Excellence-Cluster Initiative The Adaptive Mind (TAM)

TAM Data & Code Policy

V 1.0

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 our research goals but also fosters a more transparent, reproducible and more democratized open science. It also aims at providing the sources of our results for as long as possible with respect to long term archiving.

In particular, we will provide a TAM-internal storage area from where data and code can be retrieved independently from the actual researchers that worked on the respective project,. Reproducibility is to be maintained by documenting the complete workflow that lead to the published results. In addition, resources are provided to share software and maintain it independently of specific members of working groups.

The TAM researchers consent to allocate their research data and code to one another as early as possible, especially by using the TAM DataHub. Data and code is provided in a machine-readable-manner. Data- and code-producers provide their resources as early as possible by storing data on the TAM-share in the Marburg Storage Cluster (MaSC) in the first instance.

III. IMPLEMENTATION

Preregistration. If researchers have used preregistration, they are strongly encouraged to relate their accessible data and code to this preregistration index.

Internal availability and accessibility of Data and Code. Data will be stored on the TAM Data Hub during the TAM funding period. For 10 years after TAM's expiration, data and code will be kept accessible by the collaborating universities' infrastructure. Members of TAM are obliged to submit their data and code in the TAM Data Hub.

TAM Data Stewards. TAM will provide for data stewards who offer support with handling data and code and preparing it for internal interoperability and, later on, access to the public.

Publication of data. Research data - and, where appropriate, code - associated with peer-reviewed publications are distributed externally. Data will be publicly available through a public interface e.g. via subject specific repositories (e.g. in NFDI contexts, PsychData etc.), institutional repositories such as data_UMR/JLU_Data/TUDataLib and/or Zenodo and also ensures that the international community gets access to data and code that resulted in published research, including preprints.

The common data management system workflow is defined as follows:

- Reproducible management of neuroscientific data will be handled by GIN (G-Node Infrastructure; https://gin.g-node.org/). GIN is based on Open Source projects (i.e., Git, git-annex, Gogs) and includes consistent organization, annotation, and storage of data.

- Files that are uploaded in GIN will be version controlled by Git.

- Source Code development and management will be handled by GitLab, ensuring version control of all files using Git.

- With proper authorization, files can be accessed and changed by collaborators.

- Researchers will be able to decide which data and code are made available for

public access, for internal and external collaboration.

- For overall project folder organization, the GIN-Tonic research folder structure standard will be practiced (https://gin-tonic.netlify.app/; https://github.com/tonic-team/Tonic-Research-Project-Template)

- BIDS (Brain Imaging Data Structure; https://bids.neuroimaging.io/index.html) will complement GIN-Tonic for organization and description of experimentally obtained neuroscientific data.

Long-term archiving & deletion

- Data & code storage, backup and accessibility is guaranteed for 10 years after TAM's expiration by the collaborating universities' infrastructures. After that time, data may be deleted if necessary, e.g. if storage capacity of the data hub is exceeded. Data producers will be contacted to evaluate relevance of data and code. Data producers will have the option to veto deletion. In that case, data producers will be required to provide funding for storage beyond the 10 year period. Data producers are encouraged to transfer data and code to a subject-specific long term archiving institution, e.g. in NFDI contexts.