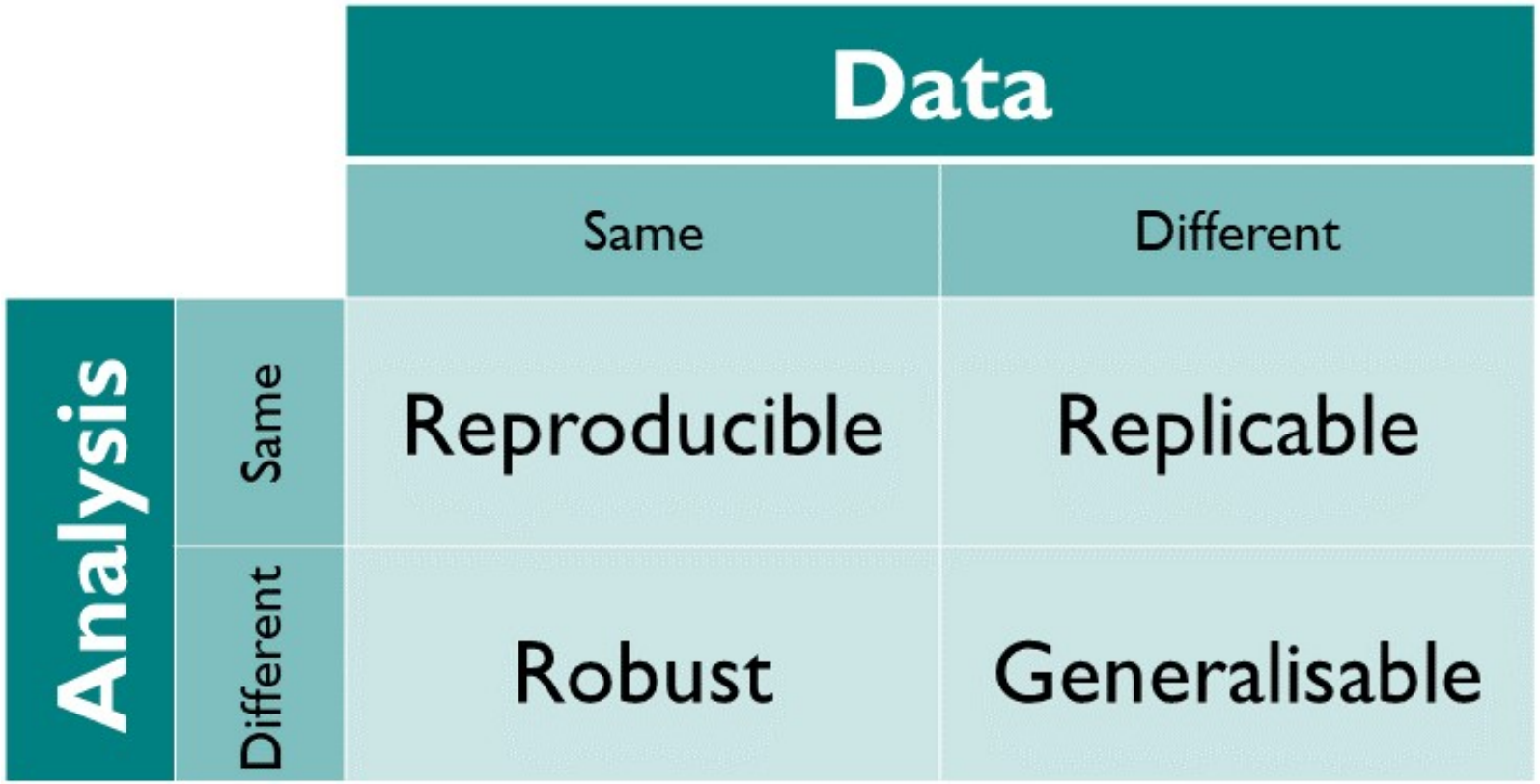
Toward an Open Science Ecosystem in Neuroimaging

Russ Poldrack
Stanford University

Transparency is essential for reproducibility



https://poldrack.github.io/talks-OpenScienceEcosystem/
https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html

“we can distill Claerbout’s insight into a slogan:

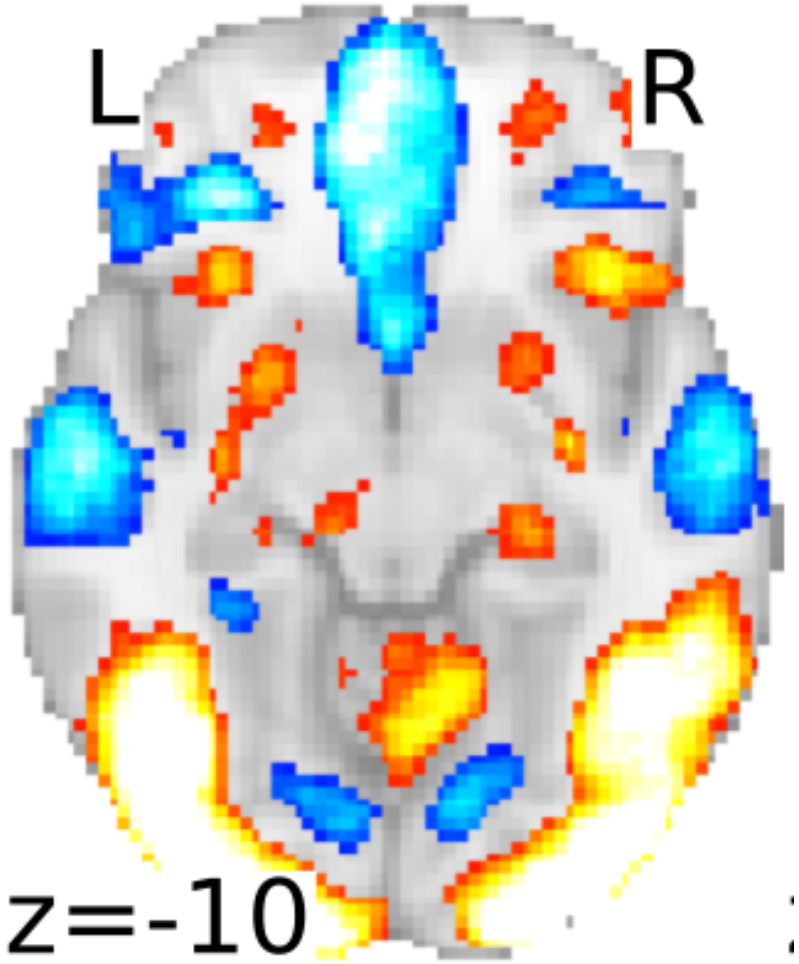
An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures..”

- Buckheit & Donoho, 1995

- Jon Claerbout

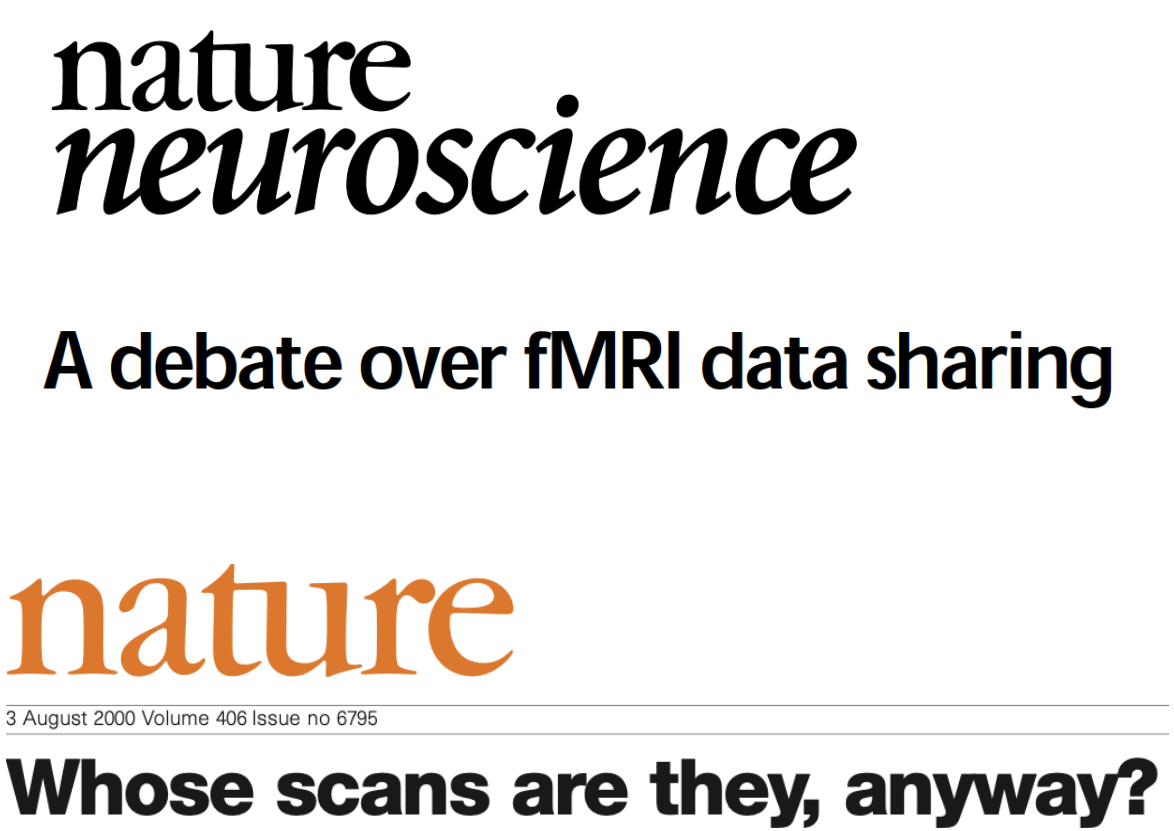
Why neuroimaging is a best-case scenario for open science

- Magnetic resonance imaging (MRI) is the primary tool for studying human brain structure and function
- MRI data are digital end-to-end
  - From MRI scanner to automated analysis
  - Usually zero/few manual analysis steps
- The field has largely converged on:
  - a standardized image format (NiFTI)
  - a ~common spatial coordinate system



A false start for fMRI data sharing

A false start for fMRI data sharing



This letter comes from a group of scientists who are publishing papers using fMRI to understand the links between brain and behavior. We are writing in reaction the recent announcement of the creation of the National fMRI Data Center (www.fmridc.org). In the letter announcing the creation of the center, it was also implied that leading journals in our field may require authors of all fMRI related papers accepted for publication to submit all experimental data pertaining to their paper to the Data Center. … We are particularly concerned with any journal's decision to require all authors of all fMRI related papers accepted for publication to submit all experimental data pertaining to their paper to the Data Center.

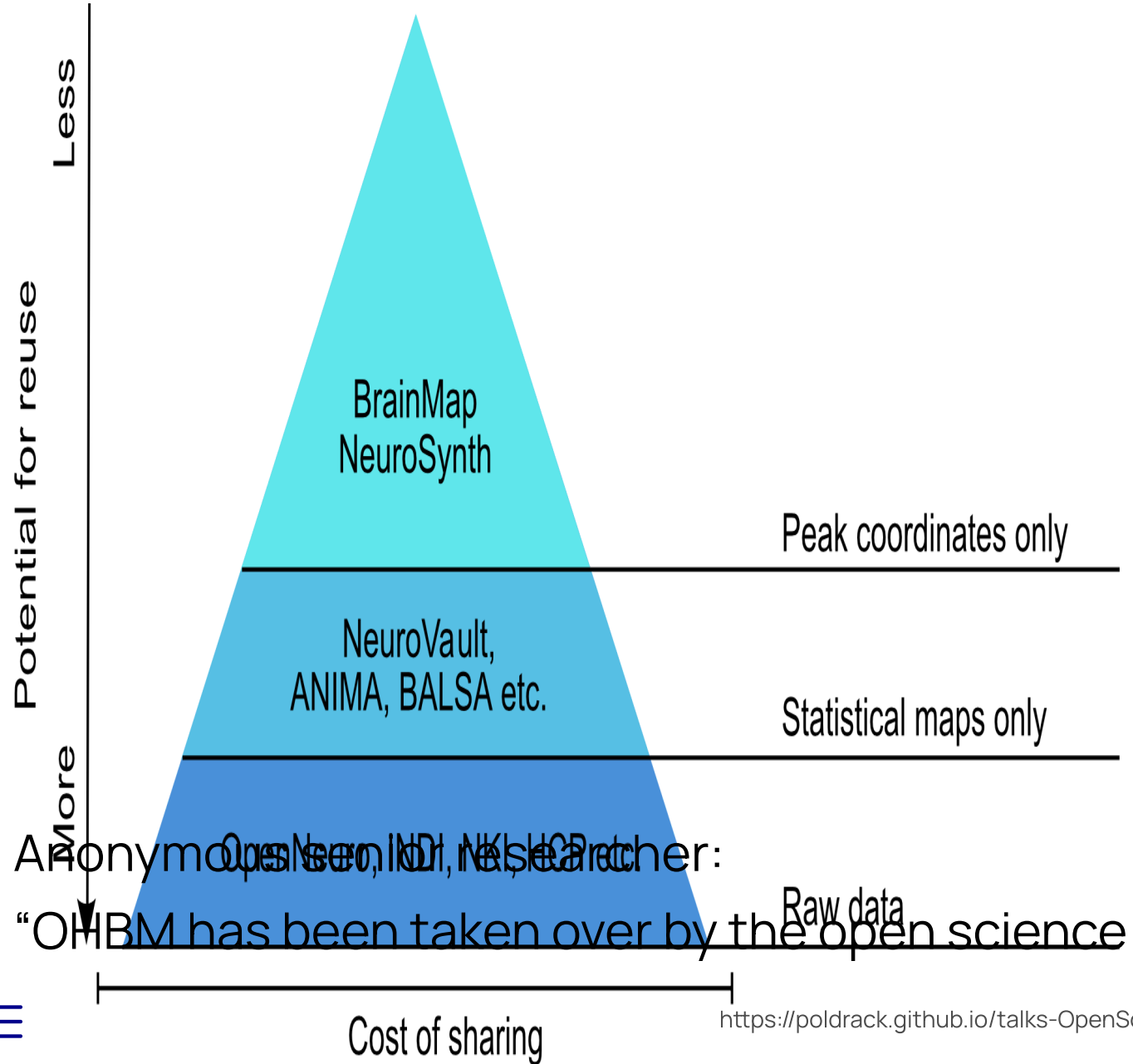
2010: The year data sharing broke in neuroimaging

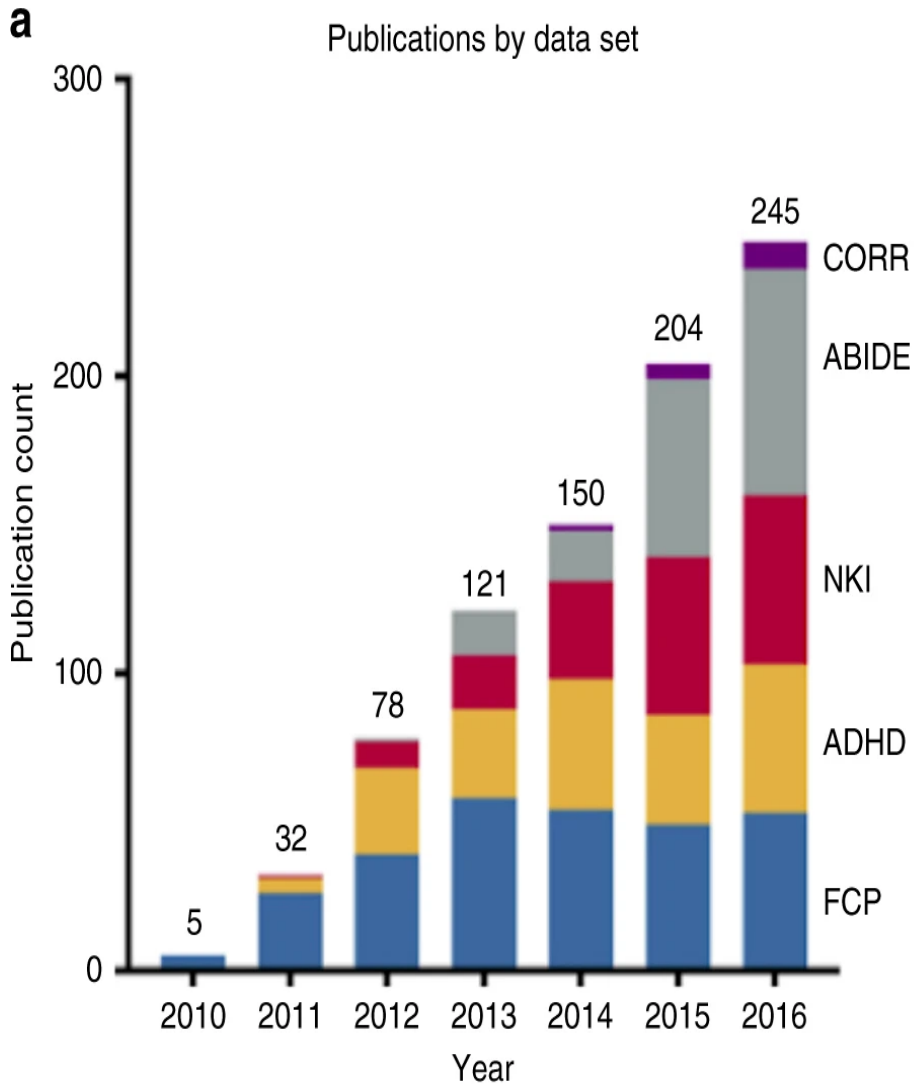
Toward discovery science of human brain function

Bharat B. Biswal[a], Maarten Mennes[b], Xi-Nian Zuo[b], Suril Gohel[a], Clare Kelly[b], Steve M. Smith[c], Christian F. Beckmann[c], Jonathan S. Adelstein[b], Randy L. Buckner[d], Stan Colcombe[e], Anne-Marie Dogonowski[f], Monique Ernst[g], Damien Fair[h], Michelle Hampson[i], Matthew J. Hoptman[j], James S. Hyde[k], Vesa J. Kiviniemi[l], Rolf Kötter[m], Shi-Jiang Li[n], Ching-Po Lin[o], Mark J. Lowe[p], Clare Mackay[c], David J. Madden[q], Kristoffer H. Madsen[f], Daniel S. Margulies[r], Helen S. Mayberg[s], Katie McMahon[t], Christopher S. Monk[u], Stewart H. Mostofsky[v], Bonnie J. Nagel[w], James J. Pekar[x], Scott J. Peltier[y], Steven E. Petersen[z], Valentin Riedl[aa], Serge A. R. B. Rombouts[bb], Bart Rypma[cc], Bradley L. Schlaggar[dd], Sein Schmidt[ee], Rachael D. Seidler[ff,u], Greg J. Siegle[gg], Christian Sorg[hh], Gao-Jun Teng[ii], Juha Veijola[jj], Arno Villringer[ee,kk], Martin Walter[ll], Lihong Wang[q], Xu-Chu Weng[mm], Susan Whitfield-Gabrieli[nn], Peter Williamson[oo], Christian Windischberger[pp], Yu-Feng Zang[qq], Hong-Ying Zhang[ii], F. Xavier Castellanos[b,j], and Michael P. Milham[b,1]

4734–4739 | PNAS | March 9, 2010 | vol. 107 | no. 10

Data sharing is becoming the norm in neuroimaging



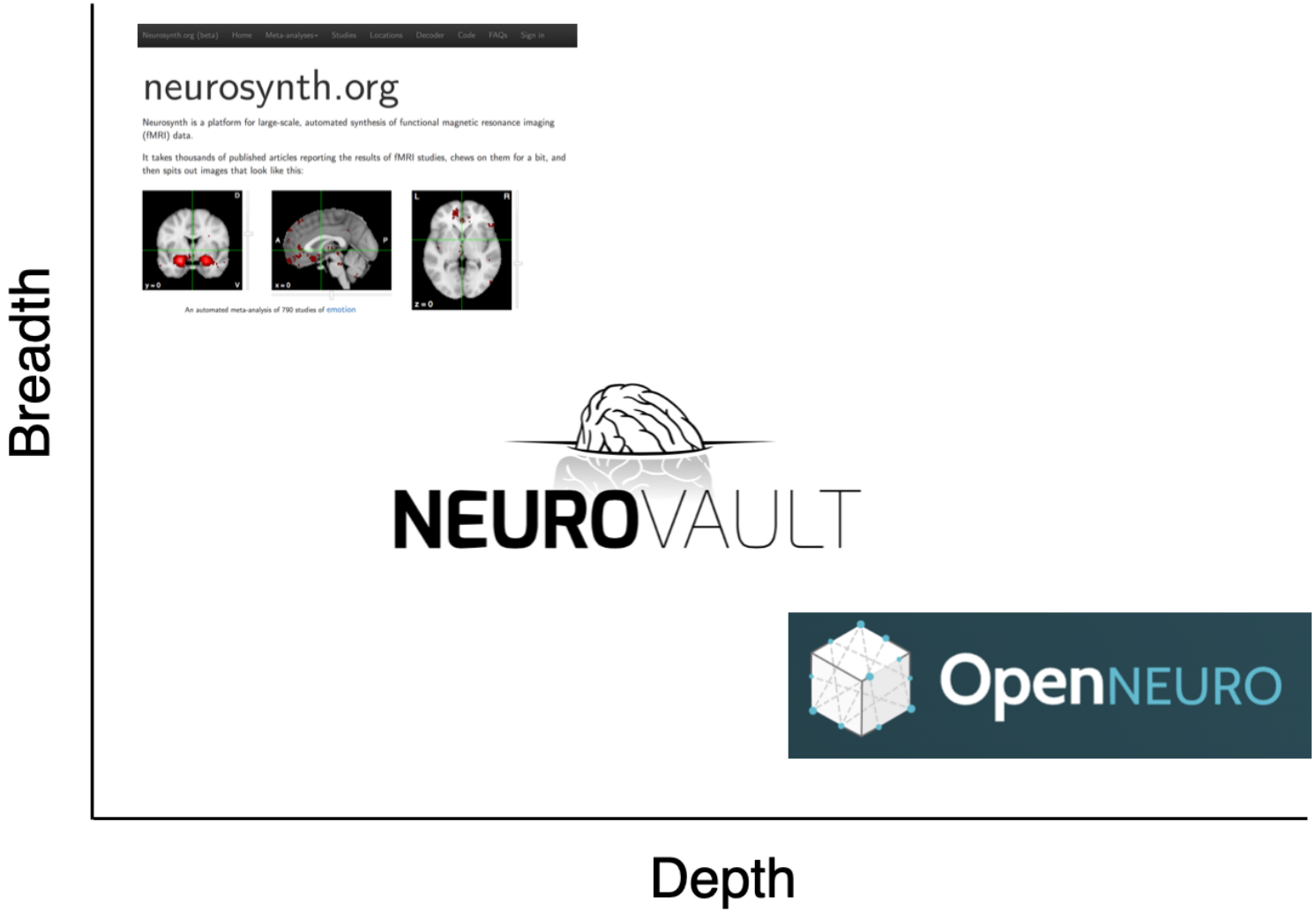


Milham et al., *Nature Communications*, 2018

Anonymous reviewer:

"OHBM has been taken over by the open science zealots!"

An open ecosystem for retrospective data sharing



- Neurosynth.org: Open database of published neuroimaging coordinates
- Neurovault.org: Open archive for neuroimaging results
- OpenNeuro.org: Open archive for raw/processed neuroimaging data

Maximally open sharing

- Data shared under maximally permissive data use agreements:
  - Neurosynth: Open Data Commons Open Database License v1.0
  - Neurovault: CC0
  - OpenNeuro: CC0
- All data available programmatically via web API as well as web page

- CC0 enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.
  - https://creativecommons.org/share-your-work/public-domain/cc0/

Neurosynth: Sharing activation coordinates

- Brain activity is reported in a (somewhat) standardized coordinate system

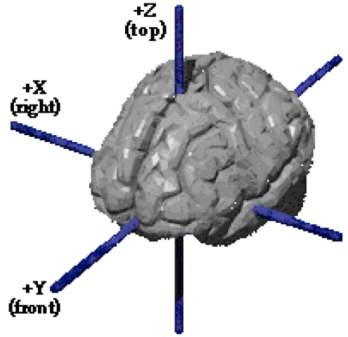


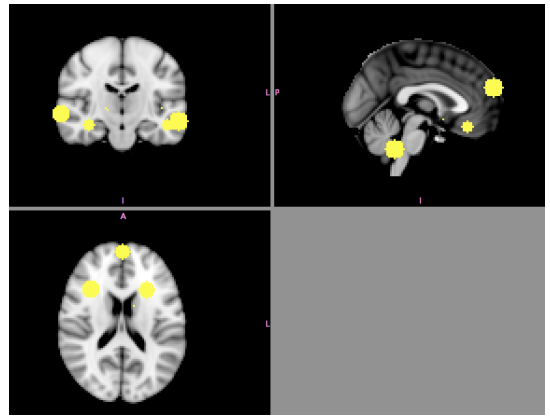
Table 1
Regions that showed a condition × time interaction in the ANOVA analysis

No.	Region	Hemisphere	BA	*x*	*y*	*z*	mm³
1	Middle/superior temporal gyrus	L	21/22/37	−52	−54	9	13257
2	Inferior frontal gyrus	L	45/46/9	−49	26	6	2781
3	Posterior cerebellum	L		−19	−79	−38	2214
4	Dorsomedial PFC	L	9/8	−11	42	47	3051
5	Left anterior PFC	L	10	−37	49	15	2025
6	Inferior parietal cortex	L	40/7	−42	−58	47	3132
7	Dorsal premotor cortex	L	6	−43	0	50	1485
8	Lingual gyrus	L	17	−10	−95	−2	378
9	Middle /superior temporal gyrus	R	21/22/37	52	−40	5	16470
10	Inferior frontal gyrus	R	45/46	51	28	6	2241
11	Posterior cerebellum	R		23	−78	−34	2808
12	Dorsomedial PFC	R	9	5	53	29	405
13	Right anterior PFC	R	10	38	42	21	5022
14	Inferior parietal cortex	R	40/7	42	−53	48	9963
15	Superior frontal gyrus	R	6/8	10	28	60	297
16	Anterior cingulate cortex	M	32	0	26	35	5076
17	Posterior cingulate cortex	M	23/31/7	0	−35	31	9612
18	Precuneus	M	7/19	1	−76	36	10044

Creating meta-analytic maps

- Automated Coordinate Extraction
  - Automatically extracts activation tables from fMRI papers for 17 journals
  - Current database has 14,371 papers (with full text)
  - 84% sensitivity, 97% specificity against manual database (SumsDB)
- Meta-analytic maps created for each paper
  - 10mm sphere placed at each focus

X	Y	Z
12	57	-6
33	21	15
24	-6	51
28	10	18



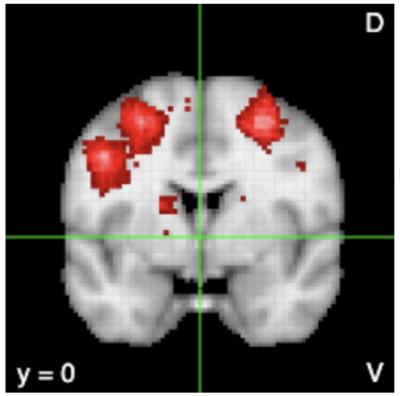
Yarkoni et al, 2011, *Nature Methods*

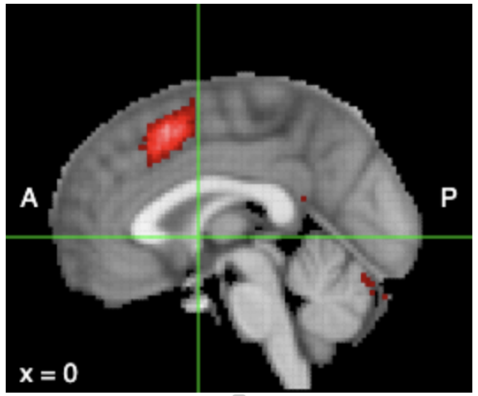
working memory

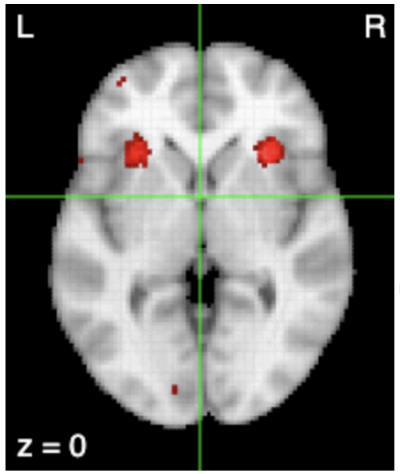
An automated meta-analysis of 1091 studies

Search for another term:

Maps | Studies | FAQs







z-score: 0

What's here?

X: 0 Y: 0 Z: 0

Layers

- working memory: association test
- working memory: uniformity test
- anatomical

Color palette: red

Positive/Negative: positive

Crosshairs

Pan/zoom

Labels

Thresholds: 0 0

Opacity: 1

Example of Neurosynth usage

Situating the default-mode network along a principal gradient of macroscale cortical organization

Daniel S. Margulies[a,1], Satrajit S. Ghosh[b,c], Alexandros Goulas[d], Marcel Falkiewicz[a], Julia M. Huntenburg[a,e], Georg Langs[f,g], Gleb Bezgin[h], Simon B. Eickhoff[i,j], F. Xavier Castellanos[k,l], Michael Petrides[m], Elizabeth Jefferies[n,o], and Jonathan Smallwood[n,o]

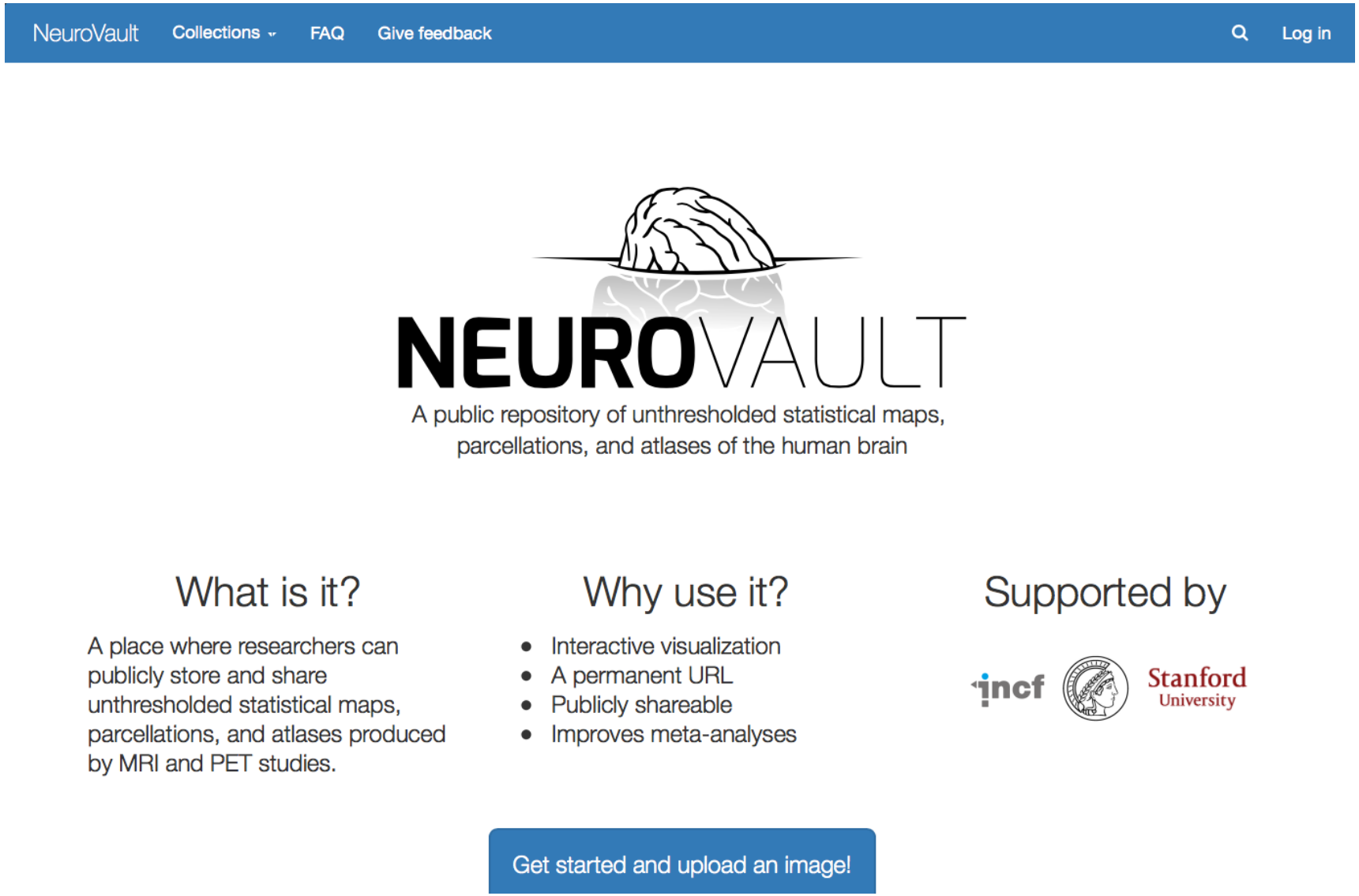
12574–12579 | PNAS | November 1, 2016 | vol. 113 | no. 44

- Identified gradients of functional organization across the cortex
- Used Neurosynth to identify the most common terms associated with each

Neurosynth impact

Neurovault: Sharing neuroimaging results

- The results of most neuroimaging studies are images with statistical estimates at each voxel
- Neurovault.org is an open archive for these results



Gorgolewski et al., 2015, *Frontiers in Neuroinformatics*

- **Collections**
  - A set of images (such as all images from a particular paper) can be uploaded as a collection
  - Each collection receives a persistent identifier

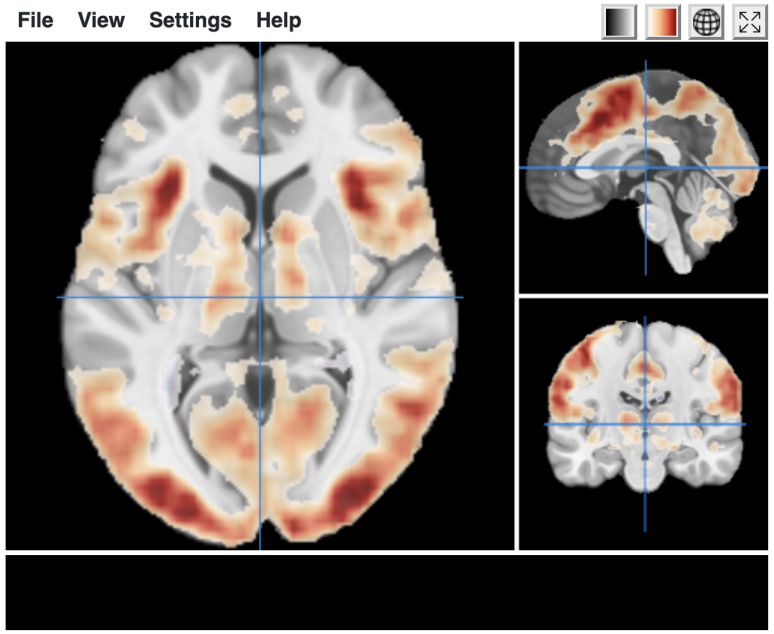
NeuroVault Collections Metaanalyses About RussPoldrack Search Search

Preprocessed Consortium for Neuropsychiatric Phenomics dataset

Related article: http://doi.org/10.12688/f1000research.11964.2

Source data:

3D View Download



Group Metadata

Show 7 entries Search:

View	ID	Name	Type
	49974	BART Accept	T map
	49975	BART AcceptParam - ExplodeParam	T map
	49976	BART AcceptParam - RejectParam	T map
	49977	BART AcceptParametric	T map
	49978	BART Accept_RT	T map
	49979	BART Control	T map
	49980	BART Explode - Reject	T map

Showing 1 to 7 of 178 entries First Previous Next Last

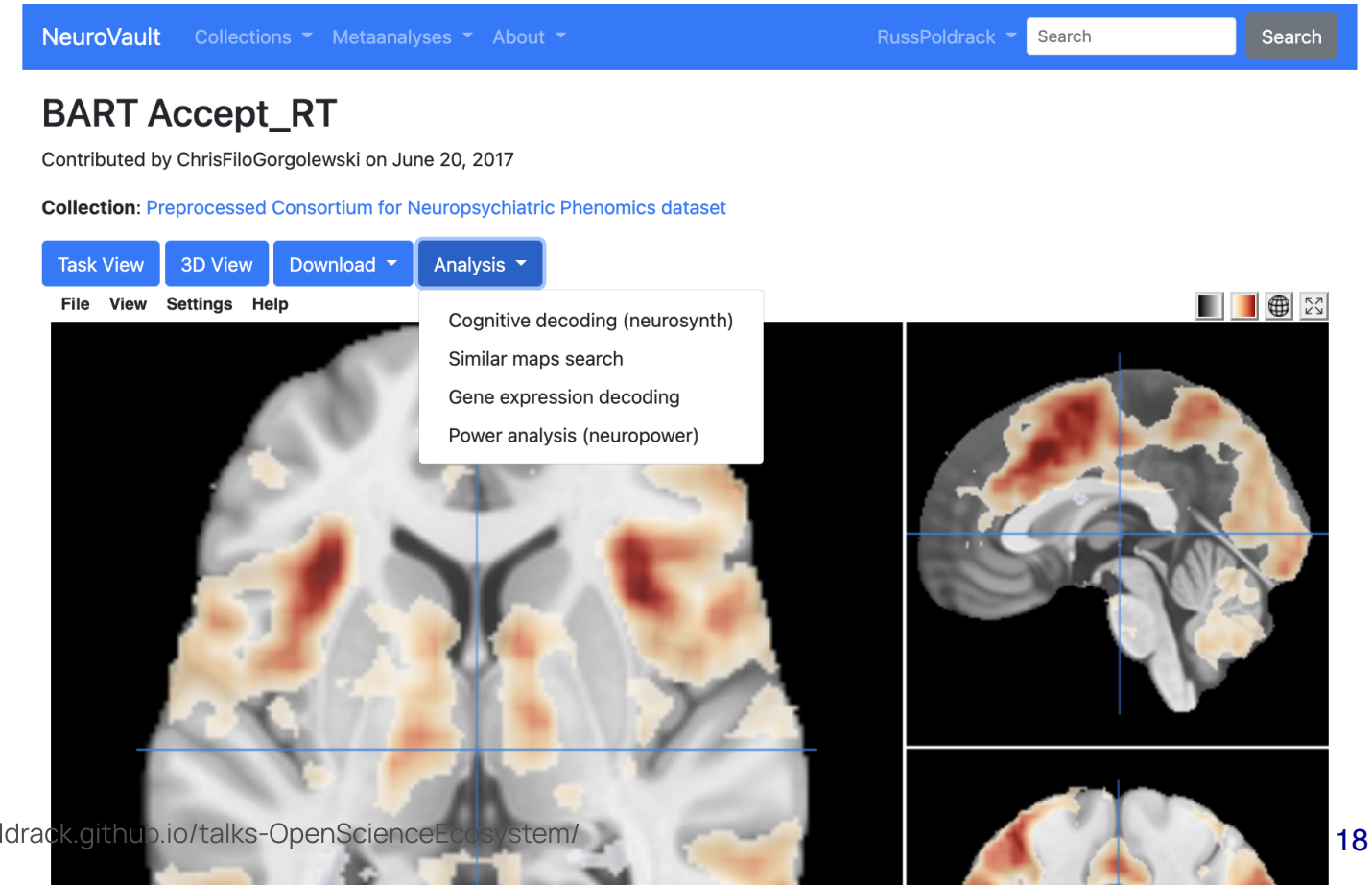
Citation guidelines

If you use the data from this collection please include the following persistent identifier in the text of your manuscript:

`https://identifiers.org/neurovault.collection:2606`

This will help to track the use of this data in the literature. In addition, consider also citing the paper related to this collection

- **Image browser**
  - Individual images can be browsed and downloaded
  - A number of analysis tools can also be applied
  - Each image also receives a persistent identifier

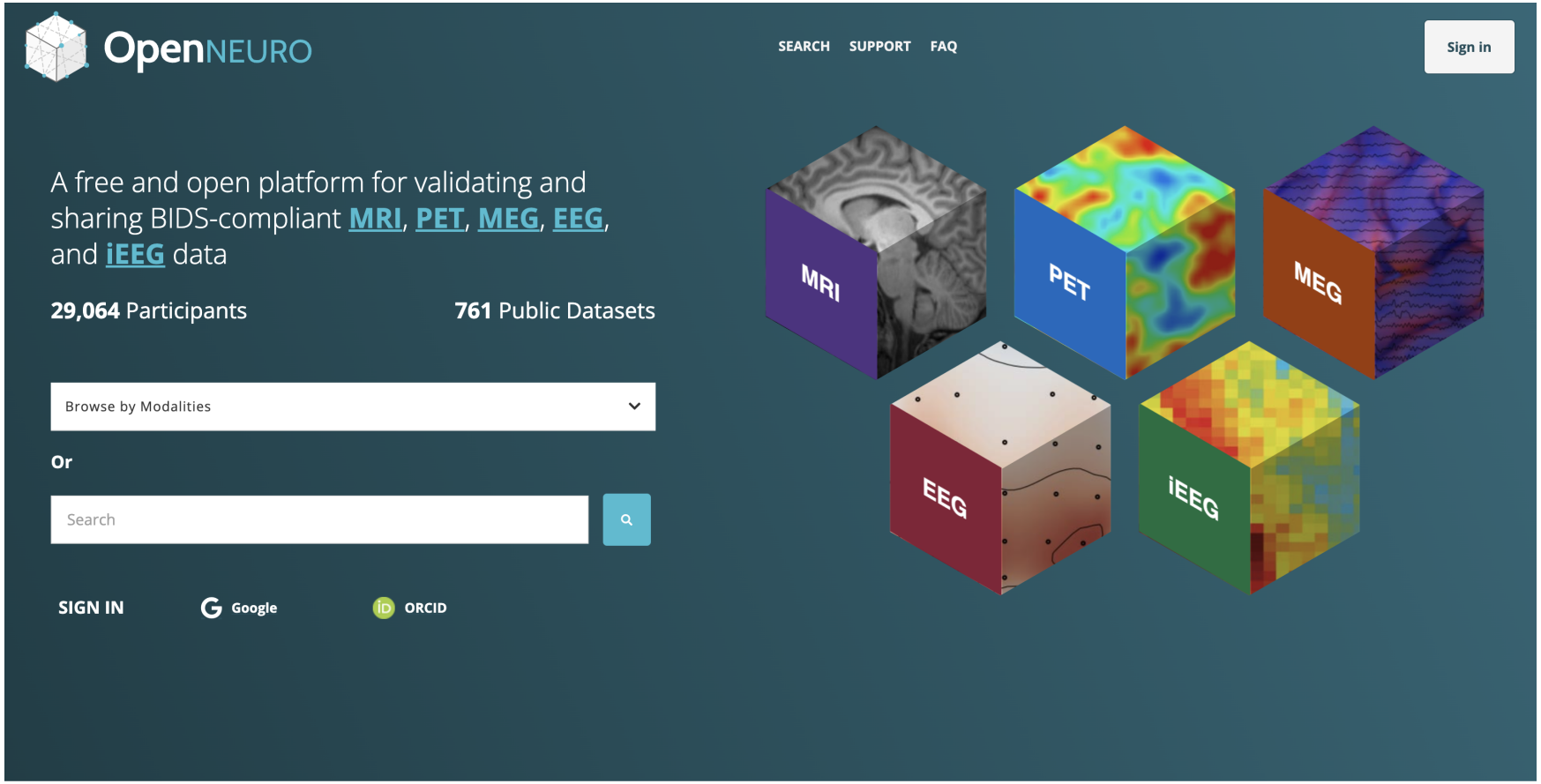


Example of Neurovault usage

Comprehensive decoding mental processes from Web repositories of functional brain images

Romuald Menuet[5,6], Raphael Meudec[1,2,3,6], Jérôme Dockès[4], Gael Varoquaux[1,2,3] & Bertrand Thirion[1,2,3]✉

Scientific Reports | (2022) 12:7050



BIDS
BRAIN IMAGING DATA STRUCTURE

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.

DataLad

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

Simply sharing data is not sufficient

It must be shared in a way that makes it useful!

It’s easy to share data badly

Data Sharing and Management Snafu in 3 Short Acts

- https://www.youtube.com/watch?v=N2zK3sAtr-4

- I received the data, but when I opened it up it was in hexadecimal
- Yes, that is right
- I cannot read hexadecimal
- You asked for my data and I gave it to you. I have done what you asked.

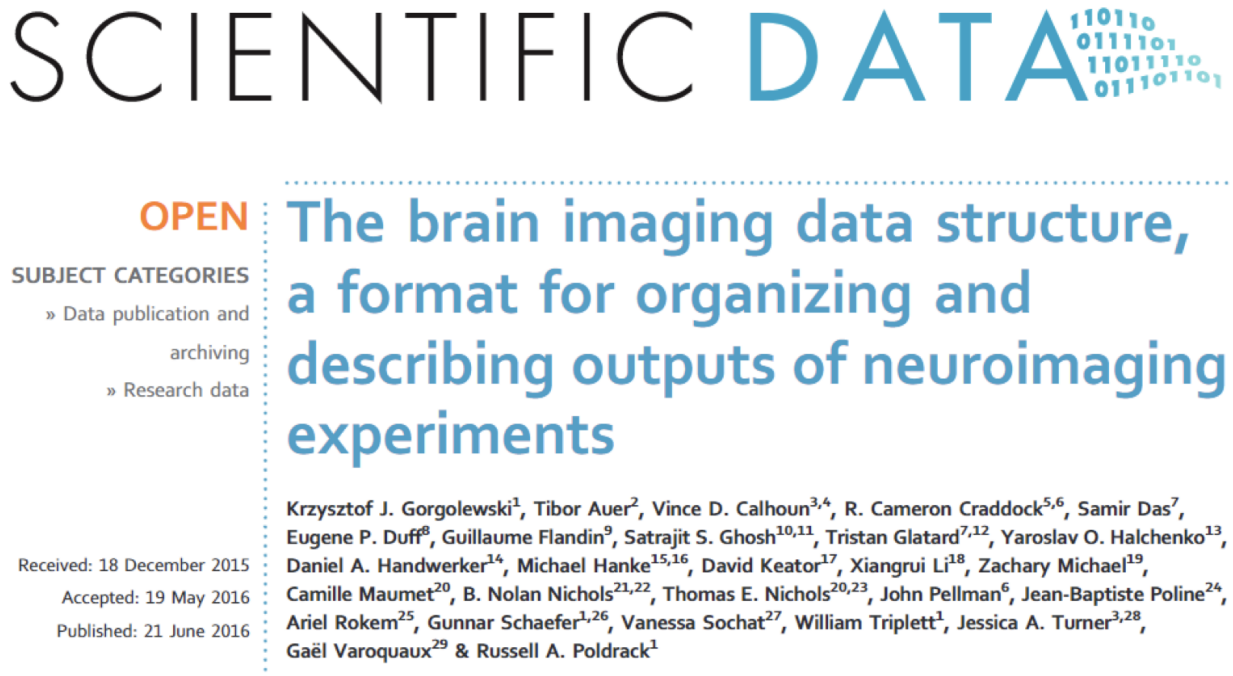
…

- Is there a guide to the data anywhere?
- Yes, of course, it is the article that is published in Science.

Brain Imaging Data Structure (BIDS)

- A community-based open standard for neuroimaging data

BIDS
BRAIN IMAGING DATA STRUCTURE


SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES
» Data publication and archiving
» Research data

The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments

Krzysztof J. Gorgolewski[1], Tibor Auer[2], Vince D. Calhoun[3,4], R. Cameron Craddock[5,6], Samir Das[7], Eugene P. Duff[8], Guillaume Flandin[9], Satrajit S. Ghosh[10,11], Tristan Glatard[7,12], Yaroslav O. Halchenko[13], Daniel A. Handwerker[14], Michael Hanke[15,16], David Keator[17], Xiangrui Li[18], Zachary Michael[19], Camille Maumet[20], B. Nolan Nichols[21,22], Thomas E. Nichols[20,23], John Pellman[6], Jean-Baptiste Poline[24], Ariel Rokem[25], Gunnar Schaefer[1,26], Vanessa Sochat[27], William Triplett[1], Jessica A. Turner[3,28], Gaël Varoquaux[29] & Russell A. Poldrack[1]

Received: 18 December 2015
Accepted: 19 May 2016
Published: 21 June 2016

The development of BIDS

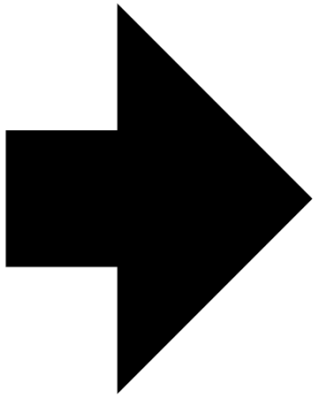
- January 2015
  - Initial stakeholder meeting at Stanford (funded by INCF)
  - Initiated development of a draft standard
- September 2015
  - Draft standard posted to BIDS web site with 22 example datasets
  - Solicited feedback from community
- June 2016
  - Published paper
- September 2018
  - BIDS-standard Github organization started

BIDS Principles

- Adoption is crucial
  - Keep it as similar to existing practices as possible
    - Don’t let technology override usability!
  - Focus on engaging the community
- Don’t reinvent the wheel
  - Use existing standards when possible
- 80/20 rule
  - Focus on the most common use cases
  - Don’t let the perfect be the enemy of the good!

From DICOM to BIDS

- dicomdir/
  - 1208200617178_22/
    - 1208200617178_22_8973.dcm
    - 1208200617178_22_8943.dcm
    - 1208200617178_22_2973.dcm
    - 1208200617178_22_8923.dcm
    - 1208200617178_22_4473.dcm
    - 1208200617178_22_8783.dcm
    - 1208200617178_22_7328.dcm
    - 1208200617178_22_9264.dcm
    - 1208200617178_22_9967.dcm
    - 1208200617178_22_3894.dcm
    - 1208200617178_22_3899.dcm
  - 1208200617178_23/
  - 1208200617178_24/
  - 1208200617178_25/



- 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/

The importance of automated validation

Summary

- 40 Files, 18.42kB
- 13 - Subjects
- 1 - Session

Available Tasks

- rhyme judgment

Available Modalities

- bold
- T1w

Your dataset is not a valid BIDS dataset.

view 1 error in 23 files

view 1 warning in 4 files

https://bids-standard.github.io/bids-validator/

BIDS Extensions

- BIDS was originally focused on structural/functional MRI data
- BIDS extension process allows extension of the standard through BIDS Extension Proposals (BEPS) initiated by the community
  - Patterned after the Python Enhancement Proposal (PEP) process

11 Completed BEPs:

BEP #	Title
BEP001	Quantitative MRI (qMRI)
BEP003	Common Derivatives
BEP005	Arterial Spin Labeling (ASL)
BEP006	Electroencephalography (EEG)
BEP007	Hierarchical Event Descriptor (HED) Tags
BEP008	Magnetoencephalography (MEG)
BEP009	Positron Emission Tomography (PET)
BEP010	intracranial Electroencephalography (iEEG)
BEP018	Genetic information
BEP030	Near Infrared Spectroscopy (NIRS)
BEP031	Microscopy

The growing usage of BIDS: An example

- MRIQC Web API
  - Crowdsourced database of MR QC metrics
  - QC metrics from ~375K unique BOLD scans and ~280K T1w scans as of June 2022
  - Publicly available: https://mriqc.nimh.nih.gov/

SCIENTIFIC DATA

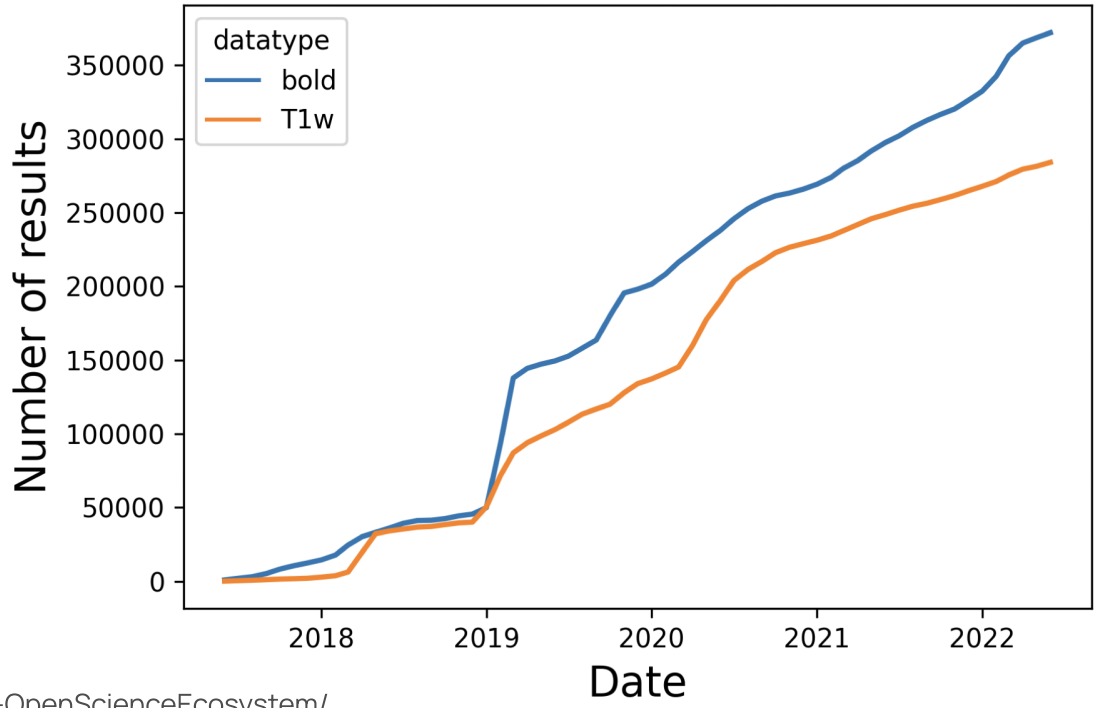
OPEN

DATA DESCRIPTOR

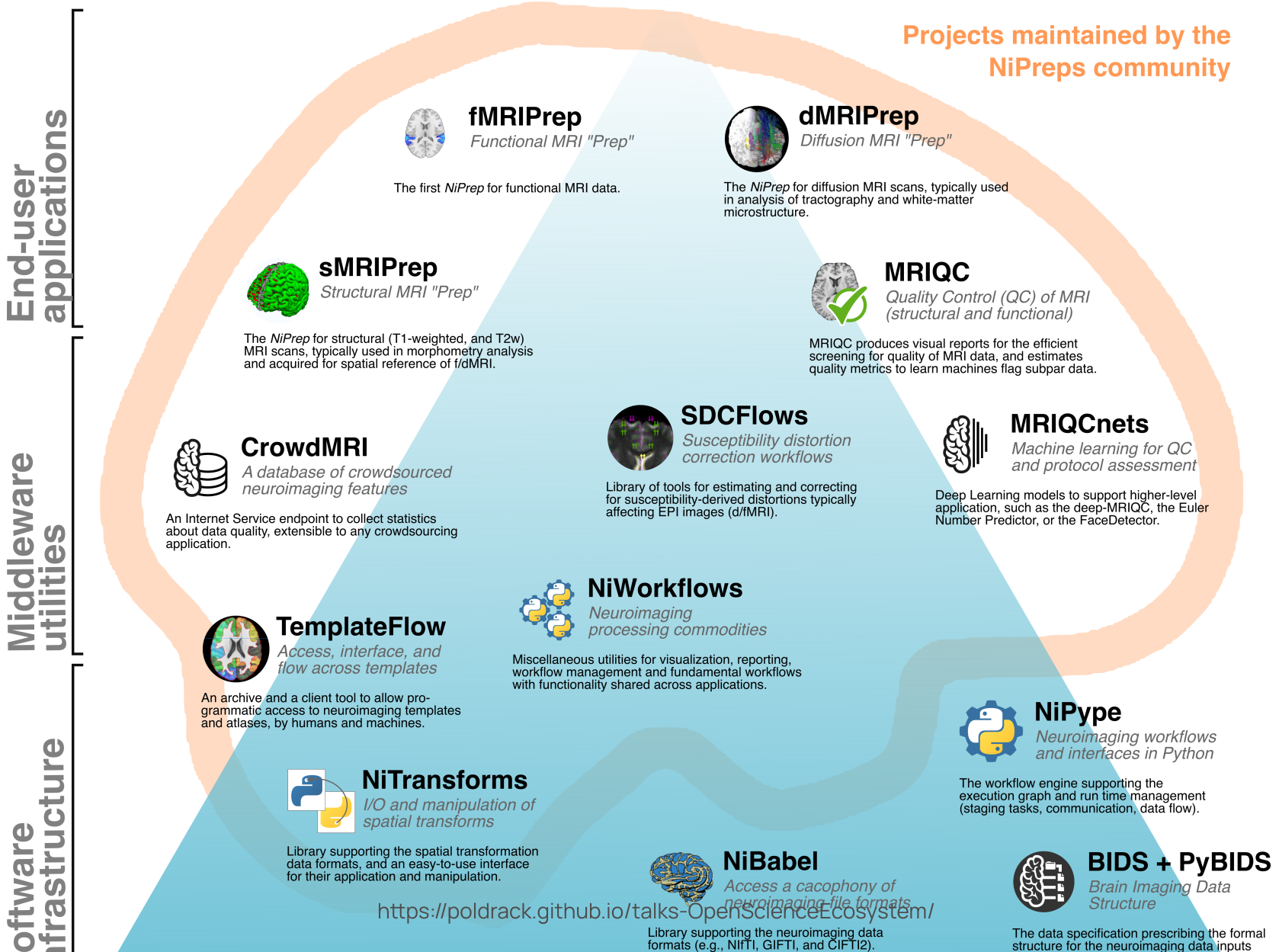
Received: 19 September 2018
Accepted: 12 March 2019

Crowdsourced MRI quality metrics and expert quality annotations for training of humans and machines

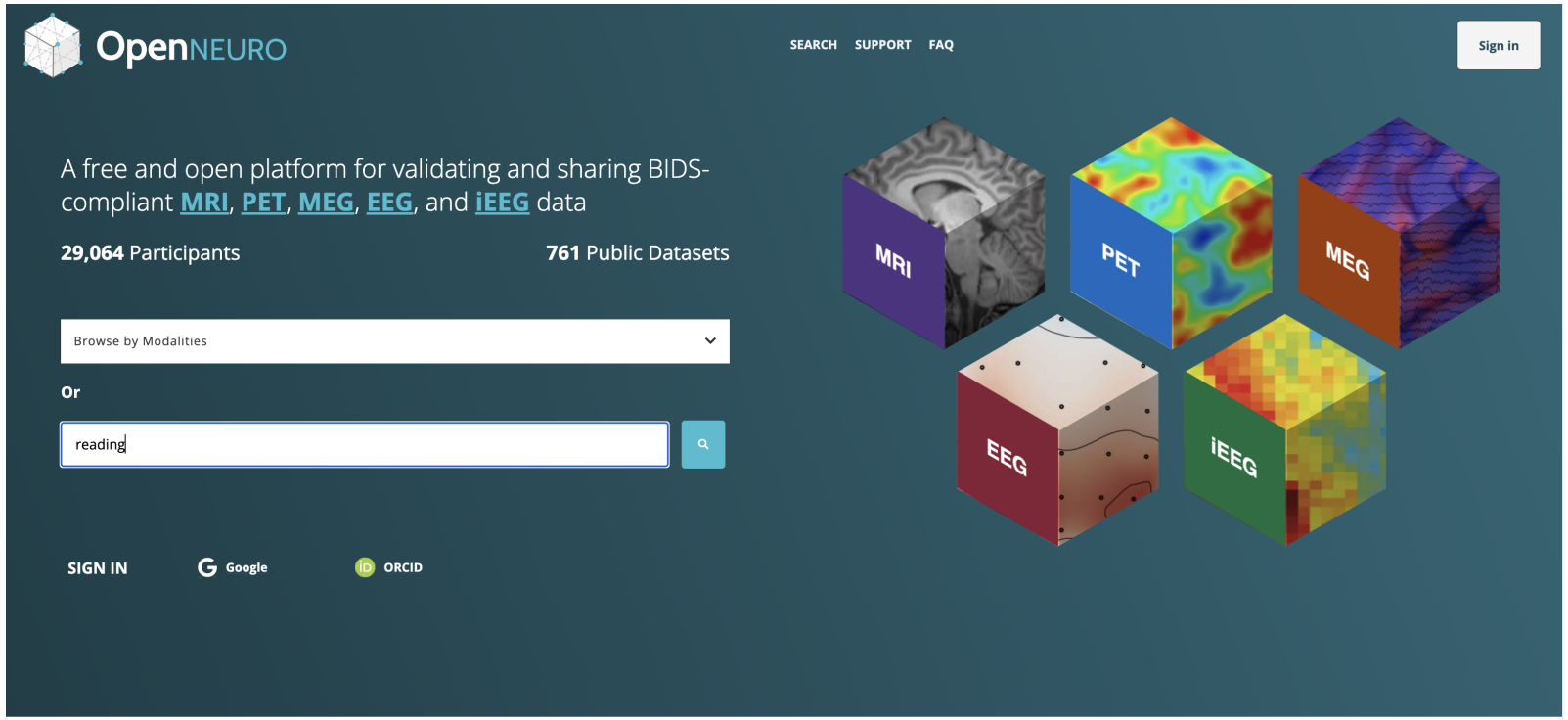
Oscar Esteban[1], Ross W. Blair[1], Dylan M. Nielson[2], Jan C. Varada[3], Sean Marrett[3], Adam G. Thomas[2], Russell A. Poldrack[1] & Krzysztof J. Gorgolewski[1]



BIDS enables a growing open-source software ecosystem



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



- A BRAIN Initiative archive for neuroimaging data
- Supports sharing of any validated BIDS dataset

BIDS

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.

DataLad

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

Search All Datasets

Keywords

Enter Keyword(s) to Search

reading ×

Modalities

- MRI
- PET
- EEG
- iEEG
- MEG

Age of Participants

Number of Participants

Diagnosis

Task

Authors / PI

Sex

Publication Date

Species

Study Type

Study Domain

Radiotracers

These filters return **194** datasets:

CLEAR ALL

KEYWORD:

reading ×

SORT BY: Relevance

The Reading Brain Project L1 Adults

Uploaded by: Chanyuan Gu on 2022-01-07 - 10 months ago | Updated: 2022-01-05 - 10 months ago

MODALITY: MRI

TASKS: read task, rest

OPENNEURO ACCESSION NUMBER: **ds003974** SESSIONS: **1** PARTICIPANTS: **52** PARTICIPANTS' AGES: **N/A** SIZE: **46.67GB** FILES: **893**

The Reading Brain Project L2 Adults

Uploaded by: Chanyuan Gu on 2022-01-11 - 10 months ago | Updated: 2022-02-01 - 10 months ago

MODALITY: MRI

TASKS: read task, rest

OPENNEURO ACCESSION NUMBER: **ds003988** SESSIONS: **1** PARTICIPANTS: **56** PARTICIPANTS' AGES: **N/A** SIZE: **63.82GB** FILES: **960**

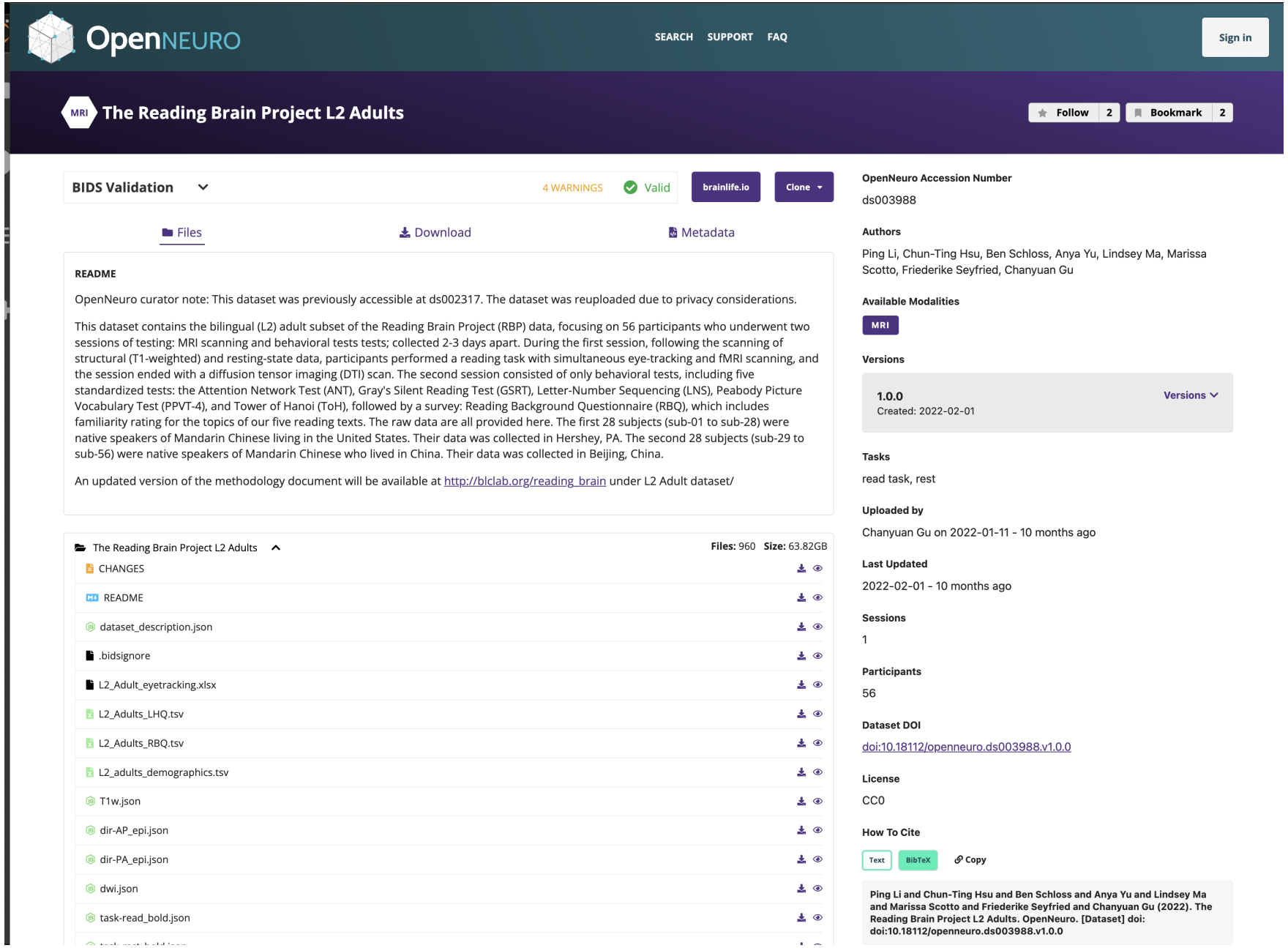
The Reading Brain Project L2 Adults

Uploaded by: Friederike Seyfried on 2021-11-08 - about 1 year ago | Updated: 2021-11-08 - about 1 year ago

MODALITY: MRI

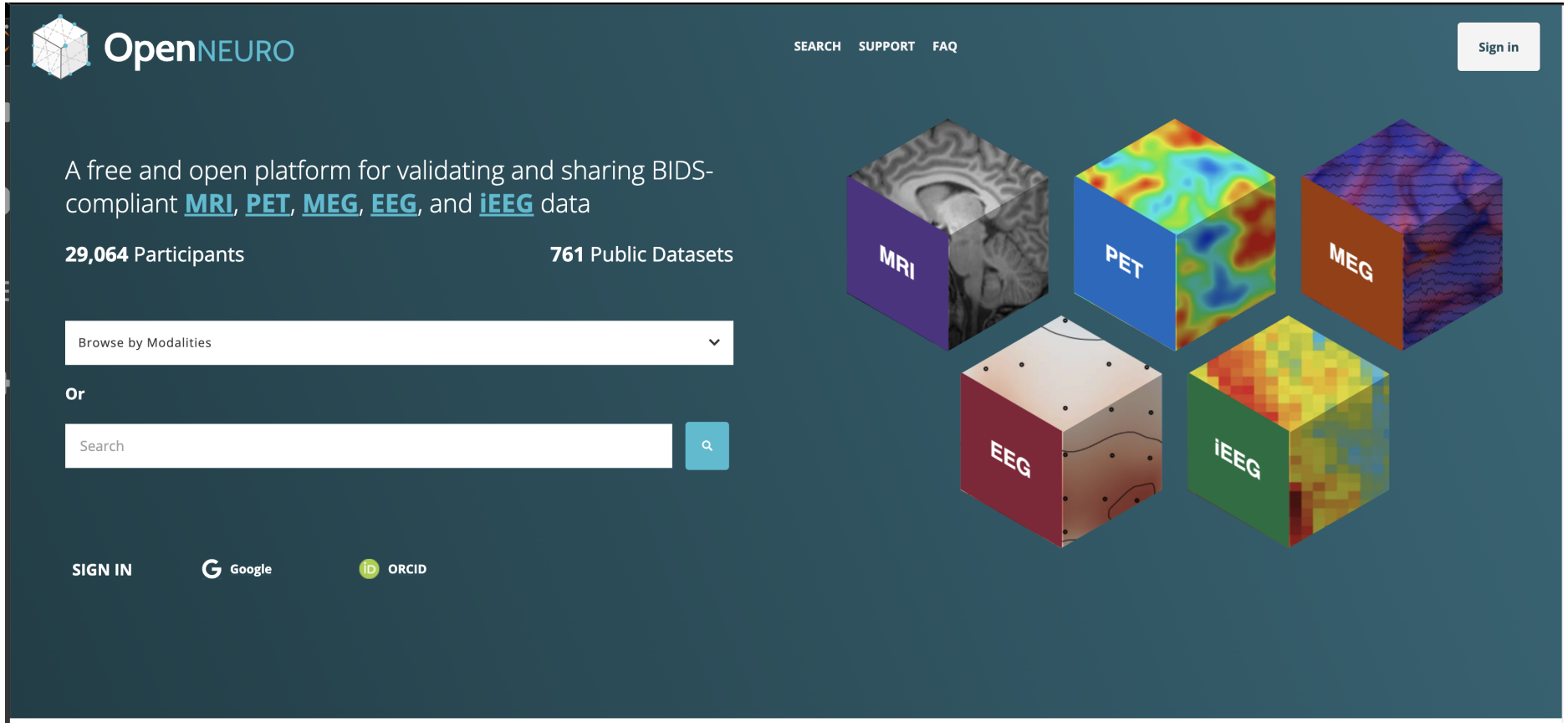
TASKS: rest, read task

OPENNEURO ACCESSION NUMBER: **ds003872** SESSIONS: **1** PARTICIPANTS: **56** PARTICIPANTS' AGES: **N/A** SIZE: **63.79GB** FILES: **960**



Each shared dataset is versioned and receives a persistent identifier (DOI)

Any valid BIDS dataset can be shared via OpenNeuro



BIDS

Validation Using BIDS

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

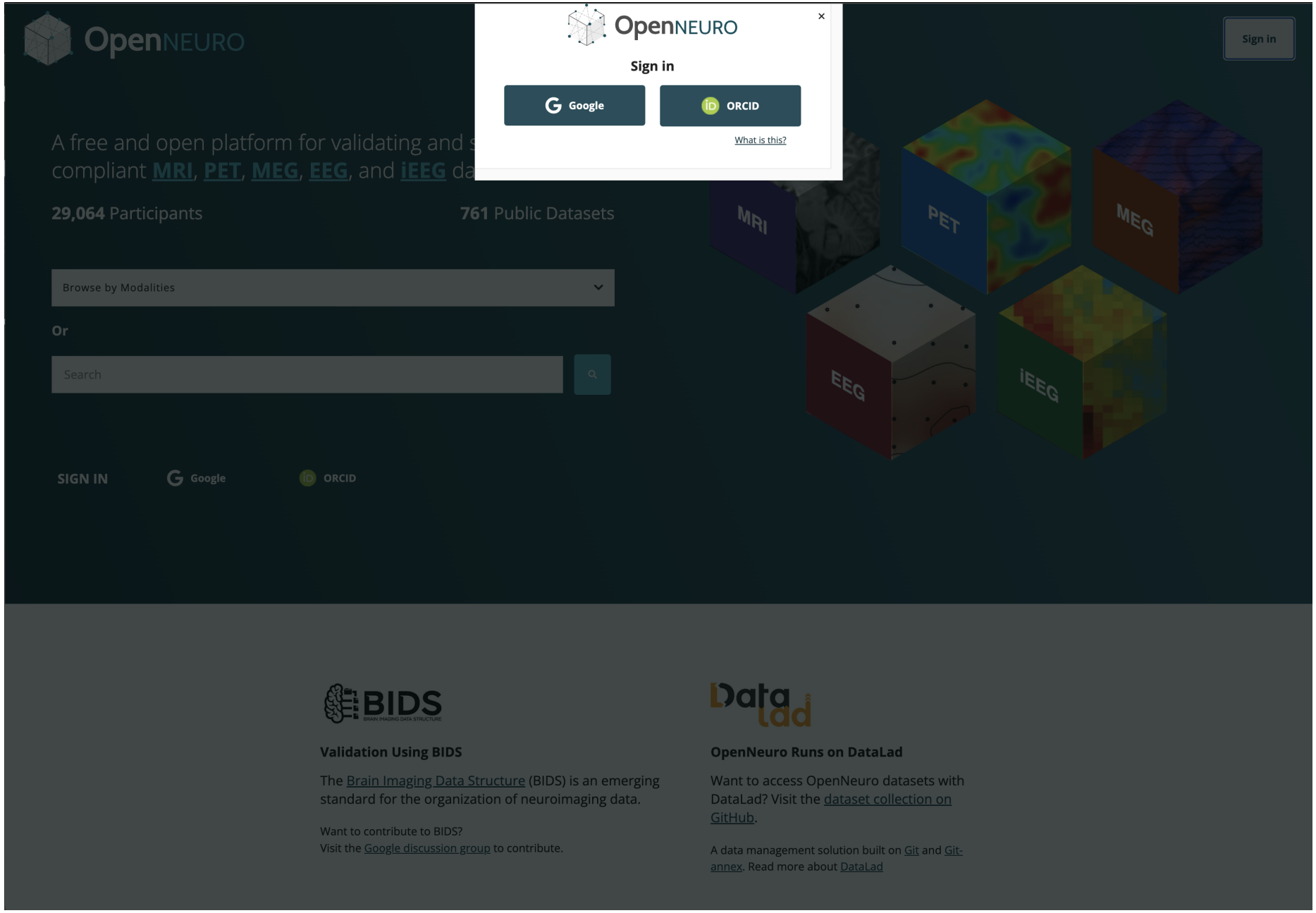
Want to contribute to BIDS?
Visit the Google discussion group to contribute.

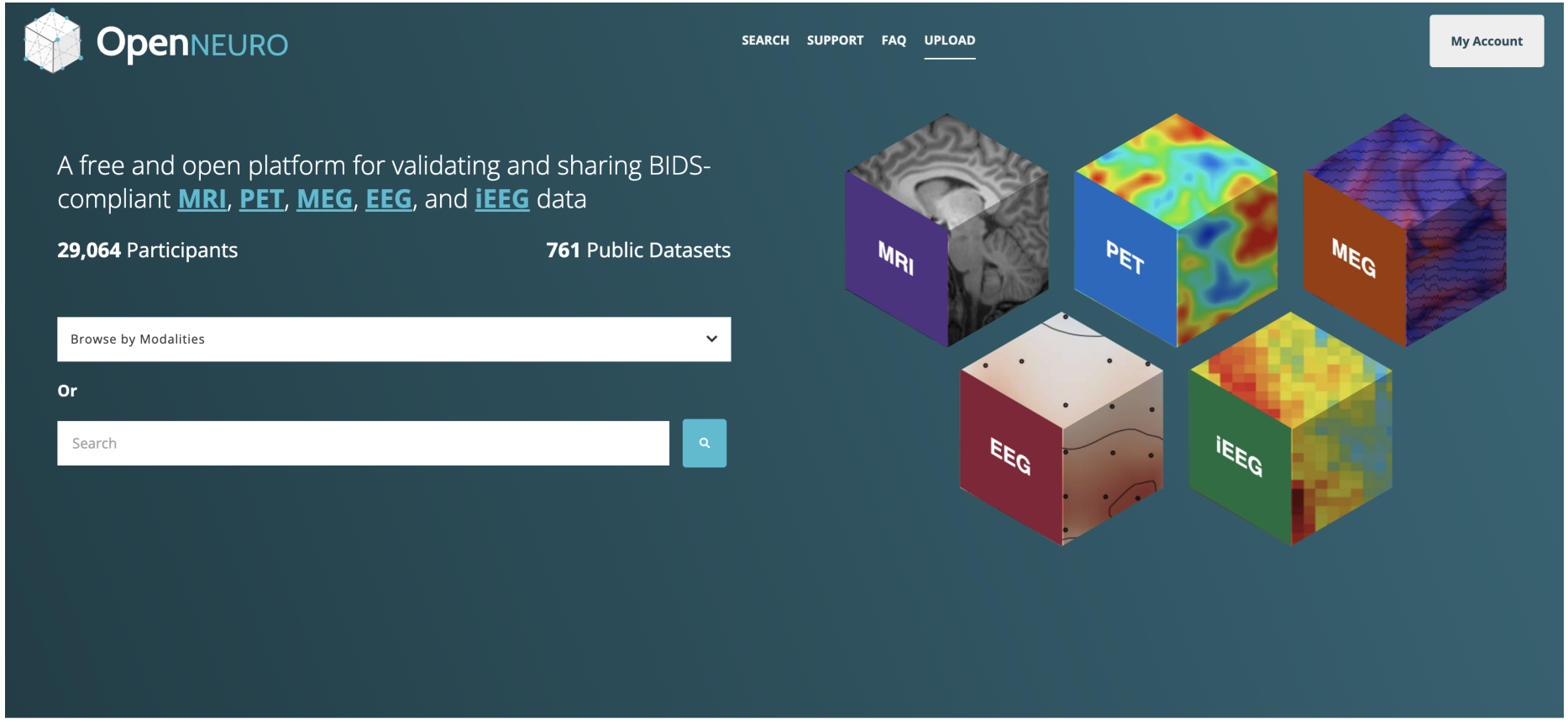
DataLad

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


OpenNEURO
Sign in
Google
ORCID
What is this?
Sign in
A free and open platform for validating and s
compliant MRI, PET, MEG, EEG, and iEEG da
29,064 Participants
761 Public Datasets
Browse by Modalities
Or
Search
SIGN IN
Google
ORCID
MRI
PET
MEG
EEG
iEEG
BIDS
BRAIN IMAGING DATA STRUCTURE
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.
Datalad
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



Validation Using BIDS

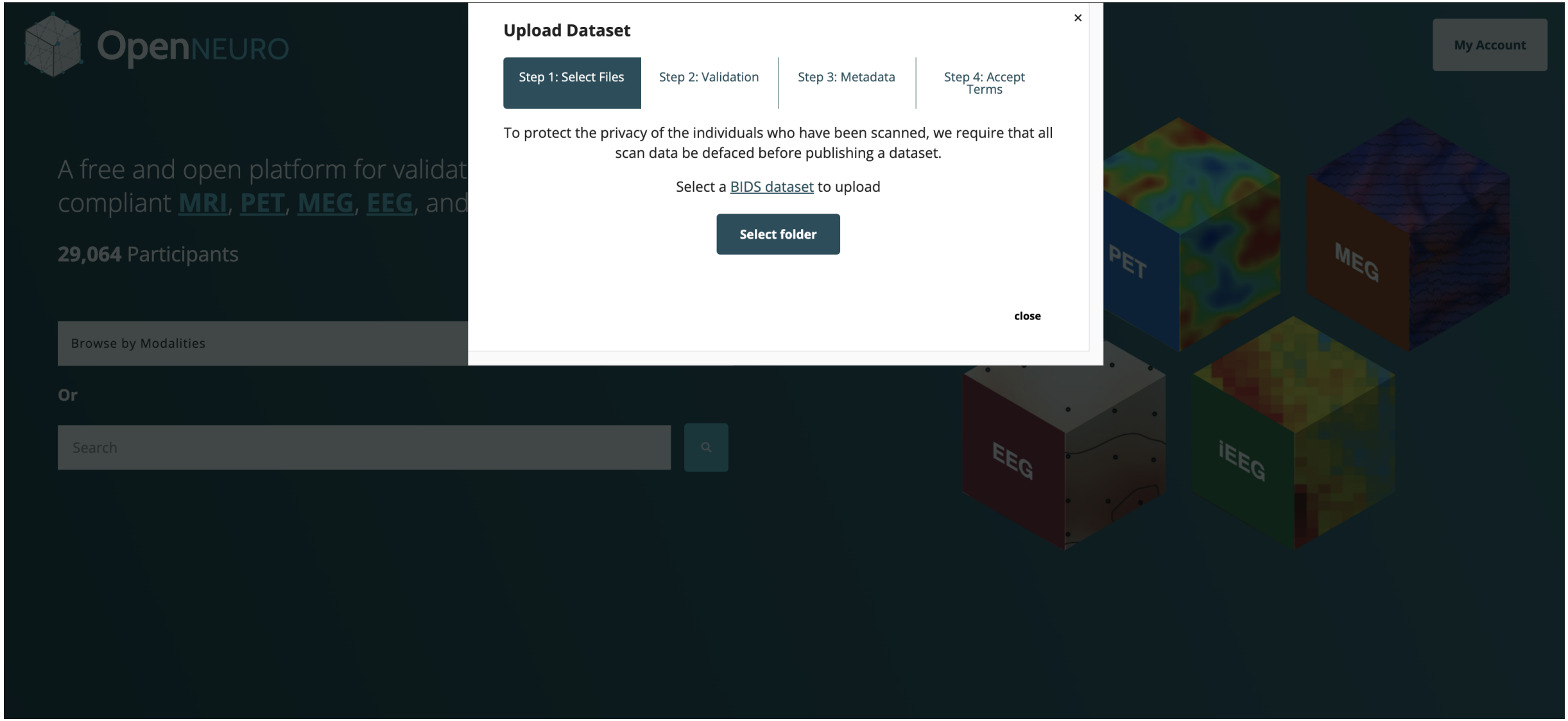
The Brain Imaging Data Structure (BIDS) is an emerging standard for the organization of neuroimaging data.

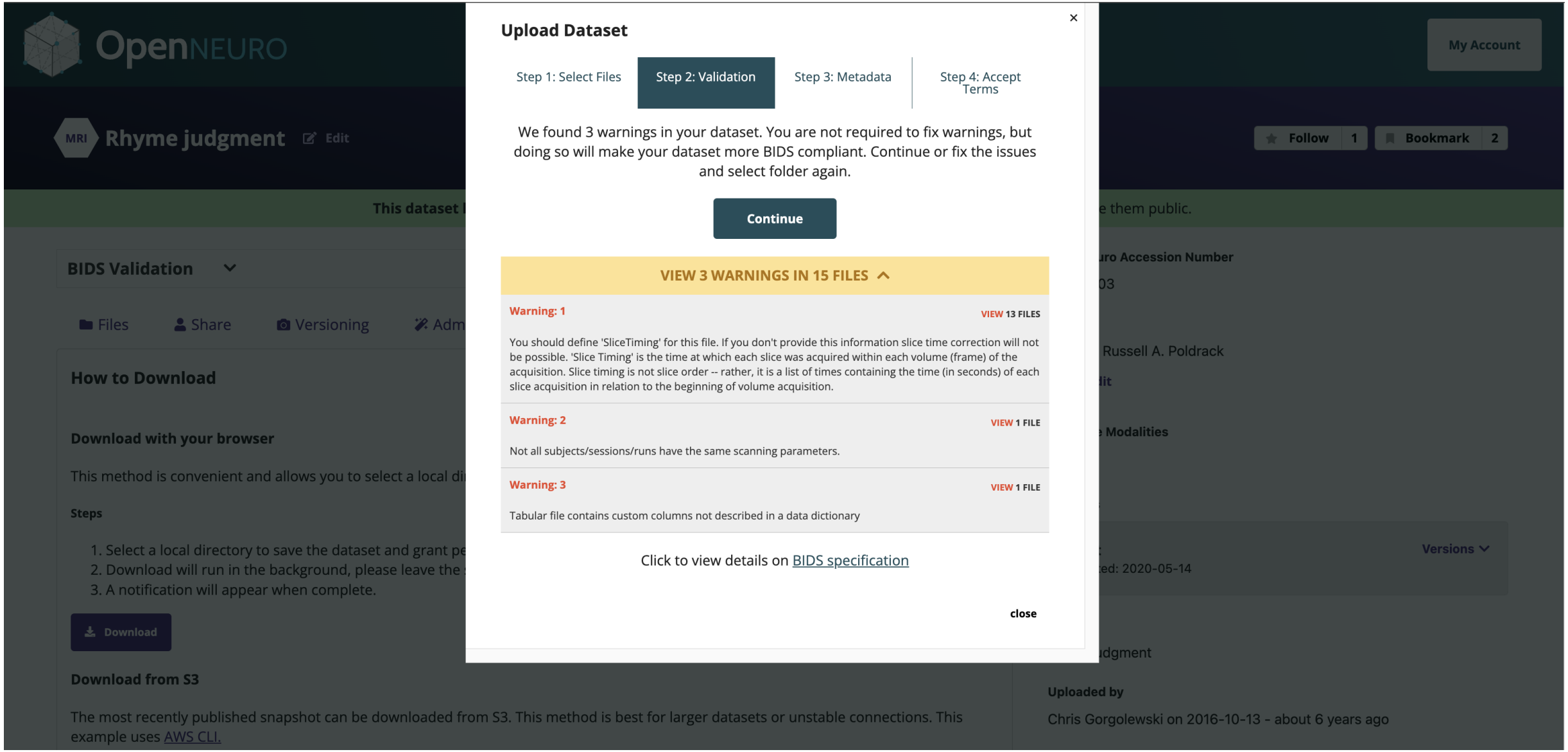
Want to contribute to BIDS?
Visit the Google discussion group to contribute.

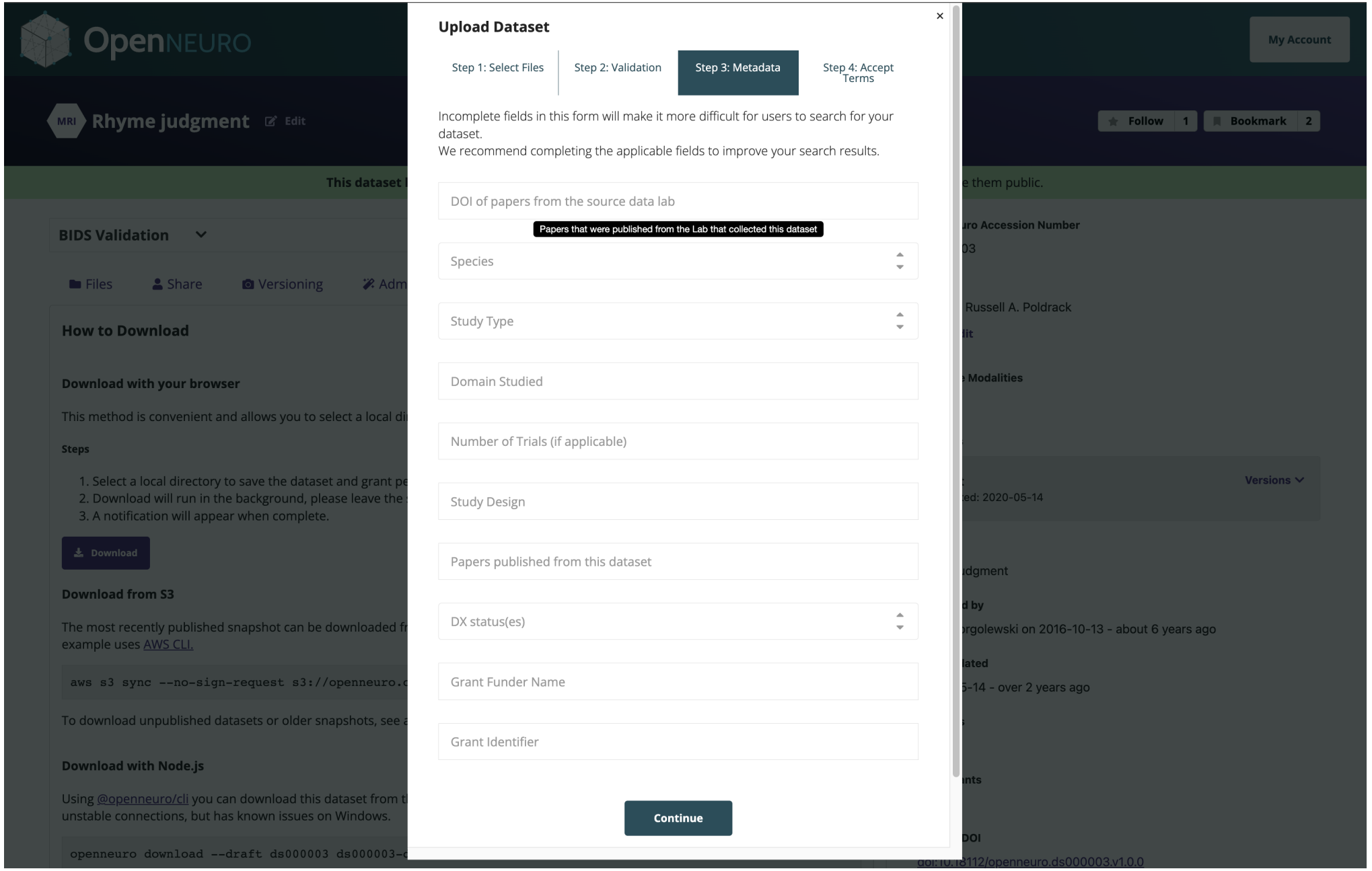
OpenNeuro Runs on DataLad

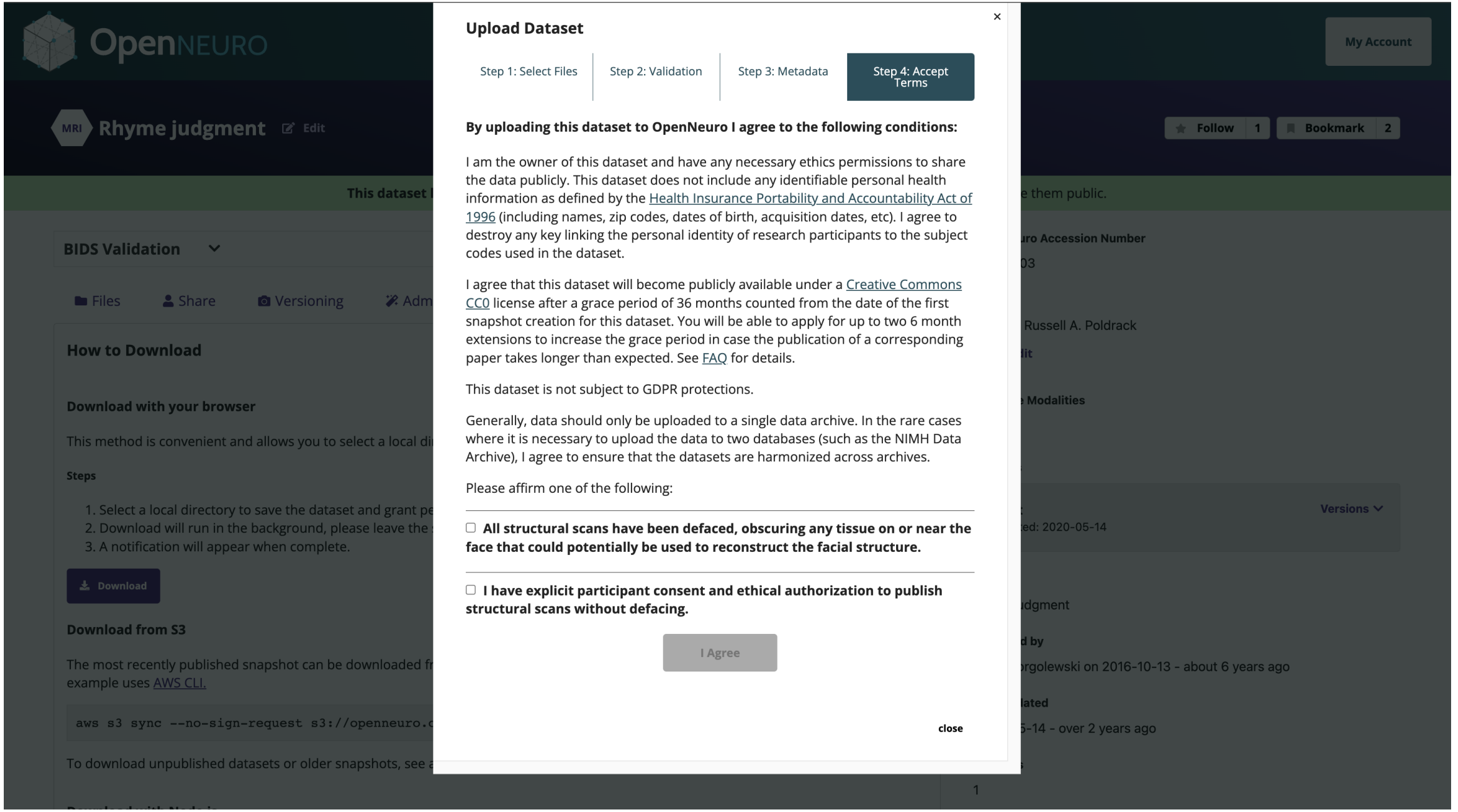
Want to access OpenNeuro datasets with DataLad? Visit the dataset collection on GitHub.

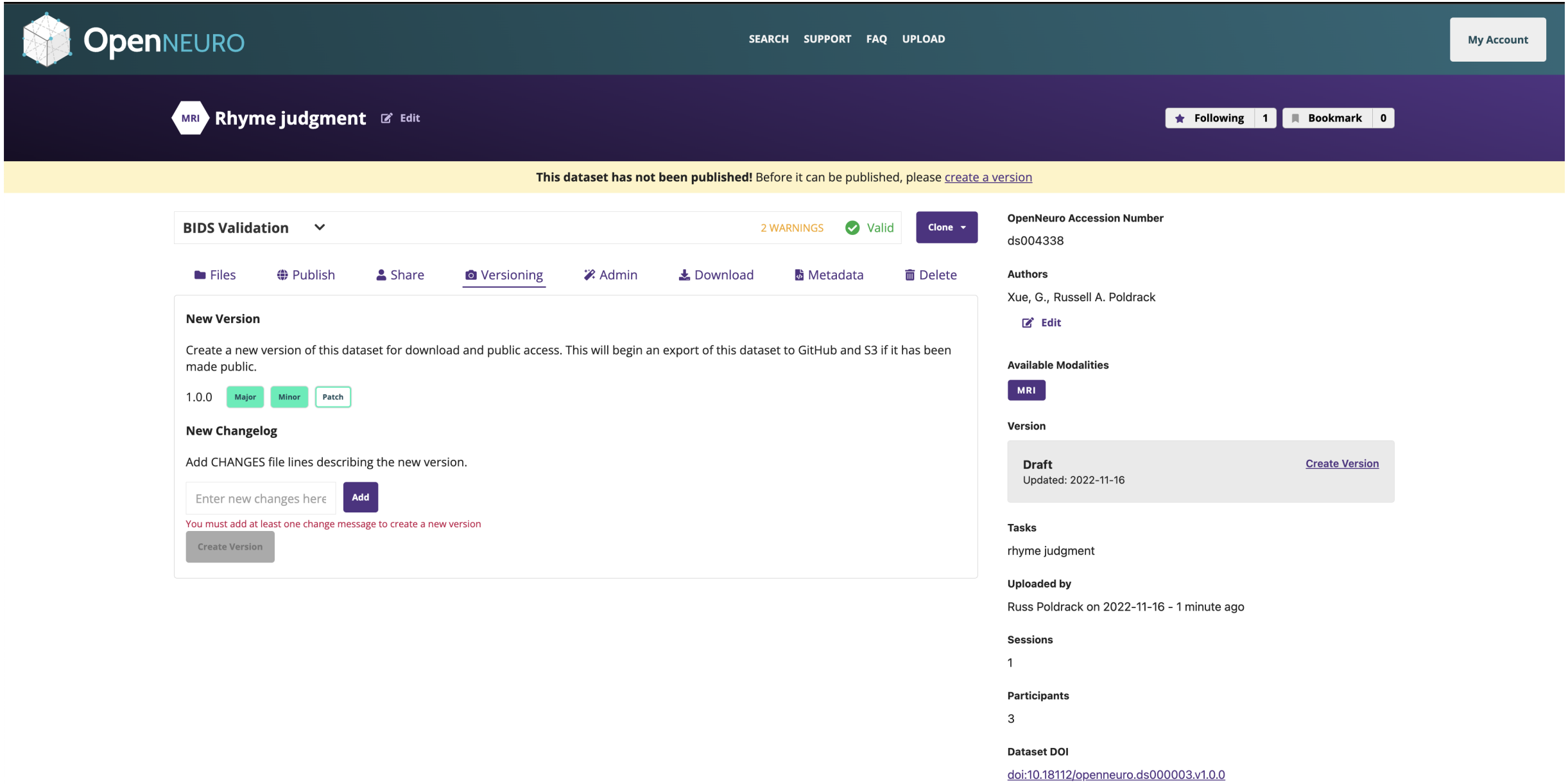
A data management solution built on Git and Git-annex. Read more about DataLad

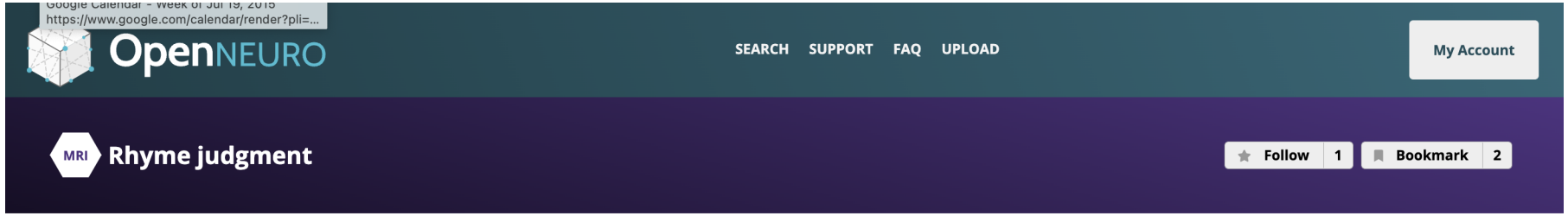

OpenNEURO
My Account
A free and open platform for validat
compliant MRI, PET, MEG, EEG, and
29,064 Participants
Browse by Modalities
Or
Search
Upload Dataset
Step 1: Select Files
Step 2: Validation
Step 3: Metadata
Step 4: Accept Terms
To protect the privacy of the individuals who have been scanned, we require that all scan data be defaced before publishing a dataset.
Select a BIDS dataset to upload
Select folder
close
PET
MEG
EEG
iEEG


OpenNEURO
My Account
MRI
Rhyme judgment
Edit
Follow 1
Bookmark 2
BIDS Validation
Files
Share
Versioning
How to Download
Download with your browser
This method is convenient and allows you to select a local di
Steps
1. Select a local directory to save the dataset and grant pe
2. Download will run in the background, please leave the
3. A notification will appear when complete.
Download
Download from S3
The most recently published snapshot can be downloaded from S3. This method is best for larger datasets or unstable connections. This example uses AWS CLI.
Upload Dataset
Step 1: Select Files
Step 2: Validation
Step 3: Metadata
Step 4: Accept Terms
We found 3 warnings in your dataset. You are not required to fix warnings, but doing so will make your dataset more BIDS compliant. Continue or fix the issues and select folder again.
Continue
VIEW 3 WARNINGS IN 15 FILES
Warning: 1
VIEW 13 FILES
You should define 'SliceTiming' for this file. If you don't provide this information slice time correction will not be possible. 'Slice Timing' is the time at which each slice was acquired within each volume (frame) of the acquisition. Slice timing is not slice order -- rather, it is a list of times containing the time (in seconds) of each slice acquisition in relation to the beginning of volume acquisition.
Warning: 2
VIEW 1 FILE
Not all subjects/sessions/runs have the same scanning parameters.
Warning: 3
VIEW 1 FILE
Tabular file contains custom columns not described in a data dictionary
Click to view details on BIDS specification
close
Russell A. Poldrack
Versions
Uploaded by
Chris Gorgolewski on 2016-10-13 - about 6 years ago


OpenNEURO
My Account
MRI
Rhyme judgment
Edit
Follow
1
Bookmark
2
BIDS Validation
Files
Share
Versioning
How to Download
Download with your browser
Steps
Download
Download from S3
Download with Node.js
Upload Dataset
Step 1: Select Files
Step 2: Validation
Step 3: Metadata
Step 4: Accept Terms
Incomplete fields in this form will make it more difficult for users to search for your dataset.
We recommend completing the applicable fields to improve your search results.
DOI of papers from the source data lab
Papers that were published from the Lab that collected this dataset
Species
Study Type
Domain Studied
Number of Trials (if applicable)
Study Design
Papers published from this dataset
DX status(es)
Grant Funder Name
Grant Identifier
Continue


OpenNEURO
My Account
MRI
Rhyme judgment
Edit
Follow 1
Bookmark 2
BIDS Validation
Files
Share
Versioning
How to Download
Download with your browser
This method is convenient and allows you to select a local di
Steps
1. Select a local directory to save the dataset and grant pe
2. Download will run in the background, please leave the
3. A notification will appear when complete.
Download
Download from S3
The most recently published snapshot can be downloaded fr
example uses AWS CLI.
aws s3 sync --no-sign-request s3://openneuro.
To download unpublished datasets or older snapshots, see a
Upload Dataset
Step 1: Select Files
Step 2: Validation
Step 3: Metadata
Step 4: Accept Terms
By uploading this dataset to OpenNeuro I agree to the following conditions:
I am the owner of this dataset and have any necessary ethics permissions to share the data publicly. This dataset does not include any identifiable personal health information as defined by the Health Insurance Portability and Accountability Act of 1996 (including names, zip codes, dates of birth, acquisition dates, etc). I agree to destroy any key linking the personal identity of research participants to the subject codes used in the dataset.
I agree that this dataset will become publicly available under a Creative Commons CC0 license after a grace period of 36 months counted from the date of the first snapshot creation for this dataset. You will be able to apply for up to two 6 month extensions to increase the grace period in case the publication of a corresponding paper takes longer than expected. See FAQ for details.
This dataset is not subject to GDPR protections.
Generally, data should only be uploaded to a single data archive. In the rare cases where it is necessary to upload the data to two databases (such as the NIMH Data Archive), I agree to ensure that the datasets are harmonized across archives.
Please affirm one of the following:
All structural scans have been defaced, obscuring any tissue on or near the face that could potentially be used to reconstruct the facial structure.
I have explicit participant consent and ethical authorization to publish structural scans without defacing.
I Agree
close
Russell A. Poldrack
Versions
1


OpenNEURO
SEARCH
SUPPORT
FAQ
UPLOAD
My Account
MRI
Rhyme judgment
Edit
Following 1
Bookmark 0
This dataset has not been published! Before it can be published, please create a version
BIDS Validation
2 WARNINGS
Valid
Clone
Files
Publish
Share
Versioning
Admin
Download
Metadata
Delete
New Version
Create a new version of this dataset for download and public access. This will begin an export of this dataset to GitHub and S3 if it has been made public.
1.0.0
Major
Minor
Patch
New Changelog
Add CHANGES file lines describing the new version.
Enter new changes here
Add
You must add at least one change message to create a new version
Create Version
OpenNeuro Accession Number
ds004338
Authors
Xue, G., Russell A. Poldrack
Edit
Available Modalities
MRI
Version
Draft
Updated: 2022-11-16
Create Version
Tasks
rhyme judgment
Uploaded by
Russ Poldrack on 2022-11-16 - 1 minute ago
Sessions
1
Participants
3
Dataset DOI
doi:10.18112/openneuro.ds000003.v1.0.0



BIDS Validation 1 ERROR Invalid brainlife.io Clone

Files View Draft Download Derivatives Metadata Deprecate Version

README

This dataset was obtained from the OpenfMRI project (http://www.openfmri.org). Accession #: ds003 Description: Rhyme judgment

Release history: 10/06/2011: initial release 3/21/2013: Updated release with QA information 2/18/2016: Updated orientation information in nifti headers for improved left-right determination

This dataset is made available under the Public Domain Dedication and License v1.0, whose full text can be found at http://www.opendatacommons.org/licenses/pddl/1.0/. We hope that all users will follow the ODC Attribution/Share-Alike Community Norms (http://www.opendatacommons.org/norms/odc-by-sa/); in particular, while not legally required, we hope that all users of the data will acknowledge the OpenfMRI project and NSF Grant OCI-1131441 (R. Poldrack, PI) in any publications.

Rhyme judgment **Files:** 57 **Size:** 391.09MB

- CHANGES
- README
- dataset_description.json
- participants.tsv
- task-rhymejudgment_bold.json
- derivatives
- sub-01
- sub-02
- sub-03
- sub-04
- sub-05

OpenNeuro Accession Number

ds000003

Authors

Xue, G., Russell A. Poldrack

Available Modalities

MRI

Versions

1.0.0 Versions
Created: 2020-05-14

Tasks

rhyme judgment

Uploaded by

Chris Gorgolewski on 2016-10-13 - about 6 years ago

Last Updated

2020-05-14 - over 2 years ago

Sessions

1

Participants

13

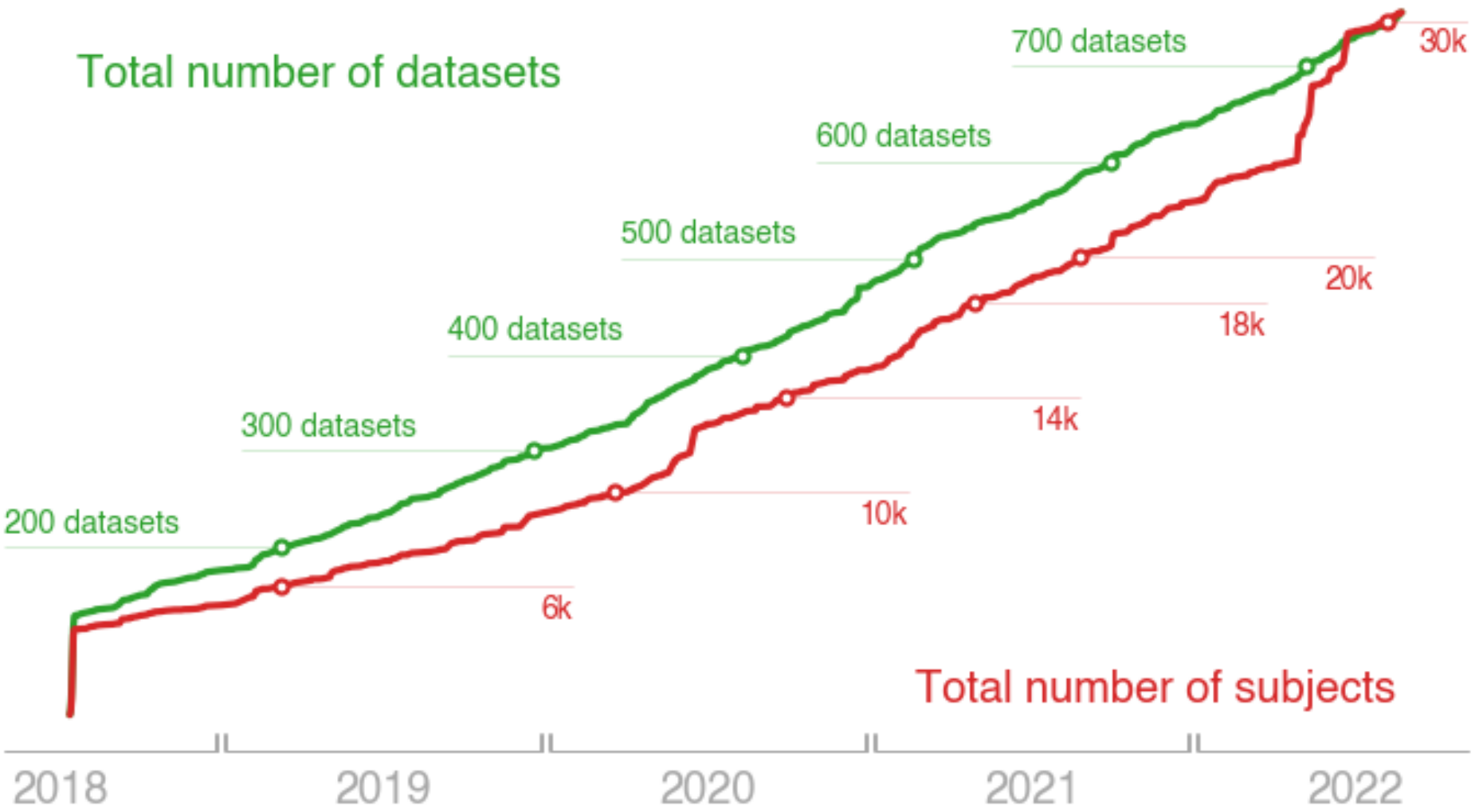
Dataset DOI

doi:10.18112/openneuro.ds000003.v1.0.0

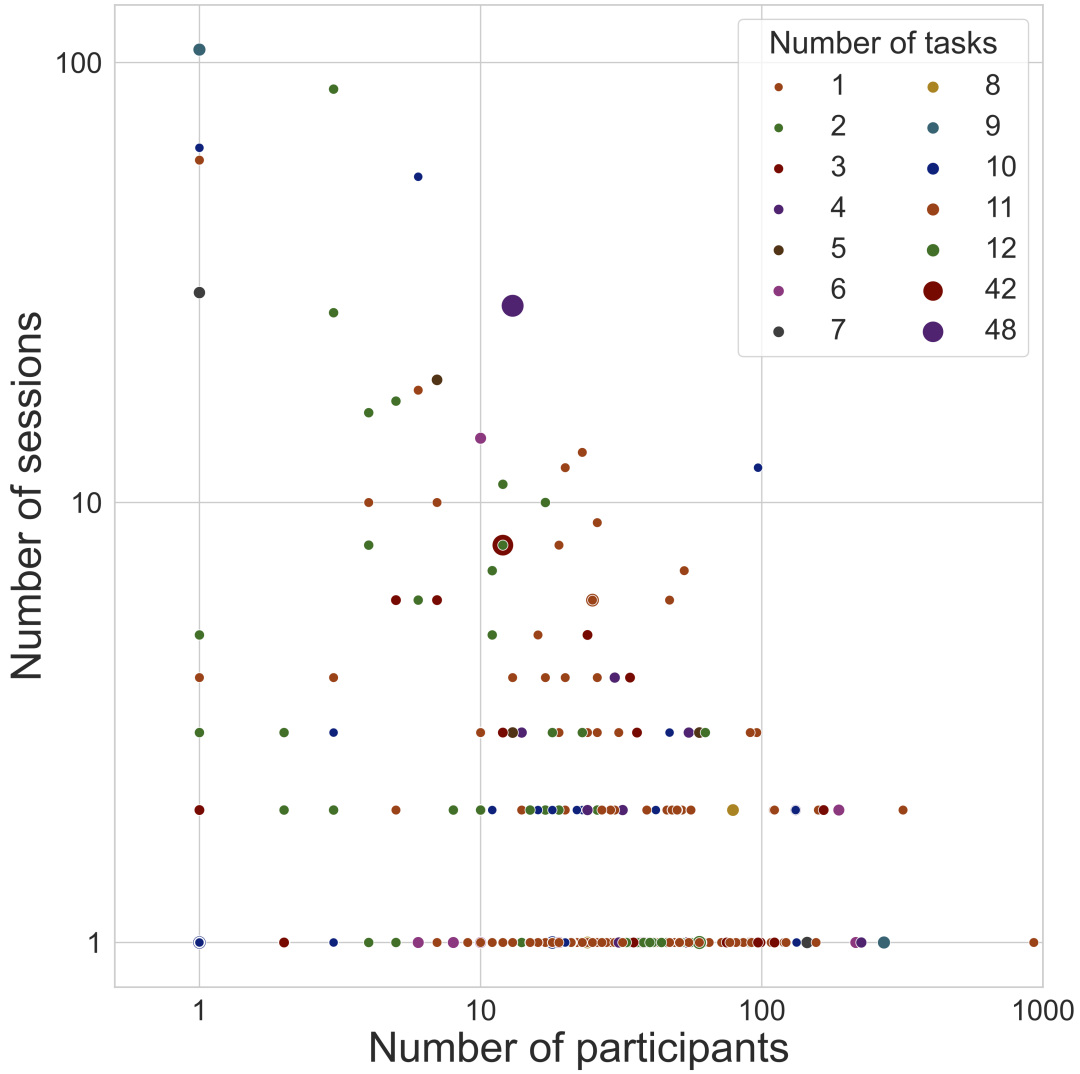
License

CC0

The growth of OpenNeuro



The diversity of OpenNeuro datasets



Datatype	#
mri - anat	597
mri - func	521
eeg	120
mri - dwi	67
meg	30
ieeg	17
beh	13
pet	11

Species	#
Human	676
Mouse	20
Rat	12
NHP	2
phantoms	1
Juvenile pigs	1
Human, Mouse	1
Dog	1
Monkey	1
Sheep	1

updated from Markiewicz et al. 2021 *eLife*

Scholarly reuse of OpenNeuro datasets

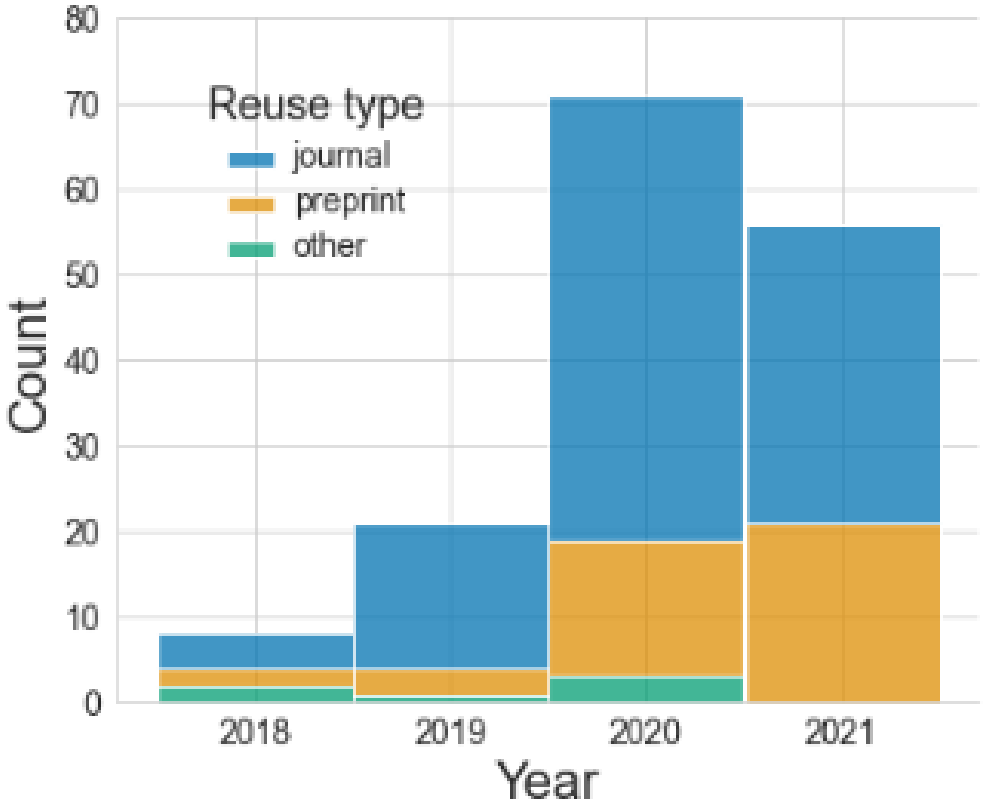
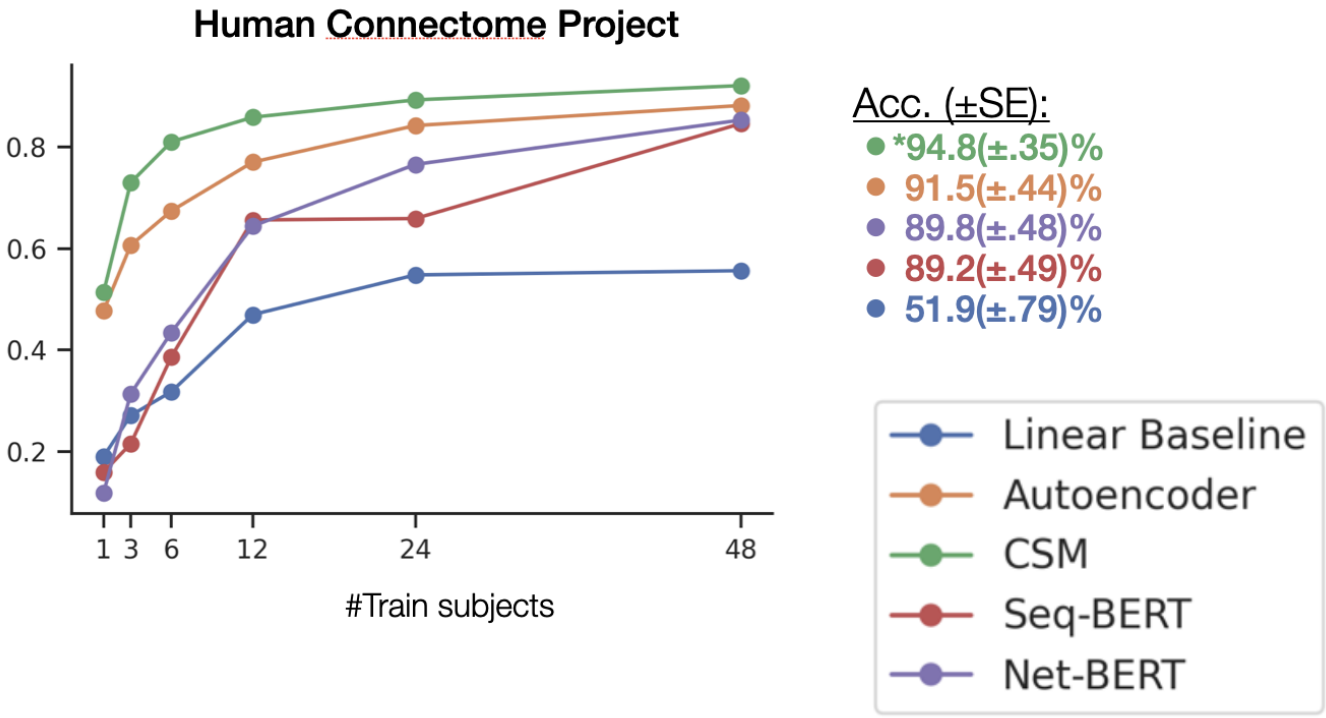


Figure 5. Published reuses of OpenNeuro datasets, split by the type of reuse. Note that the final bar includes only reuses identified through June 2021.

Markiewicz et al. 2021 *eLife*

Example of OpenNeuro reuse

- A challenge for decoding brain activity from fMRI data is that most datasets are very small
- We used OpenNeuro to train a "foundation model"
  - A pre-trained model that can be used as a starting point for decoding models on smaller datasets
- We pre-train models on broad fMRI data from OpenNeuro: 11,980 experimental runs from 1,726 individuals across 34 datasets.

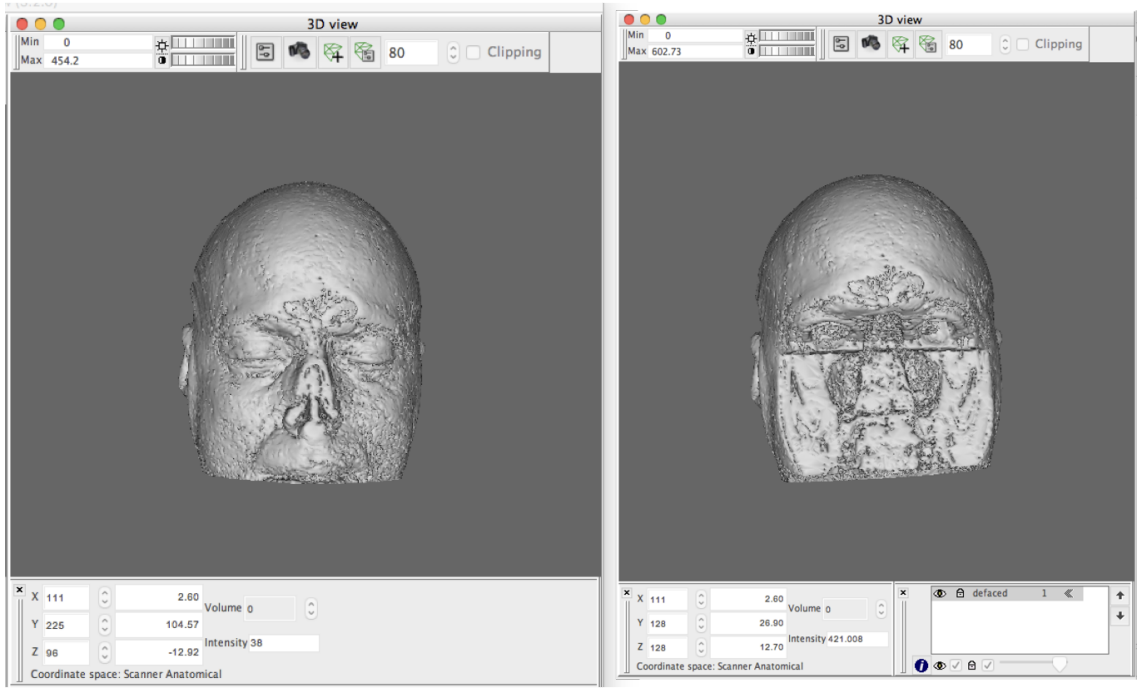


- This approach substantially increased decoding performance vs. a baseline model

Thomas, Ré, & Poldrack, 2022, *NeurIPS*

Challenges to open sharing

- All OpenNeuro MRI datasets must be *defaced*
  - To reduce risk of reidentification
- There is increasing risk that subjects might be reidentified even after defacing using advanced face recognition systems + face imputation tools (Schwartz et al., 2021)
- If the risk continues to rise, it may become necessary to move away from open sharing
  - This would be a huge loss for researchers, research participants, and the world
- We have proposed regulatory changes to protect subjects from misuse of neuroscience information in the US context (Jwa & Poldrack, 2022, *J. Law & Biosciences*)



Keys to success in neuroimaging data sharing

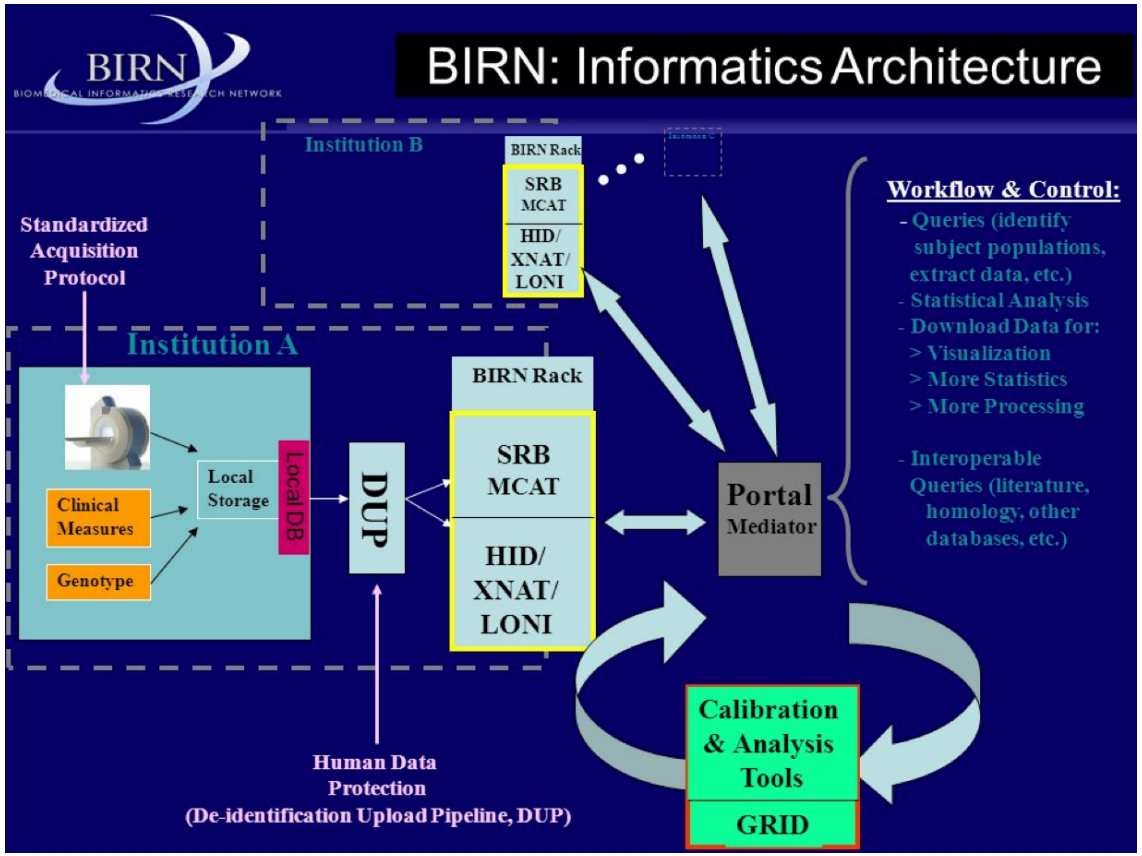
- Data are digital end-to-end
  - Minimizes manual steps in the process
- Standardized file formats and data standards
  - Makes data immediately usable by anyone
  - Reduces burden of curation and preparation
- Demonstrated scientific utility
- Numerous success stories

Lessons learned

- Community buy-in is essential
  - Mandates put in place before the community is ready can backfire
    - Unless they have overwhelmingly powerful advocates, as in genomics
  - Important that sharing advocates are members of community and eat their own dog food

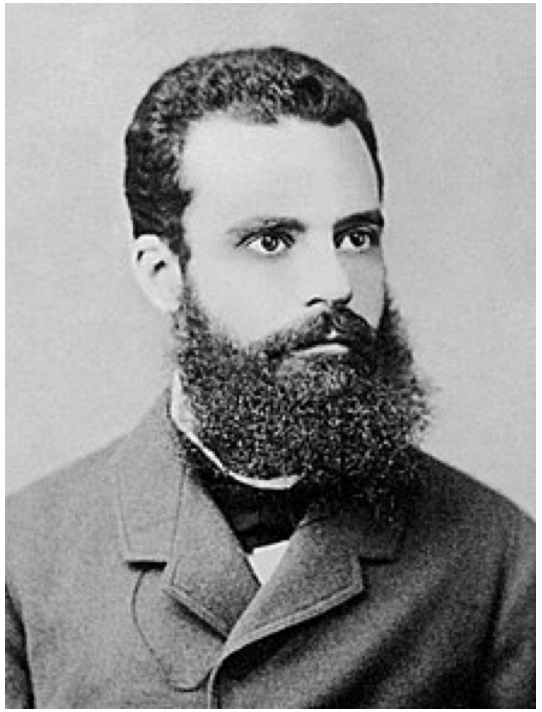
Lessons learned

- Keep it simple and as close to standard practice as possible
  - Overengineered solutions have generally failed
  - If there are more than 2 acronyms…



Lessons learned

- Don’t let the perfect be the enemy of the good
  - 20% of the effort will cover 80% of the datasets - focus on these!
  - There is a long tail of edge cases with loud advocates



Vilfredo Pareto

Conclusions

- The field of neuroimaging has built an enviable ecosystem for open science
- Infrastructure is critical to ease friction
- Community engagement has been key to adoption
- Need to keep the tools as close as possible to current practice

The Poldrack Lab

OpenNeuro Team

Funding

THE BRAIN INITIATIVE
NATIONAL INSTITUTES OF HEALTH

NSF

ljaf
laura and john arnold foundation®

NIH National Institute of Mental Health

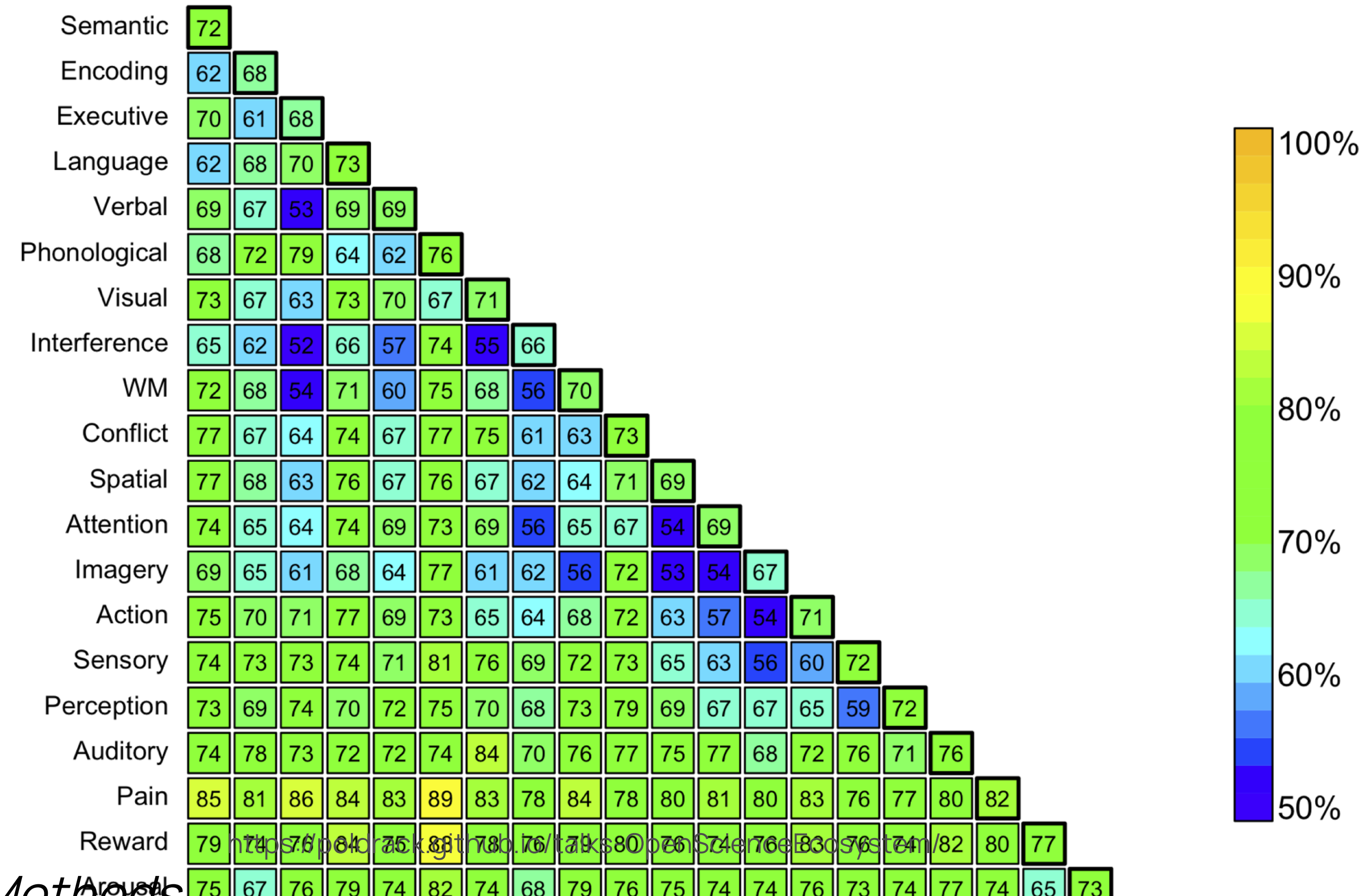
Collaborators

Meta-analytic decoding using Neurosynth

- Given 2+ terms, can determine which is most likely given the data
- Naive Bayes classifier: assumes that all features (voxels) are independent; selects the most probable class
- Can apply this to any activation map—studies, individual subjects, etc.

Yarkoni et al, 2011, *Nature Methods*

- Cross-validated classification of all studies in database
- Select 25 high-frequency terms
- Pairwise classification: how well can we distinguish between the presence of each pair of terms?



Yarkoni et al. 2011, *Nature Methods*