

# OpenNeuro: An open archive for analysis and sharing of BRAIN Initiative data

Russ Poldrack  
Stanford University



# OpenNEURO

SEARCH SUPPORT FAQ

Sign in

A free and open platform for validating and sharing BIDS-compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#) data

**30,504** Participants

**796** Public Datasets

Browse by Modalities

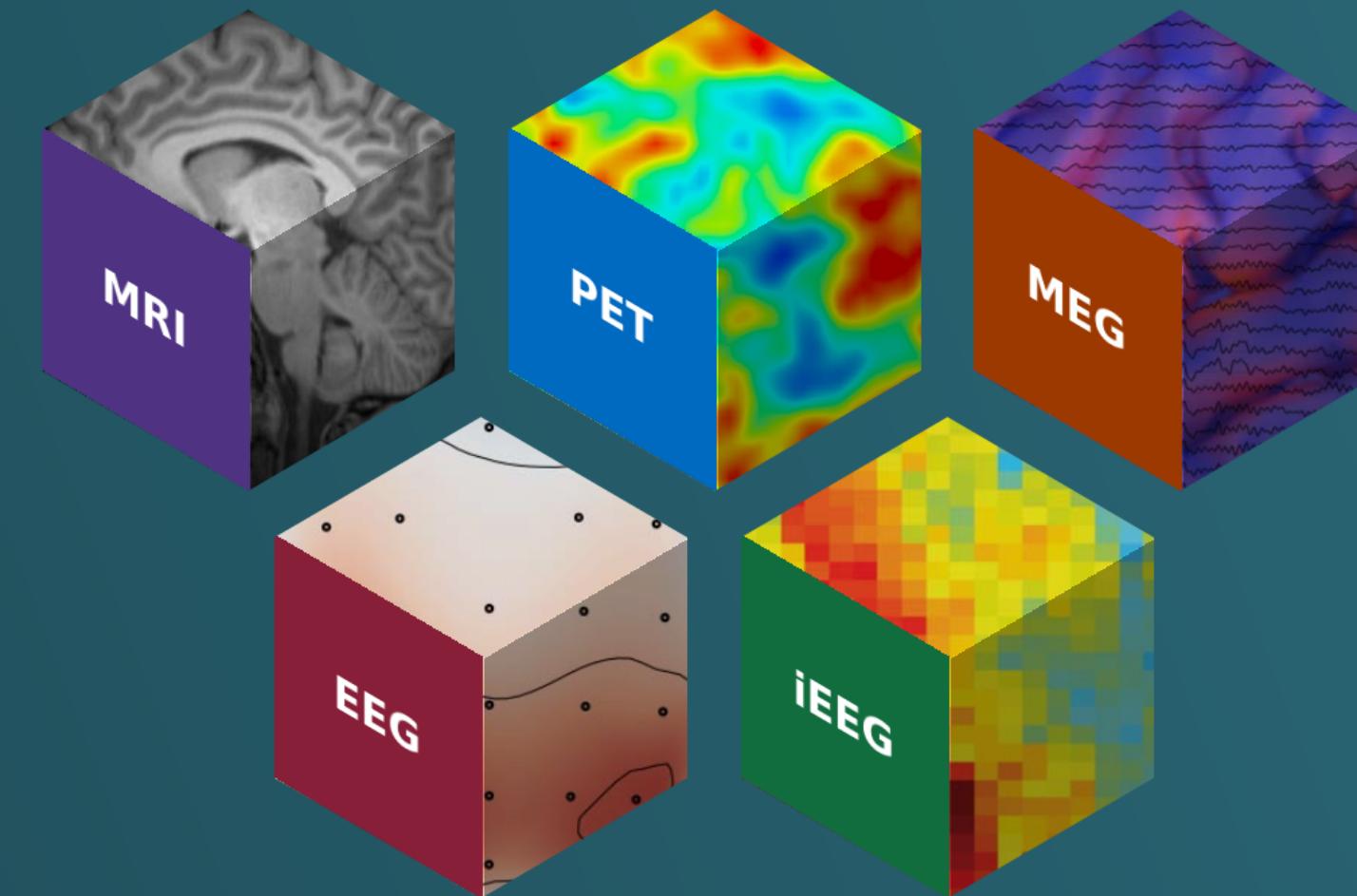
Or

Search

SIGN IN

G Google

ID ORCID



## Validation Using BIDS

The [Brain Imaging Data Structure](#) (BIDS) is an emerging standard for the organization of neuroimaging data.

Want to contribute to BIDS?

Visit the [Google discussion group](#) to contribute.

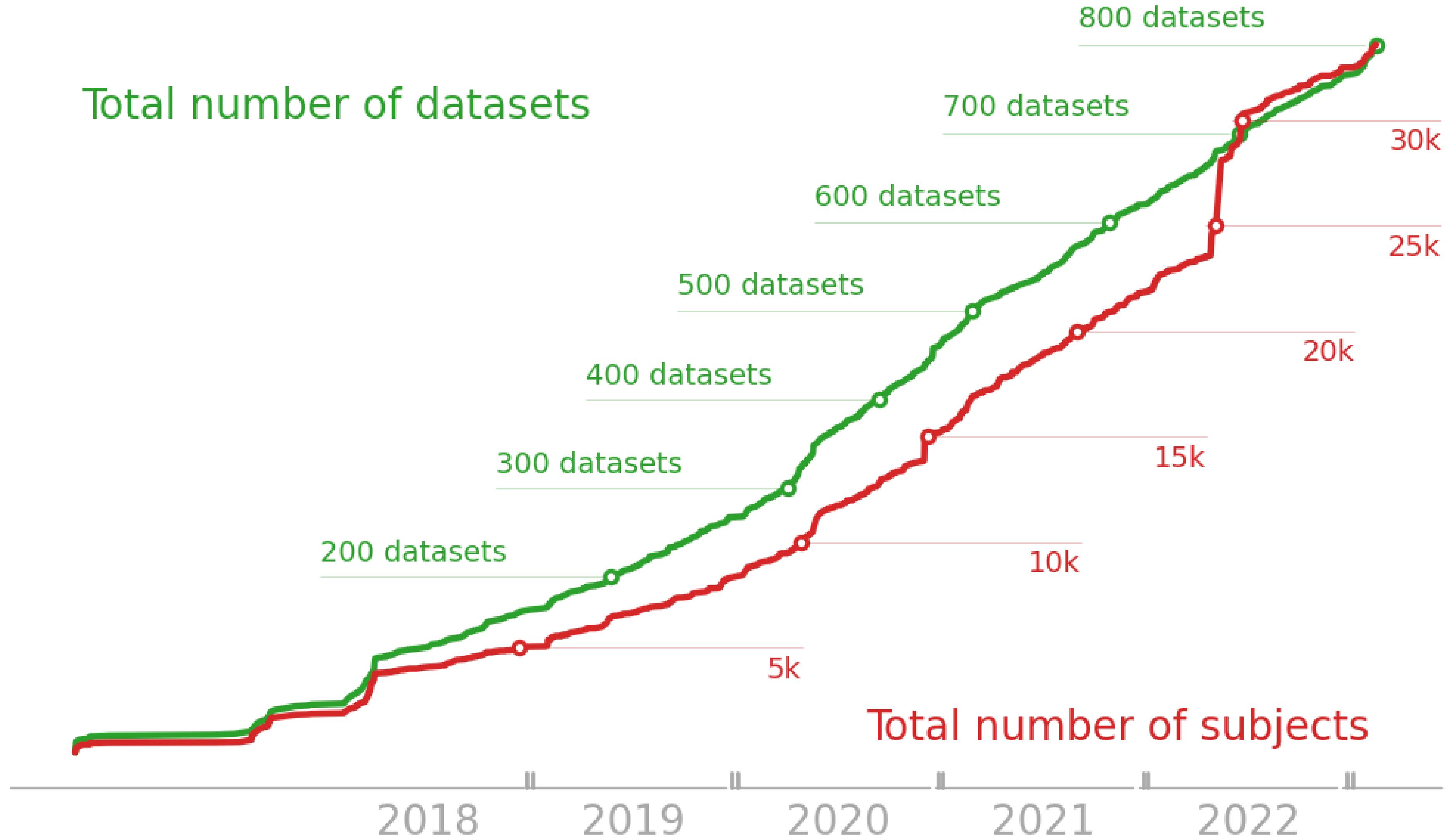
https://poldrack.github.io/talks-openneuro/



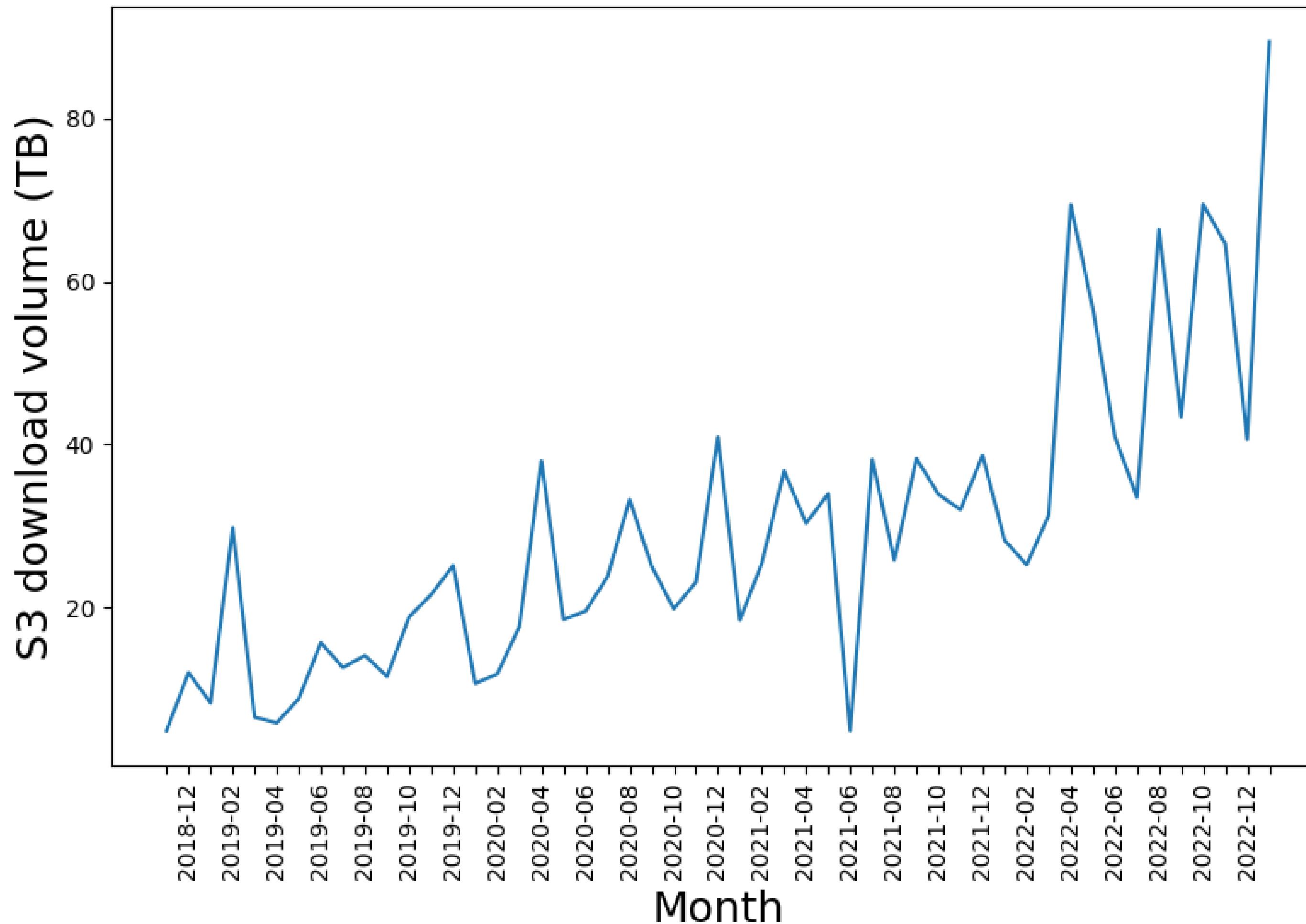
## OpenNeuro Runs on DataLad

Want to access OpenNeuro datasets with DataLad? Visit the [dataset collection on GitHub](#).

A data management solution built on Git and Git-annex. Read more about [DataLad](#)



## S3 download volume, 2018-2023



# New developments

- Renewal grant approved for funding (2023-2028)
- Migration from AWS to GCP
  - Improved performance, easier scaling, and lower cost per dataset
  - Open datasets still shared via AWS Public Datasets
- Improved performance for large datasets (now handling 2TB+)
- Published data retention/admin policies for reference in Data Sharing Plans
- Links to NEMAR for supported datasets
- Initial support for ORCID integration
- Support for fNIRS upload

# Schema-based validation for BIDS

- OpenNeuro data ingestion relies upon JavaScript BIDS validator
- Original validator built the standard structure directly into the validator code
  - Made addition of new data types quite laborious
  - JavaScript expertise is relatively rare in our community
- Work began in 2021 on defining the standard separately using a schema that
  - Three sprints to date involving the OpenNeuro team, NIMH Data Science and Sharing team, and others

# Goals of the schema-based validator effort

- Authoritative, machine-readable descriptions of BIDS concepts
  - Reduce need for proliferation of implementations, such as the PyBIDS configuration object
- Enforce consistency in specification by generating text and tables from schema
  - Unify terms reused in multiple locations in the specification
- Reduce burden of writing BEPs by eliminating the requirement for validator coding
  - Consequence: Whether a rule can be encoded in the schema or will need custom code / schema expansion is now an informal review criterion.

# What is the schema?

- A hierarchy of YAML documents in the specification repository, under src/schema.
- Three major divisions:
  - Objects (objects.\*)
    - Definitions of BIDS concepts like entities and terms like sidecar values
  - Rules (rules.\*)
    - Validatable rules, such as entity ordering or permissible/required sidecar values
  - Meta-schema (meta.\*)
    - Defines a “context” object to which rules can be applied
    - Potentially expanded to any definitions or rules related to the schema itself

```
▼ object {5}
  schema_version : 0.6.0
  bids_version : 1.8.0
  ▶ meta {2}
  ▶ objects {10}
  ▼ rules {9}
    ▶ checks {9}
    ▶ common_principles [15]
    ▶ dataset_metadata {4}
    ▶ entities [28]
    ▶ errors {27}
    ▶ files {3}
    ▶ modalities {8}
    ▶ sidecars {17}
    ▶ tabular_data {10}
```

# Integrating the schema-based validator into OpenNeuro

- Completed
  - Schema based validation can be run server side by OpenNeuro's dataset worker
- Remaining
  - Integration for client side (prior to upload) usage
  - Port OpenNeuro CLI to new JavaScript runtime (Deno) to support calling the schema validator
  - UI to control use of schema validator during upload
  - Support for dual validation with existing and schema validators
- Decisions
  - How do we handle cases when one validator passes and the other fails (for early users)?

# Sharing of derivative datasets

- Derivative data (e.g. preprocessed MRI data) are now supported through the BIDS-Derivatives extension (developed under BRAIN R24MH114705)
- One major motivation for the schema-based validator was to support the validation of derivative datasets
- In advance of this, we have been developing support for sharing of derivative datasets within OpenNeuro



## BIDS Validation ▾

3 WARNINGS

Valid

brainlife.io

Clone ▾

Files

Download

Derivatives

Metadata

## Available Derivatives

## Acknowledgements

These derivatives were generated on the [Texas Advanced Computing Center](#) Frontera computing system [1] through their [Pathways allocation](#). This work was also funded by the [NIH BRAIN Initiative](#).

[1]: Dan Stanzione, John West, R. Todd Evans, Tommy Minyard, Omar Ghattas, and Dhabaleswar K. Panda. 2020. Frontera: The Evolution of Leadership Computing at the National Science Foundation. In Practice and Experience in Advanced Research Computing (PEARC '20), July 26–30, 2020, Portland, OR, USA. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3311790.3396656>

[ds002687-mriqc](#)

## Download from S3

```
aws s3 sync --no-sign-request s3://openneuro-derivatives/mriqc/ds002687-mriqc ds002687-mriqc
```

## Download with DataLad

```
datalad install https://github.com/OpenNeuroDerivatives/ds002687-mriqc.git
```

[ds002687-fmriprep](#)

## Download from S3

```
aws s3 sync --no-sign-request s3://openneuro-derivatives/fmriprep/ds002687-fmriprep ds002687-fmriprep
```

## Download with DataLad

```
datalad install https://github.com/OpenNeuroDerivatives/ds002687-fmriprep.git
```

## OpenNeuro Accession Number

ds002687

## Authors

James R. Booth, Gillian E. Cooke, Jessica Gayda, Rubi Hammer, Marisa N. Lytle, Mark A. Stein, Michael Tennekoon

## Available Modalities

MRI

## Versions

1.2.0

Created: 2021-08-09

Versions ▾

## Tasks

VisuoSpatial Working Memory, Large Reward, Delayed Feedback, VisuoSpatial Working Memory, Large Reward, Immediate Feedback, VisuoSpatial Working Memory, Small Reward, Delayed Feedback, VisuoSpatial Working Memory, Small Reward, Immediate Feedback, Verbal Working Memory, Large Reward, Delayed Feedback, Verbal Working Memory, Large Reward, Immediate Feedback, Verbal Working Memory, Small Reward, Delayed Feedback, Verbal Working Memory, Small Reward, Immediate Feedback

## Uploaded by

Marisa Lytle on 2020-04-07 - almost 3 years ago

## Last Updated

2021-08-09 - over 1 year ago

## Sessions

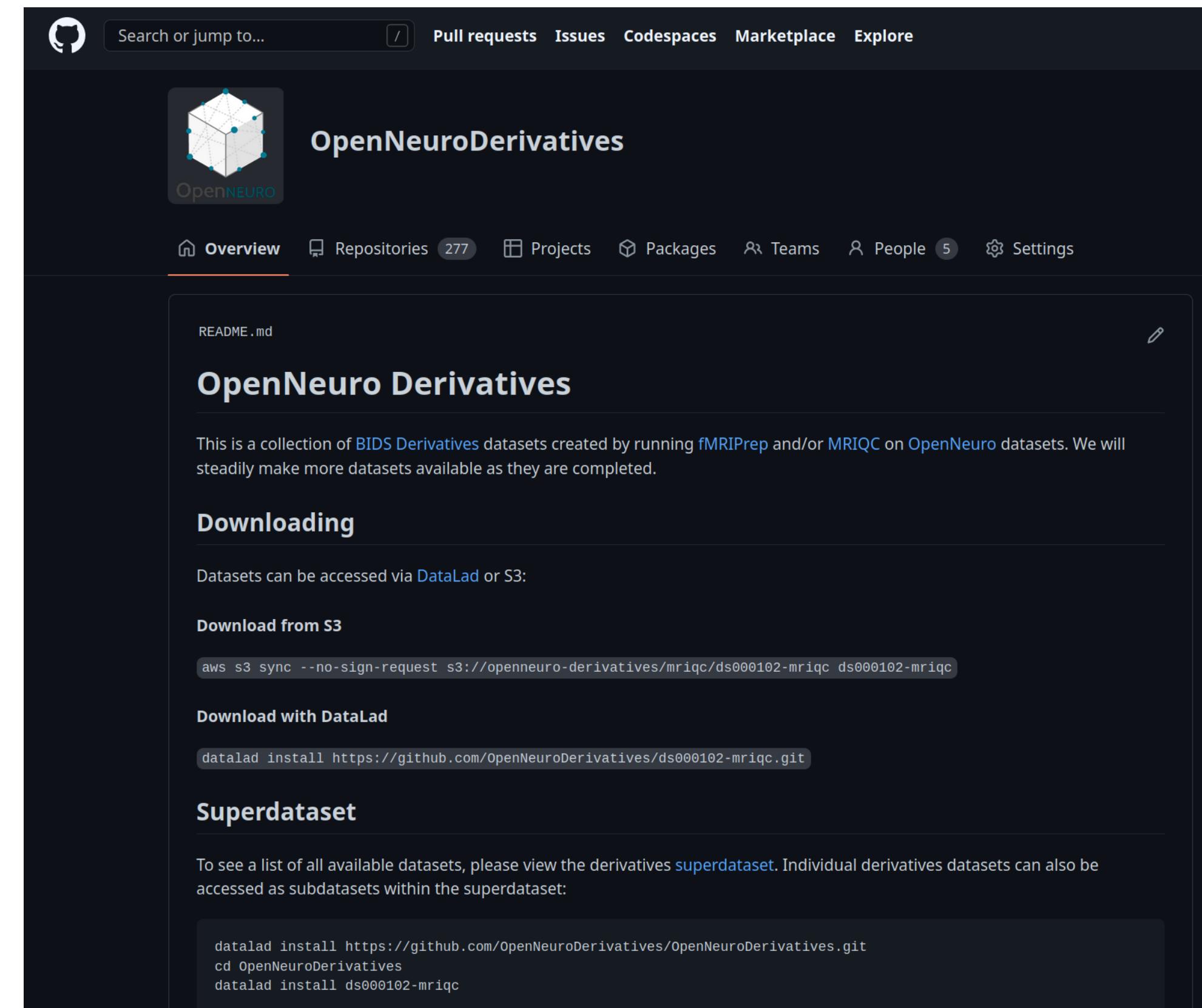
1

## Participants

24

# Preprocessing/QC of OpenNeuro fMRI datasets

- All human fMRI datasets from OpenNeuro are being run through MRIQC and fMRIPrep
  - Via a Pathways allocation on the TACC Frontera supercomputer
- To date:
  - 229/430 datasets successfully run with MRIQC
  - 65/430 successfully preprocessed with fMRIPrep
- Derivatives openly available via S3 or Datalad



<https://github.com/OpenNeuroDerivatives>

# Alignment with new NIH Data Sharing Policy

“we highlight the importance of researchers considering whether, in choosing where and how to make their data available (if not already specified by an FOA or funding NIH ICO expectation), access to scientific data derived from humans should be controlled, even if de-identified and lacking explicit limitations on subsequent use...SACHRP identified concerns regarding re-identification of otherwise de-identified data, and indeed technological advances and increasing interoperability among data resources, while providing opportunities for new analyses, present identifiability concerns that are widely acknowledged.”

Final NIH Policy for Data Management and Sharing (<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>)

# Potential solution

- We are considering adding a click-through agreement for all users requiring agreement to not attempt reidentification
- This would cause friction because automated download would require an API key for each user
  - vs. unauthenticated downloads from S3/datalad at present

# Neuroethics supplement

- A neuroethics supplement has supported Dr. Annie Jwa, a legal scholar with expertise in neuroethics
- Jwa & Poldrack (2022, Journal of Law and the Biosciences) argued for development of regulatory protections against misuse of neuroscience data
  - “Neuroscience Information Non-discrimination Act”
- With funding from Stanford HAI, we are working with Sanmi Koyejo to examine potential for adversarial perturbations of structural MRI data to disable reidentification attempts

Journal of Law and the Biosciences, 1–25  
<https://doi.org/10.1093/jlb/lSac025>  
Original Article



## Addressing privacy risk in neuroscience data: from data protection to harm prevention

Anita S. Jwa \* and Russell A. Poldrack

Department of Psychology, Stanford University, 450 Jane Stanford Way, Building 420, Stanford, CA 94306, USA

\*Corresponding author. E-mail: anniejwa@stanford.edu

### ABSTRACT

A recent increase in the amount and availability of neuroscience data within and outside of research and clinical contexts will enhance reproducibility of neuroscience research leading to new discoveries on the mechanisms of brain function in healthy and disease states. However, the uniquely sensitive nature of neuroscience data raises critical concerns regarding data privacy. In response to these concerns, various policy and regulatory approaches have been proposed to control access to and disclosure of neuroscience data, but excessive restriction may hamper open science practice in the field. This article argues that it may now be time to expand the scope of regulatory discourse beyond protection of neuroscience data and to begin contemplating how to prevent potential harm. Legal prohibition of harmful use of neuroscience data could provide an ultimate safeguard against privacy risks and would help us chart a path toward protecting data subjects without unduly limiting the benefits of open science practice. Here we take the Genetic Information Non-Discrimination Act (GINA) as a reference for this new legislation and search for answers to the core regulatory questions based on what we have learned from the enactment of the GINA and the merits and weaknesses of the protection it provides.

# The Poldrack Lab



# OpenNeuro Team



## Funding



## Collaborators



