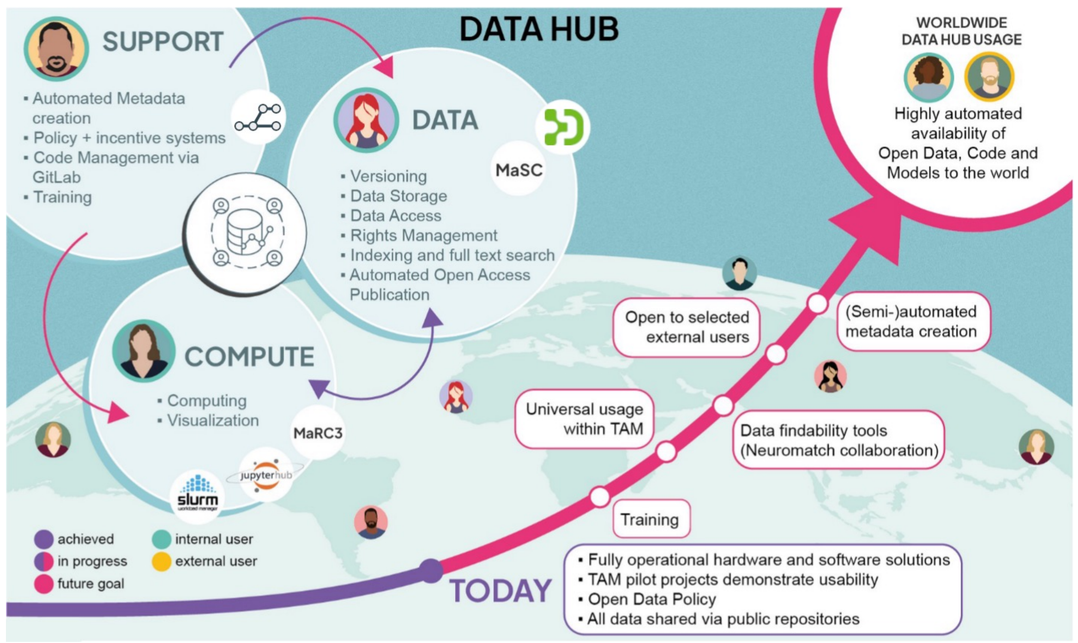
The DataHub: Your Gateway to Effective and Efficient High-Performance Computing in Psychology and Neuroscience Research

The adaptive Mind
Funded 2021-2025 by HMWK

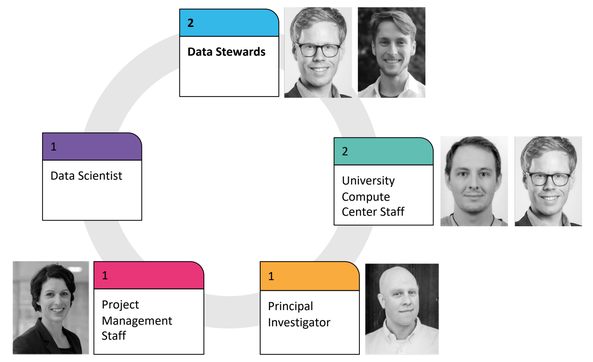
Nick Augustat, Stefan Lenze, Tobias Ortwein, Ortrun Brand, Bernd Nicklas, Dominik Endres
Philipps-Universität Marburg

datahub@uni-marburg.de
DataHub Support

Vision, team & collaboration



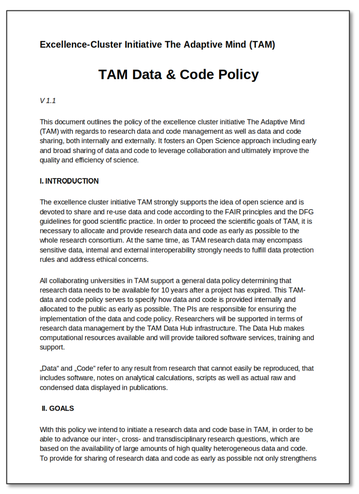
DataHub team



Collaboration

nfdi Nationale Forschungsdaten Infrastruktur · HeFDI · Neuromatch · incf enabling open and FAIR neuroscience · Philipps Universität Marburg · Servicezentrum digital gestützte Forschung

Policy, training & support


Excellence-Cluster Initiative The Adaptive Mind (TAM)

TAM Data & Code Policy

V 1.1

This document outlines the policy of the excellence cluster initiative The Adaptive Mind (TAM) with regards to research data and code management as well as data and code sharing, both internally and externally. It fosters an Open Science approach including early and broad sharing of data and code to leverage collaboration and ultimately improve the quality and efficiency of science.

I. INTRODUCTION

The excellence cluster initiative TAM strongly supports the idea of open science and is devoted to share and re-use data and code according to the FAIR principles and the DFG guidelines for good scientific practice. In order to proceed the scientific goals of TAM, it is necessary to allocate and provide research data and code as early as possible to the whole research consortium. At the same time, as TAM research data may encompass sensitive data, internal and external interoperability strongly needs to fulfill data protection rules and address ethical concerns.

All collaborating universities in TAM support a general data policy determining that research data needs to be available for 10 years after a project has expired. This TAM-data and code policy serves to specify how data and code is provided internally and allocated to the public as early as possible. The PIs are responsible for ensuring the implementation of the data and code policy. Researchers will be supported in terms of research data management by the TAM Data Hub infrastructure. The Data Hub makes computational resources available and will provide tailored software services, training and support.

„Data" and „Code" refer to any result from research that cannot easily be reproduced, that includes software, notes on analytical calculations, scripts as well as actual raw and condensed data displayed in publications.

II. GOALS

With this policy we intend to initiate a research data and code base in TAM, in order to be able to advance our inter-, cross- and transdisciplinary research questions, which are based on the availability of large amounts of high quality heterogeneous data and code. To provide for sharing of research data and code as early as possible not only strengthens

Principles of the TAM policy

- Use shared hardware, services and support.
- Share data and code as early as possible.
- Practice FAIR principles and common standards.
- Make results more transparent and reproducible.
- Prepare for long-term accessibility and publish.
- ***To do***: frictionless workflow for international students subject to dual-use concerns

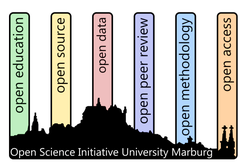
Enhance collaboration and re-use of scientific work.

TAM data and code policy

The adaptive Mind
The DataHub User Manual

- Access & onboarding
- DataHub workflow
- Service descriptions
- Data & code management
- Git and GitLab tutorial
- Recorded workshops
- FAQ & troubleshoot

DataHub user manual



Workshops and Summer Schools:

- Fostering open science practices
- **Brainhack Summer School**

Register now!
Brainhack 2025

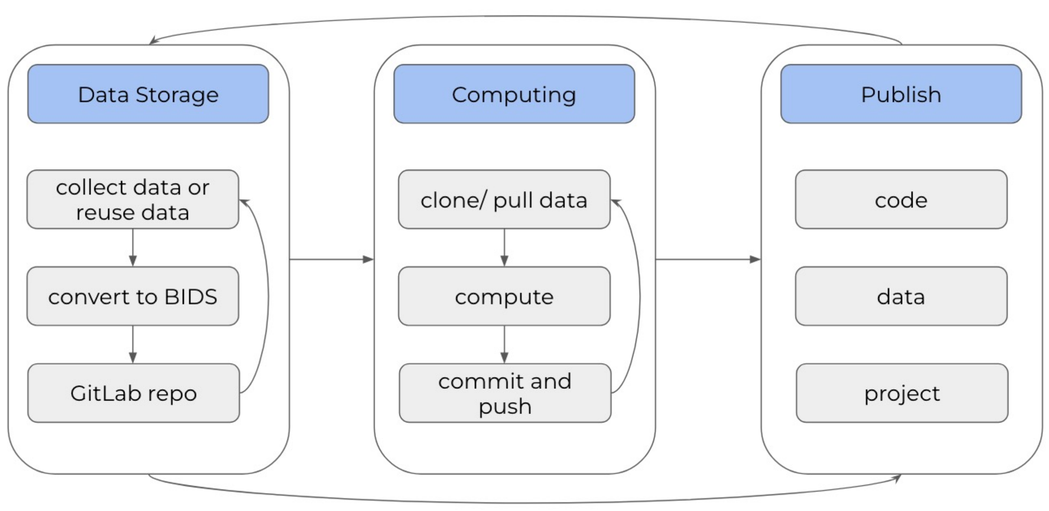
SFB 135/NOWA DYNAMIC
Neuroscientific Workflow Assistance

Recorded tutorials, materials and instructions, websites and JupyterBooks on:

- Git & GitLab,
- Clean and reusable rode
- Continuous integration, code review, etc.

Workflows & standards

Common workflow



Tonic research folder template

Available from TAM-GitLab

tonic-single-project

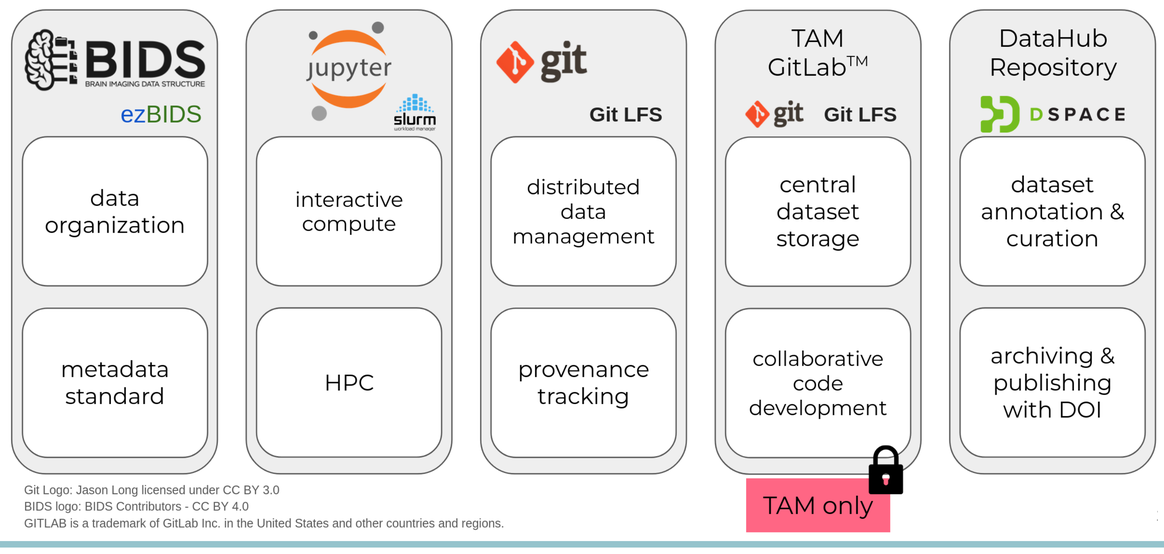
```
01_project_management
  01_data_management_plans
  02_preregistration
02_material_and_methods
  01_study-protocol
  02_code
  README_MM.md
03_data
  README_DATA_STORE.md
04_figures
  README_FIGURES.md
05_dissemination
  01_reports_conferences
  02_manuscripts
  03_other
  README_DISSEMINATION.md
LICENSE-CC-BY
README.md
```

BIDS_dataset

```
source
raw
  sub-01
    ses-01
      anat
      beh
      micr
derivatives
  pipeline_X
CHANGELOG.md
dataset_description.json
LICENSE-CC-BY
README.md
```

BIDS starter kit

Technical services



Git Logo: Jason Long licensed under CC BY 3.0
BIDS logo: BIDS Contributors - CC BY 4.0
GITLAB is a trademark of GitLab Inc. in the United States and other countries and regions.

DataHub status and latest features

Technical services

- HPC integration: Production
- ***JupyterHub***: Production (new hardware and profiles)
- ***TAM GitLab***: Production (advanced access management)
- ***Git + Git-LFS workflow integration***: Production
- TAM DataHub repository: Pilot
- ***GitLab - DSpace publication pipeline***: Planned
- ***ezBIDS***: Test

Support services

- User manual (GitLab pages): Production
- Support queue (request tracker): Production
- Curation service (DSpace & GitLab): Pilot
- ***Data protection policy***

Technical resources & architecture

MaRC3 total resources

- 51 compute nodes,
- 3264 CPU cores
- 26x AMD EPYC 7702P, 64 cores, 2 GHz
  23x AMD EPYC 7713, 64 cores, 2 GHz
  2x AMD EPYC 9334, 2x32 cores, 2.7 GHz
- 256-1024 GB RAM
- 12x V100S, 16x A100 40 GB, 70x A100 80 GB, 6x A40, 4x L40S
- 1.7-7.2 TB local, 134 TB shared scratch, 134 TB home storage
- 25 GBit Ethernet network (storage: 200 GBit)

TAM nodes in **MaRC3**

- 14 nodes / 896 cores + 4 GPUs (A100 80GB)
- 4 nodes / 192 cores (CPU-only)

TAM storage in **MaSC**

- 703 TB total usable capacity