Untangling the Black Box of Neurosciences

Part #1 – Data Structure (Antoine Jacquemin)

Part #2 – Data Sharing (Agustina Aragón Daud & Tobias Muganga)

Open-Science Day 2027

Poster #5 & #6

Slides & Posters at:

Supervisor: Athena Demertzi, PhD. & Christophe Phillips, Ir. PhD.

GIGA-CRC Human Imaging

LIÈGE université Sciences Appliquées

JÜLICH Forschungszentrum

fnrs FREEDOM TO RESEARCH

LIÈGE université GIGA CRC Human Imaging

Untangling the Black Box of Neurosciences

Part #1 - COFITAGE, a Multimodal, Longitudinal, BIDS-Compliant Neuroimaging Dataset

Open-Science Day 2027

Poster #5

Antoine Jacquemin
PhD student, Ir.

Supervisor: Christophe Phillips, Ir. PhD.
Development in data acquisition and modeling
GIGA-CRC Human Imaging

a.jacquemin@uliege.be

LIÈGE université Sciences Appliquées

fnrs FREEDOM TO RESEARCH

LIÈGE université GIGA CRC Human Imaging

in vivo neurosciences data

Modalities

Neuroimaging flavors

- ✓ MRI
  - Anatomical (Fig. 1)
  - Quantitative
  - Diffusion-weighted imaging
- ✓ PET
  - Amyloid-sensitive
  - Tau-sensitive

Phenotype flavors

- ✓ Demographic, psychological and lifestyle
  - Age, sex, coffee, sport etc.
- ✓ Genetics
  - APOE genotyping
- ✓ Cognition & cognitive reserve
  - 20+ tests (Fig. 2)
- ✓ Polysomnography (sleep-related metrics)

Study

Cohort

✓ 120 participants

Sessions

✓ Baseline
✓ 2-year follow-up
✓ 7-year follow-up

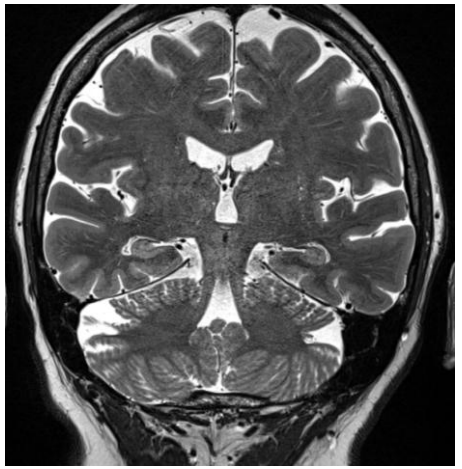
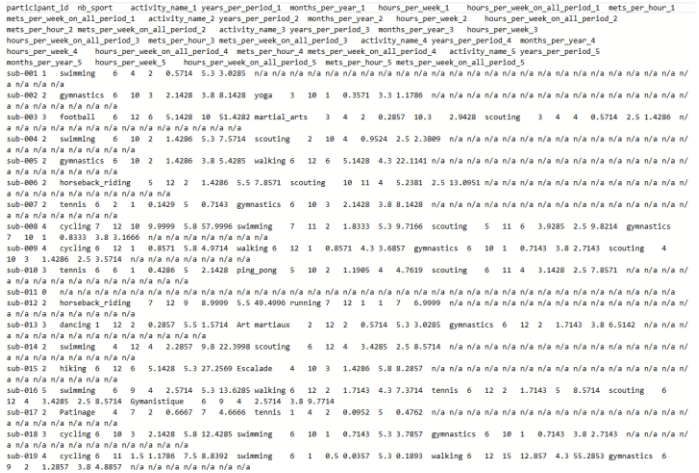


Fig. 1



participant_id nb_sport activity_name_1 years_per_period_1 months_per_year_1 hours_per_week_1 hours_per_week_on_all_period_1 mets_per_hour_1
mets_per_week_on_all_period_1 activity_name_2 years_per_period_2 months_per_year_2 hours_per_week_2 hours_per_week_on_all_period_2
mets_per_hour_2 mets_per_week_on_all_period_2 activity_name_3 years_per_period_3 months_per_year_3 hours_per_week_3
hours_per_week_on_all_period_3 mets_per_hour_3 mets_per_week_on_all_period_3 activity_name_4 years_per_period_4 months_per_year_4
hours_per_week_4 hours_per_week_on_all_period_4 mets_per_hour_4 mets_per_week_on_all_period_4 activity_name_5 years_per_period_5
months_per_year_5 hours_per_week_5 hours_per_week_on_all_period_5 mets_per_hour_5 mets_per_week_on_all_period_5
sub-001 1 swimming 6 4 2 0.5714 5.3 3.0285 n/a n/a n/a

The Standard Data Structure

BIDS
BRAIN IMAGING DATA STRUCTURE

BIDSification

Data Curation

- ✓ Conversion to BIDS format (Fig. 3)
- ✓ Harmonized BIDS-compliant naming conventions
- ✓ Conversion of proprietary file formats to open formats

Anonymization

- ✓ Defacing of anatomical MRI scans (Fig. 4)
- ✓ Removal of identifying metadata (subject IDs, dates)

Optimal storage (Fig. 5)

FAIR principles

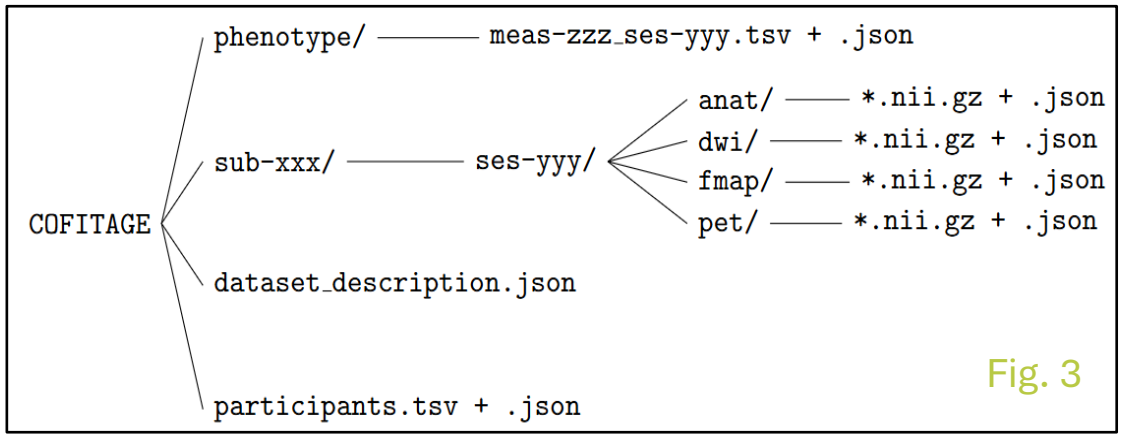


Fig. 3

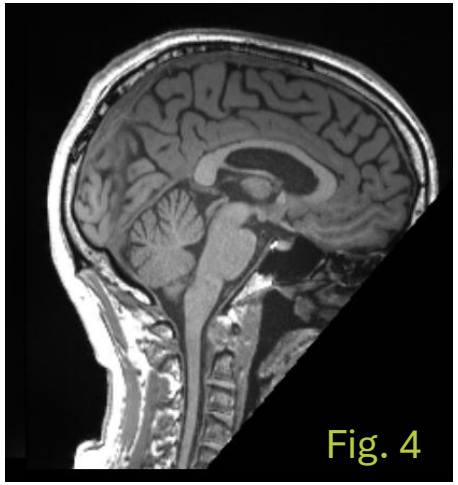


Fig. 4

Data storage Fig. 5

377	64.6	GB
61,913	16,709	files
4,545	610	folders

Real Case Scenario (aka The Data Structurer's Horror List) 🎃

The Horror List

- 👻 **The Phantom:** a participant with data that don't actually exist
- 🕳️ **The Vanishing:** a participant missing data that should exist
- 🌀 **The Multiverse:** multiple versions of the "raw" data
- 💾 **The Treasure Hunt:** one modality, N external disks
- 🆔 **The Identity Crisis:** each modality has its own participant IDs
- 🐺 **The Wolf in Raw's Clothing:** derived data labeled as "raw"
- 🖨️ **The Liar's Dictionary:** participant IDs don't match the dictionary
- 🤫 **The Silent Exclusion:** data excluded, no explanation
- 🎭 **The Shapeshifter:** participant IDs change between sessions
- 🎁 **The Trojan Horse:** mystery data from a "helpful" friend

inspired by true events....

Take-Home Message

The Horror Path	The Happy Path
Acquisition → Processing → Structure	Acquisition → Structure → Processing
Horror List items 🎃	Horror List: empty ✅
Reformat everything later	BIDS from day one

Data structured early is data you'll thank yourself for later.

Untangling the Black Box of Neurosciences

Part #2 - Sharing FAIR Neuroimaging Data with DataLad and Dataverse: A Practical Workflow

Open-Science Day 2027

Poster #6

Agustina Aragón Daud & Tobias Muganga
PhD students

Supervisor: Athena Demertzi, PhD.
Physiology of Cognition Lab
GIGA-CRC Human Imaging

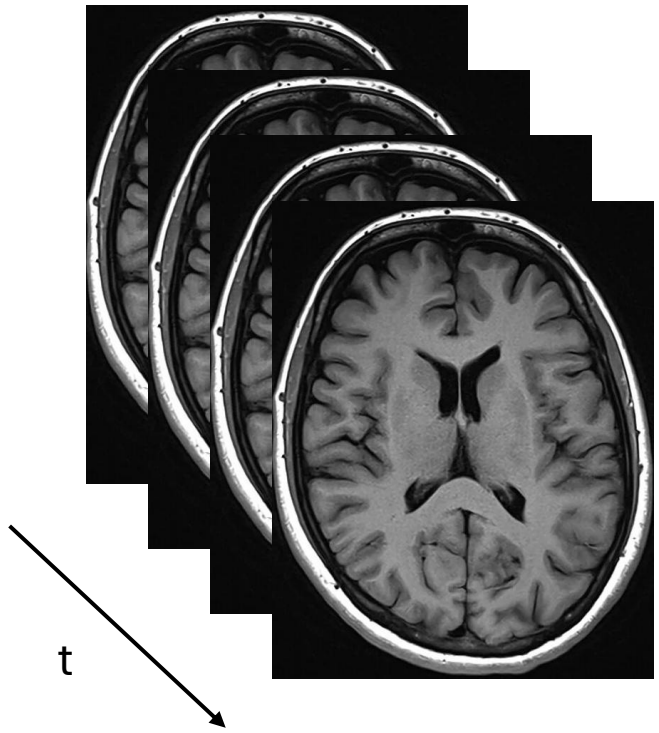
a.aragondaud@uliege.be

JÜLICH Forschungszentrum

fnrs FREEDOM TO RESEARCH

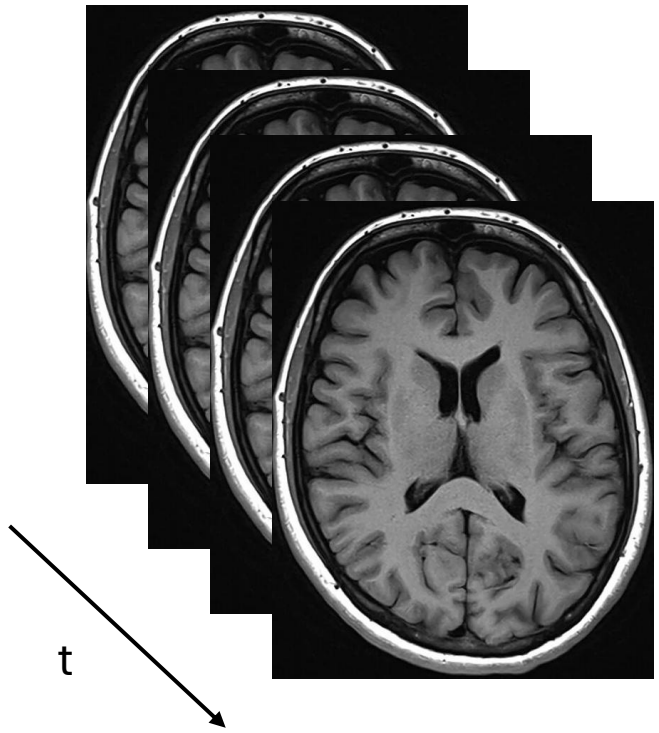
LIÈGE université GIGA CRC Human Imaging

Why sharing neuroimaging data openly?



- Acquiring neuroimaging data is costly
- Larger pooled samples increase statistical power
- Helps address the reproducibility crisis

Which are the difficulties when sharing neuroimaging data openly?



~10 GB

Neuroimaging data is heavy

x N participants

~500 GB

x Derivatives

~1 TB

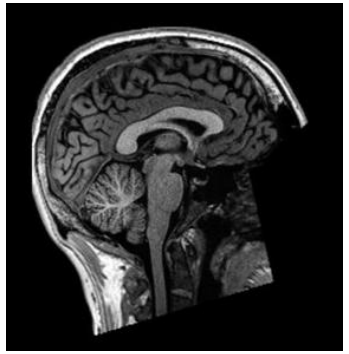
How to share neuroimaging data openly?

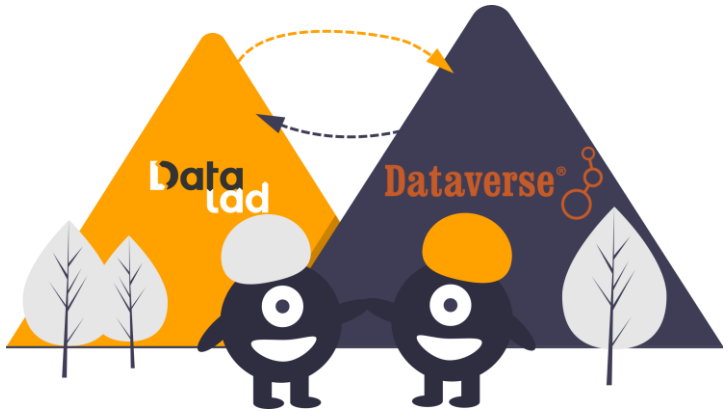
Institutional repository

Dataverse

LIÈGE université

Share data files

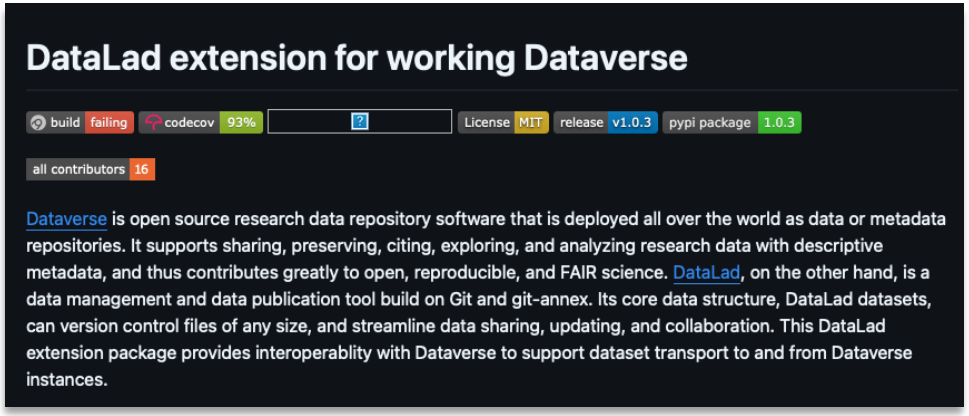




Version control for data

Datalad

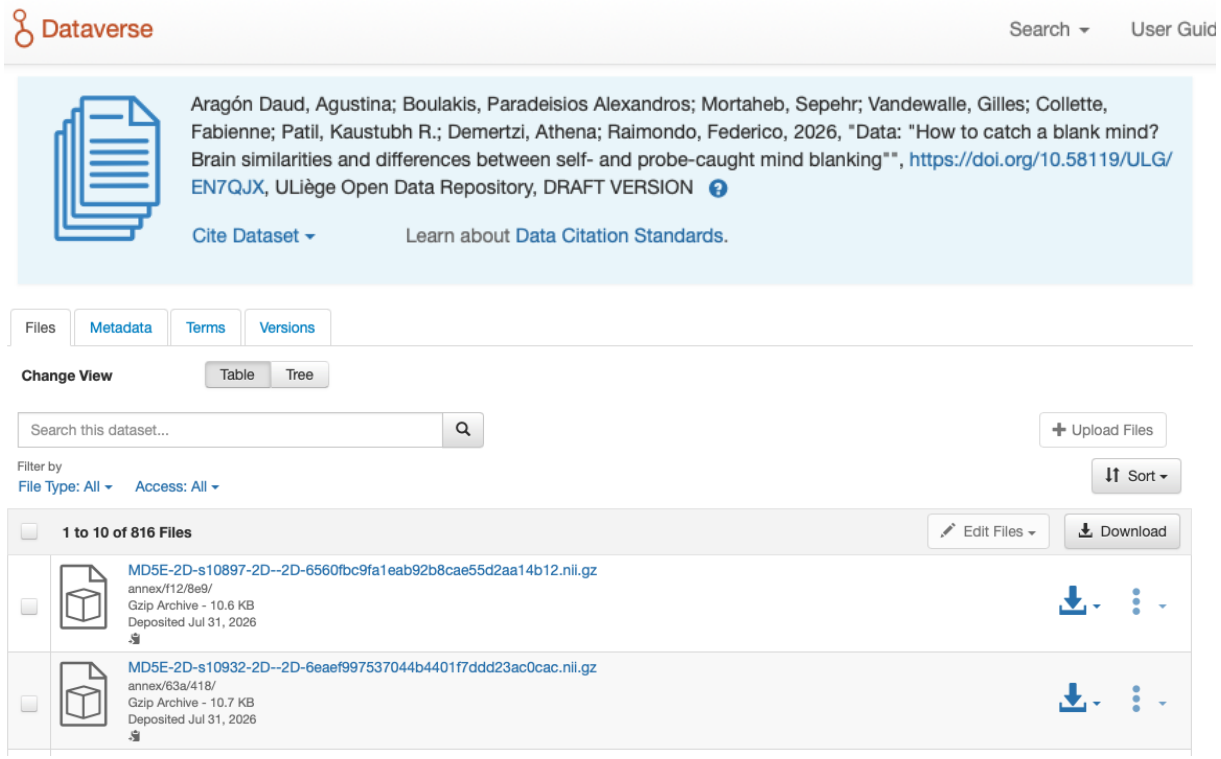
Share structure

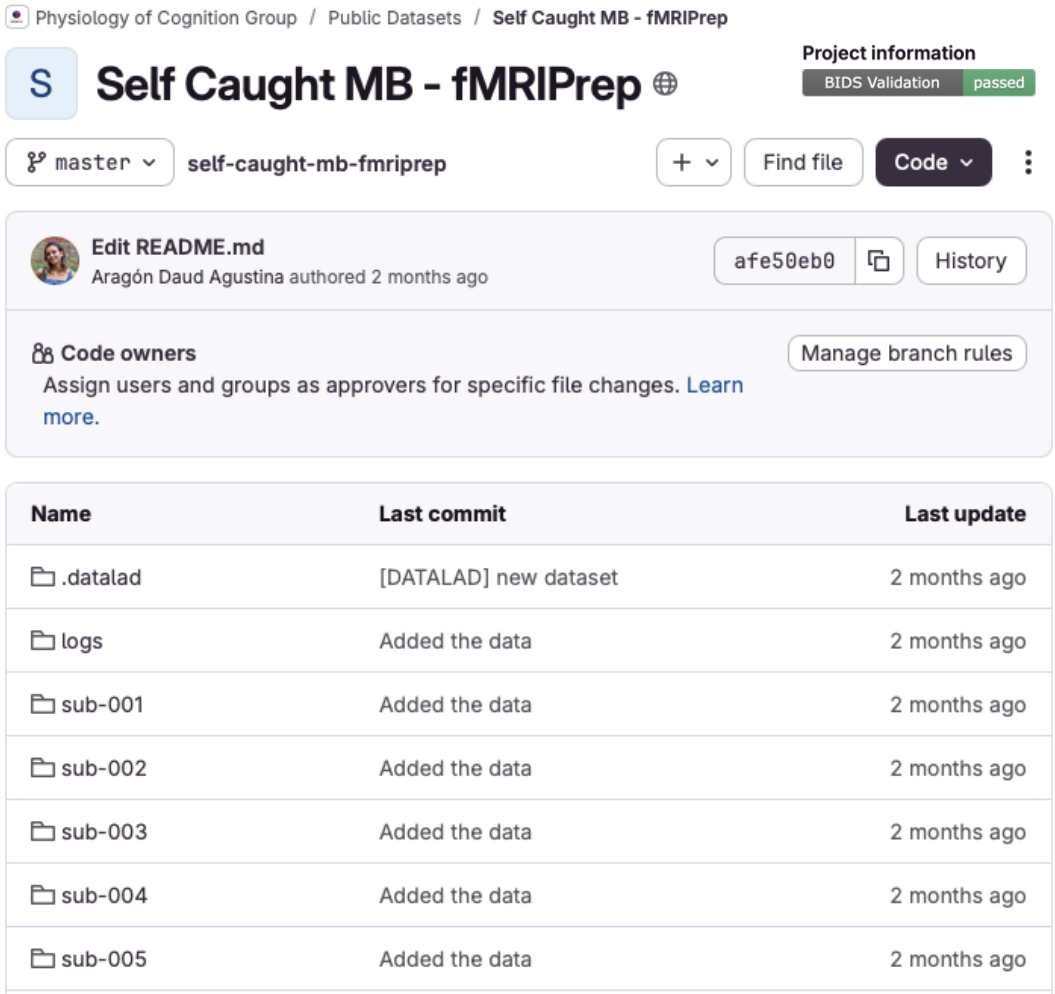


git

my_dataset/
participants.tsv
sub-01/
anat/
sub-01_T1w.nii.gz
func/
sub-01_task-rest_bold.nii.gz
sub-01_task-rest_bold.json
dwi/
sub-01_dwi.nii.gz
sub-01_dwi.json
sub-01_dwi.bval
sub-01_dwi.bvec
sub-02/
sub-03/
sub-04/

How to share neuroimaging data openly?





DataLad user case demo



Benefits of combining DataLad & Dataverse





✓ Trustable institutional infrastructure

✓ Dataset has a DOI — **F**INDABLE

✓ Selective download (only what you need)

✓ Command-line access (reproducible) — **A**CCESIBLE

✓ Compliance with standard format — **I**NTEROPERABLE

✓ Full version history tracking — **R**EUSABLE

✓ Free and built on open-source tools

PHYSIOLOGY OF COGNITION LAB

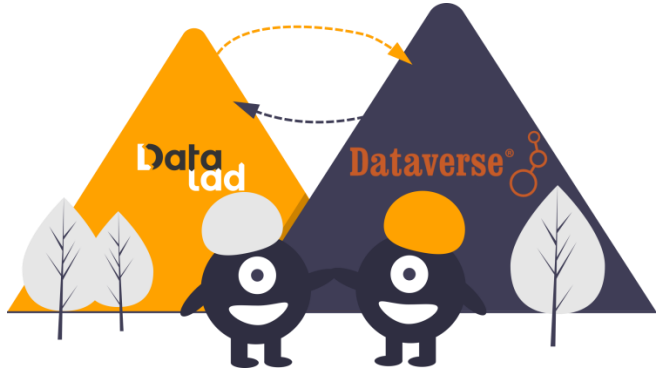
LIÈGE université GIGA CRC Human Imaging

FREEDOM TO RESEARCH

JÜLICH Forschungszentrum

ORBI

Slides & Posters



Datalad

Thanks!