

Data management and stewardship in environMENTAL

Providing a FAIR data infrastructure

Sven Twardziok

environMENTAL seminar 3

06.07.2022



Aus Forschung wird Gesundheit

Data Management and Bioinformatics Services

Digital Health Center, BIH@Charite



Ulrike Taron
Data manager



Marcel Jentsch
Data manager



Philipp Strubel
Software
developer



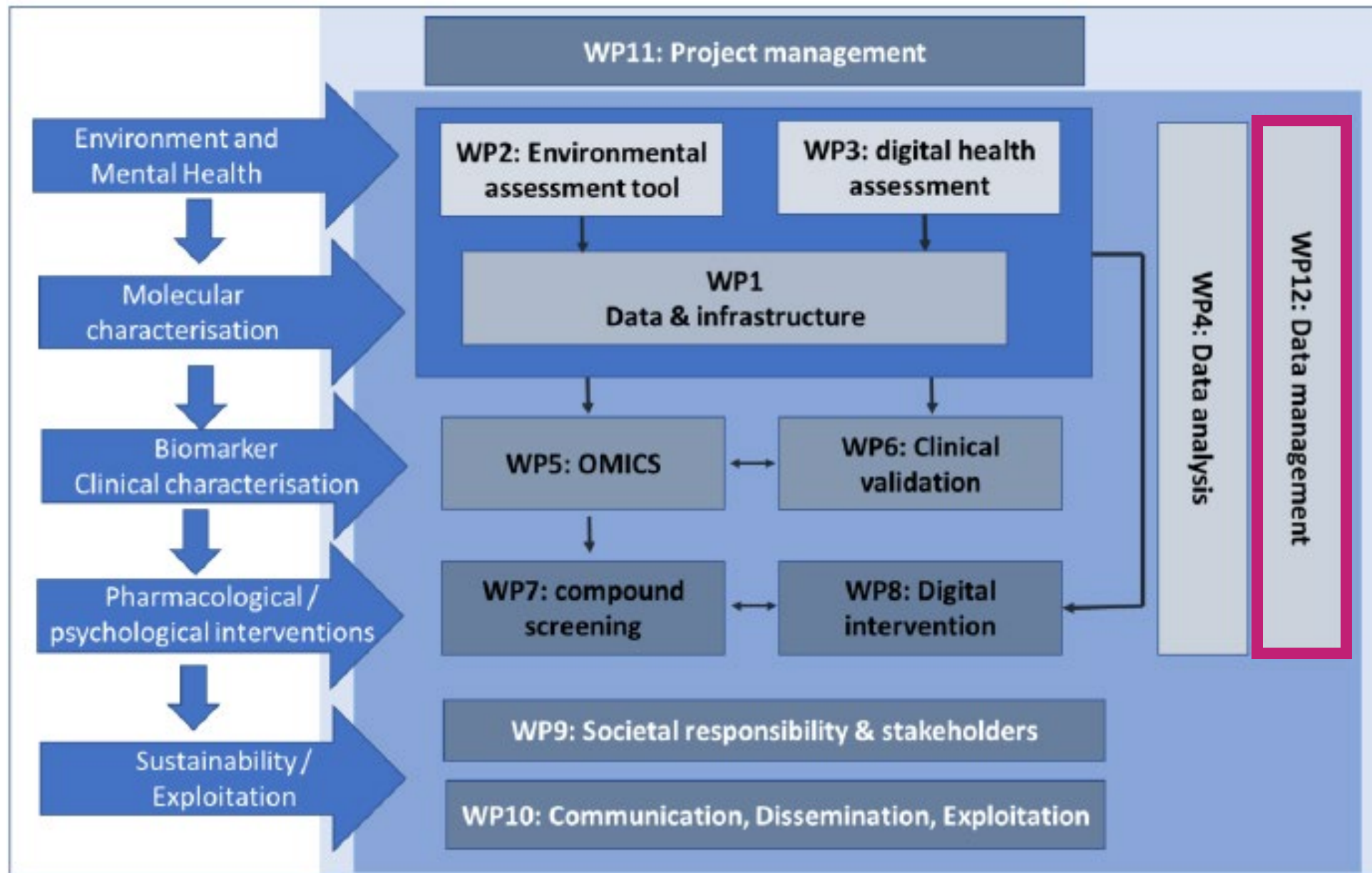
Joshua Waack
Apprentice



**Valentin
Schneider-Lunitz**
Bioinformatician



Sven Twardziok
Bioinformatician,
group leader



WP12: Data Management

Objectives

1. Project wide data management following the FAIR principles
2. Providing a data infrastructure at BIH
3. Distribution of samples for analysis and integration of deep phenotyping data

Deliverables:

- D12.1 Data management plan (M6)
- D12.3 Provision of data infrastructure (M12)
- ...

Data Management Plan (DMP)

A DMP describes the data management life cycle for the data to be collected, processed and/or generated. As part of making research FAIR, a DMP should include information on¹:

- handling of research data during and after the end of the project
- what data will be collected, processed and/or generated?
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project).

¹ Guidelines on FAIR Data Management in Horizon 2020

Big picture

General concepts

Involvement of end-users - participatory approach that takes account of gender differences and involves monitoring and reflecting on ethical and societal aspects

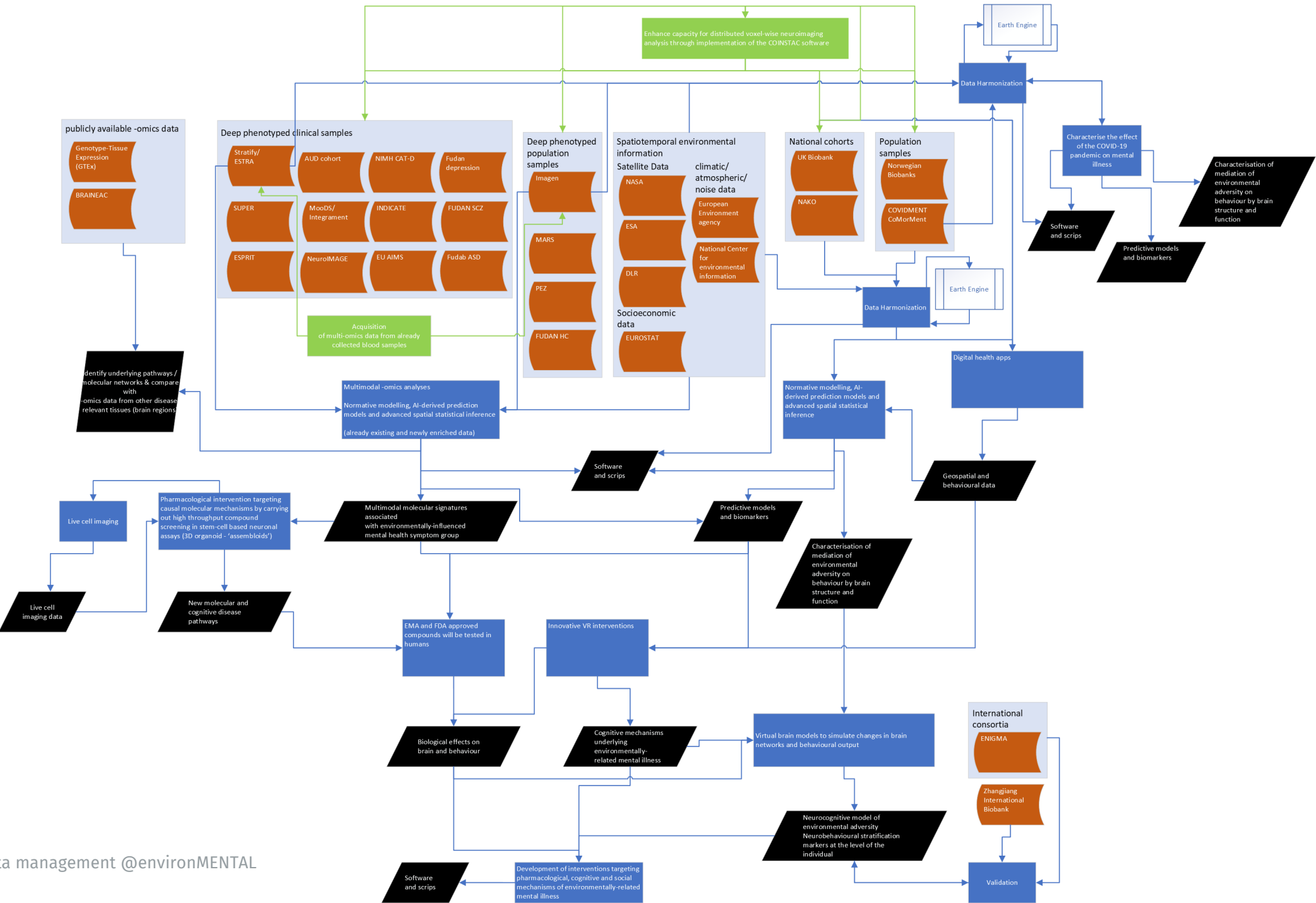
Educational interventions in the form of consultations and reports emphasising recommendations for societal action and digital health programmes will be carried out at every step of the project through workshops, seminars and discussions with public health agencies and experts and local service providers to disseminate our findings

Data access has been confirmed by each of the cohorts and ethical approval for comparable studies has already been granted to us

Computational facilities are provided by the UiO Services for Sensitive Data (TSD)

Berlin Institute of Health (BIH), designed and approved for sensitive data according to GDPR

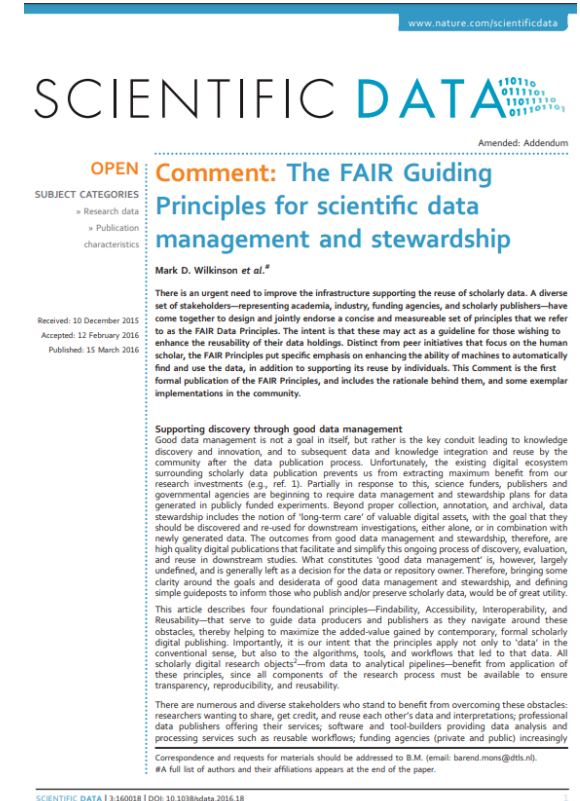
Contracts, Templates & Documents



FAIR

Findable, Accessible, Interoperable, Reusable

- FAIR principles contain 15 rules to **define good data management**
- **Goal:** share data in machine readable way (aka. **Fully AI-Ready**)
- Focus on reuse of research data but also applicable for medicine
- FAIRness can be easily validated by self-assessment or automatic validation



Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

FAIR

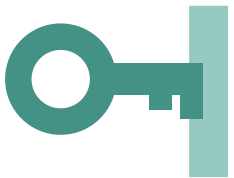
Findable, Accessible, Interoperable, Reusable



Findable: store metadata in public catalogues and use persistent identifiers for data/metadata



Interoperable: use common file formats and apply standard vocabularies and ontologies



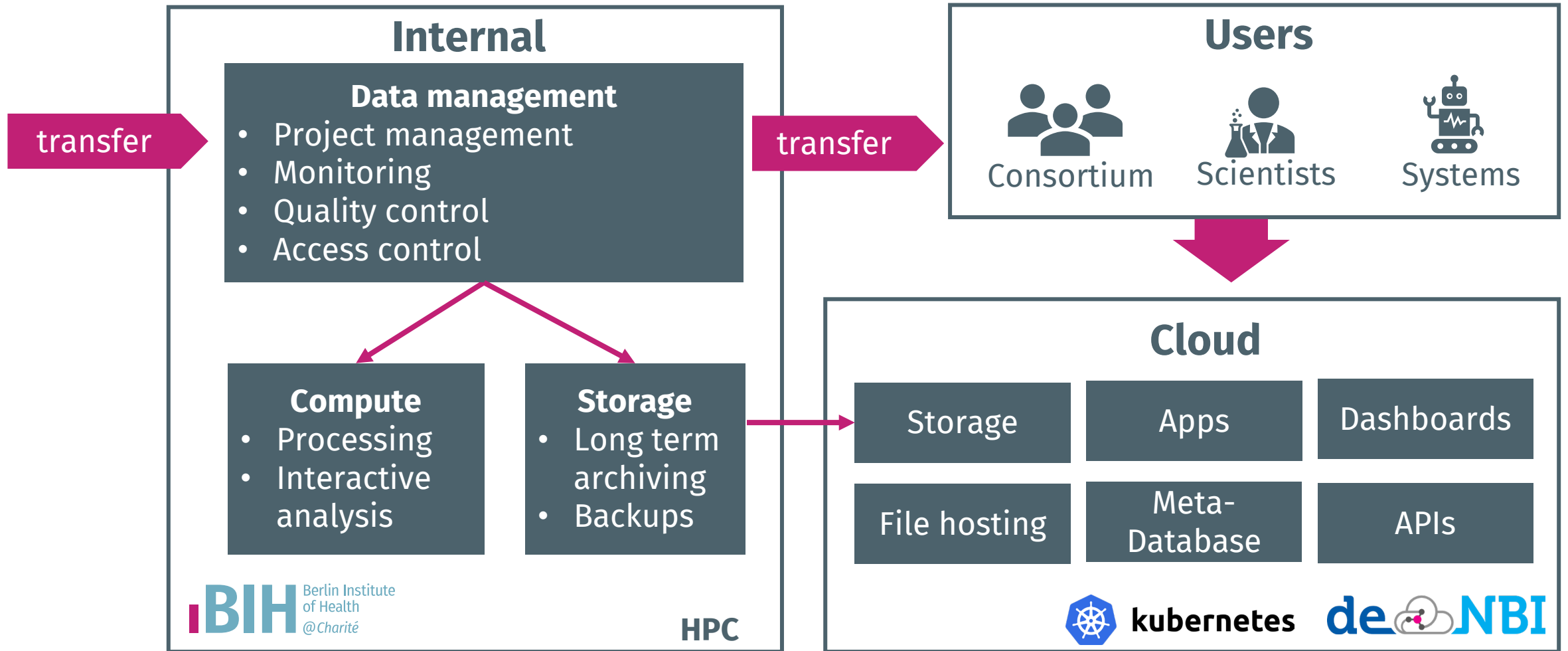
Accessible: store data in long term storage and make data accessible (controlled)



Reusable: provide rich information about data creation and data usage information

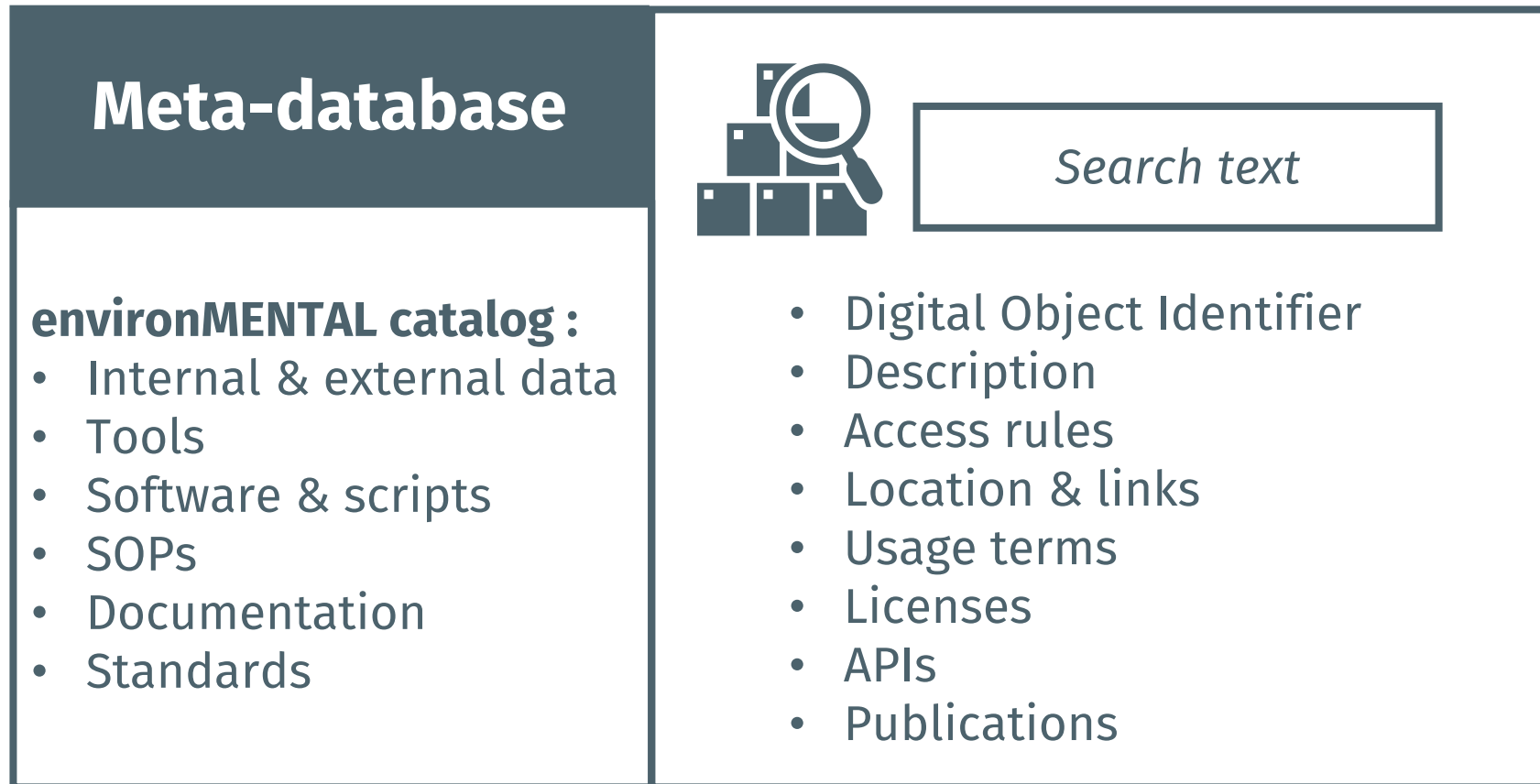
Data Infrastructure @BIH

FAIR data infrastructure



Meta Database

Findable & Reusable



PsyShare: File Hosting Service

Accessible & Reusable

- Synchronize and share files
- Standardized and secure data infrastructure
- Compliant with the legal framework and GDPR
- Low-barrier & user friendly
- Project management and social apps

Data
transfer



Nextcloud

BIH Cloud Infrastructure



LS LOGIN

File
sharing

GUI

API

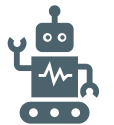
Users



Consortium



Scientists

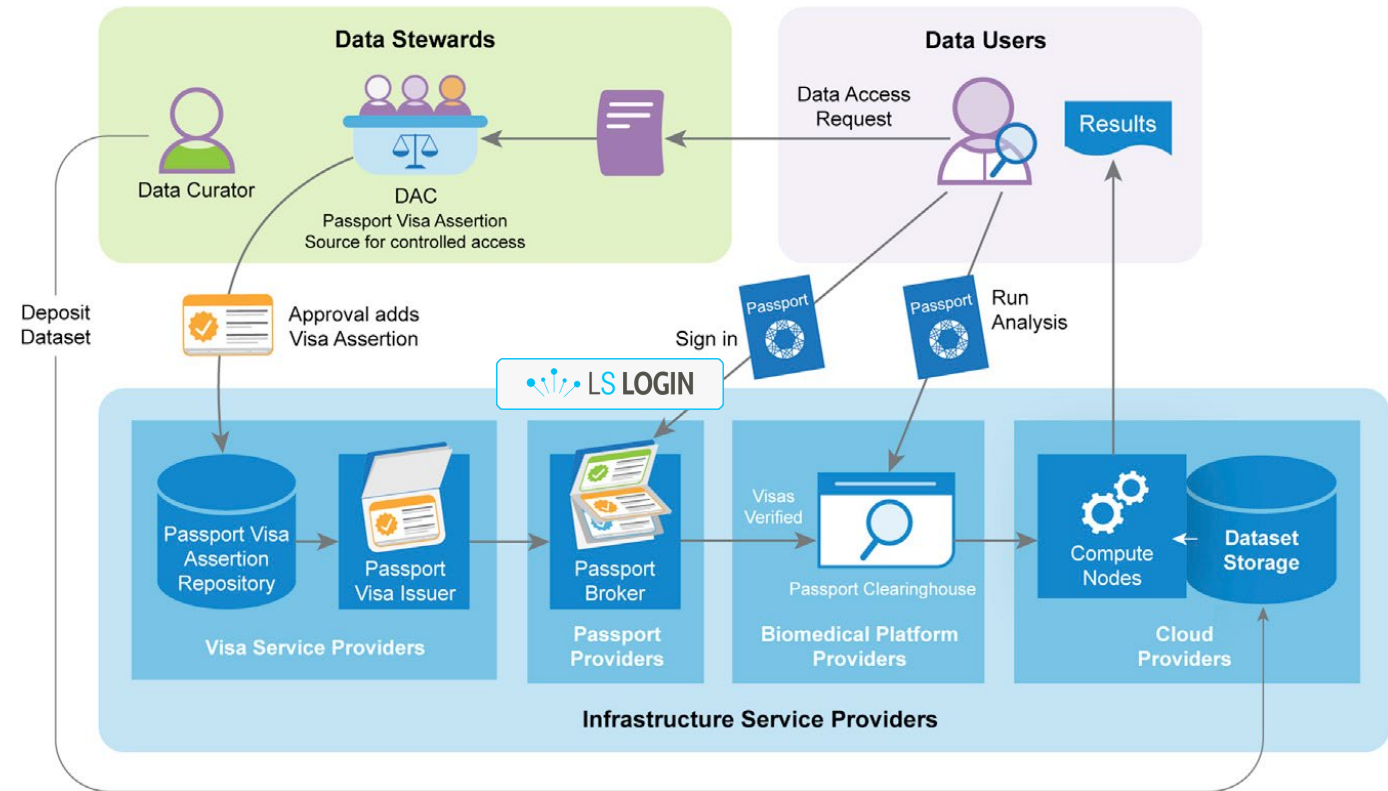


Systems

Federated Access Control Accessible & Reusable

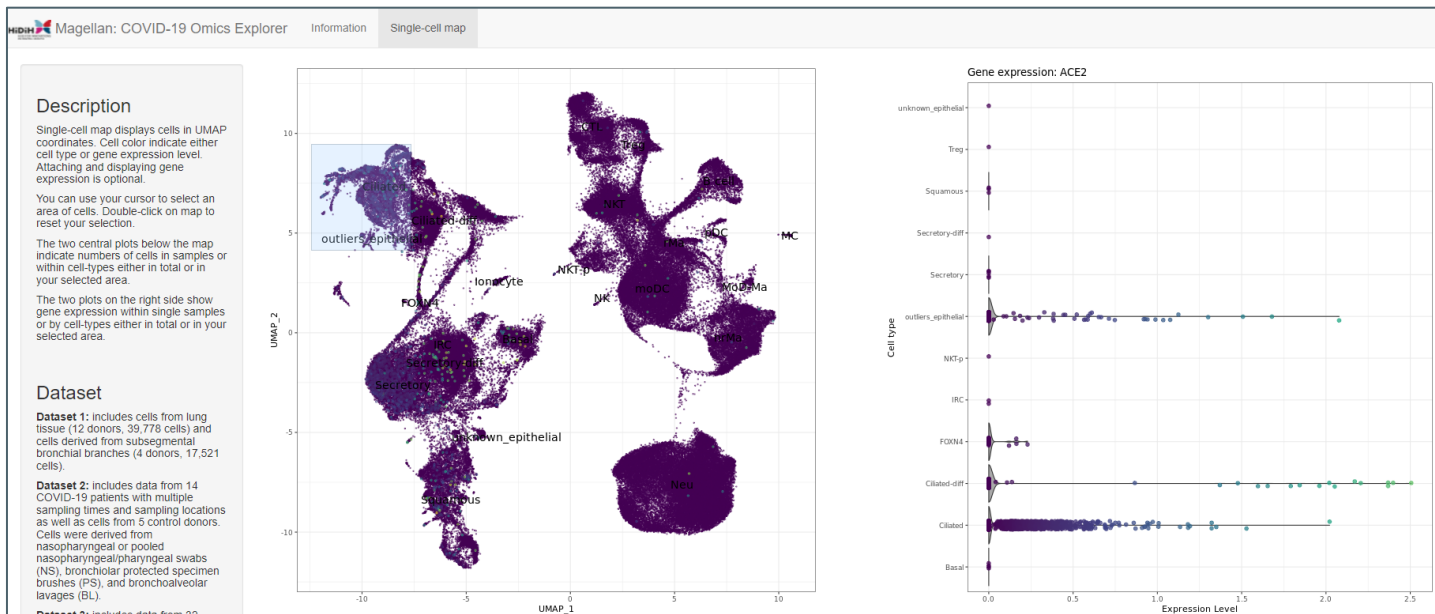
Life-Science AAI + GA4GH passports :
manage users and access in a federated framework

- Federated login system developed by EOSC-life and ELIXIR
- GA4GH Passports define standard for setting roles and access rights



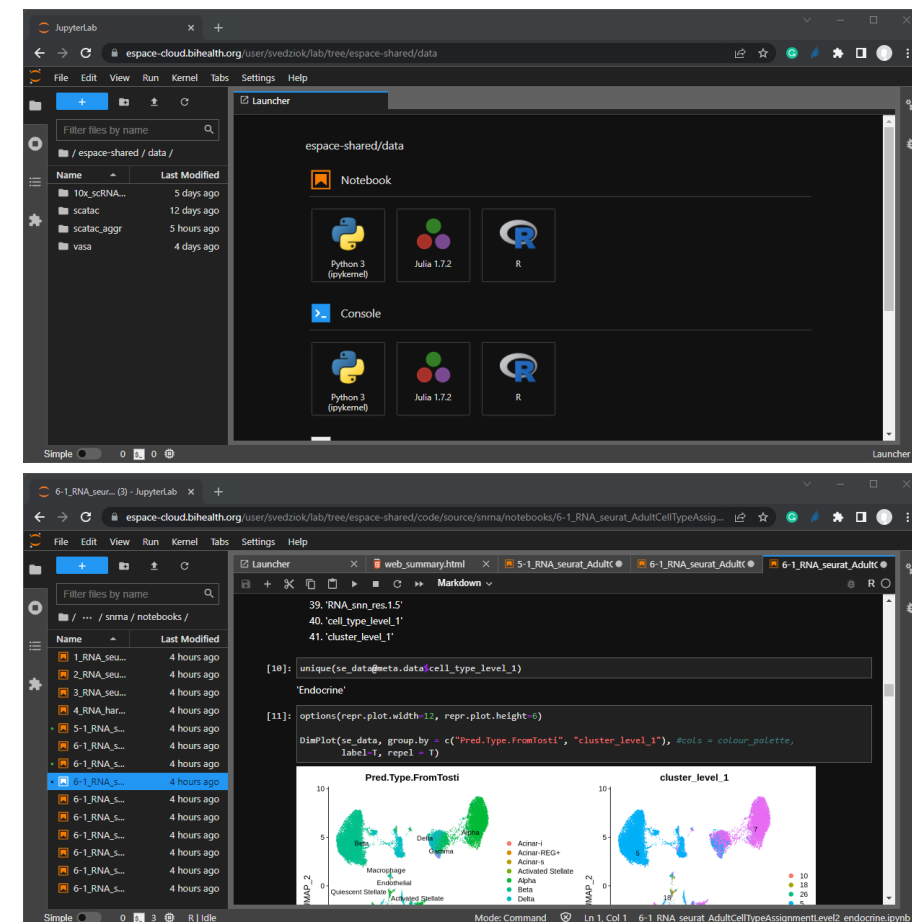
Craig Voisin, et. Al., GA4GH Passport standard for digital identity and access permissions, Cell Genomics, Volume 1, Issue 2, <https://doi.org/10.1016/j.xgen.2021.100030>

Cloud Workspaces Interoperable



<https://digital.bihealth.org/>

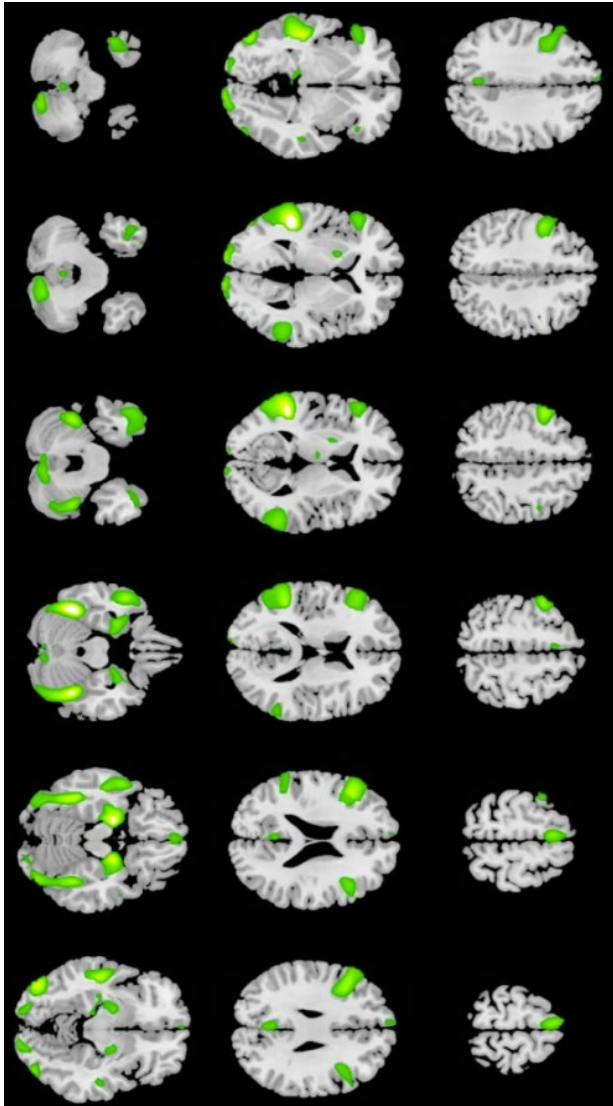
Chua, R.L., Lukassen, S., Trump, S. *et al.* COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol* **38**, 970–979 (2020).
<https://doi.org/10.1038/s41587-020-0602-4>



<https://espace-cloud.bihealth.org>

Horizon 2020 project for “Expression and Spatial analysis Pancreas Atlas Consortium Europe”,
ESPACE

IMAGEN, STRATIFY and ESTRA @BIH



European research projects

Longitudinal neurobiological basic research

Influence of biological, psychological, and environmental factors during adolescence on brain development and mental health

IMAGEN: over 2000 adolescents and their parents

STRATIFY: over 400 young adults with major depression, schizophrenia, alcohol use disorders and eating disorders

ESTRA: eating disorders and comorbid mental illnesses

Demographics, neuropsychological assessments, medical questionnaires, MR neuroimaging and genomics

Data collected over a 10 year period

Covid-19 Follow-Up (IMACOV/STRATICO)

Image source: [nature.com/articles/s41562-020-0846-5](https://www.nature.com/articles/s41562-020-0846-5)

Summary

- Goal: manage all data according to FAIR principles
- Open data management plan
- Infrastructure at BIH:
 - IMAGEN, STRATIFY&ESTRA cohorts
 - Long term data archiving
 - PsyShare: file hosting service
 - Cloud workspaces (Jupyter, Apps, APIs)
- Connection to external infrastructures (e.g. EBRAINS)

Vielen Dank

Sven Twardziok

Group leader at the Digital Health Center

Berlin Institute of Health (BIH)

Charité - Universitätsmedizin Berlin

Berlin Institute of Health (BIH)

Digital Health Center | Health Data

Kapelle-Ufer-2 | 10117 Berlin | Germany

Phone: +49 30 450 543105

www.bihealth.org

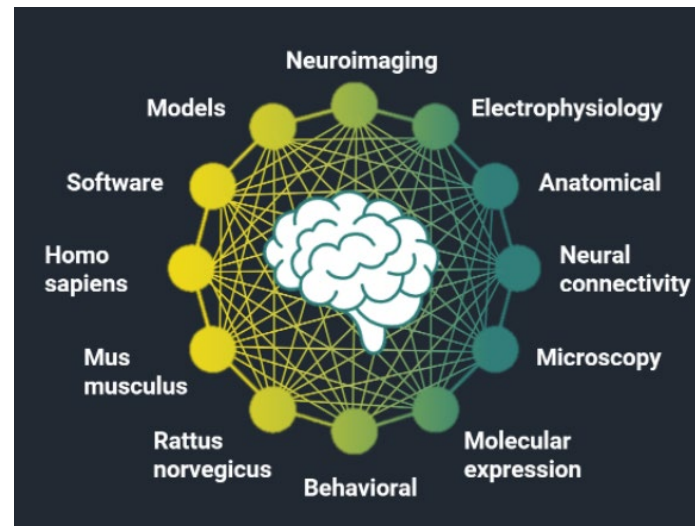



Aus Forschung wird Gesundheit



EBRAINS European shared digital research infrastructure

1. Open to integrating brain research
2. Developed and powered by the EU-funded Human Brain Project
3. Provides digital tools and services
 - Data and Knowledge
 - Atlases
 - Simulations
 - other



Search (e.g. brain or neuroscience) 

CATEGORIES

Project	124
Dataset	878
Model	109
Software	166
Contributor	1804

> FILTERS [Reset](#)

MODALITY

<input type="checkbox"/> microscopy	490
<input type="checkbox"/> neuroimaging	415
<input type="checkbox"/> anatomy	400
<input type="checkbox"/> histology	373