Medical Data Analytics: Human Intracerebral EEG Platform (HIP) and Medical Informatics Platform (MIP) (SC5) – status at M42 (D4.13 SGA3)

Figure 1: Medical Data Analytics

AI generated impression of Medical Data Analytics Services.

Project Number:	945539	Project Title:	HBP SGA3

Document Title:	D4.13 Medical Data Analytics: Human Intracerebral EEG Platform (HIP) and Medical Informatics Platform (MIP) (SC5) - status at M42
Document Filename:	D4.13 (D44) SGA3 M42 RESUBMITTED 231207
Deliverable Number:	SGA3 D4.13 (D44)
Deliverable Type:	Report
Dissemination Level:	PU = Public
Planned Delivery Date:	SGA3 M42 / 30 SEP 2023
Actual Delivery Date:	SGA3 M42 / 27 SEP 2023 (resubmitted 07 Dec 2023)
Author(s):	Birgit SCHAFFHAUSER, CHUV (P27); Evita Mailli ATHENA (P133)
Compiled by:	Birgit SCHAFFHAUSER, CHUV (P27)
Contributor(s):	Olivier DAVID, Anthony BOYER, Maciej JEDYNAK, AMU (P78)/UGA (P125), contributed to Section 3 and the Annex Philippe RYVLIN, Manuel SPUHLER, Samira NAJIH-BESSON, Viktoria BASTIC-SCHMIDT, Pauline DUCOURET, CHUV (P27), contributed to Sections 2, 3 and the Annex Carolina CIUMAS, Nathalie CASATI, Florian SIPP, CHUV (P27), contributed to Section 3 and the Annex Viktor JIRSA, Marmaduke WOODMAN, AMU (P78), contributed to Section 3 Jean-Philippe LACHAUX, Florian SIPP, Benjamin BONTEMPS, UCBL (P108), contributed to Section 3 Thanasis KARABATSIS, Maria - Olympia KATSOULI, ATHENA (P133) contributed to Section 2 and the Annex Iosif SPARTALIS, AUEB (P4), contributed to Section 2
WP QC Review:	Martha Elisabeth BRIGG, UIO (P81)
WP Leader / Deputy Leader Sign Off:	Jan BJAALIE, UIO (P81)
T7.4 QC Review:	N/A
Description in GA:	A fully operational HIP platform, providing all services needed to fully manage, share and analyse iEEG data within EBRAINS according to the HIP consortium charter, with ≥ 200 fully curated iEEG datasets.
Abstract:	The EBRAINS Medical Data Analytics Service aims to enhance accessibility of data sharing and federation in the fields of brain health, clinical neuroscience, and medicine. It offers secure solutions for both federated and centralised data processing, contributing to the digital transformation and global advancement of medical research. Within the EBRAINS framework, two platforms play a crucial role in addressing the practical requirements and obstacles related to harmonising, sharing, federating, and analysing intricate brain health data. These platforms are the Medical Informatics Platform (MIP) and the Human Intracerebral EEG Platform (HIP). This status report presents an overview of the accomplishments and ongoing developments in both platforms, highlighting their contributions in the field. This document is a follow up from Deliverable D4.9 (D40) from January 2022.
Keywords:	Federated and centralised infrastructures, data interoperability, code-visits-data, data model, data privacy, human iEEG data, 3-tier architecture, In-Browser-Apps, BIDS data standard, data harmonisation, brain health, clinical neuroscience, medical data, Kubernetes, federated learning, Secure multi-party computation (SMPC)
Target Users/Readers:	Clinical researchers, computational neuroscientists, HBP/EBRAINS users, consortium members, funders, general public, policymakers, students

Table of Contents

Table of Tables

Table of Figures

History of Changes made to this Deliverable (post Submission)

Date	Change Requested / Change Made / Other Action
27 Sep 2023	Deliverable submitted to EC
01 Dec 2023	Resubmission with specified changes requested in Review Report Main changes requested: • Change 1 (A revised deliverable should properly reflect the state of MIP at M42. The document refers to MIP v7.1. as latest release. This part should be extended at least with a summary of features added with MIP v8.0 and a respective link reference to D4.16 to point to more detailed information.) • Change 2 correction of link to MIP7.1 release
06 Dec 2023	Revised draft sent by WP to PCO. Main changes made, with indication where each change was made: • Change 1 (see Section 2.1.3 MIP Releases –MIP7, MIP8) • Change 2 (see footnote in section 2.1.3 MIP Releases –MIP7, MIP8) • Change 3 correction of link to MIP deployment in Table 2 • Change 4 correction of link to MIP Kubernetes Deployment in section 2.1.2
07 Dec 2023	Revised version resubmitted to EC by PCO via SyGMa

1. Introduction

The **EBRAINS Medical Data Analytics Service** addresses the need for harmonising, sharing, federating, and analysing complex brain-health data. It offers innovative solutions for promoting collaboration and accelerating progress in brain health, clinical neuroscience, and medicine. The service provides accessible solutions for federated and centralised data processing, with online access to technologies and algorithms that necessitate minimal user and IT requirements. Two unique platforms, the **Medical Informatics Platform (MIP)** and the **Human Intracerebral EEG Platform (HIP)**, were co-designed with HBP developers, researchers, and users to address the needs and challenges of the clinical neuroscience community.

The MIP, building on efforts made in previous project phases of the Human Brain Project (HBP), is an optimised and secure open-source research tool for data federation and federated data analytics. The pursued approach overcomes the dualism of data sharing versus privacy, enhancing research potential and the value of clinical information that cannot be centralised due to institutional or national regulations. The MIP's defined governance framework and standardised harmonisation rules enable hospitals and clinical research institutions to leverage, explore, and analyse medical data, unlocking the potential of unexploited clinical data while respecting privacy, ethics, and legal requirements.

The HIP is an open-source European platform dedicated to centralising Human iEEG data, which are recorded in a limited number of clinical centres for patients with drug-resistant focal epilepsy or after deep brain stimulation. Developed in the last phase of the HBP, the platform offers solutions for storing, processing and analysis of iEEG data, along with guidance and scientific pipelines to promote interoperability through data curation and harmonisation. It provides accredited clinical researchers with unprecedented access to ethics and GDPR-compliant resources, in the long-run with data collated and curated in neurophysiological iEEG databases from over 70 clinical centres worldwide. Integration with other EBRAINS services, such as the Knowledge Graph, The Virtual Brain (TVB), and HPC centres, enhances the user experience and simplifies iEEG and neuroimaging data curation, processing, and analysis, which will be of enormous benefit for advancing clinical science in this field.

2. The Medical Informatics Platform - MIP

2.1 Medical Informatics Platform maintenance and upgrade

The MIP offers tools for enhancing knowledge, diagnosis, early prediction, and research for innovative treatment of brain diseases. As a federated data processing and analysis system, it is ideal for exploring decentralised, anonymised, and harmonised medical data from various sources. Current federations mostly relate to brain science, clinical research, and patient care, intending to facilitate collaborative research. The main principle of the MIP is "Code visits Data," where standard and machine learning algorithms are executed in selected nodes/hospitals, ensuring privacy-aware federated queries. Users can access aggregated findings with no possibility of accessing individual-level data. In the last phase of the Human Brain Project, work on the MIP aimed at strengthening the hospital network, maintaining, and servicing existing instances, and establishing a sustainable roadmap beyond the end of HBP by upgrading its analysis engine. It also explored privacy methods like SMPC (Secure Multi Party Computing) and DP (Differential Privacy) for enhanced privacy guarantees.

2.1.1 Evolution and scalability of the MIP

2.1.1.1 Technology Stack

The MIP went through two cycles of renewal during the last period of the HBP. While some major improvements were made for the stability and deployment automations of version MIP6, also new technology (specifically Kubernetes) and a new analysis engine were introduced in the MIP from release MIP7 onwards, which resulted in substantial development requirements for the MIP.

2.1.1.1.1 Improvements based on the implementation of a Virtual Private Network (VPN)

The MIP security system includes a secure VPN layer, data volume encryption, Ubuntu 22.04 (an open-source Linux-based operating system) and Debian GNU[1] (another Linux distribution composed of free and open-source software) operating systems: The implementation of a VPN at the transition period to the Specific Grant Agreement 3, the last stage of the Human Brain Project (SGA3)brought improvements like strong authentication, over-the-top encryption of data channels, and outbound hospital connections. Improved configuration for the reverse-proxy server[2] and EBRAINS Identity and Access management (i.e., EBRAINS accreditation and Keycloak[3] authentication) ensures secure programmatic access for each user.

2.1.1.1.2 Central Proxy and Firewall

No federated MIP is exposed directly to the public Internet. The central nodes act as a firewall and proxy, ensuring data security, traceability, and audit trails: Analyses are dispatched, distributed and coordinated by the central nodes, which also render the aggregate results, never compromising the data of an individual MIP. All analyses and outputs are systematically recorded and associated to the user’s credentials, providing full traceability of data access and data analyses.

2.1.1.1.3 Secure storage

Virtual isolation of different data volumes is used for secure storage, data reside in the federated nodes.

2.1.1.1.4 Automation tools

A major technical improvement for deployment of the MIP in comparison to earlier versions of the tool, was the implementation of a semi-automated procedure for installations. Vagrant[4], an automation tool in virtual machine environments, was utilised for various system execution aspects, reducing manual interventions. The usage of Vagrant was explicitly useful for MIP6, in the Kubernetes implementation of the MIP, similar automation will be achieved for MIP7+.

2.1.1.1.5 The MIP Gateway

Starting with the MIP6.5 release, the MIP Gateway was introduced as a new component in the system architecture. This modular structure greatly improves interoperability and flexibility. The gateway enables support for multiple analytics engines, an agnostic front end, and new data/analysis visualisation outputs. It also enhances state management and separation of concerns. Additional

---

[1] DEBIAN GNU https://www.debian.org/

[2] Reverse-proxy server: a reverse proxy server is a mediator situated inside a private network, behind the firewall. It works behind the scenes and guides client requests to the right destination on the server side. A proxy server adds an extra layer of organization and control, ensuring a seamless network traffic flow between clients and servers.

[3] Keycloak https://www.keycloak.org/

[4] Vagrant: https://www.vagrantup.com

features include generalization, better testability, and a GraphQL implementation with comprehensive API documentation and shared data structure.

Technical documentation: **MIP gateway**[5]

2.1.1.1.6 *The MIP User Interface (UI)*

User Interface and User Experience (UX) are crucial for determining the acceptance of a solution by end-users. While many services cater to the needs of advanced users (data scientists, statisticians, etc.) only, the MIP UI (see Figure 2: The MIP 7.1.0 User Interface), has been especially developed for the clinical community (clinicians, clinical scientists, etc.), an asset for reaching the MIP`s target audience.

The user-friendly design of the UI enables the adoption of advanced statistical analytics based on the integrated statistical methods and predictive machine-learning algorithms for data exploration and modelling, diagnosis and research in clinics, notably without the requirement of having expert knowledge in programming or design of algorithms. Sharing of experiments and workflows between users is enabled; authorisation and security is applied uniformly across layers. The available capabilities and easily accessible design attracted community interest and led to new collaborations.

Specific improvements include:

MIP6.5: A new formula component has been implemented in the front end, allowing users to enter formulas for selected variables. These formulas enable the calculation of additional variables based on user-defined expressions with well-defined semantics and widely accepted transformations available in the R statistical software package. This functionality is used for Descriptive Statistics and Logistic Regression, providing options such as transformations on continuous variables (log, exp, centre, and standardisation functions) and interactions between pairs of continuous variables. The formula component has been designed for simplicity to reduce user errors when manually entering formulas.

In the development environment, the front end now integrates "**StoryBook**[6]" a library that lists components in isolation. This integration allows developers and data scientists to have a comprehensive overview of the visualisation library offered by the MIP and better shape associated data types.

MIP7+: Enhanced visualisations and performance:

- Enhancements of the core descriptive statistic calculations, providing better insights into the available datasets were put in place; new histograms and descriptive analysis algorithms were made available; new versions of several algorithms (Pearson Correlation, TTest-independent, Anova Two-way, and Logistic Regression Cross Validation) were integrated.

UX enhancements:

- The interactive user guide was updated, a new visualisation library (**Bokeh.js**[7]) was integrated, and the export of results in PDF, csv and json was enabled.

Data Management:

- Standardisation of categorical variables was performed; the stability of the data ingestion pipeline was enhanced; some Quality Control Tool features were re-designed.

Technical documentation: **MIP Federated Algorithms** [8]

---

[5] MIP Gateway https://mip-front.gitbook.io/mip-gateway-doc/
[6] Storybook https://storybook.js.org/
[7] bokehJS https://docs.bokeh.org/en/2.4.3/docs/user_guide/bokehjs.html
[8] MIP Federated Algorithms https://github.com/madgik/exareme/tree/22.2.3/Exareme-Docker/src/mip-algorithms

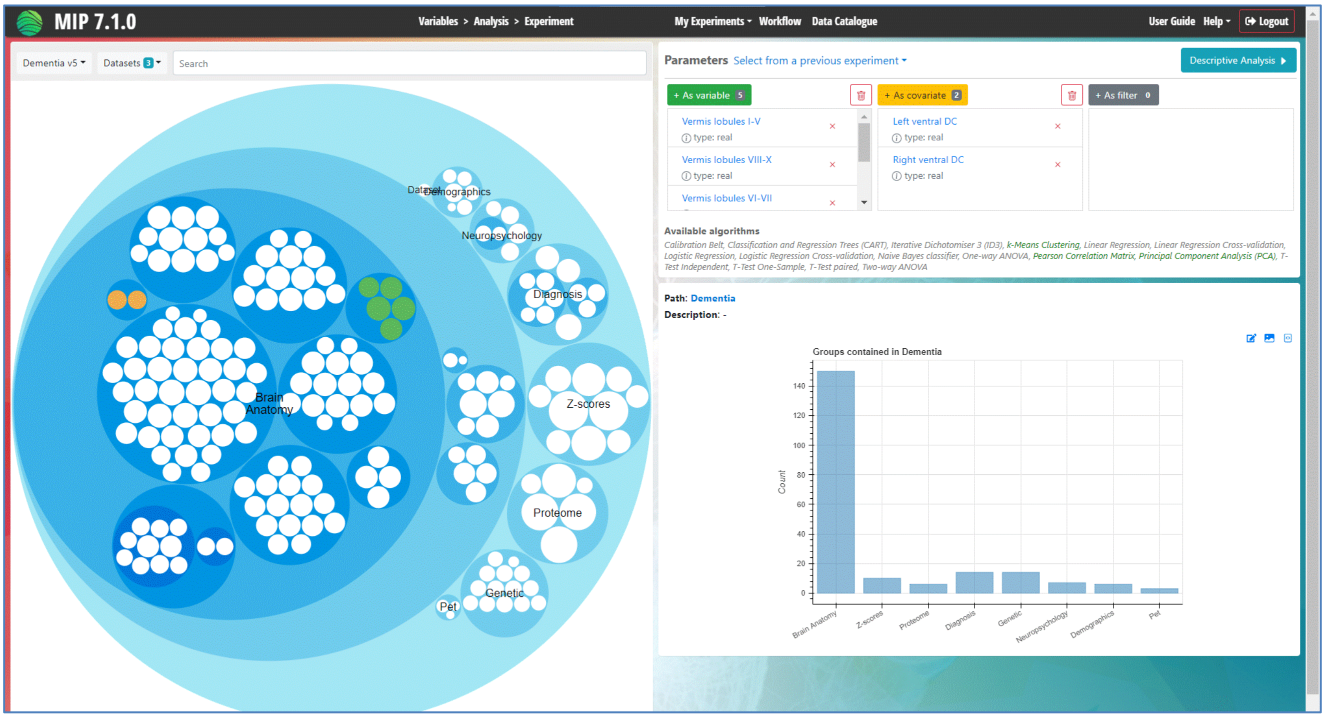

Figure 2: The MIP 7.1.0 User Interface

The MIP user interface displays three main parts, the relevant data-model, displaying all the available data including the hierarchical structure of the data (bubble structure on the left), the selected parameter for the experiment (variables, covariates and filters) including the available algorithms, top right and the visualisation of the selected parameters or experiment (bottom left).

2.1.1.1.7 The next generation of the MIP - Kubernetes

Tackling scalability improvements and technology upgrades, we introduced Kubernetes in the MIP. **Kubernetes**[9] (K8s), is a scalable open-source system for managing containerized applications across multiple hosts, especially relevant in a federation-set-like in the MIP. It offers basic mechanisms for the deployment, maintenance, and scaling of applications and provides a consistent API for running multiple containers across a cluster of nodes. The new MIP infrastructure is set-up as a self-managed Kubernetes cluster, allowing control over cluster configuration and enabling efficient management of containerized workloads. The set-up allows for customisation, scaling, and integration with other tools, simplifying deployment processes and improving application resilience.

2.1.1.2 The new MIP analysis engine

A new analysis engine was implemented in the MIP, from MIP7+, enlarging the potential of the MIP with regards to analysis, integration of analysis packages and federated learning. Exareme 2 is a cohesive solution safeguarding privacy and security within a virtual network of remote nodes. It is designed with Federated Analysis as its main focus. The system utilizes industry-proven solutions where applicable. It is modular, with independent components communicating through well-defined interfaces, allowing for easy extendibility and faster deployment of features such as new algorithms and data wrangling functionality. The engine operates on a core concept of splitting algorithm implementation into two components: User Defined Functions (UDFs) and Algorithm Flow. UDFs are small functions that run inside the database and provide pointers to the resulting data. The Algorithm Flow manages the execution of UDFs, determining when, how, and where they should run. The core components of the engine include nodes that perform database actions like creating tables, inserting data, and executing Python UDFs. Additionally, a Control Service is responsible for scheduling algorithm execution by queuing tasks on the nodes for execution.

[9] Kubernetes: https://github.com/kubernetes/kubernetes

Extensive testing under different demands (data volume, remote nodes, concurrent users) resulted in various improvements, e.g., warranting the proper initialisation and allocation of memory resources for the database within the Kubernetes environment, thus enhancing the stability and performance in Kubernetes deployments. Diagnostics and benchmarking have been conducted to optimise the Secure multi-party computation (SMPC) engine, achieving a balance between robust security measures and efficient execution of computations. All components of Exareme 2 are distributed as Docker images, making it easier to manage and deploy. The orchestration system no longer uses Docker Swarm and instead utilizes Kubernetes, which allows for staging upgrades, monitoring, health checks, and network security through VPN. Exareme 2 has a large ecosystem of tooling and provides APIs for programmability and offers operations-related functionalities such as backup, monitoring, and system health.

2.1.1.3 Scalability enhancements

The MIP network expansion in SGA3 required increased system capacity and faster response times. The MIP infrastructure on the EBRAINS Research Infrastructure, is utilising a secure VPN for network connectivity, and providing common, over-the-top and firewall-friendly network topologies for data transport, remote access, connectivity and infrastructure interoperability across all modules. Automation tools have been developed to scale up federations, manage servers remotely, and streamline installation and configuration processes. An API-coupled middleware allowing the abstraction of interfaces, for achieving a modular architecture and incorporating asynchronous communication for horizontal scaling, extending the overall system throughput, utilising established Advanced Message Queuing Protocols (AMQP), was introduced to the MIP.

The MIP federated architecture (up to release MIP6.5), was set up on docker-based virtual machines (Ubuntu 20.04), tailored and optimised for deployment at the Swiss National Supercomputing Centre (CSCS). Starting from release MIP7.0 (Ubuntu 22.04) the MIP moved to an orchestrator-based infrastructure, implementing a self-managed Kubernetes cluster, also on CSCS. By employing Kubernetes, the MIP achieves improved scalability, while at the same time optimising and reducing its resource footprint.

Great attention has been given to the interoperability and portability of the service, which is especially relevant for carrying the MIP into the future. The current MIP infrastructure consists of around 120 VMs on EBRAINS and 30 remote nodes in Europe (16 additional nodes being processed for set-up at time of deliverable submission), with each federation having central and pusher VMs connected to their federated nodes, which are holding the anonymised data of the institution and an instance of the analysis engine (see Figure 3: The MIP Network Topology).

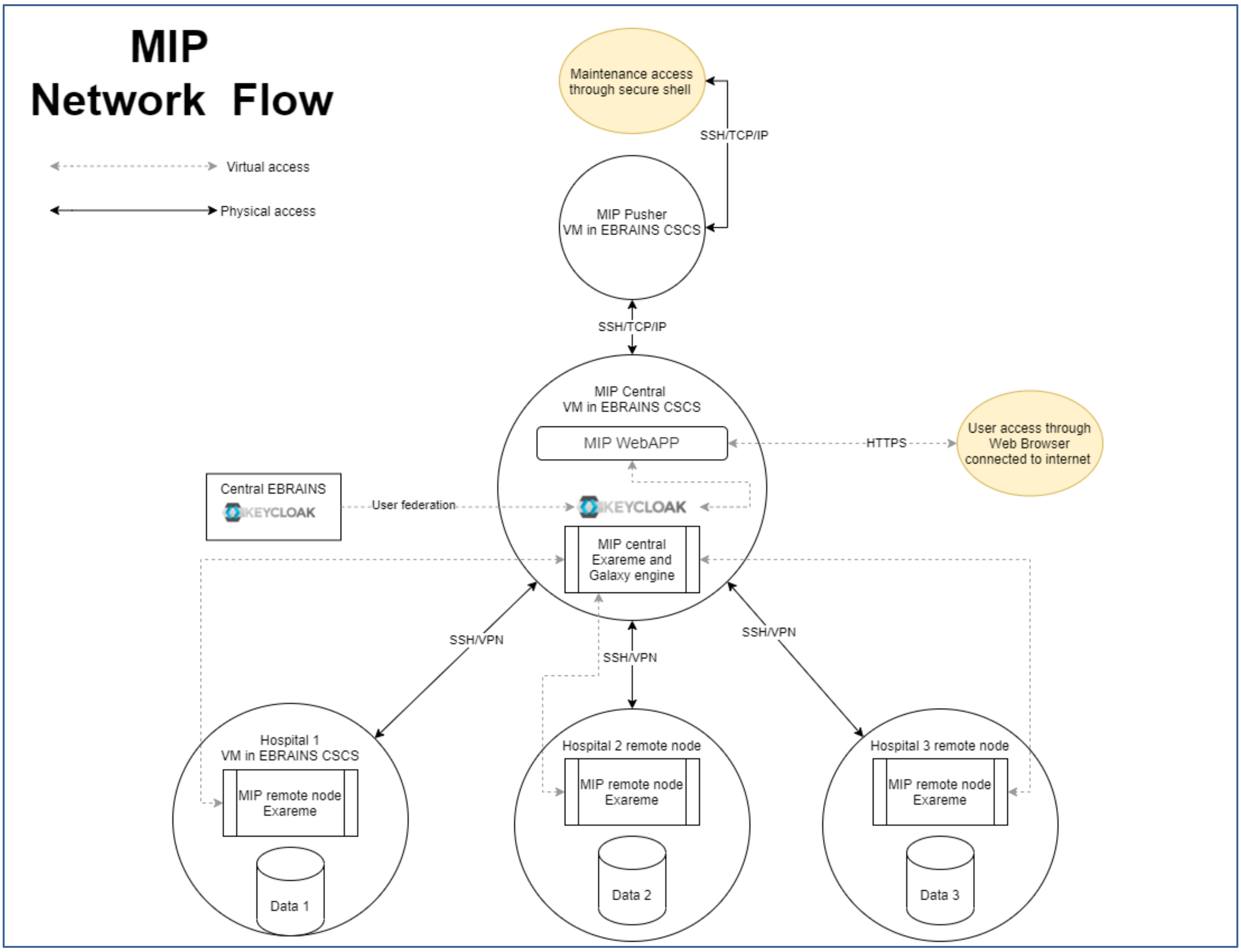

Figure 3: The MIP Network Topology

Represented is a classical technical set-up of a hybrid federation, entailing a central MIP node and remote nodes either deployed on premise (hospital) or on CSCS. All relevant components are indicated, physical or virtual access is displayed.

The deployment procedure has been optimised and modified to incorporate the latest deployment choices for the MIP. These options include setting up a remote node within the institution's infrastructure or, alternatively, utilising a virtual machine (VM) hosted on the EBRAINS RI.

Technical documentation: MIP Kubernetes Deployment[10]; MIP EXAREME[11].

2.1.2 MIP Federation Models

In SGA3, a new option called EBRAINS embedded federations or embedded MIP instances (VMs) has been introduced as an alternative to the traditional model of MIP federation. This allows users to use a secure virtual machine pre-installed within EBRAINS, which is especially helpful for those facing challenges with accessing or providing IT infrastructure and support. Each hospital that chooses this option receives its own dedicated EBRAINS VM with restricted access for their accredited users.

There are now three federation models available: federating several remote nodes, federating several VMs on EBRAINS (embedded federation), or a hybrid federation (see Topology combining both remote and hosted nodes within the same federation.

[10] MIP Kubernetes Deployment: https://github.com/HBPMedical/mip-deployment/
[11] MIP EXAREME https://github.com/madgik/exareme/tree/24.1.2/Federated-Deployment

2.1.3 MIP Releases –MIP7, MIP8

Since the former MIP6.5 release, which brought along the possibility to include additional variables in experiments, calculated in an ad-hoc fashion through user-defined formula expressions, as well as visualisation enhancements in core descriptive statistics, efforts were provided to further testing and strengthening the MIP in technically demanding scenarios. The new analysis engine Exareme 2 was integrated.

New algorithms:

The primary objective was to develop exact algorithms with cross-validation functionality (Linear Regression, Logistic Regression, Naive Bayes) within a test-driven environment, exploring diverse datasets and scenarios. We also explored and implemented alternative averaging strategies and added the user required longitudinal data analysis capabilities. Further, we started to pursue a different federated strategy known as the **fedAvg method** (Federated Averaging). This approach aims to expand the algorithm repertoire, albeit with a trade-off in precision, work continues.

In the MIP7 release an **automated way of metadata consolidation** was introduced.

Previously, data managers used scripts to unify metadata across federation nodes in order to ensure consistency of data. In this version, when a hospital is added or removed from the federation, resulting in changes to the metadata, the engine automatically handles the consolidation process.

By automating the metadata consolidation, the analysis engine streamlines the management and synchronization of metadata across the federation nodes. This improvement simplifies the workflow for data managers and reduces the potential for errors or inconsistencies in the metadata information.

Technical documentation: MIP7.1 Release[12]

The final release of the MIP at M42, **MIP 8.0.0**, introduces significant updates across various aspects of the platform. The release emphasises enhanced data security and privacy with a first integration of differential privacy through the Secure Multiparty Computation (SMPC) cluster, offering a global guarantee for federated analytics. This addition ensures cryptographic guarantees for obfuscating local, partial results. The release also brings notable improvements to its analysis engine, Exareme2, deprecating the former Exareme1 Galaxy workflow engine. Algorithmic advancements include new analysis algorithms, such as Naïve Bayes with integrated cross-validation, K-means clustering, and Support Vector Machine (SVM) and the capability for analysis of longitudinal data. Further, we expanded our model training capabilities with the federated averaging approach, where the central node's model parameters are updated via the obtained trained parameters of the local models.The release expands monitoring and reporting capabilities through Kubernetes-related packages, using Prometheus for container monitoring, Grafana for resource visualisation, and Loki for log aggregation and search. MIP 8.0.0 is described in detail in Deliverable **"D4.16 release of MIP 8.0.0 (SC5)"**.

Technical documentation: MIP8.0.0 Release[13]

2.1.3.1 Integration Tests

We developed **pytest-mip**[14] (published on **Zenodo with a DOI**[15]), an open-source tool for testing and reporting the status of federations using MIP 7+ releases. pytest-mip uses selenium-python to create robust test automation suites that interact with the MIP front end in a headless Chrome browser. The tests are written in Python and leverage widely accepted open-source frameworks like pytest and pytest-html. pytest-mip is encapsulated in a Docker container for easy deployment in fixed computing environments. A GitHub Action workflow automates the process of building the container

[12] MIP7.1 RELEASE https://github.com/HBPMedical/mip-docs/tree/7.1.0
[13] MIP8.0.0 RELEASE https://github.com/HBPMedical/mip-docs/tree/8.0.0
[14] Pytest MIP: https://github.com/HBPMedical/pytest-mip
[15] Zenodo DOI https://doi.org/10.5281/zenodo.7801386

image, running the tests, and generating reports. The workflow runs daily and updates the overall test status badge in GitHub.

The reports are automatically pushed to a GitHub repository, presenting the test results in a table format, including execution time, HTML report links, and the test status for each federation.

2.1.4 MIP Data Management and Quality Control

Significant efforts have been made to optimise the MIP's data management and quality control functionalities, enabling end-users to independently curate data. It supports users in processing, integrating, organising, and enriching data from diverse sources, while ensuring compliance with hospital and EU data privacy rules, including GDPR, within the MIP.

In the federated MIP, datasets in a specific federation adhere to a standardised metadata schema known as the Common Data Element (CDE) or Data Model. The CDE file contains a list of distinct variables. The process begins with extracting the CDE from each dataset in csv format, merging them into a standardised variable dictionary, and then harmonizing the data to build the data model.

For this purpose, the MIP offers user-support tools for simplified data preparation (see Figure 5: The MIP data architecture and flow):

- The **MIP Data Catalogue**[16]: A web-based application to explore and manage MIP Data Models and Common Data Elements (CDEs) pre-defined by the community or use case group.
- MRI parallel neuromorphometrics pipeline: A Python wrapper for running the Statistical Parametric Mapping SPM12 pipeline in parallel with Matlab on multiple CPU cores.
- The MIP Data Quality Tool (**MIP-DQC Tool**[17]**,** see Figure 4: The MIP Data Quality Control Tool GUI): Stand-alone software for exploring, validating, and transforming datasets based on CDEs before uploading to the MIP. It has both a Command Line Interface (CLI) and a Graphical User Interface (GUI). Originally developed for Linux, a lighter Windows version has been developed for wider usage.

The MIP-DQC Tool GUI version has the following functionalities:

- Validating hospital tabular data (csv) against a Data Model in the MIP Data Catalogue, generating a validation report with overall statistics.
- Validating longitudinal data by checking the presence and uniqueness of certain columns (VisitID and SubjectID) throughout the dataset.
- Data cleaning functionality based on the received data validation.
- Inference of a dataset schema, producing a schema file in a **Frictionless json**[18] or in the MIP Data Catalogue's excel format.
- Validating Excel files for seamless insertion into the MIP Data Catalogue during Data Model creation.
- Performing schema mapping of hospital datasets to a specific Common Data Element (CDE) schema.
- Generating DICOM MRI validation and statistical reports based on meta-data headers.
- Tabular data validation includes the option to download CDE metadata directly from the MIP Data Catalogue's API.

---

[16] MIP Data Catalogue: https://datacatalogue.hbpmip.link/pathologies
[17] MIP Data Quality Control tool: https://github.com/HBPMedical/DataQualityControlTool
[18] Frictionless data: https://frictionlessdata.io/

- Schema mapping is performed by the **MIPMap engine**[19], which runs in the background as a Docker container.

A time-consuming aspect is reaching agreement among federation partners on Common Data Elements (CDEs), including variable definitions, scales, and measurements. The Data Quality Tool supports the subsequent steps and is guided by the MIP Data Manager, who plays a central role in coordinating collaborative efforts.

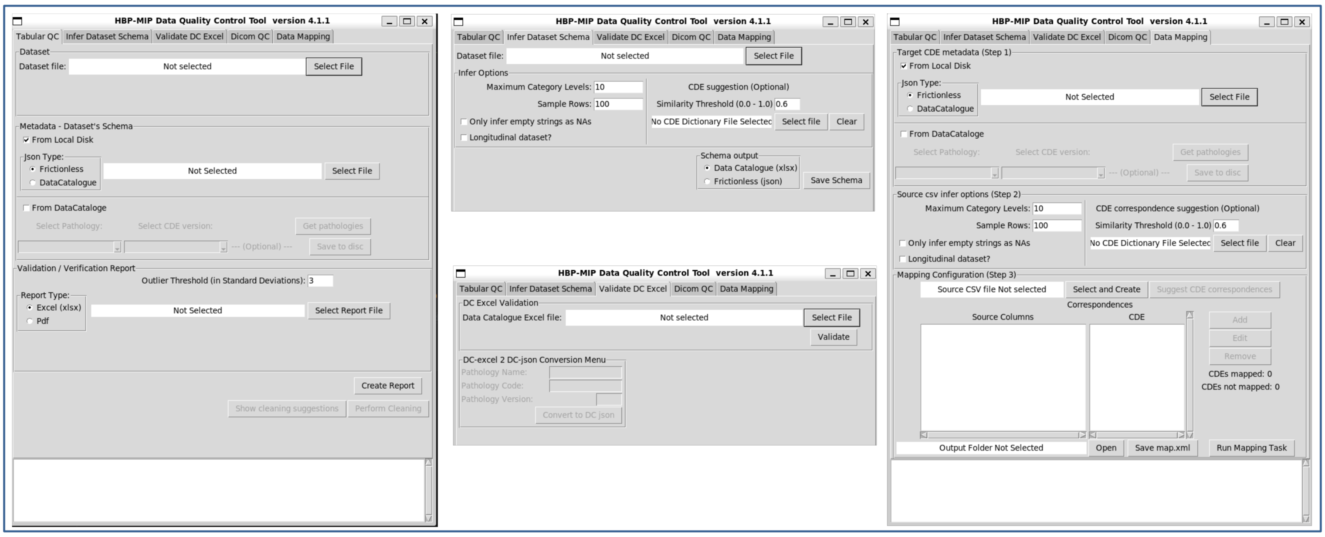

Figure 4: The MIP Data Quality Control Tool GUI

Layouts of the graphical interface of the DQC tool, illustrating the four main functionalities of the application: a) Hospital dataset validation/correction, b) Dataset Schema inference, c) Data Catalogue Excel file validation, and d) Schema mapping design and execution.

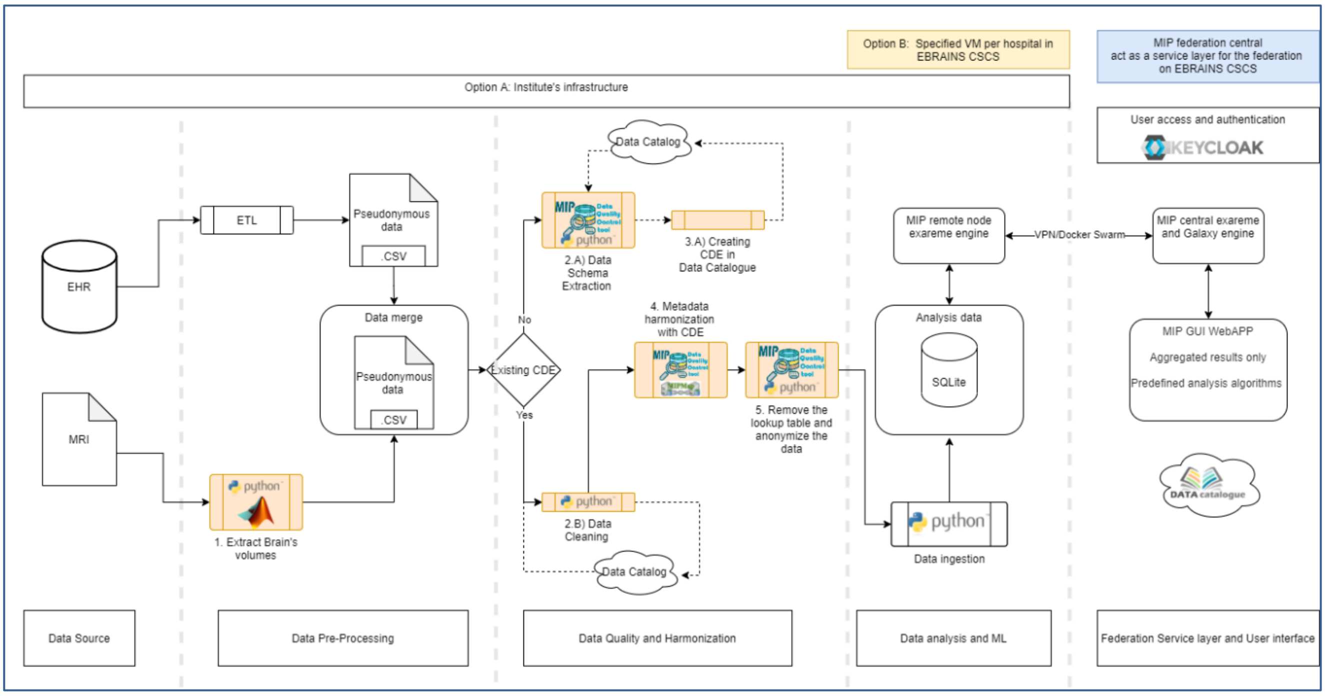

Figure 5: The MIP data architecture and flow

HR - electronic health record, MRI - magnetic resonance imaging, ETL - data integration (extract, transform, load), CDE - common data elements, ML - machine learning, GUI - graphical user interface, VM - virtual machine. Data pre-processing: Extract data from EHR records and produce pseudonymised data in .csv format. Optional Step1: Extract brain volumes from MRI images and merge with data EHR record data. Data Quality and Harmonisation: Prepare CDE. If CDE exists, follow Steps 2B/4/5. If CDE needs preparation, perform Steps 2A and 3A, then as before. Data Analysis

[19] Mipmap engine: https://github.com/HBPMedical/MIPMap

and ML: Upload anonymised data to the remote node or dedicated VM on EBRAINS CSCS. Perform Data Analysis via the Federation Service Layer. GUI: Use predefined algorithms, retrieve aggregated results.

2.1.5 The MIP clinical network, federations and collaborations

During the last phase of the HBP, new collaborations were formed, including a partnership with EBRAINS and the scientific committee of the European Academy of Neurology (**EAN**[20]) to federate datasets from various neurological conditions.

The expansion of the MIP network was one major goal towards the end of the HBP and to date spans 45 hospitals in 16 European countries (see Figure 6: The MIP Clinical Network to date). Currently, four other institutions have signed agreements and await installation, concrete discussions with five additional centres are ongoing.

As a first EAN use case, supported by the European Stroke Organisation (ESO[21]), we launched the federation of national stroke registries from seven EU countries (Switzerland, France, Austria, Greece, Ireland, Denmark, Italy), encompassing half a million patient records. This Stroke federation is a signpost, not only because it involves leading stroke experts across the EU, National stroke registries, overseen by governmental bodies, typically cannot centralize data from other registries outside their country, hindering cross-country data sharing. A data model, currently under review, was developed by the MIP data managers in collaboration with leading investigators of each registry, dataset mapping is underway. The infrastructure of the federation was prepared, more centres will be added upon signature of legal documents.

In SGA3, significant enhancements have been made to the MIP. These improvements, along with extensive collaboration efforts, are expected to foster a thriving MIP community and increase its adoption. The MIP has become a secure, robust, reliable, and user-friendly tool for health data federation, as evidenced by the interest of European consortia and prominent academic organizations such as the EAN, ESO, and European Reference Networks.

The MIP joined two IMI (Innovative Medicines Initiative) projects, IMI **SOPHIA**[22] (obesity, September 2021) and IMI **EPND**[23] in the field of neurodegenerative diseases (Nov 2021), and a Horizon Europe project in the field of emergency medicine, **eCREAM**[24] (Nov 2022). In EPND and eCREAM MIP federations will be set-up, processes towards achieving these goals are underway.

[20] European Academy of Neurology https://www.ean.org/
[21] European Stroke Organisation https://eso-stroke.org/
[22] SOPHIA https://imisophia.eu/
[23] EPND https://epnd.org/about
[24] eCREAM https://ecreamproject.eu/

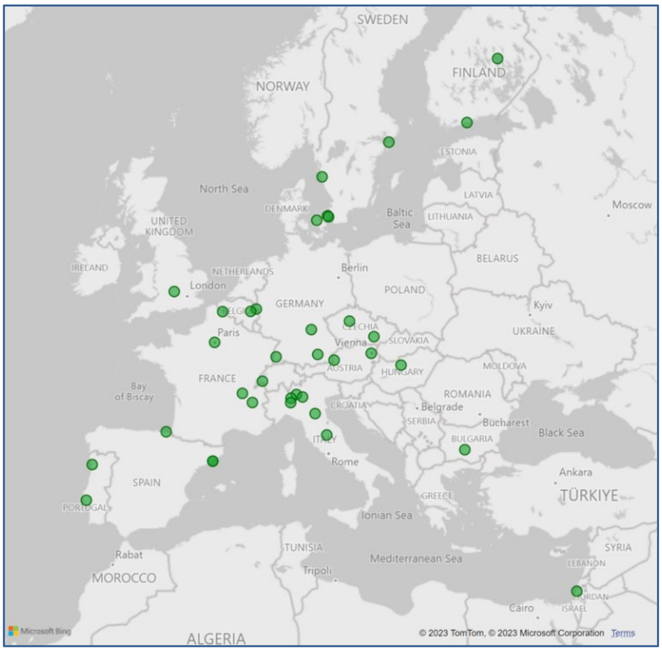

Figure 6: The MIP Clinical Network to date

Deployment of the MIP in 45 centres in 16 countries in either local or federated set-ups, across Europe.

2.1.6 MIP Data Governance

Working across institutional and national boundaries poses challenges to transparent governance models and data interoperability. The MIP offers a solution for transnational data federation with an established legal framework. Efforts were made in SGA3 to semi-automate data pre-processing and standardisation. For details see ANNEX Figure 13: The MIP Data Governance flow, where data governance and data flow in the MIP are depicted. Each institution remains the data controller of its dataset, responsible for compliance, curation, and authorisation. The MIP and EBRAINS provide tools and guidance for data controllers.

2.2 MIP exploitation – The PUBLIC MIP

The MIP is a service, accessible to accredited users only. To fully exploit the MIP and raise potential interest from users, the Public MIP was implemented.

The **Public-MIP** [25] is available on the EBRAINS website to any scientist via EBRAINS accreditation. No installation, no download is required. Once accredited, any person can analyse online all data placed in the public MIP and test the available MIP analytical tools.

The Public MIP mainly contains synthetic data (epilepsy, dementia, mental health and TBI, used for testing purposes). Recently, anonymised real-world data from the Leenaards Memory Centre (LMC) at the University Hospital of Lausanne were made available to the research community. Data from a Traumatic Brain Injury study (CREACTIVE) are in preparation for publishing, the respective agreement is being signed.

The MIP is also part of the **HBP Technology Catalogue**[26], where a comprehensive overview of the service is provided.

[25] Public MIP https://hbpmip.link/

[26] HBP Technology Catalogue https://www.humanbrainproject.eu/en/collaborate-hbp/innovation-industry/technology-catalogue/medical-informatics-platform/

2.3 The MIP EBRAINS integration

The EBRAINS Research Infrastructure is consolidating results and services generated within the HBP to secure their future availability. To comply with the EBRAINS integration requirements, work was done integrating **matomo analytics**[27] with the MIP. Further, we have implemented the integration of **metricbeat**[28] within the virtual machines (VMs) of the MIP infrastructure. This integration allows us to collect and send crucial metrics and monitoring data from the VMs to a centralised location. By leveraging metricbeat, we gain valuable insights into the performance, resource utilisation, and overall health of the VMs. This information enables us to effectively monitor and optimise the public MIP infrastructure, ensuring its stability, reliability, and efficient operation.

2.4 The MIP Looking Forward

In SGA3, the MIP was enhanced by significant upgrades, adding to its functionalities and benefits:

- Strengthened data privacy guarantees by incorporating **SMPC (secure multi-party computation)** and differential privacy into its analytical engine.
- Introduction of the MIP gateway, enabling **interoperability** of its front end to establish connections with different analytical engines, i.e., Exareme I and Exareme II or Datashield.

These enhancements solidify the MIP's position as a leading, freely available, and open-source platform for clinical researchers to federate their health data. It offers distinct advantages for integration into future projects, facilitating improved exchange and access to various health data types, supporting both primary and secondary data usage.

- Proof of concept stage: addition of a **Differential Privacy module**, which adds randomised noise to data to prevent the (re-)identification of individuals.
- Enhancement and enforcement of security over the whole cloud infrastructure with SSO (Single Sign-On) and deployment of the full security solution on all servers.
- Release of a MIP local version to be easily installed by end-users on their own professional laptop.
- Interoperability of data with other open standards; e.g., EBRAINS Knowledge Graph, which is promoting the open Metadata Initiative openMINDS.

  Deployment automation by the use of e.g., Terraform[29], an infrastructure as code tool and Ansible[30], a tool for simple but powerful automation for application deployment, updates on servers, configuration management, intra-service orchestration, etc., is being pursued.

---

[27] Matomo https://matomo.org/
[28] Metricbeat https://github.com/elastic/beats/tree/master/metricbeat
[29] Terraform https://www.terraform.io/
[30] Ansible https://www.ansible.com/

The Human Intracerebral EEG Platform - HIP

To provide a perspective on the progress made with the **HIP**[31] development, it is important to note that at the start of this last phase of the HBP in April 2020, no other service within the project was equipped to handle or manage sensitive clinical data. The necessary infrastructure requirements, security measures, and legal framework were not yet established. However, after 42 months of dedicated effort, we have successfully created an innovative platform and overcome numerous challenges along the way. Various components and workflows were developed and integrated. In the past six months, we were able to transition the HIP prototype into a first production platform. This allowed the first users to access the platform and use test-data to thoroughly test the platform's functionalities. The current phase is crucial as it involves further comprehensive testing and refinement of the implemented concepts with input from end-users and stakeholders. Their feedback is invaluable in refining and further developing the HIP concept.

Intracerebral EEG data poses challenges due to its limited availability, complexity in collection, and the requirement of domain-specific knowledge and IT resources for pre-processing, processing, and analysis. As a result, there is a scarcity of established standards, methodologies, and also platforms that enable the sharing and collaborative utilisation of this type of data. However, in 2019, the international community took a significant step forward by introducing a data sharing format known as BIDS-iEEG. Building upon this development, we utilised the BIDS-iEEG format as a foundational element to store data on the Human Intracerebral EEG Platform (HIP) and process them with open-source software integrated on the platform.

2.5 The HIP management and frameworks

2.5.1 The HIP - key benefits

- The world's first repository of its kind, supporting visualisation, processing, and analysis of intracerebral EEG data.
- Delegated compute processing, resulting in improved performance and efficiency.
- Increased accessibility for application developers, as apps can be tested and used without the need for installations.
- Data curation and adherence to standards to enhance interoperability and future data analytics potential.
- Preloaded pipelines and workflows of commonly adopted methodologies, serving as scientific guides and enabling continuous automation support.
- Enhanced international data interoperability, accessibility, availability, and affordability of high-quality imagery, curation, and analysis tools.
- Adoption of the Brain Imaging Data Structure (BIDS) as a common standard for organising and describing data and metadata, with the ability to accommodate other data structures and formats.

2.5.2 The HIP collaborative network

A survey was conducted among 78 Stereoelectroencephalography (SEEG) centres across Europe, Asia, and Oceania to gauge their interest in joining a unified network of iEEG data providers for the development and utilization of a new platform like the HIP. The survey revealed that 95% of the specialized centres expressed their interest in participating in this consolidated network. Many of

[31] The Human Intracerebral EEG Platform - HIP https://thehip.app/login

these centres are already engaged with the Human Brain Project (HBP) through their partnership with the ERN EpiCARE network. Additionally, several other centres are affiliated with active consortia such as SEEGMIP, EPINOV, and F-TRACT.

To date, 14 centres from 9 European Countries have been granted access to the HIP (see Figure 7: First adopters of the Human intracerebral EEG Platform). More centres will be added to the platform after legal requirements are fulfilled.

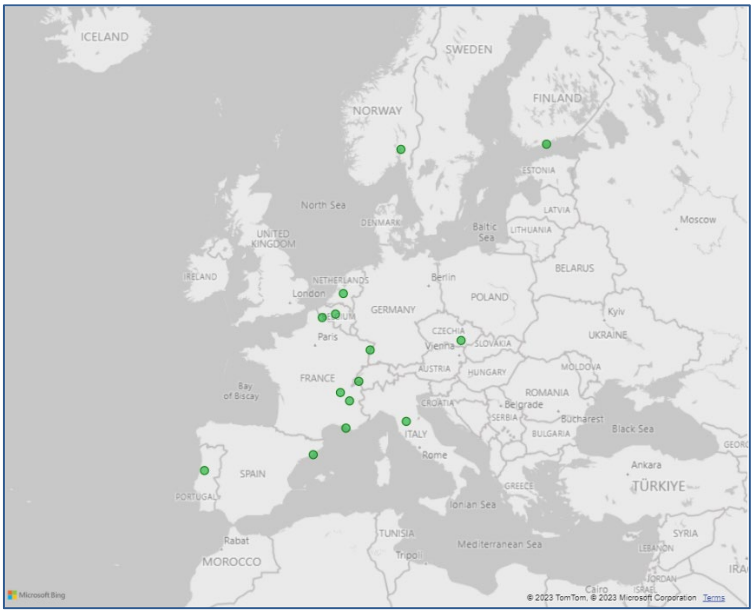

Figure 7: First adopters of the Human intracerebral EEG Platform

First centres were joining the HIP for beta testing of the platform. The HIP networks entail now centres in nine European countries and is expected to continuously grow.

2.5.3 The HIP legal and ethical framework

The HIP is an EBRAINS Service. The use of the platform, including its content, tools, and services, is subject to the EBRAINS Terms of Use and Policies[32]. While an EBRAINS account[33] is required to get access to the HIP, some HIP-tailored agreements between users and the CHUV have been drawn up.

To ensure effective collaboration, governance, and adherence to best practices, the **HIP Consortium Charter** has been developed. This charter provides guidance on various aspects, including decision making, rules for citation and publication. While it is not a legally binding document, collaborators are expected to sign it and adhere closely to its principles.

The **HIP Data Transfer Agreement** is a detailed document that outlines the legal framework and format and types of research data that can be uploaded to the HIP. This agreement is signed between a participating centre and the Centre Hospitalier Universitaire Vaudois (CHUV) before any data transfer can take place. It is mandatory for centres wishing to use the HIP platform as data providers.

The HIP **Terms of Service** (ToS) was established, Users need to approve the ToS set by the CHUV prior to any access to the platform.

The **HIP Registry** serves as an international database coordinated by the CHUV Project Leader and Management Team. It provides a regulatory framework for collecting and storing the data, fragmented across institutions and countries. It complies with the legal and ethical rules of CHUV, as well as the standards set by the Centre Opérationnel des Biobanques et Registres (COB) and includes both retrospective and prospective data collected during clinical practice and research. Included data are pseudonymised and covered by appropriate ethical and patient consent protocols.

[32] EBRAINS Terms of Use and Policies https://www.ebrains.eu/page/terms-and-policies
[33] Request EBRAINS ACCOUNT https://www.ebrains.eu/page/sign-up

Research data are largely covered by the **CogniEEG** and **CogEpiStim** protocols specifically designed for the HIP. Other iEEG research datasets, both prospective and retrospective, can also be integrated into the HIP registry, provided they meet the legal and ethical requirements for data reuse and transfer.

Risk Minimisation: The HIP registry employs different storage mechanisms to ensure data privacy. Raw iEEG and neuroimaging data are stored in the private space of the HIP platform, with each participating centre having its own access-protected repository. This is also the case for behavioural and cognitive research data collected during iEEG cognitive tests. Clinical data, however, which provide pseudonymised information about patients` epilepsies and comorbidities, are stored in a dedicated electronic case report form (eCRF) outside the HIP platform. These data are stored in the RedCap eCRF, either locally in the participating centres or on the CHUV RedCap server for centres without local access. This framework of separate storage for raw data and clinical data aims to minimise the risks of data privacy breaches.

2.5.4 The HIP use and accreditation modalities

Compliance with national and international laws and regulations, including privacy protection, intellectual property rights, and ethical considerations, is a priority for the HIP. Access to the platform is strictly limited to authorised users, with accreditations initially managed by the HIP Leadership and oversight provided by the Data Governance Steering Committee. The process for obtaining an EBRAINS account and the fine-grained accreditation process are being developed, with technical design and implementation underway.

2.6 The HIP design and architecture

The HIP offers a comprehensive and secure platform for scientists, physicians, and researchers to curate, share, and analyse intracerebral EEG data online. The platform was designed with a strong emphasis on integrating leading open-source software in a unique way that complies with regulatory and legal requirements. Users can access all applications, including desktop applications, directly in their web browser without any special IT setup. The HIP enables programmatic and automated use of applications that support scripting.

2.6.1 The HIP high-level overview

- Privacy-aware distributed compute storage and remote application and display servers to facilitate secure collaboration.
- Compliance with ethics and international regulatory frameworks.
- Cross-border sharing of pseudonymised iEEG data, study and ethics protocols, analytical tools, and workflows through browser-based applications.
- Security measures and closed-circuit collaborative functions to prevent data exfiltration.

The HIP platform follows a three-tier architecture[34] methodology, serving as a platform-as-a-service solution. The development and integration strategy prioritized a seamless user experience, leading to client and server-side applications, tools, and services being adapted for execution and operation on remote server instances (see Figure 8: System context diagram for the HIP). This approach combines storage, compute, and desktop experience within browser-based web applications while maintaining compatibility with standard data transfer protocols. The modular and interconnected architecture enables resources to be hosted independently in different cloud and HPC environments, programmatically accessed via the HIP API functions, and leverages both CPU and GPU performance.

[34] HIP Architecture https://github.com/HIP-infrastructure/hip-doc#architecture

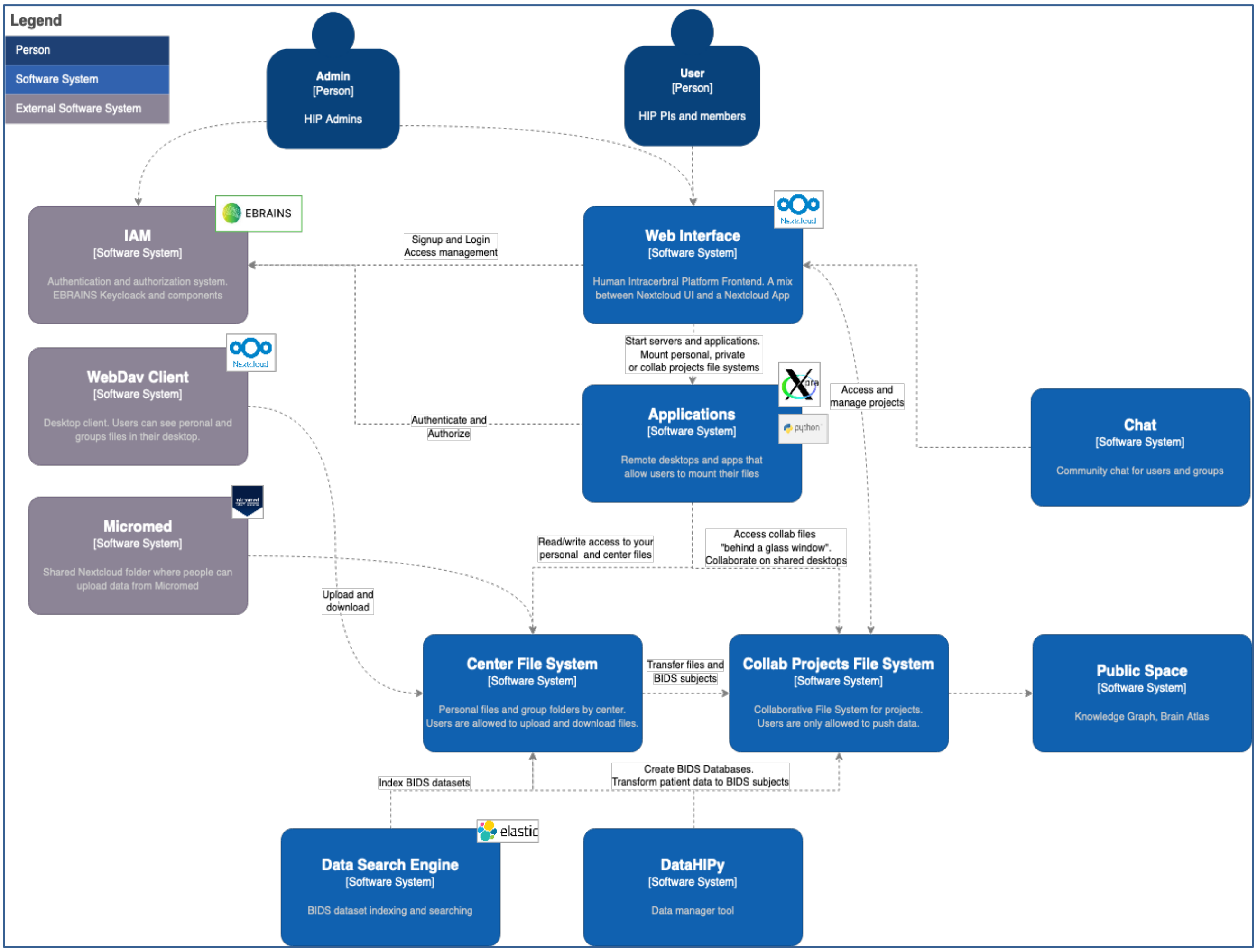

Figure 8: System context diagram for the HIP

Overview of the HIP, displaying various technologies and components, and their communication with each other.

The HIP establishes a privacy-enabling data-sharing ecosystem, comprising distinct spaces:

- **Personal Space**: A secure, access-controlled area where data providers can upload, store, and process their pseudonymised or anonymised data. This serves as the entry point to the HIP. All "personal spaces" of accredited users of a specific centre are bundled in the "**Centre Space**", that is accessible to all users of that centre.
- **Collaborative ("collab") Space**: An access-restricted area where curated and pseudonymised data can be shared with other accredited HIP users for approved data analysis. Moving data from the Personal Space to the Collaborative Space requires the authorization of the corresponding data's controller, who also determines which research studies and publications can access the data in this space.
- **Public Space**: A space for fully anonymised iEEG data, accessible to any scientist and referenced or searchable in the EBRAINS Knowledge Graph or other open access repositories. Data can be moved from the Personal or Collaborative Space to the Public Space only by the corresponding data's controller after anonymisation and obtaining patient consent for making their data public.

2.6.2 The HIP access and security

The HIP ensures top-tier security measures by implementing **Nextcloud**[35] as its data storage solution. It meets industry standards such as Health Insurance Portability and Accountability Act (HIPAA) and Health Information Technology for Economic and Clinical Health (HITECH) compliance, as well as being fully compliant with the General Data Protection Regulation (GDPR). Nextcloud offers File Access Control, comprehensive monitoring, and auditing capabilities. Data transfer is protected by SSL/TLS encryption. Data at rest can be encrypted using AES-256 encryption with server-based or custom key management, and there is an option for end-to-end encryption on the client side.

2.6.3 The HIP storage and compute

The HIP is designed to provide a storage environment with horizontally distributed components; a major focus was put on security and data safety. To mitigate the risks of data breach or loss, the HIP adopts a strategy of dividing different spaces into separate storage instances that can be federated. This federation capability allows for the distribution of multiple instances for each space, also enabling easier growth and expansion while minimising the risk of data corruption.

As part of the ICEI (Interactive Computing E-Infrastructure) project, the HIP provided access to compute and storage resources offered by the **Fenix infrastructure**[36]. Fenix is a network comprising five European supercomputing centres. Collaboration is facilitated on the HIP through standardised File Transfer Protocols (FTPs) and the utilisation of Virtual Desktop Infrastructure environments. To ensure secure data transport, Virtual Private Network (VPN) and other networking technologies are employed.

Currently, the HIP is hosted at the Swiss Supercomputing Centre (**CSCS**[37]) and **Exoscale**[38], a European provider of secure cloud infrastructure services based in Switzerland.

2.6.3.1 The HIP repositories and backend

HIP repositories: All git repositories are hosted on GitHub with automatic synchronisation on the GitLab instance on the EBRAINS CSCS platform. The Docker repository is available on GitLab.

HIP backend: Continuous integration (CI) ensures that docker images are automatically built upon pushing updates. An app-in-browser setup is implemented, utilising submodules for apps. GPU/CPU integration takes place on the EBRAINS CSCS platform, employing a custom docker container orchestrator framework. Key components include the xpra server[39] for remote desktop access, the pm2 process manager[40], and the caddy web server[41]. Various base images (the starting point for most container-based development workflows) and jupyterlab-desktop are utilised. A Dockerfile template is provided to facilitate easy integration of apps by partners and users.

2.6.3.2 The HIP frontend and gateway

The **HIP frontend** is a web application created using React[42] and integrated into Nextcloud to take advantage of its functionalities such as authentication, authorisation, user quotas, chat, and file system browsing for the Private Space and Centre Space. The interface combines elements from both Nextcloud and the web app, utilising the Nextcloud API and the HIP gateway. In the

---

[35] Nextcloud https://nextcloud.com/industries/healthcare/
[36] Fenix RI: https://fenix-ri.eu/
[37] CSCS https://www.cscs.ch/
[38] Exoscale https://www.exoscale.com/
[39] XPRA : https://xpra.org/
[40] PM2 : https://pm2.keymetrics.io
[41] CADDY : https://caddyserver.com/
[42] REACT : https://react.dev/

Collaborative Space, file browsing is facilitated by a custom metadata browser that provides access to metadata while preserving data privacy.

The **HIP Gateway** serves as a hub for Services APIs communication, facilitating interaction among various apps, resources, and components. It allows for the programmatic selection of datasets and ensures their availability at the appropriate workflow step. It sends an index request to the Elasticsearch instance of the HIP Data Search Engine. This indexes the dataset with the information available in the dataset summary metadata file and is described in more details in section 2.7.1.1.

A state machine is employed to enforce the correctness of session and app sequences, enabling start, stop, and destroy functionalities. User states, including sessions, apps, and credentials, are stored in Redis[43], serving as a messaging and caching service.

2.6.3.2.1 Creation of a new collaborative project

When a user initiates the creation of a new "Collab" project in the HIP frontend, the process unfolds as follows: The request for collab dataset creation is sent from the frontend to the HIP Gateway via the REST API. The HIP Gateway, in turn, interacts with EBRAINS IAM services to establish a new group specific to the project, along with assigning different roles to the group members. To set up the initial structure for the project, the HIP Gateway executes a command using DataHIPy (see section 2.7.1.2.1), creating a structured directory. This directory serves as the foundation for organising and managing project-related data within the collab.

The HIP Collab entails encrypted storage on CSCS, accessible to the HIP collab via Network File Sharing (NFS), a protocol that allows to share directories and files with other Linux clients over a network. This storage is securely mounted in Desktops using ghostfs (Container image that includes a distributed file system tailored to the HIP).

2.6.4 The HIP user documentation

To assist end users in accessing and utilising the platform, the **HIP user documentation**[44]**,** hosted in a GitHub repository, has been meticulously crafted.

- **Starter guides** aid users in accessing the platform, creating a HIP account with relevant accreditations, understanding data controller responsibilities, and adhering to FAIR data principles. It also helps users connect to the HIP portal, access different services (such as data uploading and storage, BIDS conversion tools, private and collaborative spaces, app library, and processing pipelines), as well as utilise internal tools like the wiki and support/assistance.
- **Data storage and sharing guides** provide instructions on preparing data (eligibility, pseudonymisation, curation, and uploading conditions), using dedicated tools (web-based uploader, WebDAV[45] or client software for Micromed[46] data transfer), and harmonising data according to the BIDS standard through the integrated BIDS conversion tool. Furthermore, it explains how to transfer validated BIDS datasets to collaborative spaces and share them with other HIP users within a research protocol-associated working group.
- **Data analysis guides and use-case tutorials**, facilitate data exploration, processing, and result sharing. These guides cover initiating private or collaborative working sessions, launching apps/pipelines from the HIP software library, and sharing analytical tools and outcomes. The HIP integrates pre-existing iEEG tools, each with its own documentation, accessible via integrated links.

The HIP user documentation is a continuously evolving resource that will adapt alongside the platform's maturity.

---

[43] Redis: https://redis.io/
[44] HIP User Documentation: https://github.com/manikbh/HIP-doc
[45] WebDAV: http://www.webdav.org/
[46] Micromed System Plus: https://micromedgroup.com/download/systemplus-evolution-user-manual/

2.7 Data handling on the HIP

2.7.1 The HIP standardised procedures and workflows

Efforts have been focused on developing App Deployment Frameworks based on containerisation technologies like Docker[47], aiming to integrate applications and services provided by the HIP development teams. This also sets the stage for future collaborative development of the platform with external users. As mentioned above, on the HIP, the BIDS is the common standard for organising and describing data and metadata. Thus, procedures and workflows around the BIDS data structure were developed for the platform.

2.7.1.1 BIDS-indexation and search

Indexing a BIDS-iEEG dataset in the HIP platform involves the following steps: Users perform actions in the HIP frontend, such as creating a BIDS dataset or importing/converting files. These actions trigger requests sent to the HIP Gateway via the REST API. The HIP Gateway executes the necessary commands within the DataHIPy container and generates a summary of the dataset. This summary provides a concise overview, enhancing organisation and management within the HIP.

The process for **searching for BIDS-EEG datasets** is as follows: Users initiate a “BIDS dataset search query” from the HIP frontend, which is transmitted to the HIP Gateway through the REST API as a “dataset search request”. The HIP Gateway formats and forwards the query to the Elasticsearch instance of the HIP Data Search Engine. The Elasticsearch engine processes the query and returns paginated search results to the HIP Gateway. Subsequently, the HIP Gateway sends the search results back to the HIP frontend for user review and further exploration.

2.7.1.1.1 Import of subject directories from indexed BIDS datasets into the private space

When a user wants to add new files to a project's BIDS dataset within the HIP, they start by clicking "add new files" in the project. This action triggers a “dataset file import request”, which is sent to the HIP Gateway through the REST API. Upon receiving the request, the HIP Gateway executes a DataHIPy[51] command to import the subject directory and update relevant tables and files in the BIDS dataset of the project. An updated dataset is generated. The HIP Storage and Compute components handle the storage and computational aspects of the imported files, ensuring secure and efficient processing in the HIP environment.

2.7.1.2 The HIP data formats, tool development and integration

Former work done within the HBP to implement an infrastructure to manage and process iEEG data, led to the development of the specification of the **BIDS (Brain Imaging Data Structure)** for iEEG (**BIDS-iEEG**) format (international consortium involving UGA) (Holdgraf et al., Sci Data, 2019)[48], the **BIDS Manager**, a BIDS-iEEG data management tool led by AMU (Roehri et al., Neuroinformatics, 2021)[49] and the generation of a BIDS-iEEG dataset (**CCEP database**[50] accessible through ebrains.eu, led by UGA/AMU).

We integrated a **suite of analytical tools for iEEG data analysis**, allowing to organise the data (SEEG or anatomical imaging) into the BIDS-iEEG folder architecture, for further processing (SEEG or anatomical imaging), and visualisation of the results. The apps can be instantiated on remote cloud infrastructures. The tools are encapsulated within individual Docker containers and have direct

[47] Docker: https://www.docker.com/
[48] BIDS-iEEG format (P2033): https://doi.org/10.1038/s41597-019-0105-7
[49] BIDS Manager(P2978): https://doi.org/10.1007/S12021-020-09503-6
[50] CCEP database https://search.kg.ebrains.eu/instances/Dataset/ebe50517-41d5-4029-9355-04f1e49e23c8

access to the GPU of the hosting machine. Collaborators can conveniently access these integrated apps by launching sessions on the HIP directly through their web browser.

The available apps on the HIP primarily consist of open-source software capable of reading data stored in accordance with the **BIDS standard**. However, if necessary, the HIP can also support data upload or import using different data structures and formats as required by specific apps. These diverse apps enable the creation of workflows for processing iEEG data.

Interoperability among the apps is achieved through the utilisation of the **BIDS formalism**. It allows for the standardised storage of raw or pseudonymised data and centralises the main outputs of different apps in a dedicated location (the derivatives folder of the BIDS dataset). Certain apps like **Brainstorm** maintain their own database on the HIP to preserve pertinent information.

2.7.1.2.1 DataHIPy

We developed **DataHIPy** (Tourbier et al., 2023)[51] a dockerised open-source tool to support the management of iEEG data on the HIP following the BIDS standard. DataHIPy leverages the **PyBIDS** (Yarkoni et al., 2019)[52] and the BIDS Manager libraries to provide a set of utility commands that enable a specific workflow within the HIP platform. First, a BIDS dataset is created by importing and converting files into formats accepted by the Brain Imaging Data Structure standard. A dataset summary metadata file is generated in JSON format, serving as an index for the dataset in the HIP Data Search Engine. Additionally, a "Collab" Project dataset is created, structured according to the YODA principles and including a BIDS dataset. Subject folders from BIDS datasets in the Private Space are then imported into the project dataset within the Collaborative Space, ensuring seamless collaboration and sharing of data, maintaining structure and integrity of the original BIDS dataset.

2.7.1.2.2 The HIP data search engine

The **HIP Data Search Engine** comprises an Elastic[53] stack (**Elasticsearch, Logstash, and Kibana)** (containerised and orchestrated using Docker compose). It facilitates indexing and searching of BIDS datasets within the HIP. The containerised DataHIPy tool aids the management of neuroimaging data within the HIP. It facilitates the creation of BIDS datasets, file importing and conversion into the appropriate file naming, organisation, and formats based on the **BIDS Manager library**. Interaction with BIDS datasets is supported using the **PyBIDS library**, providing a dataset metadata summary in JSON format. Services offered by the HIP Gateway interact with DataHIPy and the HIP Data Search Engine. The new Frontend-integrated browser/manager for BIDS datasets enables Users to query available BIDS datasets, create new BIDS datasets, and edit participants' demographic information through the BIDS participants.tsv file and files can be imported and converted to BIDS-accepted formats. This streamlined tool provides a user-friendly interface for efficient data handling.

2.7.1.2.3 Integration of visualisation tools

The CHUV team successfully developed an innovative solution within the HIP: **3D rendering in virtual environments**. This advancement enables users to visualise data within the HIP apps, simulating a local environment. The 3D rendering functionality serves various purposes, such as data quality control, data preparation, and exploration of processing results (see Figure 9: Use of the HiBoP app for 3D visualisation).

2.7.1.2.4 Pre-processing tools on the HIP

The HIP offers a variety of options for pre-processing imaging and iEEG data, allowing to enhance and optimise research in real time (see also Table 1: The HIP Library – In-Browser Apps).

- **Efficient Data Conversion**: Using the dependable **dcm2niix** programme, one can easily convert data from Dicom to Nifti format.

---

[51] DataHIPy https://doi.org/10.5281/zenodo.7692220
[52] PyBIDS https://doi.org/10.21105/joss.01294
[53] ELASTIC: https://www.elastic.co/

- **Cropping the Field of View**: Use the capabilities of the **3D slicer** and **MRIcroGL** to accurately crop the field of view, ensuring focused and accurate three-dimensional analysis.
- **Data in Multiple Modes Co-registration**: Using cutting-edge tools such as **3D Slicer, NiftyReg**, or **FSL** (see Figure 10: Co-registered CT with implanted electrodes & the T1 image - visualisation with FSL), one can seamlessly co-register imaging data from various sources such as CT, PET, SPECT, or MRI. Accurate alignment is required for thorough analysis.
- **Comprehensive Reconstruction**: Using **Freesurfer**, one can receive a complete reconstruction of white and grey matter, allowing for extensive exploration and analysis of brain regions.
- **Precise iEEG Annotation**: Use of **Anywave** and **Brainstorm** to annotate iEEG data, allowing for a more detailed analysis and interpretation of neural activity.
- **Extraction of Task-Related Cognitive Brain Networks**: Use **Frites** to extract task-related cognitive brain networks, allowing for the investigation of functional connectivity and neural dynamics.
- **Extraction of Frequency Components**: Using **Localiser**, one may extract frequency components from raw iEEG recordings, revealing insights into unique brain oscillations and their relevance.
- **Visualise Task-Related iEEG Components**: Use **HiBoP** to visualise task-related iEEG components, providing an intuitive and comprehensive understanding of brain activity during specific tasks.
- **Accurate Electrode Reconstruction**: With the help of **Gardel, Slicer, and INTRAnat**, ensure exact electrode reconstruction, allowing for accurate localization and analysis of brain impulses (see e.g., Figure 11: Reconstruction of all implanted electrodes in Slicer 3D using the Milano pipeline).
- **Anonymization of Structural Data**: Protect participant privacy by using **MRIdeface** to anonymize structural data while adhering to ethical standards.
- **Individual Connectome Models**: Using **TVB**, create individual connectome models to examine personalised brain networks and their functional features.
- **BIDS Format Conversion**: Using **MRIcroGL** and **BIDS-manager**, easily convert data into the Brain Imaging Data Structure (BIDS) format, assuring compatibility and interoperability with other neuroimaging tools and pipelines.

With these robust features, the platform offers a comprehensive set of tools to improve the pre-processing and analysis of imaging and iEEG data, enabling meaningful and rigorous neuroscience research.

Figure 9: Use of the HiBoP app for 3D visualisation

Visualisation of the experimental data of a visualisation task on a patient Brain with the corresponding site activity for three conditions.

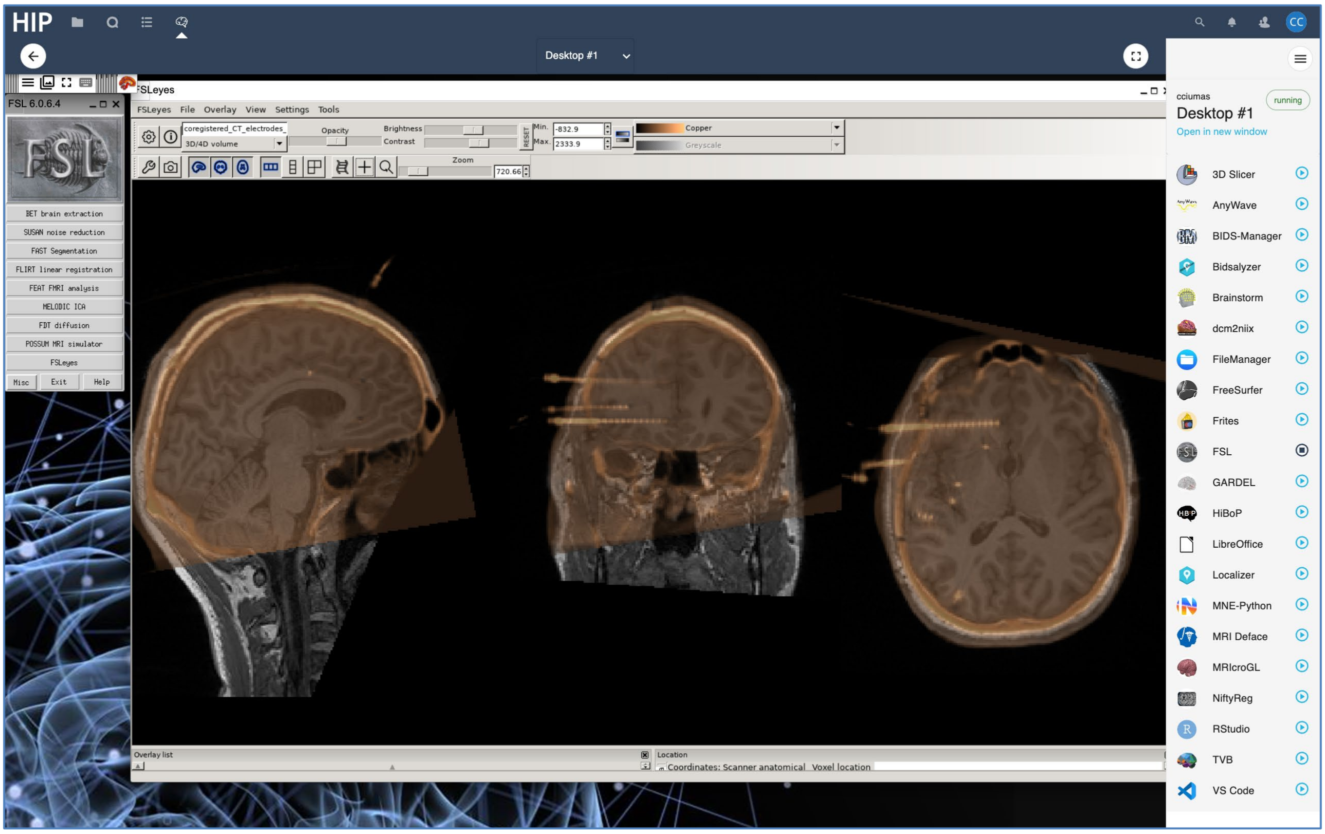

Figure 10: Co-registered CT with implanted electrodes & the T1 image - visualisation with FSL

Various imaging modalities from different scanners can be difficult to superimpose without proper co-registration parameters. FSL offers Rigid Body transformation to facilitate this process. The figure shows the CT image taken immediately after electrode implantation (a mandatory step to exclude any occurrence of bleeding), superimposed on the 3D T1 image obtained during the pre-surgical workup prior to the implantation.

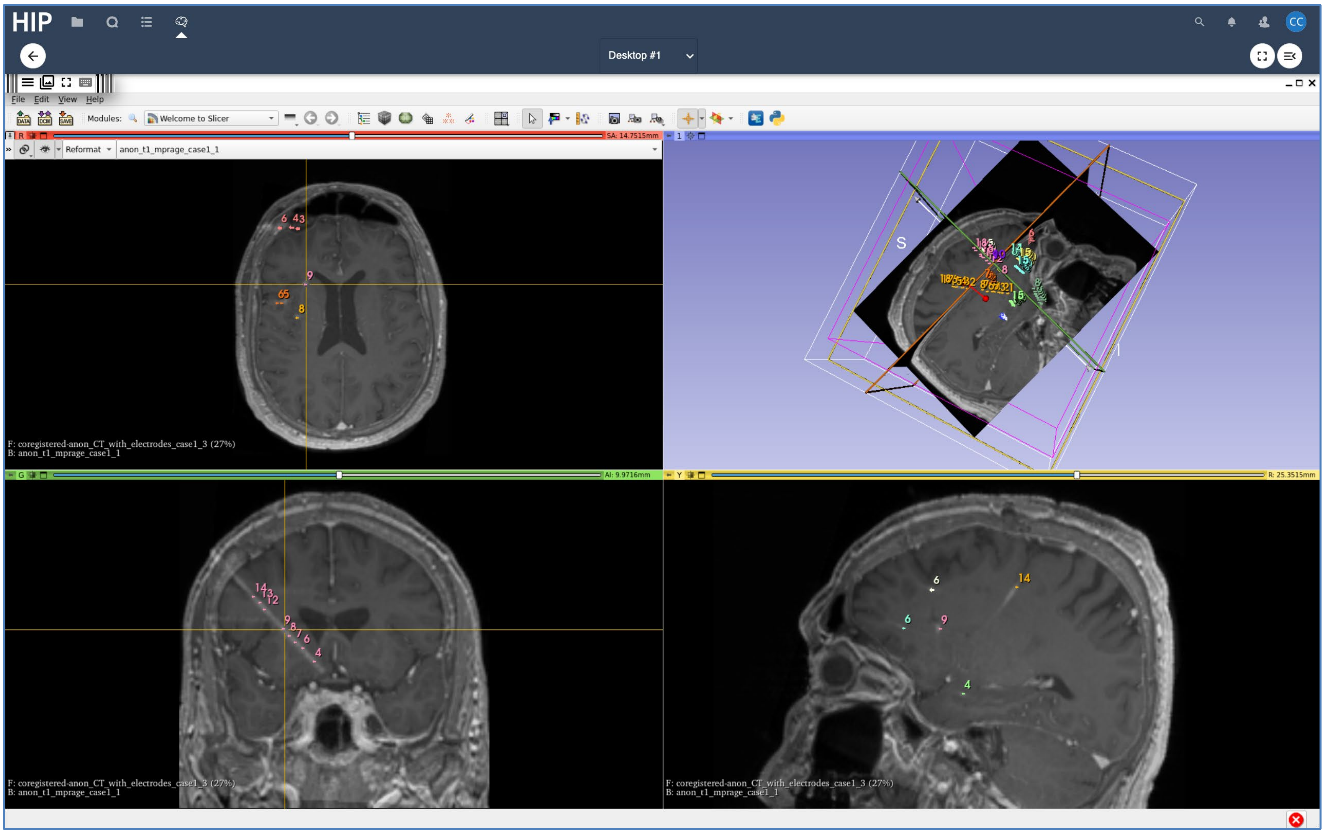

Figure 11: Reconstruction of all implanted electrodes in Slicer 3D using the Milano pipeline

After aligning all images and providing the implantation scheme, users mark up the electrodes to identify their true trajectory. The marking ranges from 1 (deepest probes) to 5, 8, 10, 15, or 18 probes, depending on the electrode used. Typically, 15 to 18 electrodes are implanted, allowing clinicians to locate the iEEG contact triggering a response and precisely define its location within the brain on the 3D T1 image.

2.8 Integration with other EBRAINS Services

The HIP is developed in parallel with other EBRAINS services, such as the knowledge graph, data atlases and simulation services. Integration with these services enables users to access data in a coherent fashion and use them with simulation and analysis tools. This integration focuses on two major efforts: inclusion of the Virtual Brain (TVB) package on the HIP platform and integration of F-TRACT data in the atlas services. The inclusion of the F-TRACT data in the EBRAINS atlas services demonstrates how intracerebral data can be usefully included in the EBRAINS platform to benefit the larger neuroscience community. The integration of TVB on HIP serves two roles. First, because TVB is a flexible framework for whole brain simulation which includes tools to interface to different EBRAINS components, it can serve in EBRAINS as a functional hub between simulation and data services, which can then enter the HIP, linking EBRAINS data and wider simulations services to HIP workflows. Secondly, it paves the way for introducing other EBRAINS-native simulation and analysis tools into a human-data friendly environment. In summary, the key condition for interoperability between HIP and the wider EBRAINS community is alignment of the anatomical and functional neuroimaging data, and the F-TRACT data and TVB package demonstrators together pave the way for this enabling fruitful collaboration on a wider range of scientific and clinical problems than previously possible.

In the following, the efforts on interoperability will be described in more detail (see Figure 12: Interoperability on HIP with workflows, including TVB modelling): the workflows coordinate use of multiple apps on HIP, the various apps allow constructing a patient brain model optionally including data from the EBRAINS atlas. The TVB epilepsy workflow constructs a parametric map of the epileptogenic zone based on the individual patient brain optionally making use of HPC resources and producing maps usable in other HIP tools, and finally F-TRACT data integrated in KG.

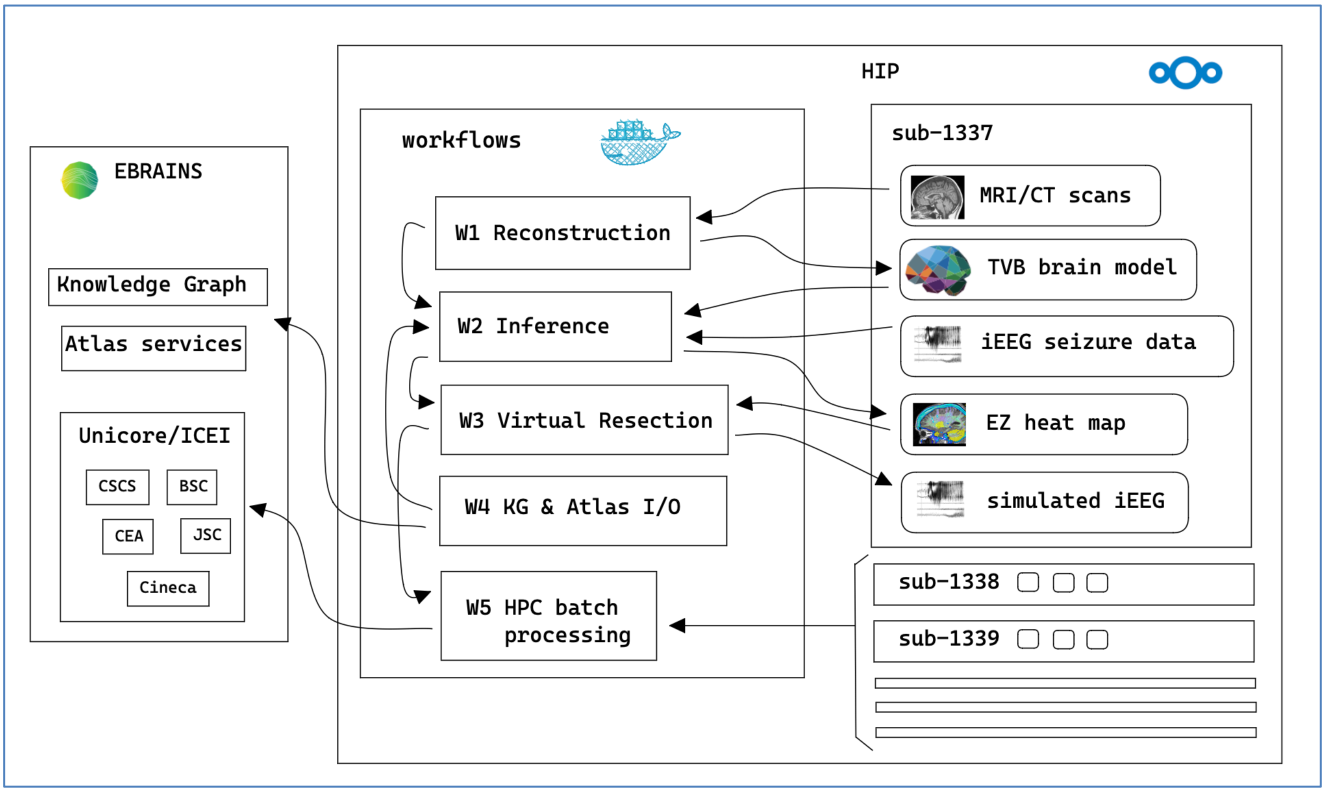

Figure 12: Interoperability on HIP with workflows, including TVB modelling

In this diagram the structural and functional workflows available on HIP are shown, illustrating the interoperability with the wider EBRAINS platform. Key aspects of this interoperability are (a) the ability for workflows to query the KG and Atlas services (b) make use of ICEI HPC resources for suitably de-identified data and (c) stage data between HIP and the EBRAINS platform as required for diverse use cases. The different demonstrator workflows foreseen already target modelling clinical epilepsy but pave the way for future user workflows.

2.8.1 The HIP structural and functional processing workflows

In addition to the suite of preprocessing apps provided on HIP, several workflows are available for performing complex processing of multimodal intracerebral EEG (iEEG) neuroimaging datasets. Frequently acquired alongside iEEG data are supplementary structural images such as T1, T2 and diffusion-weighted images, which serve to identify cortical and subcortical structures and connectivity. In order to relate the activity observed on iEEG electrodes to specific brain structures and connectivity, an image with the electrodes, T1-weighted MRI or CT scan, are also used.

While these different images can be worked with one by one in the individual apps on HIP, it is useful to put the data together to construct a single structural model fo the patient's brain and electrodes. This is the first structural workflow provided on HIP, as part of the TVB app, which has as outputs the set of cortical and subcortical geometries, the connectomes (via MRtrix3) and forward model (via Brainstorm/OpenMEEG) for the iEEG electrodes. Importantly, attention is paid that the reconstruction of these outputs occurs within a single reference frame such that the results are immediately usable for analysis and simulation.

Processing of functional iEEG can start with signal processing for instance using the MNE-Python app to extract relevant data features such as time frequency power densities, or statistics on connectivity as performed in the Frites app. Because the electrodes are implanted directly in the brain, these analyses on iEEG are typically performed in the sensor space only. On HIP, the model inversion workflow in the TVB app, can be used to take the data features obtained with MNE or Frites and estimate parametric maps over the source space be it regions of interest or the full cortical surface. In the example workflow provided, the model inversion produces a parametric map of epileptogenicity, which can be exported as a Nifti image and visualized as an overlay in other tools such as 3D Slicer or HiBop.

For datasets which are suitably de-identified, users can obtain an allocation for ICEI Fenix HPC resources, allowing the above workflows to bring significantly more compute resources to bear to the scientific or clinical question. This can be particularly effective for cohort level analyses or methodology testing, and the resources are straightforward to integrate into the workflows.

Finally, in addition to workflows which bring EBRAINS data and services to HIP, a first workflow by AMU on bridging the HIP to EBRAINS has fed the EBRAINS Atlas services (Service category 2, SC2) with SEEG data as if processed on the HIP using the CCEP manager. These data represent the F-TRACT atlas (f-tract.eu), which provides from CCEP unique information on long-range fibres' directionality and dynamics. These data are integrated using the OpenMinds framework of the EBRAINS Knowledge Graph. This work was performed into two phases.

2.8.2 Integration of F-Tract in the Multilevel Human Brain Atlas

The F-Tract ERC project has collected SEEG recordings from approximately one thousand drug-resistant epileptic patients. A subset of this data has already been integrated into the EBRAINS Knowledge Graph, specifically the (EBRAINS - CCEP database of the Medical Informatics Platform[54]; DOI: 10.25493/SV5Z-FSB). Integration of this large dataset into the HBP was performed through interactions between WP4 and WP1, as well as interactions between the Atlas Services and Medical Data Analytics Service (SC5) within Work Package 4. This integration was performed in two steps.

First, as part of our work in WP1 on iEEG-based connectomes, we generated a human brain atlas from the F-TRACT cohort, considering 613 adult subjects. This atlas contains a collection of connectivity features, including conduction axonal delays, synaptic constants and velocities obtained with the dynamic causal modelling (DCM) method described in [Lemaréchal et al. BRAIN, 2021 – P3210]. Furthermore, this atlas has been generated in four versions, yielded by application of two independent conditions. The first condition determines whether only early responses are considered. Applying it, focuses the atlas onto direct connections. In the alternative version late responses are also considered. The second condition removes or not those data that might be related to epilepsy-originating abnormalities. We projected this atlas onto a number of parcellations, including the Juelich parcellation 2.5, thereby allowing for referring it to other components of the multi-scale atlas. The first version of the F-TRACT atlas in the knowledge graph is accessible by searching for "f-tract": https://search.kg.ebrains.eu/instances/dc0750b4-1dd1-42d8-9e50-7f2fa97db085

Second, the integration of all F-TRACT information into the multi-scale atlas, under the Juelich cytoartchitectonic parcellation scheme, was taken over by the data curation team in Juelich (JUELICH). During the transfer from AMU to JUELICH, it appeared that the Juelich parcellation 2.5 was no longer supported by the data curation team. AMU thus generated a new version of the F-TRACT atlas including the Juelich parcelation 3.0. This was performed during July and August 2023, including its publication on the knowledge graph. Currently the data are still under embargo, awaiting clearance from the EBRAINS curation team. The F-TRACT connectivity maps are thus integrated into the online [55]. In addition to providing connectivity patterns from the SEEG-derived probabilistic atlas, the viewer will enable users to explore properties of connections between specific brain parcels, such as the speed of neuronal signal transmission. The F-TRACT data descriptor was integrated into the Atlas Viewer through OpenMinds. ,

[54] CCEO database https://search.kg.ebrains.eu/instances/Dataset/ebe50517-41d5-4029-9355-04f1e49e23c8
[55] Multilevel Human Brain Atlas https://ebrains.eu/service/human-brain-atlas/

2.9 The HIP Looking Forward

- The development of the platform will continue throughout the phase and beyond the end of the HBP. Currently ongoing tasks are:
- Framework development for supporting auto-deployment of apps-in-browser for development teams and, once in production, for onboarding of external tools, brought in by end-users.
- Optimisation of file transfer throughput to load data in remote apps.
- BIDS manager conversion (Common Workflow Language, CWL)
- Apps like the CCEP manager, which enables the processing of iEEG data and extraction of Cortico-Cortical Evoked Potentials (CCEP) features, are being prepared for installation on the HIP. Specific use cases, including SEEG electrode localization, epileptic seizure analysis, and connectivity analysis from electrical stimulation, are being designed and will be fully documented in the production phase.
- AMU-UGA is re-implementing the F-TRACT database processing pipeline with a dedicated BIDS-app called the CCEP manager.
- As the HIP integrates more advanced tools and workflows, there is a growing demand for an enhanced user interface. The current prototype UI will evolve into an advanced HIP User Interface, developed collaboratively with end-users to meet their needs.

Storage and compute operations

- Security: A prospective integration with the EBRAINS service for sensitive data, the Health Data Cloud (HDS)[56] will be investigated. Synergies in respect of security, data safety and high availability modalities, as required for an EBRAINS integrated service will be identified, benefitting both the HDS and HIP in their maturation.
- Storage facilities will be geographically distributed across multiple countries and data centres to ensure adaptive ethics and legal compliance. A Proof-of-Concept setup is currently being pursued with the CEA[57] (Commissariat à l'énergie atomique et aux énergies alternatives) in France and will be expanded to include the JSC [58](Jülich Supercomputing Centre) in Germany.
- Compute operations can, if requested, in the future be localised to specific European regions to adhere to ethical guidelines and compliance regulations.

In future we plan to build on, advance and integrate existing technologies in continued joint collaboration between MIP/HIP & HDC teams towards an interoperable network of platforms with different functional scopes tailored to their user communities. We plan to also package and offer the entire suite of these technologies in a way compatible with the EOSC Interoperability Framework and the EHDS. Compliant with data protection legislation and other local, national and European regulations, we will jointly develop a complete accountability toolset, e.g., automated contract services and logging, thereby supporting cloud hosted workflows across computer centres in different countries.

[56] Health Data Cloud (HDS) https://www.healthdatacloud.eu/
[57] CEA https://www.cea.fr/english/Pages/Welcome.aspx
[58] JSC https://www.fz-juelich.de/de/ias/jsc

3. Annex

In this Annex is some additional information about the platforms, some summary tables for relevant links.

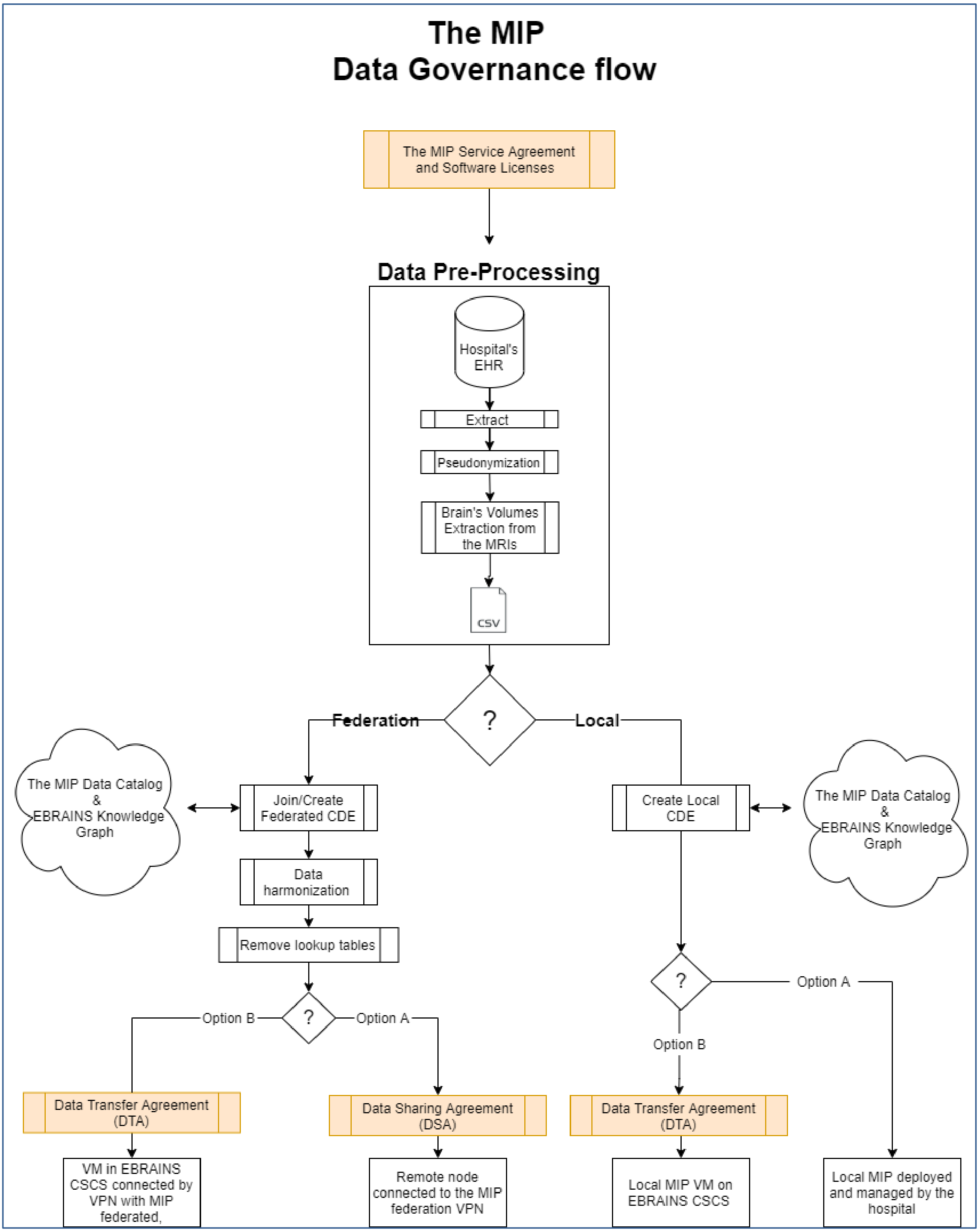

Figure 13: The MIP Data Governance flow

The diagram shows the different possible set-ups of the MIP, i.e., federated and local deployments. Relevant legal contracts and data processing steps are indicated.

Table 1: The HIP Library - In-Browser Apps (software/tools)

App integrated	Short description / LINK
Anywave	Matlab-based software made for processing EEG, MEG, ECoG and SEEG. Developed by AMU (HBP, France) https://meg.univ-amu.fr/wiki/AnyWave
Brainstorm	Matlab-based software made for processing EEG, MEG, ECoG and SEEG. Developed by University of South Carolina (USA) and Montreal Neurological Institute (Canada). https://neuroimage.usc.edu/brainstorm/Introduction
Localizer	Application to process raw brain recordings to obtain the frequency responses using a combination of band pass filter and Hilbert Envelope. https://github.com/CRNL-Eduwell/Localizer
HiBoP	Unity-based software made for visualisation of SEEG data. Developed by UCBL (HBP, France). https://github.com/hbp-HiBoP/HiBoP
FSL	Community-based software for imaging visualisation. https://www.slicer.org/
MRIcroGL	Compiled software made for the visualisation of MRI. Developed by University of South Carolina (USA). https://www.nitrc.org/projects/mricrogl/
BIDS_Manager	Python-based software made for creating and managing BIDS-iEEG databases: Developed by AMU (HBP, France). https://github.com/Dynamap/BIDS_Manager
Micromed	Software to review SEEG recordings. https://micromedgroup.com/download/systemplus-evolution-user-manual/ SystemPlus EVOLUTION™ User Manual - Micromed Group
dcm2niix	A script to convert MRI data. https://github.com/rordenlab/dcm2niix
Freesurfer	Software made for processing of MRI data. Developed by Harvard University (USA). https://surfer.nmr.mgh.harvard.edu
Filemanager	PCManFM-Qt is a Qt-based file manager which uses GLib for file management. It was started as the Qt port of PCManFM, the file manager of LXDE. https://github.com/lxqt/pcmanfm-qt
GARDEL	GARDEL is a tool for automatic segmentation and labelisation of SEEG contacts. https://meg.univ-amu.fr/wiki/GARDEL:presentation
LibreOffice	LibreOffice is a free and powerful office suite, and a successor to OpenOffice.org (commonly known as OpenOffice). https://www.libreoffice.org
MNE-Python	Open-source Python package for exploring, visualising, and analysing human neurophysiological data: MEG, EEG, sEEG, ECoG, NIRS, and more. https://mne.tools
Frites	Frites is a Python toolbox designed for assessing information-based measures on human and animal neurophysiological data (M/EEG, Intracranial) for the discovery of cognitive brain networks. https://github.com/brainets/frites
MRIDeface	This package contains an algorithm for removing identifiable facial features (eyes, nose, and mouth). https://surfer.nmr.mgh.harvard.edu/fswiki/mri_deface
Niftyreg	NiftyReg is a command that allows to perform rigid, affine and non-linear registration of 2D and 3D images stored as Nifti or Analyze (nii or hdr/img). https://github.com/KCL-BMEIS/niftyreg
RStudio	RStudio is an integrated development environment (IDE) for R and Python. https://posit.co/products/open-source/rstudio
VSCode	Visual Studio Code is a lightweight but powerful source code editor which runs on your desktop. https://code.visualstudio.com
3D Slicer	Slicer is a free, open-source software for visualization, processing, segmentation, registration, and analysis of medical, biomedical, and other 3D images and meshes; and planning and navigating image-guided procedures. https://www.slicer.org/
TVB-HIP-APP	https://github.com/ins-amu/tvb_hip
IntrAnat	IntrAnat is a free database and visualisation software for Intracranial Electroencephalographic (iEEG) data. https://f-tract.eu/software/intranat it
Brainvisa	BrainVISA provides a complete, modular, infrastructure for neuroimaging software. It helps organising heterogeneous software and data and provides a common general graphical interface for users. https://brainvisa.info
Matlab	https://mathworks.com/products/matlab.html A programming platform designed specifically for engineers and scientists to analyze and design systems and products.
TRCAnonymizer	https://github.com/floriansipp/TRCAnonymizer A software allowing to check if your Micromed data files are anonymised and provides the means, through a simple interface, to do so.
Apps under preparation for integration	
CCEP Manager	https://ccepmanager.readthedocs.io

Application listed in Table 1Table 1 were installed on the HIP and subjected to thorough testing. The performance of each software has been carefully evaluated, and feedback from end-users and tool developers has facilitated the adaptation of functionalities based on user requirements and the fine-tuning of additional plugins.

Table 2: Documentation and other references

Technical Documentation	
Main MIP Technical Documentation Repository	https://github.com/HBPMedical/mip-docs
MIP Deployment	https://github.com/HBPMedical/mip-deployment/
MIP Federated Algorithms	https://github.com/madgik/exareme/tree/22.2.3/Exareme-Docker/src/mip-algorithms
MIP Gateway	https://mip-front.gitbook.io/mip-gateway-doc/
Other relevant links for the MIP	
The Medical Informatics Platform (MIP) - website	https://ebrains.eu/service/medical-informatics-platform/ https://www.ebrains.eu/health-research-platforms/health-platforms/work-with-health-data-2
The PUBLIC MIP	https://hbpmip.link/
Redolfi A, et al., 2020: Medical Informatics Platform (MIP): A Pilot Study Across Clinical Italian Cohorts. Front. Neurol. 11:1021. **P2409**	https://doi.org/10.3389/fneur.2020.01021
The Medical Informatics Platform (MIP): Analysis and options for its exploitation in the healthcare market	https://sos-ch-dk-2.exo.io/public-website-production/filer_public/82/4a/824ac0ed-d2f6-495f-b560-c5647838c46a/upm_hbp_mip_final_dt.pdf
Projects and Partnerships beyond HBP	
IMI Sophia	https://www.imi.europa.eu/projects-results/project-factsheets/sophia
IMI EPND	https://www.imi.europa.eu/projects-results/project-factsheets/epnd
eCREAM	https://ecreamproject.eu/
European Academy of Neurology	ean.org - Welcome to the European Academy of Neurology
European Stroke Organisation	https://eso-stroke.org/
Integration with other EBRAINS services - Dataset examples of the MIP in the EBRAINS Knowledge Graph	
CCEP database of the Medical Informatics Platform	https://search.kg.ebrains.eu/instances/Dataset/ebe50517-41d5-4029-9355-04f1e49e23c8; DOI: 10.25493/SV5Z-FSB
Example for MIP Metadata representation in the EBRAINS Knowledge Graph	https://search.kg.ebrains.eu/instances/a526f0c6-cbd4-4864-ae3a-87ad2cb7496a; DOI: 10.25493/2XWW-K4Y
Relevant links for the HIP	
Technical Documentation Repository	https://github.com/HIP-infrastructure
HIP User Documentation	https://github.com/HIP-infrastructure/hip-doc https://github.com/manikbh/HIP-doc
Human Intracerebral EEG Platform (HIP) Website	On HBP: https://www.humanbrainproject.eu/en/medicine/human-intracerebral-eeg-platform/ and on EBRAINS: https://www.ebrains.eu/health-research-platforms/health-platforms/work-with-health-data-2
Nextcloud	https://nextcloud.com/industries/healthcare/
Holdgraf et al., Sci Data, 2019 **P2033**	https://www.nature.com/articles/s41597-019-0105-7
CCEP Data Base	https://search.kg.ebrains.eu/instances/Dataset/ebe50517-41d5-4029-9355-04f1e49e23c8
F-tract	https://f-tract.eu/
Health Data Cloud (HDC)	https://www.healthdatacloud.eu/
General links relevant for the EBRAINS Medical Data Analytics Service	
EBRAINS Website	https://ebrains.eu/
EBRAINS General Terms and Policies	https://www.ebrains.eu/page/terms-and-policies