Sharing FAIR Neuroimaging Data with DataLad and Dataverse: A Practical Workflow

Agustina Aragón Daud[1-3], Tobias Muganga[1,3,4], Federico Raimondo[3-4], Athena Demertzi[1,2,5]

[1] University of Liège, GIGA Research, GIGA-CRC Human Imaging Unit, Physiology of Cognition Lab, Belgium
[2] Fund for Scientific Research (FNRS), Brussels, Belgium
[3] Institute of Neuroscience and Medicine, Brain and Behaviour (INM-7), Research Centre Jülich, Jülich, Germany
[4] Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[5] Psychology and Neuroscience of Cognition Research Unit, University of Liège, Liège, Belgium

a.aragondaud@uliege.be

Introduction

Why sharing neuroimaging data openly?

Sharing data openly is particularly important in neuroimaging research as:

- Acquiring MRI data is costly in time, money, and coordination effort.
- Larger, pooled datasets increase statistical power and allow findings to generalise across populations.
- Open data also directly addresses the reproducibility crisis in neuroscience and psychological sciences, as other researchers can access, verify, and build on published findings.

What are the difficulties when sharing such data?

Neuroimaging are very heavy. Uploading files to a repository and downloading them manually is error-prone, inefficient, and often impractical at this scale.

Here we present a workflow combining **DataLad** and **Dataverse** to share a neuroimaging dataset in a way that is efficient, reproducible, and compliant with FAIR principles (Findable, Accessible, Interoperable, Reusable).

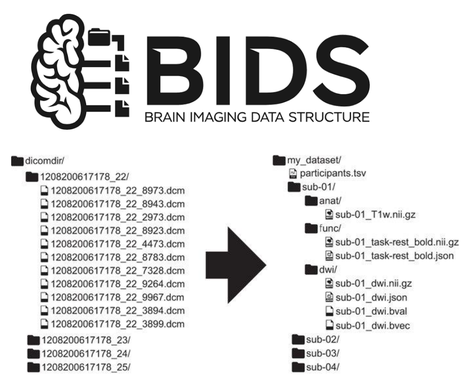
Workflow

STEP 1: Prepare the data

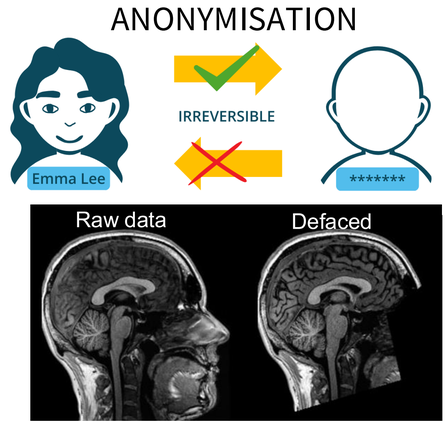
Data should be structured following BIDS (Brain Imaging Data Structure), a community standard for organizing neuroimaging files that makes the dataset immediately understandable and usable by others without needing to contact the authors.

Before sharing, data needs to be anonymized: it cannot be possible to trace back the participant identity. This involves defacing MRI images and removing all files and notes that could be used to re-identify participants.

FORMATTING



ANONYMISATION



STEP 2: Share structure of the data with DataLad & the data files in Dataverse

Store the structure of the dataset

DataLad is a data management tool built on Git and git-annex. It treats a dataset like a Git repository: every change is tracked, the full history is preserved, and the structure is readable by anyone.
It separates the structure of the dataset (metadata in GitLab) from the content of the files (Dataverse). This means a user can clone the full structure without downloading a single data file, then retrieve only what they need.

Store data files

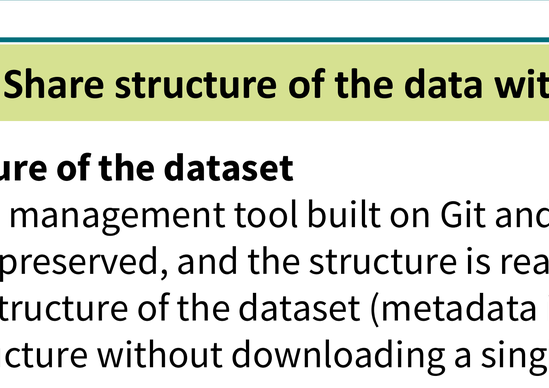
Dataverse is an open-source research data repository deployed by institutions worldwide, including ULiège. The dataset receives a DOI, making it permanently citable and findable.



The two are linked via the datalad-dataverse extension, which handles the connection between the GitLab repository and the Dataverse storage automatically.

Result





- **Download data structure:** `datalad clone <gitlab.uliege.be/dataset_url>`
- **Download files:** `datalad get <specific file or all>`
- **Delete files:** `datalad drop <specific file or all>`

Conclusion

Combining DataLad and Dataverse provides a practical, scalable solution for sharing large research datasets openly:

- **Findable**: Dataset has a DOI via Dataverse
- **Accessible**: Command-line access; selective download of only needed files; no manual clicking
- **Interoperable**: Standard, open format; works with any tool
- **Reusable**: Full version history, documented, citable and licensed

This workflow is not specific to neuroimaging. DataLad works with any type of data and any file format.

Datalad

Extension