KR8.4** and **KR8.5**
- Deploy the MIP (WP8.1, WP8.2, WP8.3) which contributed to **KR8.2** and **KR8.3**
- Integrate the projects selected through 2018 SP8 CEoIs (WP8.7, WP8.8, WP8.9, WP8.10) which was delayed until November 2018 for administrative reasons independent of SP8

In terms of outputs, the M12 status can be summarised as follows:

- MIP consolidation, which is a continuous activity, has already delivered the anonymisation MIP module fulfilling KR8.1, while limited outputs are yet available regarding KR8.4 and KR8.5. The latter shall be provided with the next MIP releases. It should be noted however that significant changes in the work plan will affect KR8.5.
- MIP deployment has progressed very satisfactorily, fulfilling most of M12 milestones and key performance indicators, and significantly contributing to KR8.2 and KR8.3. We are very confident that all M24 expected targets will be met, though the lack of SGA3 support to the MIP entails a significant risk that new hospitals will be discouraged to install the MIP.

In terms of Key Results, it is worth distinguishing the three first Key Results from the two others:

KR8.1, KR8.2 and KR8.3 are primarily operational and directed towards deploying the MIP in a large network of hospitals (KR8.2) in order to gather patients' data from more data providers across more brain diseases (KR8.3) while fulfilling all ethics and data privacy requirements (KR8.1). Most of these activities were performed through WP8.1, WP8.2 and WP8.3 according to the SGA2 work plan. In particular, a number of documents and SOP, including the MIP deployment package, the DPIA and the Data Governances Steering Committee charter, were developed, validated and implemented. In parallel, WP8.2 and WP8.3 have been very active in preparing the installation and installing the MIP in a number of new hospitals. Accordingly, most milestones and key performance indicators were met at M12 (9 new hospitals with a MIP installed, 11 more having signed a MIP installation agreement and about 10,000 patients' datasets in MIP local). WP8.5 has participated to KR8.1 by developing the anonymisation module which ensures that all data used in the MIP federated nodes are duly anonymised and thus not covered by GDPR regulation. WP8.8 and WP8.10, which were selected from the 2018 Calls of Expression of Interest (CEoIs), have only entered into force in November 2018, but have already contributed to the network of data providers by increasing the number of MIP-equipped hospitals within their own consortium. WP8.8 has also brought new data, i.e. Human intracerebral EEG data (iEEG), for which a specific sharable data format has been developed according to the international BIDS format (BIDS-iEEG). Finally, in WP8.3, T8.3.3 undertook a pilot study within the framework of its assessment of the requirements, costs and potential benefits of adding new data types into the MIP. This pilot study has focused on data provided by wearable devices

We anticipate that KR8.1, KR8.2 and KR8.3 will continue to progress smoothly during the second year of SGA2 and will reach all related targets, including the number of MIP-equipped hospitals and MIP patients' datasets (30 hospitals, 30,000 datasets covering four brain diseases). These developments shall be enriched by the work of the four WPs selected through CEoIs (WP8.7, WP8.8, WP8.9 and WP8.10) which only started a few months ago. WP8.7 and WP8.8 will provide TVB and iEEG data from patients with dementia and epilepsy, respectively, while adding new features to the MIP data factory. WP8.9 will deliver two important additional components, the "Ontologies updated and curated" and "Ontology lookup service (OLS-NEURO) for the HBP infrastructure". WP8.10 will bring data from patients with mental health disorders into the MIP.

KR8.4 and KR8.5 correspond to new MIP developments, including new analytical tools to enable federated analyses of multidimensional longitudinal data (KR8.4) and advanced automatised extraction of clinically relevant data from hospitals' electronic health records (EHRs) (KR8.5). These developments, which are being primarily performed in WP8.4 and WP8.5, are ongoing with yet

limited outputs at M12. The delay in producing these outputs primarily reflect changes in the strategy developed for enhancing the MIP following the SP8-SGA2 kick-off meeting held in Geneva-in July 2018. Extensive discussions with old and new users of the MIP led to a decision by all SP8 WP leaders that the scope for some Component/Outputs should evolve. In particular, the development and implementation of three EHR wrappers for hospitals was considered inadequate. SP8 members agreed that patients' data of interest were more likely to be generated from existing structured clinical research databases maintained in each hospital, rather than directly from EHR where data are usually less well structured (free text) and not curated. Directly querying hospitals' EHR systems also raises major regulatory and data privacy issues that are likely to result in hospitals' refusal to proceed. It was thus decided to reduce this activity to the development of a single EHR wrapper as a Proof of Concept, and to spend more efforts on connecting the MIP to eCRF such as RedCap. In that regard, it was decided to replace the XNAT connector by the Loris connector, since most of XNAT's functionalities are covered by LORIS, while the later will enable integrating data from REDCAP sources.

Despite the above limitations and/or delay regarding KR8.4 and 8.5, MIP analytical capabilities have been further enhanced by new algorithms. The scientific Workflow Engine component is under development, being already integrated with the Federated Processing Engine. The metadata catalogue component was also added, enabling users to explore MIP's data. For the remainder of SGA2, we plan to develop LORIS connector, so that MIP users can explore and annotate MRI scans, enable support for Longitudinal data, and further enhance analytical tools with the introduction of new algorithms. Moreover, in the second half of SGA2, we will develop a data cleansing tool for tabular data that will be giving recommendations for value corrections that can be useful to hospitals' personnel.

Most importantly, a new enhanced and consolidated MIP4.0 shall be delivered in June 2019 with a plan to provide a fully consolidated MIP 5.0 by the end of SGA2. The collaboration between SP8 historical partners involved in MIP development and consolidation have considerably improved during the last few months and shall enable a fast-track achievement of these most important SP8 outputs.

However, we wish to stress that the current plan for SGA3, which does not include any support for the MIP, is starting to have a very negative impact on hospitals aware of this situation, some of which have recently expressed strong reservations for investing in a MIP installation that might stop its activity in 12 months. To counteract this issue, we are actively working on a business model that would enable to establish a sustainable public-private partnership to continue developing and deploying the MIP.