

The Open Science Movement

Maria Panagiotopoulou (ECRIN),
Birgit Schaffhauser (CHUV)

November 3, 2022
BOLOGNA, IT



Funded by
the European Union

eCREAM project
N. 1010557726

Data Sharing – The Way Forward

Growing need to **increase capacity to share health data** to accelerate progress in clinical sciences and medicine - numerous initiatives aim at collating big data.

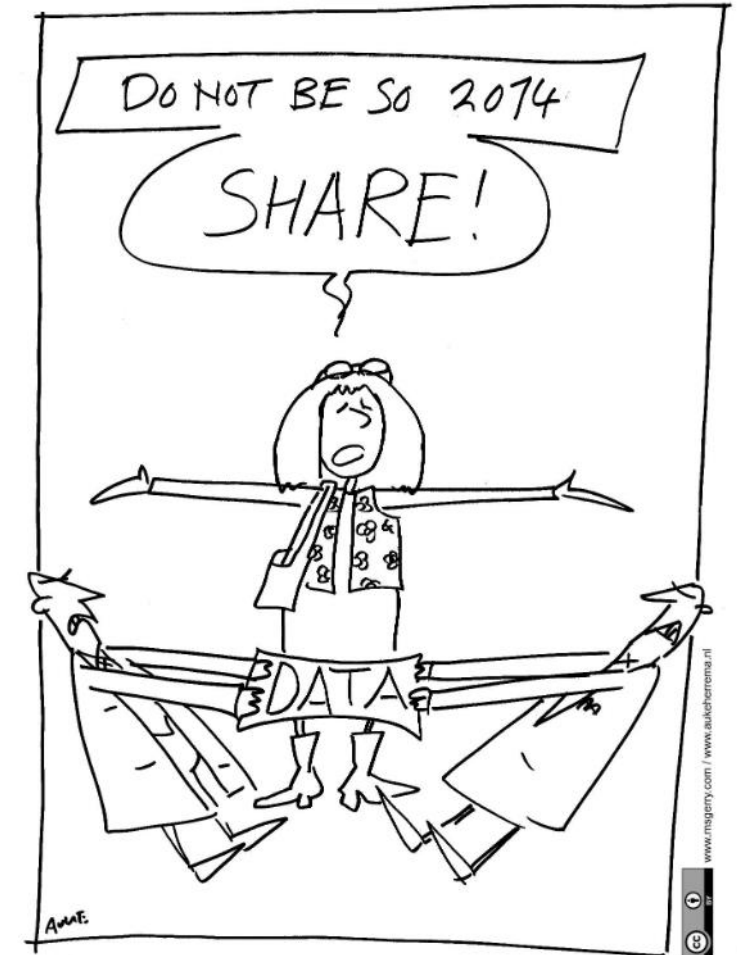
General understanding

- Solving current and future challenges in modern healthcare, **data, computational power and analytics expertise need to be combined**.
- Huge need for **high-quality big data** in the context of machine learning and AI.
- Demand on **integrated access to high-quality health data** from scientific communities, health care providers and policy makers.

Data privacy and ethics policies impose severe restrictions on health data access and sharing - delineate the **boundaries for development** of new solutions.

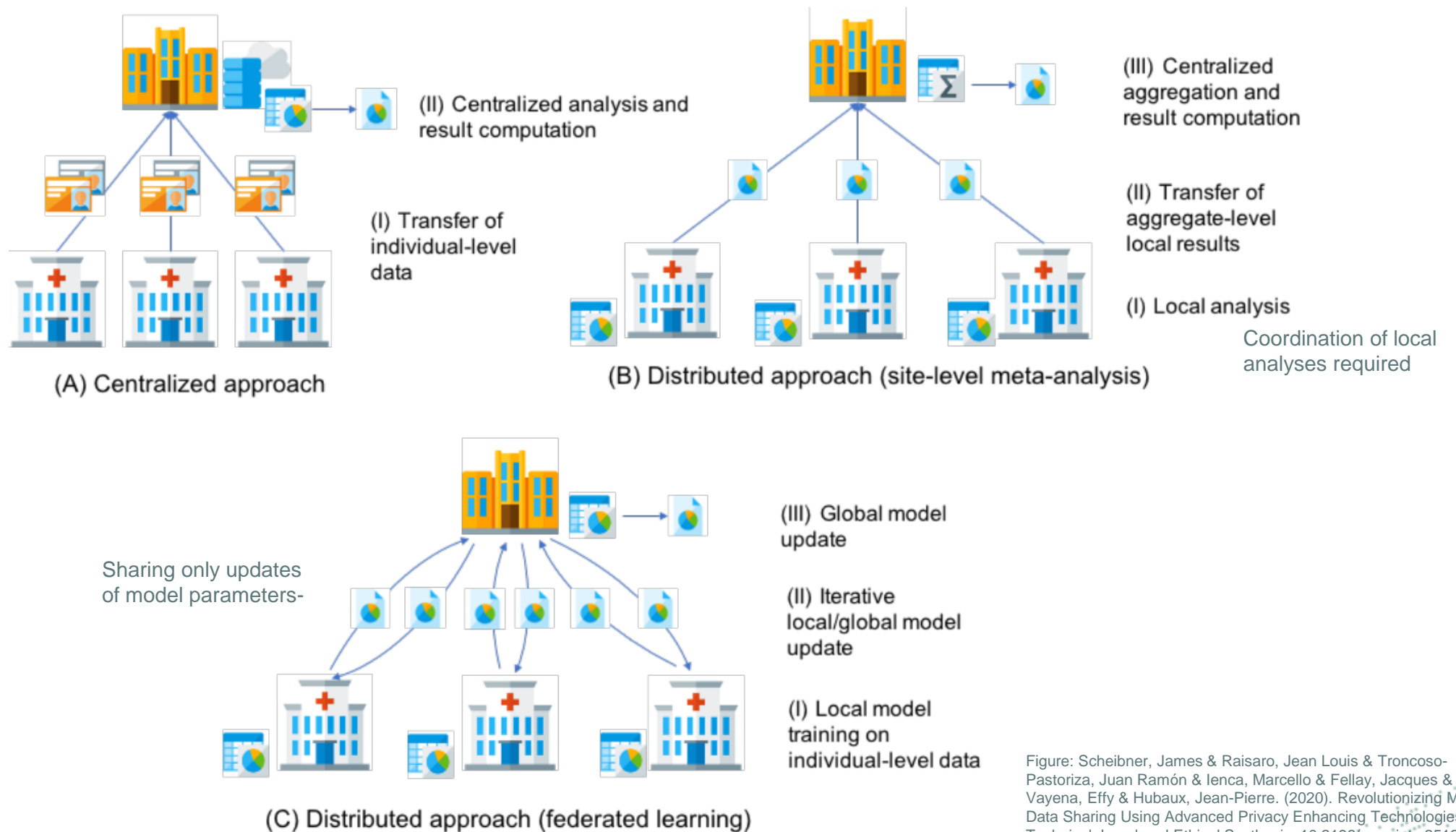
Promotion of Data Sharing is essential

- Implementation of standardization and harmonization rules enables collaborations and reduces burden of unexploited data
- Data interoperability and re-usability increase reproducibility and reduces duplication of efforts
- Improvement of data management, productivity and quality of health studies



Source: <https://twitter.com/ChrisBanks/status/524829741787127808/photo/1>

Data Sharing - Federated and Centralized Approaches



Data Sharing – Federated and Centralized Approaches

Take home message - no solution fits all purposes – no solution is the best - advantages and disadvantages in all the complementary environments

Leverage synergies, mix and match

Relevant for all approaches – **FAIR PRINCIPLES**
(**F**indable, **A**ccessible, **I**nteroperable, **R**e-usable)

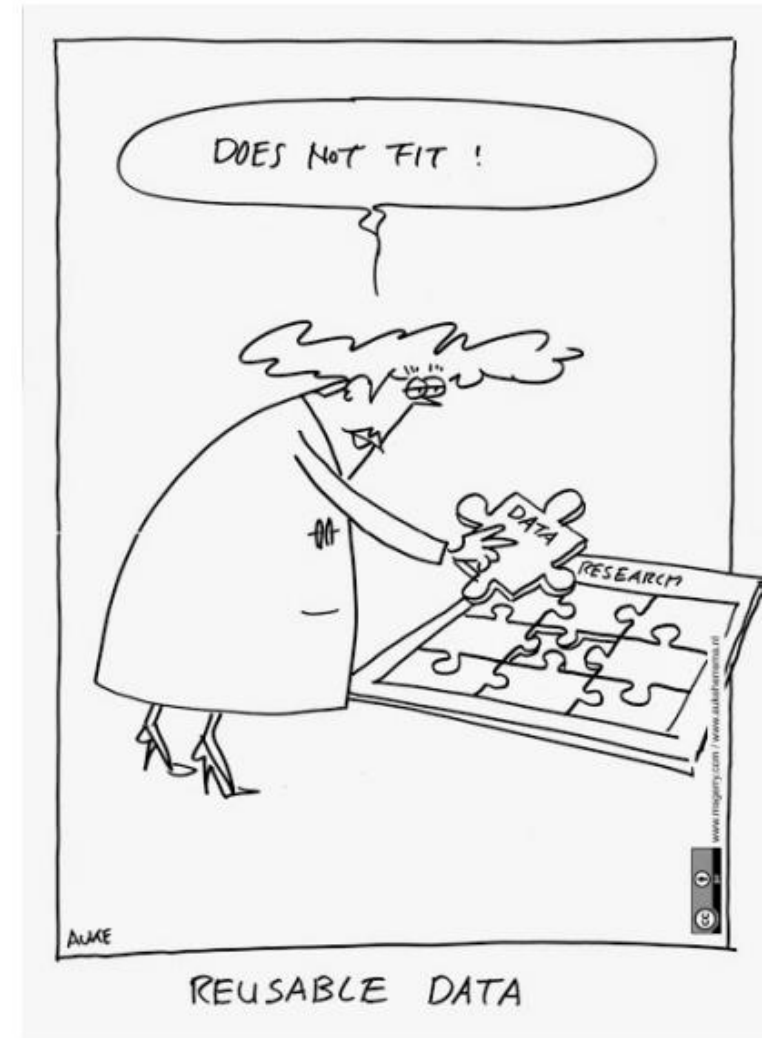


Image by A. Herrema & H. Bouwteam (CC-by)

Reference: [Microsoft PowerPoint - 3_ethics.pptx \(fosteropenscience.eu\)](#)



Data Sharing in eCREAM

Three main aims:

- 1) Develop new technical solutions to extract clinical information from structured and unstructured data contained in different electronic patient files
- 2) **FAIRify** the established databases for clinicians, researchers, health policymakers and citizens, respecting the European and national legislations
- 3) Pilot the exploitation of the established databases in two use cases:
 - i) assessment of ED propensity to hospitalise a patient, and ii) development of a dashboard to be used by citizens and policymakers to improve the quality of care in ED

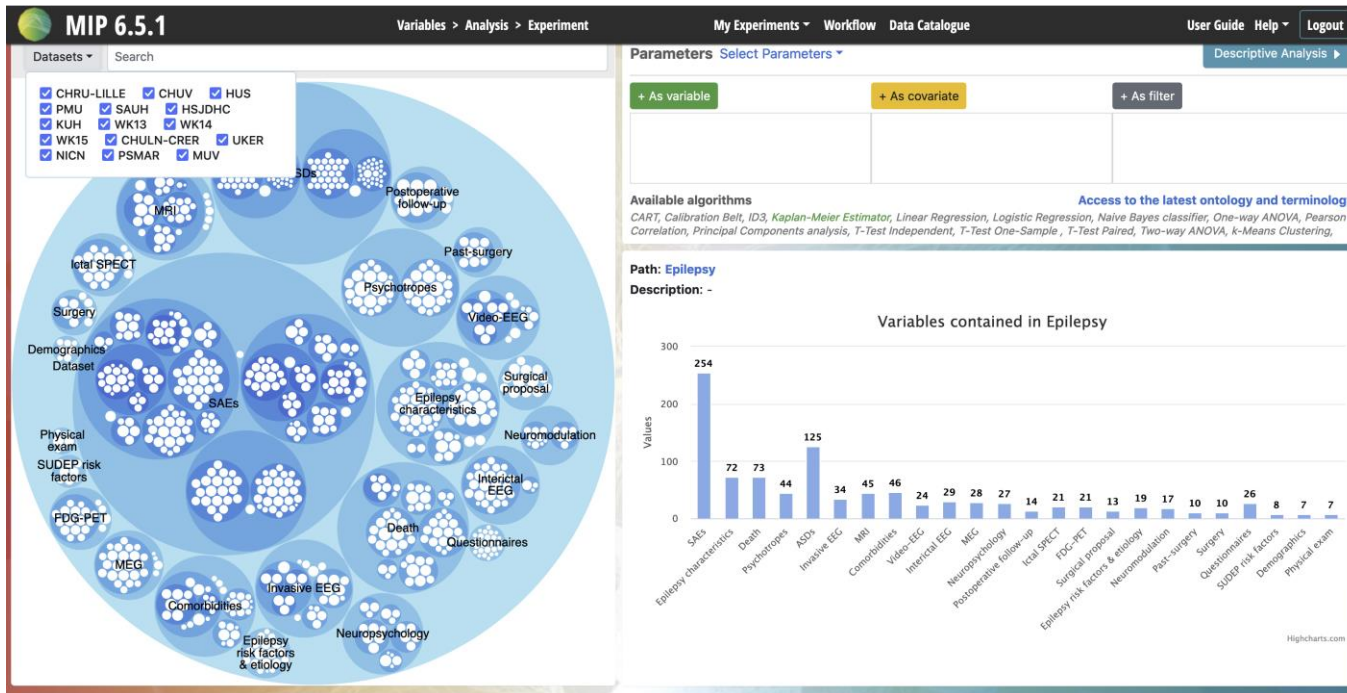
FAIRifying the eCREAM clinical databases



Major Impact: Increased **use and valorisation** of health data due to improved **data interoperability - standardization of meta knowledge** (metadata, ontologies and reference repositories) and **clinical data** from different clinical services and multiple countries.

Synergies and integration of outcomes with relevant European initiatives like **EHDS**, **EOSC-Life** and **HRIC** (European Health Research and Innovation Cloud).

Medical Informatics Platform - MIP



- Development since 2013 within the Human Brain Project (HBP)
- Privacy-preserving, **federated** data processing and analysis system
- Leverage and **re-use de-centralized real world data** and research cohort datasets
- Exploration of **anonymized, harmonized** data (datamodel, CSV and JSON format)
- **Integrated standard descriptive and analytical statistical tools** and machine learning algorithms
- Data reside at original site of storage
- Web-based user interface tailored for clinicians
- **MIP Data Quality Tool** supports data management (aligning datasets with datamodels)

eCREAM MIP Federation

Central MIP infrastructure hosted on the EBRAINS RI*, runs federated web-based queries, executes algorithms simultaneously in each node, aggregates results.

* EBRAINS RI, the HBP legacy, is a sustainable research infrastructure with the aim of advancing neurosciences, medicine, computing and brain-inspired technologies.

Principle: **Code visits Data**

Minimally invasive in IT, ethics and legal context through **virtual appliance technology**

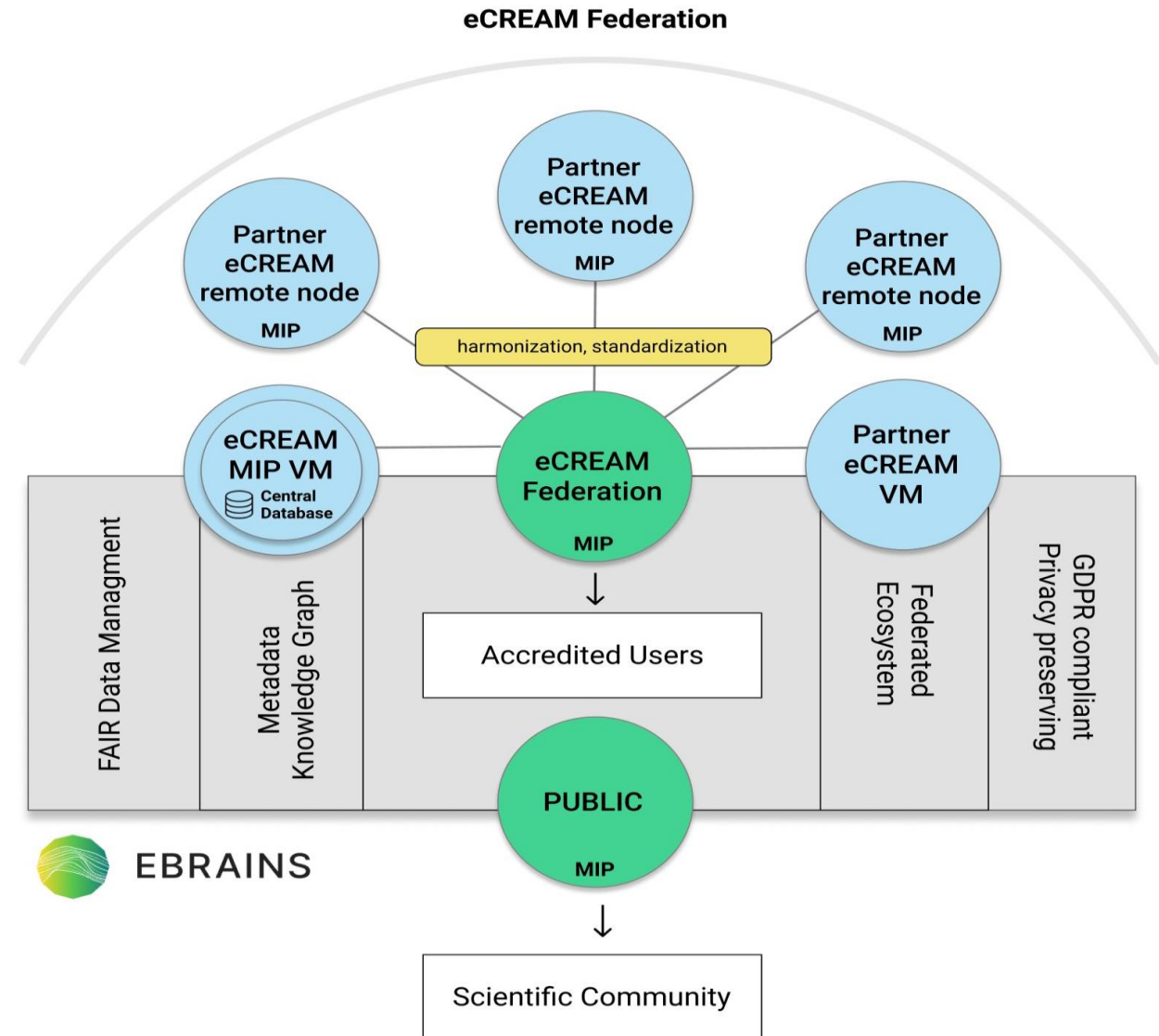


Image: Overview of eCREAM federation set-up, including Public-MIP, eCREAM MIP and the federated nodes integrated in EBRAINS

eCREAM Clinical Databases

Three databases of enormous value for the scientific community.



Clinical notes

Database for Clinical Notes: free-text notes on patients, EDs in 5 EU countries in 5 languages. Anonymised and enriched (age, sex, hospitalisation, diagnosis, and outcome at hospital discharge). Subset of clinical notes annotated by expert physicians (comorbidities, concomitant therapies, specific diagnostic test results, patient performance status, and symptom characteristics). **Unique for the NLP (Natural Language Processing) community.**



ED visits

Clinical Database of ED visits: structured EHR data on patients arriving at the ED. 25 EDs involved in data collection. Clinical data derived from currently used EHRs and administrative sources (patients' demographic and clinical characteristics, priority code assigned at ED triage, eventual hospitalisation, and 30-day vital status). **Available to authorised researchers only. Federation with other sources through the Medical Informatics Platform (MIP) envisaged.**



Patients' characteristics

Database on Patients' Characteristics: data on all patients visiting the participating EDs (date and time of arrival, age, sex, priority code assigned at triage, main acute symptoms at arrival, diagnosis at ED discharge, and final hospitalisation). **Anonymized Data will be uploaded onto the public MIP to be exploited by scientific community.**

eCREAM MIP Federation

Re-use of the eCREAM databases (DB)

Public-MIP:

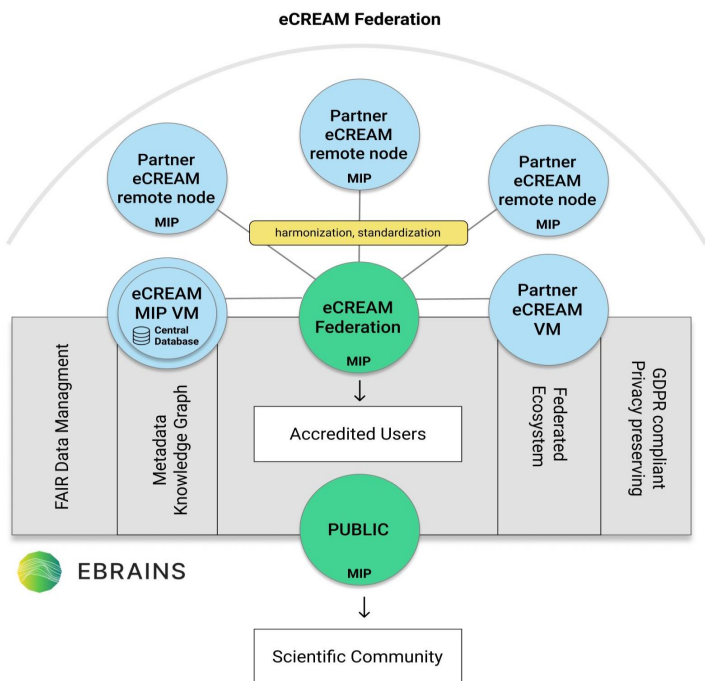
- accessible for any EBRAINS accredited user (personal or institutional accreditation)
- subset of the anonymized Clinical DB and Patients` Characteristics DB available to a broader community, including patient and citizen advocacy groups
- eCREAM steering committee defines complete list of variables to be shared
- non-expert users able to perform analyses, opening data to students and communities with limited access to clinical data

eCREAM-MIP:

- anonymized Clinical DB
- access restricted to a defined set of users (MIP - EBRAINS Identity and Access management, monitored and recorded)

eCREAM Federation:

- hospitals wishing to federate their data with the eCREAM-MIP invited to join
- installation of the MIP at each participating site
- anonymizing and harmonising datasets according to the eCREAM datamodel
- User trainings and support



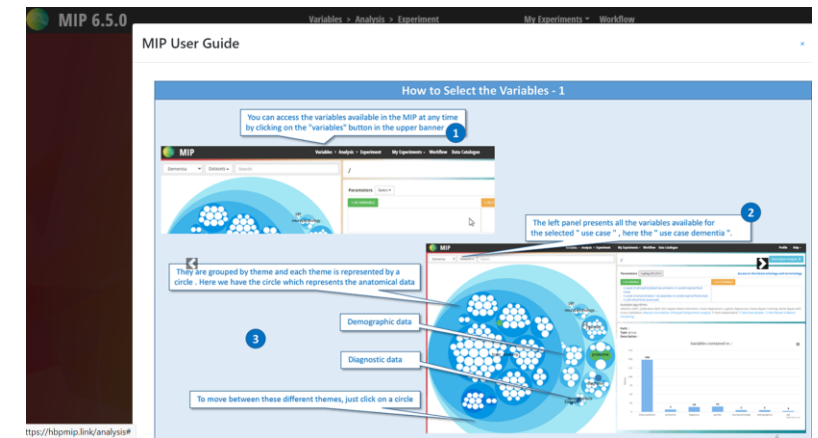
eCREAM MIP Federation

Some Data Privacy and Safety principles:

- All data must be **ethics cleared for scientific re-use**
- All data must be **anonymized**
- MIP user interface does **not allow to access, explore, manipulate, copy or download the data (CSV files)**
- Federated queries provide **aggregated findings calculated on at least 10 subjects**, prohibiting to perform indirect re-identification of individual subjects by reconstructing patient profiles through a specific series of queries
- **Updates of datasets require replacement of the entire dataset**, also relevant for e.g. management of consent withdrawal
- Users of the MIP eCREAM federation are **authorized and accredited** to query the federated datasets
- **Metadata will be open** to any accredited EBRAINS user
 - Users of the Public MIP will see the eCREAM data model, describing all its variables.
 - The **EBRAINS Knowledge Graph** developed by the Human Brain Project, will expose the same metadata information plus other relevant metadata (e.g. numbers of available patient records).



Federations





THANK YOU
FOR YOUR
ATTENTION

Birgit Schaffhauser
birgit.schaffhauser@chuv.ch



Funded by the
European Union