Co-funded by the European Union

Data sharing in EBRAINS

Curation task leder: Ida Aasebø

EBRAINS Data services leader and
HBP Infrastructure Director: Jan Bjaalie

Innlegg til utredningsgruppen - FAIR-utredningen

EBRAINS
Curation Services

What is EBRAINS?

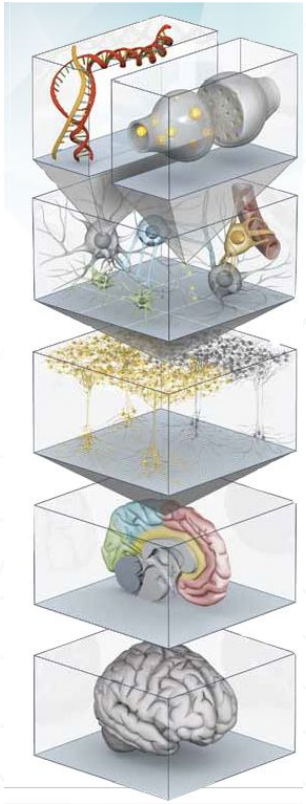
Co-funded by the European Union

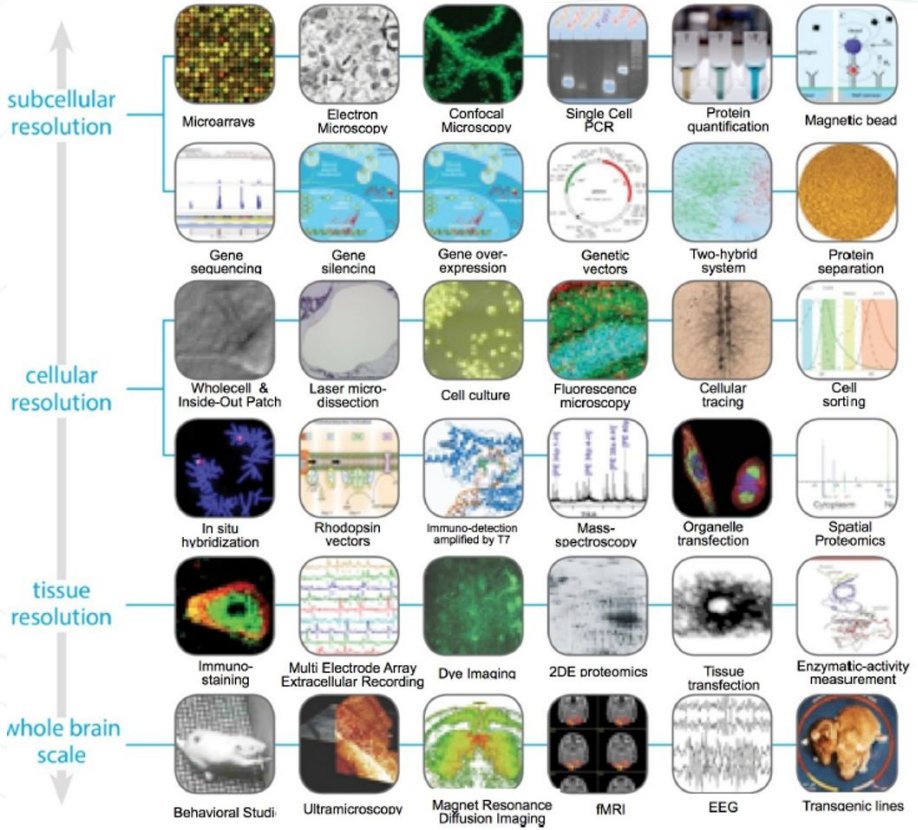
EBRAINS – European distributed infrastructure for brain and brain-inspired research

- A Research Infrastructure on the ESFRI Roadmap (European Strategy Forum on Research Infrastructures) since June 2021
- Developed by the EU Flagship Human Brain Project, launched as a Research Infrastructure in 2019
- Operated by a consortium of partners in Europe
- Central Hub in Brussels (EBRAINS AISBL, legal entity established by 7 institutions, including UiO)
- Providing tools and services assisting scientists in their research: collecting, analysing, sharing and integrating brain data, and performing modeling and simulation of brain function

Heterogeneous neuroscience datasets

Co-funded by the
European Union

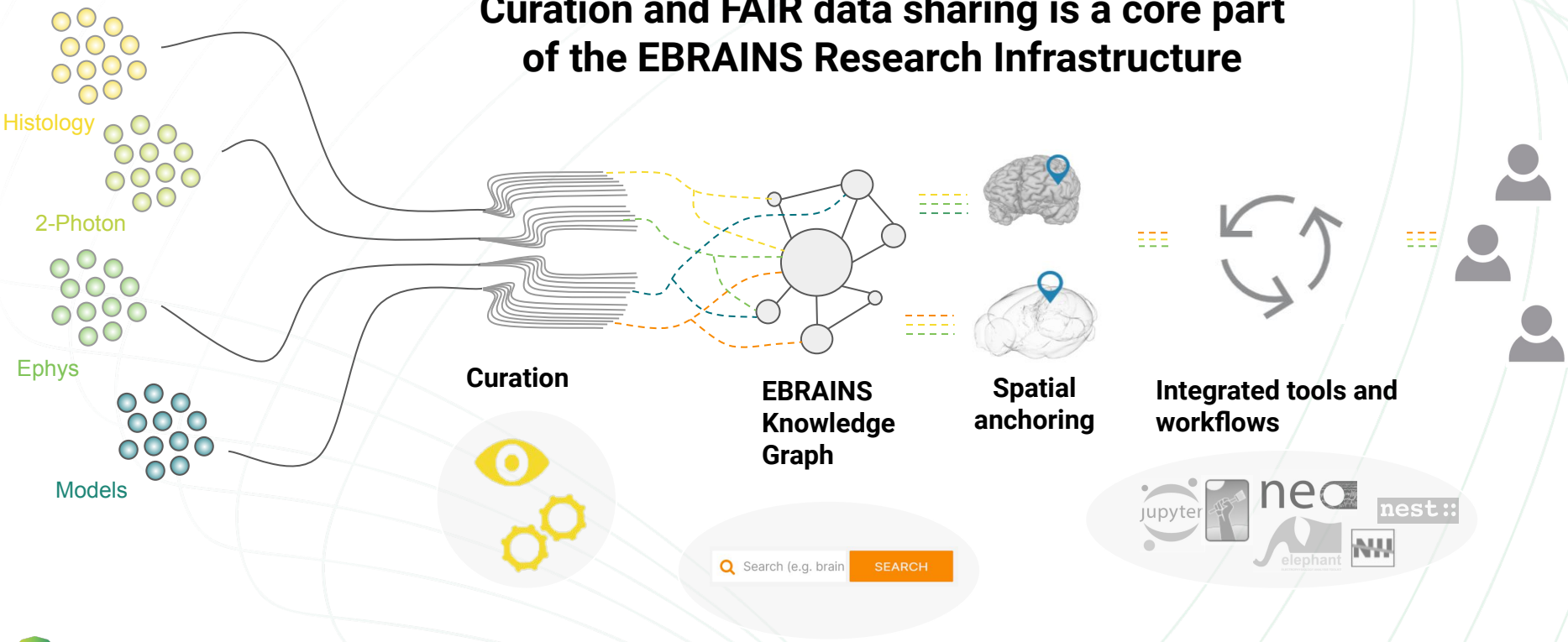




What we do

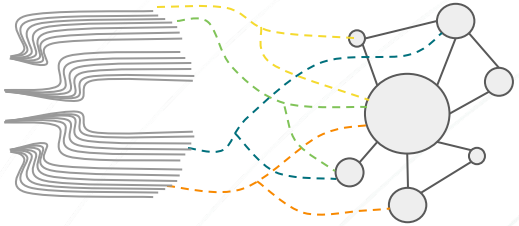
Co-funded by the European Union

Curation and FAIR data sharing is a core part of the EBRAINS Research Infrastructure



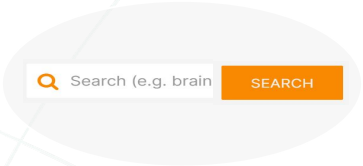
EBRAINS Knowledge Graph Search

Co-funded by the European Union

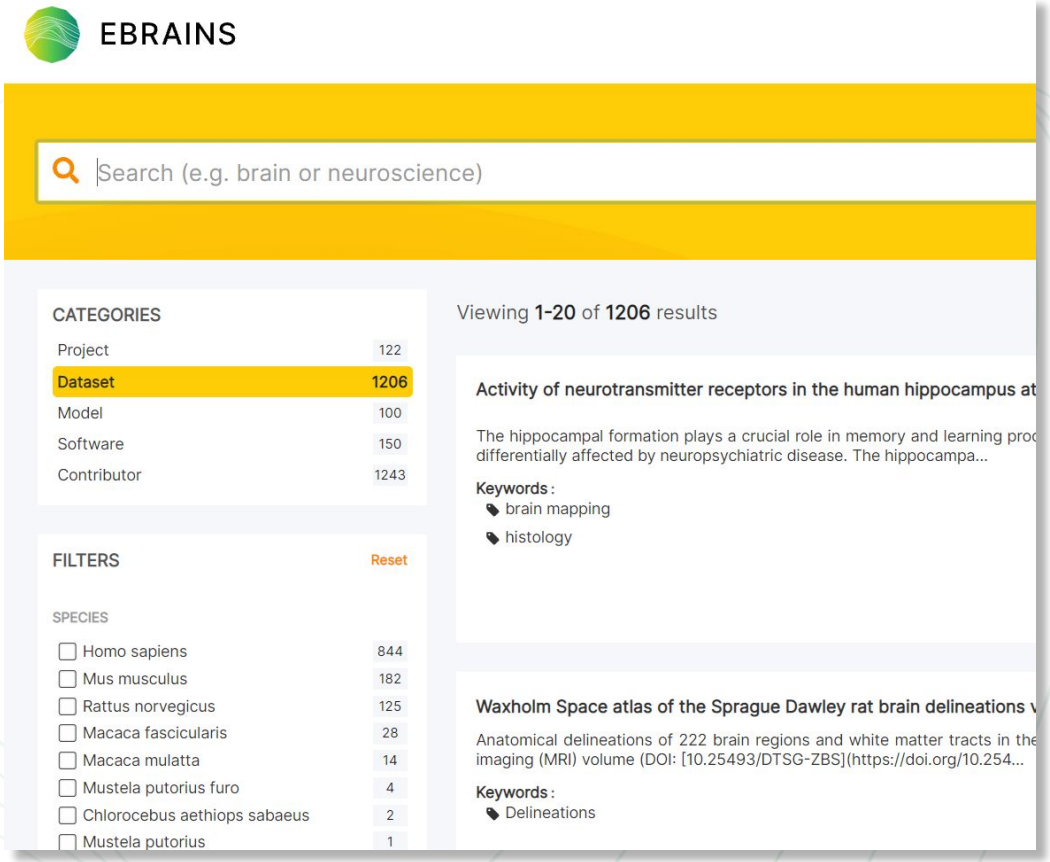


Curation

EBRAINS Knowledge Graph

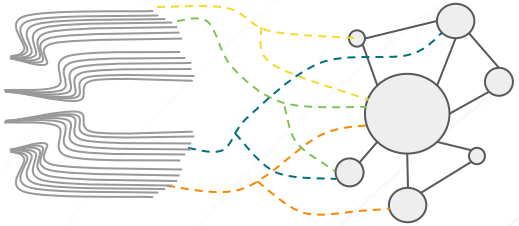


https://search.kg.ebrains.eu/



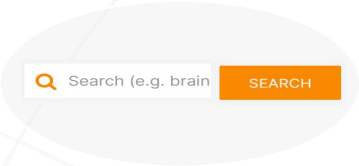
EBRAINS Knowledge Graph Search

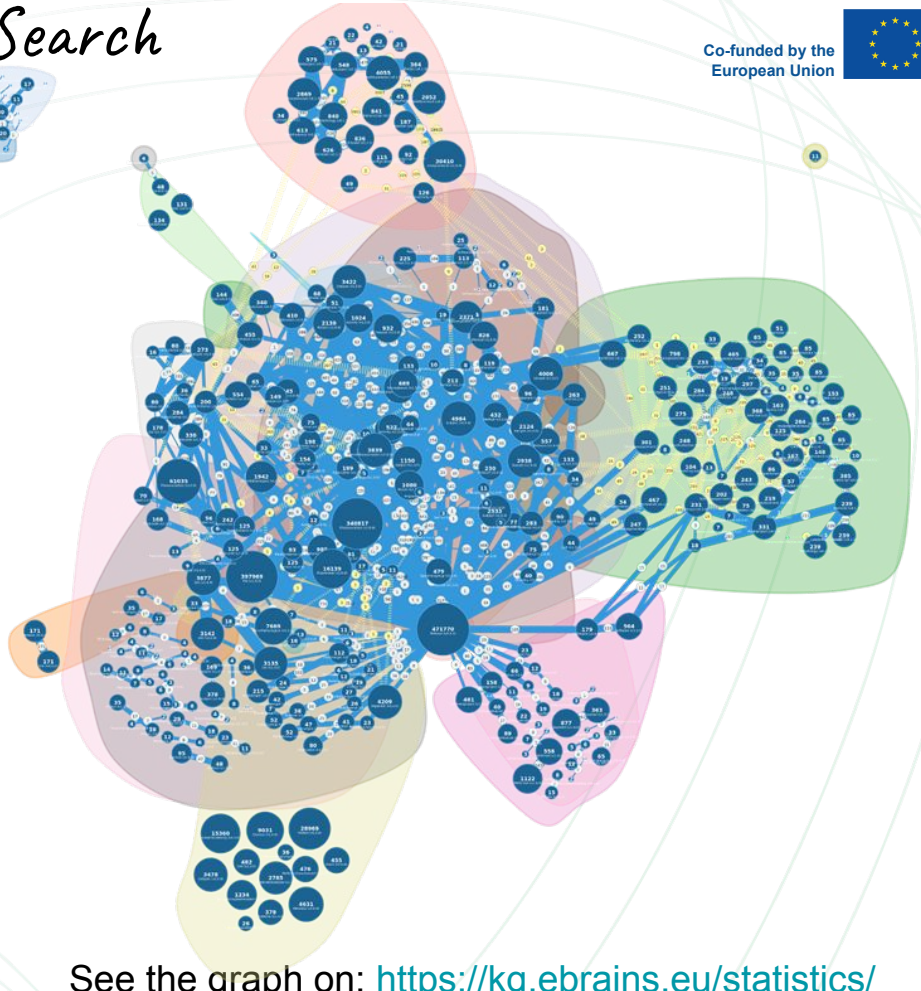
Co-funded by the European Union



Curation

EBRAINS Knowledge Graph

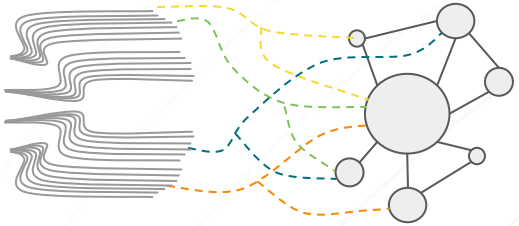




See the graph on: https://kg.ebrains.eu/statistics/

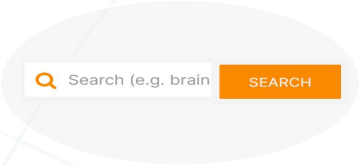
Curation request

Co-funded by the European Union

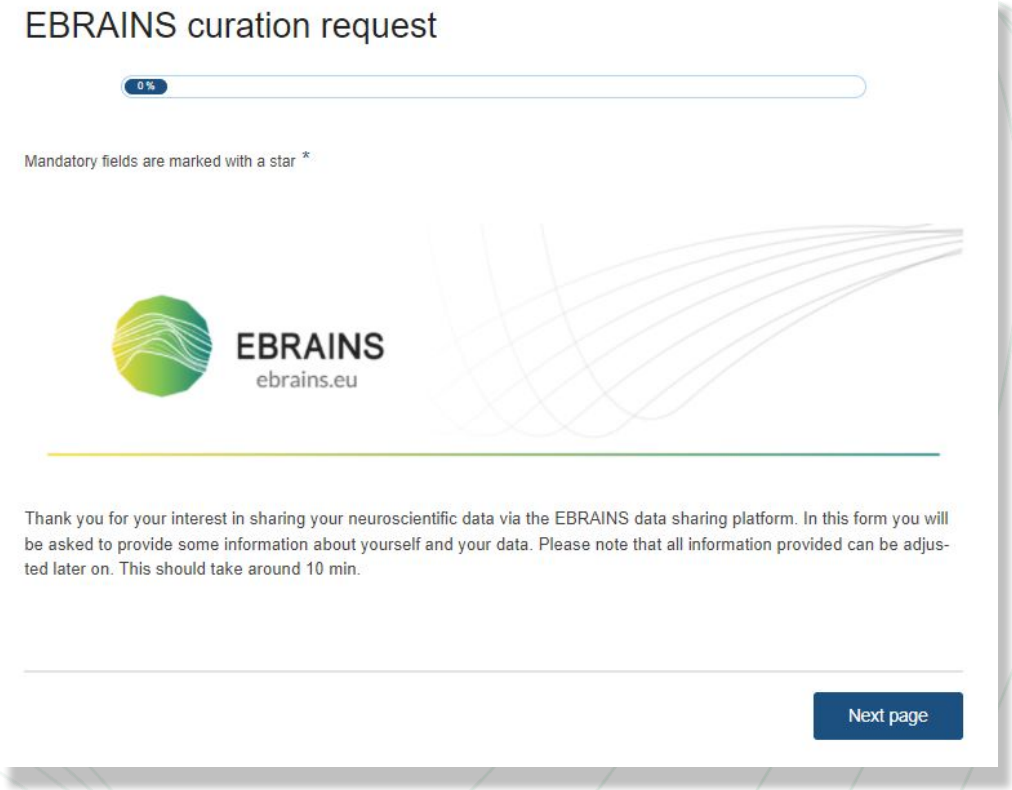


Curation

EBRAINS Knowledge Graph



To initialize data sharing on EBRAINS, submit a curation request at:



Curation workflow 3.0

Co-funded by the European Union

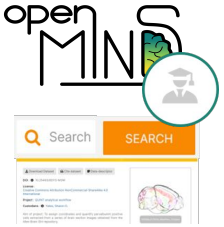
Track selection: **Opt-in / opt-out** where specifics (benefits and costs) are outlined for each researcher providing data

Choice

1 Common curation track

Submit metadata via an online form, or a python package (coming soon)

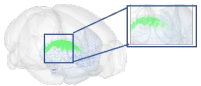
- **Faster**
- **Simpler**
- **Minimal level of FAIR**
- **Low effort**
- **Allow later enrichment**



Recommended for the average researcher

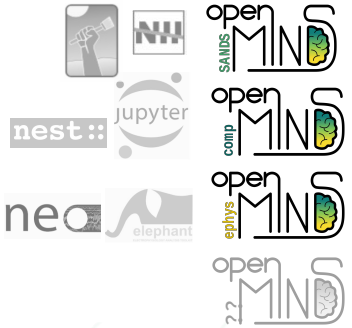
2 Advanced curation track

Currently programmatic submission of metadata (Python- openMINDS generator on Github)



- **Slower**
- **More complex**
- **High level of FAIR**
- **Tool compatibility**
- **Allow later enrichment**

For deeper integration with tools and workflows



Recommended for **large research groups OR groups with data management resources**

EBRAINS Communication work

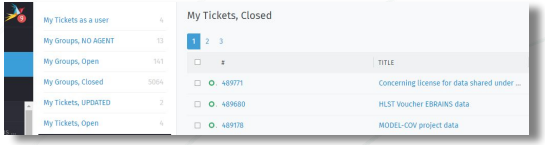
Co-funded by the European Union

- Data-sharing became **mandatory** in the HBP gradually as the project developed.
- The **reactions** from the researchers to this was quite diverse, but with a large proportion responding positively. And especially positive to the **assistance received** by the curation team in this process since they frequently were quite uninformed.
- As the **funders and large journals** began to **recommend** data sharing we have began to receive a regular flow of external data providers
- **Communicating** the benefit of a data sharing infrastructure based on **optimizing FAIR**, has resonated well in initiating a **culture to share.**

EBRAINS Communication work

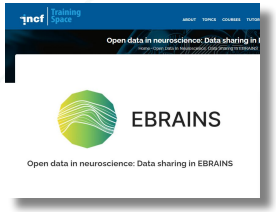
Co-funded by the European Union

Ticketing system:



Documented feedback and dialogue with single researchers (> 5000 tickets)

Digital courses:



E.g. https://training.incf.org/course/open-data-neuroscience-data-sharing-ebrains

Infographics and web work:

See our Curation collab: https://wiki.ebrains.eu/bin/view/Collabs/data-curation

Active social media presence:

@EBRAINScuration

LinkedIn

Active conference activities: talks and boots

EBRAINS Data and Knowledge Team 2021

Co-funded by the European Union

UiO, Norway

Camilla Hagen Blixhavn
Anne Marte Kvello
Eszter Agnes Papp
Ingrid Reiten
Ulrike Schlegel
Maaike van Swieten
Ida Aasebø
Trygve Leergaard
Jan Bjaalie

FZ Jülich, Germany

Anna Hilverling
Stefan Köhnen
Sara Zafarnia
Lyuba Zehl
Timo Dickscheid

KI, Sweden

Tom Gillespie
Mathew Abrams

CNRS, France

Florent Bonnier
Elodie Legouée
Peyman Najafi
Andrew Davison

DMU, UK

William Knight
Damian Eke
Peter Gierow
Simisola Akintoye

EBRAINS KG team, Belgium

Alexis Florian Durieux
Gilles Dénervaud
Ioannis Tsanaktsidis
Oliver Schmid

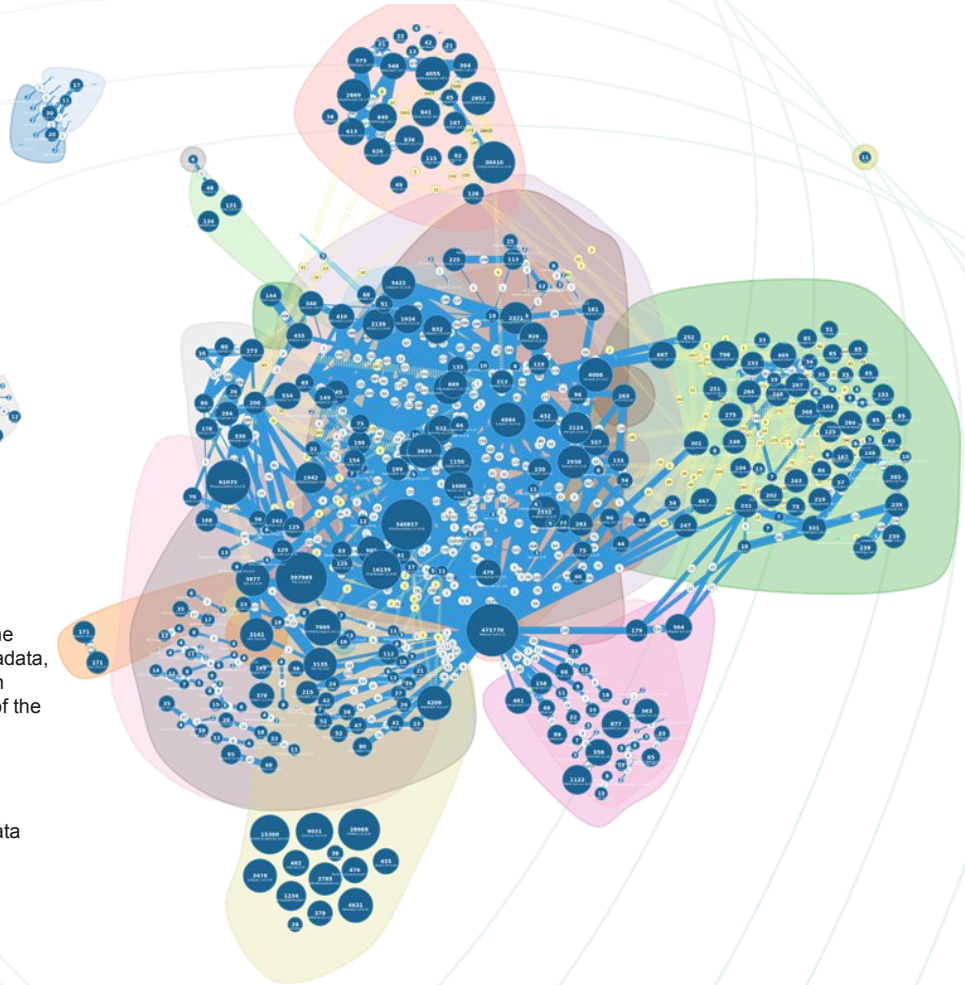
Co-funded by the
European Union

Backup-slides

Benefits of a Knowledge Graph:

- Less performance lags
- Highly scalable, reliable and distributed
- Enables complex relationships between heterogeneous data

A Knowledge Graph is a noSQL (not only SQL" database. They store relationships in a different way than relational databases.



EBRAINS KG:

Spring Boot

Spring boot is the technology for the service -> it provides all the business logic such as normalization and validation of the metadata, defining and resolving the identifiers, building the links between instances, and many things more. This is the "business logic" of the graph.

Primary store

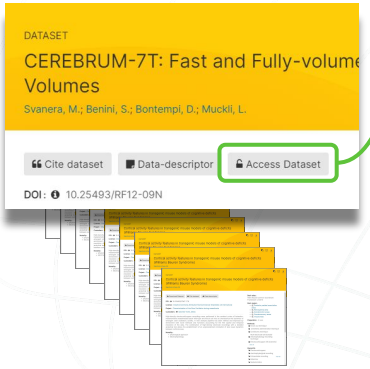
ArangoDB

ArangoDB is the underlying database storing the actual metadata

Human Data Gateway for pseudonymized datasets

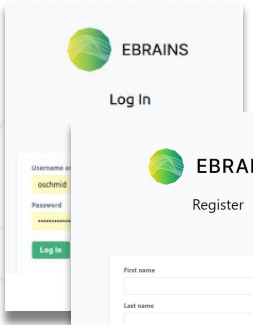
Pseudonymized human data can be shared through the Human Data Gateway (HDG)

EBRAINS Knowledge Graph Search. - "Controlled Access"

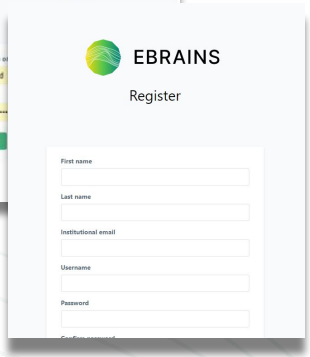


1 **Identify yourself**

..log in

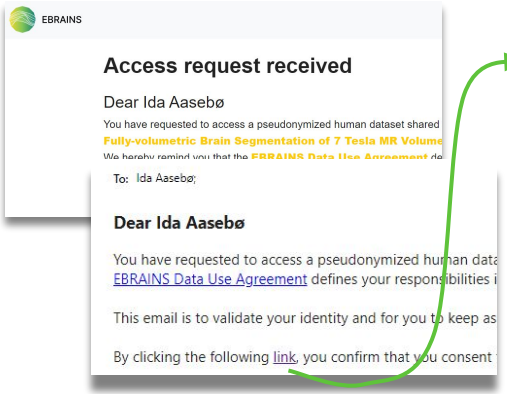


..or register



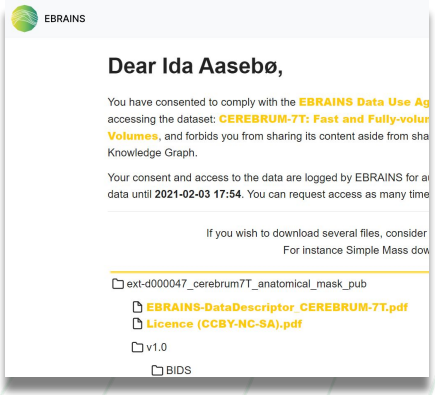
To consent to EBRAINS Access Policies

2 **Accept conditions**



..email verification step to accept conditions of the Data Use Agreement

3 **Access files**



Access is time restricted and logged

Workflow satisfies GDPR compliance requirements for pseudonymized datasets.

EBRAINS

EBRAINS - EBRAINS Knowledge Graph Search

KG Query API - Programmatic access

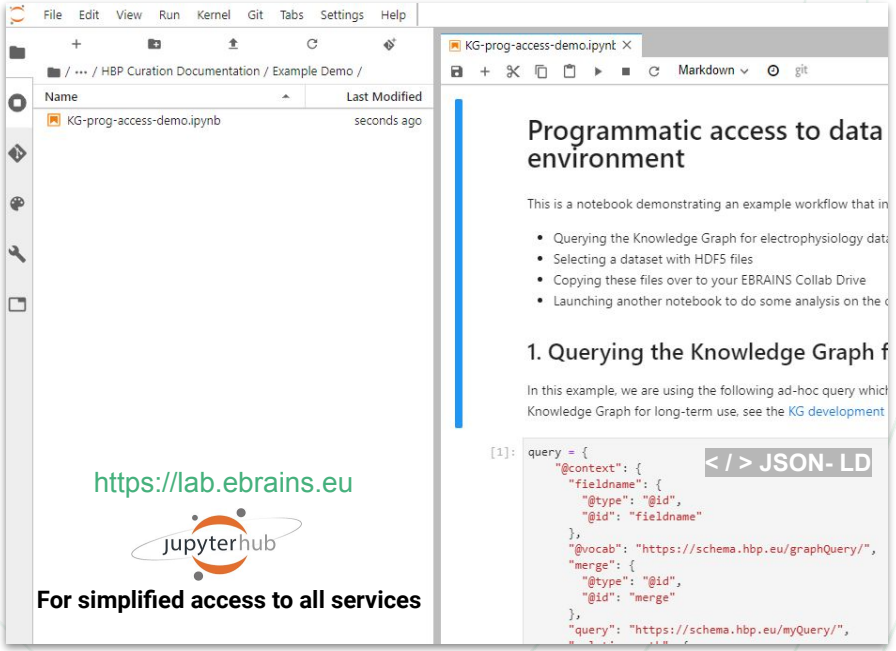
You need:

- **EBRAINS account**
- **Register and request credentials**

To use:

1) **Specify your EBRAINS KG query**
2) **Download datasets and metadata from your query target / collection**

Please see:
KG instructions: https://kg.ebrains.eu/develop.html
API:
https://kg.humanbrainproject.eu/apidoc/swagger-ui.html?url=/apispec/spring%253Fgroup%253D00_external#/query45api



and contact kg@ebrains.eu for help getting onboard.

Different datasets within a particular research domain:

Slow brain wave analysis with multiple methods (ECoG and Calcium imaging
- investigating sleep, anesthesia, and the transition to wakefulness)

Use Case SGA2-SP3-002 KR3.2: Integrating multi-scale data and the output of simulations in a reproducible and adaptable pipeline.
https://wiki.ebrains.eu/bin/view/Collabs/slow-wave-analysis-pipeline
Gutzen, De Bonis, Pastorelli, Capone, De Luca, Denker, Grün, Paolucci & Davison.
https://www.youtube.com/watch?v=uuAiY6HScM0

Use- case based on two example datasets from EBRAINS

https://search.kg.ebrains.eu/instances/Dataset/2ead029b-bba5-4611-b957-bb6feb631396
https://search.kg.ebrains.eu/instances/Dataset/71285966-8381-48f7-bd4d-f7a66afa9d79

Pipeline can be used for the dataset collections of electrophysiology and two-photon microscopy users

Pipeline executed in:

lab.ebrains.eu

jupyterhub

Using the following tools:

elephant

neo

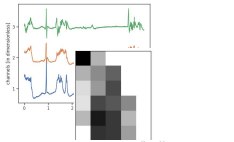
snakemake

Analytical pipeline

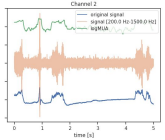
Modality specific

Data entry

Preprocessing

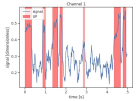


E.g:
- Configuration
- Roi_selection
- Background_removal
- Normalization
- Frequency filter



Research **domain** specific

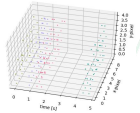
Trigger Detection



E.g.
- Transformation
- Detection

Research **subdomains** pecific

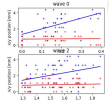
Wavefront Detection



E.g:
- Trigger filter
- Wave filter

Research **question** specific

Wavefront characterisation



E.g.
- Velocity
- Direction

The design of the pipeline aims at interfacing a variety of general and specific analysis and processing steps in a flexible modular manner.

The modules of the pipeline makes it possible to adapt to diverse types of data (e.g., electrical ECoG, or optical Calcium Imaging recordings) and to different analysis questions.

This makes the analyses a) more reproducible and b) comparable amongst each other since they rely on the same stack of algorithms and any differences in the analysis are fully transparent.

EBRAINS