

Toward an Open Science Ecosystem in Neuroimaging

Russ Poldrack
Stanford University

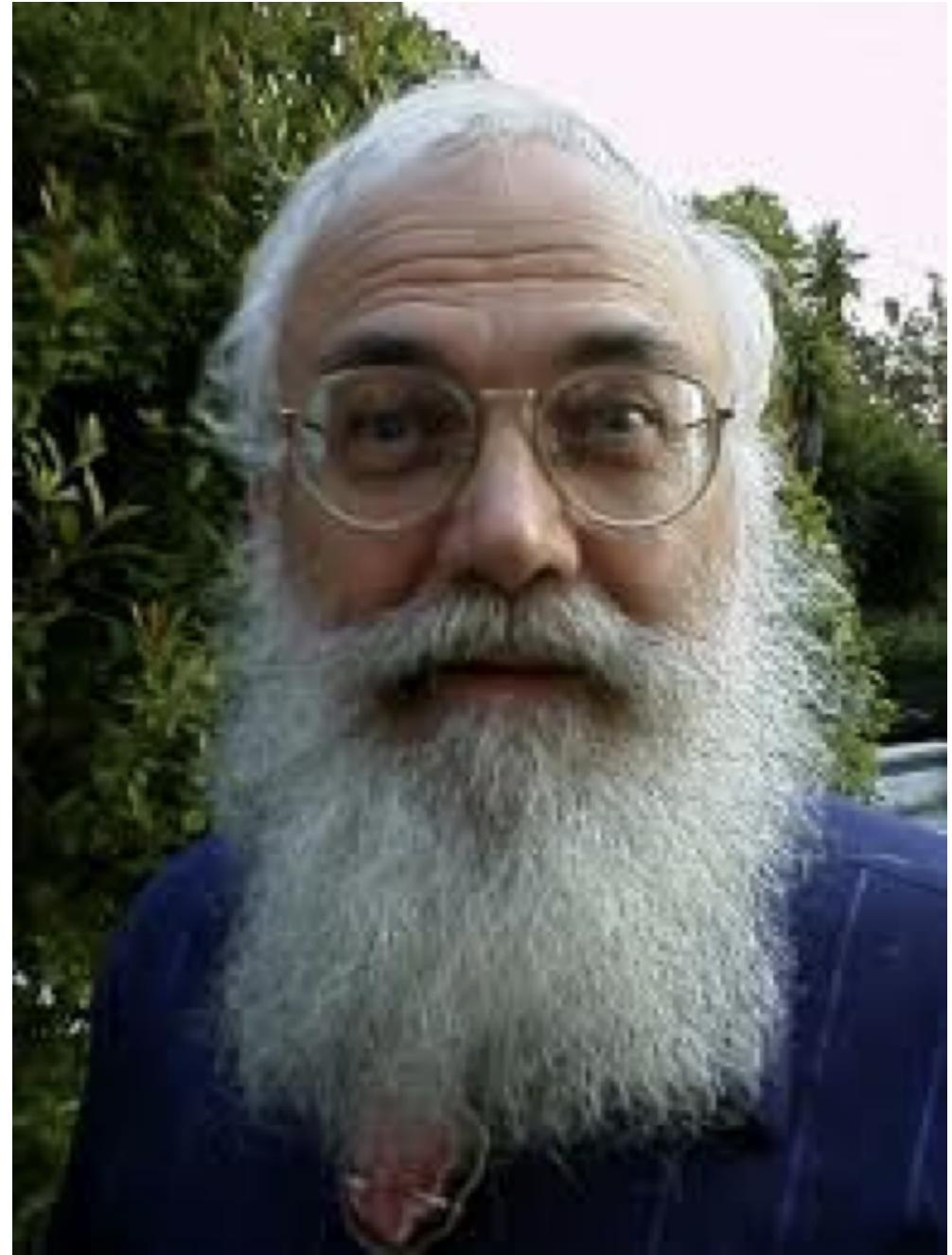
Transparency is essential for reproducibility

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

“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..”

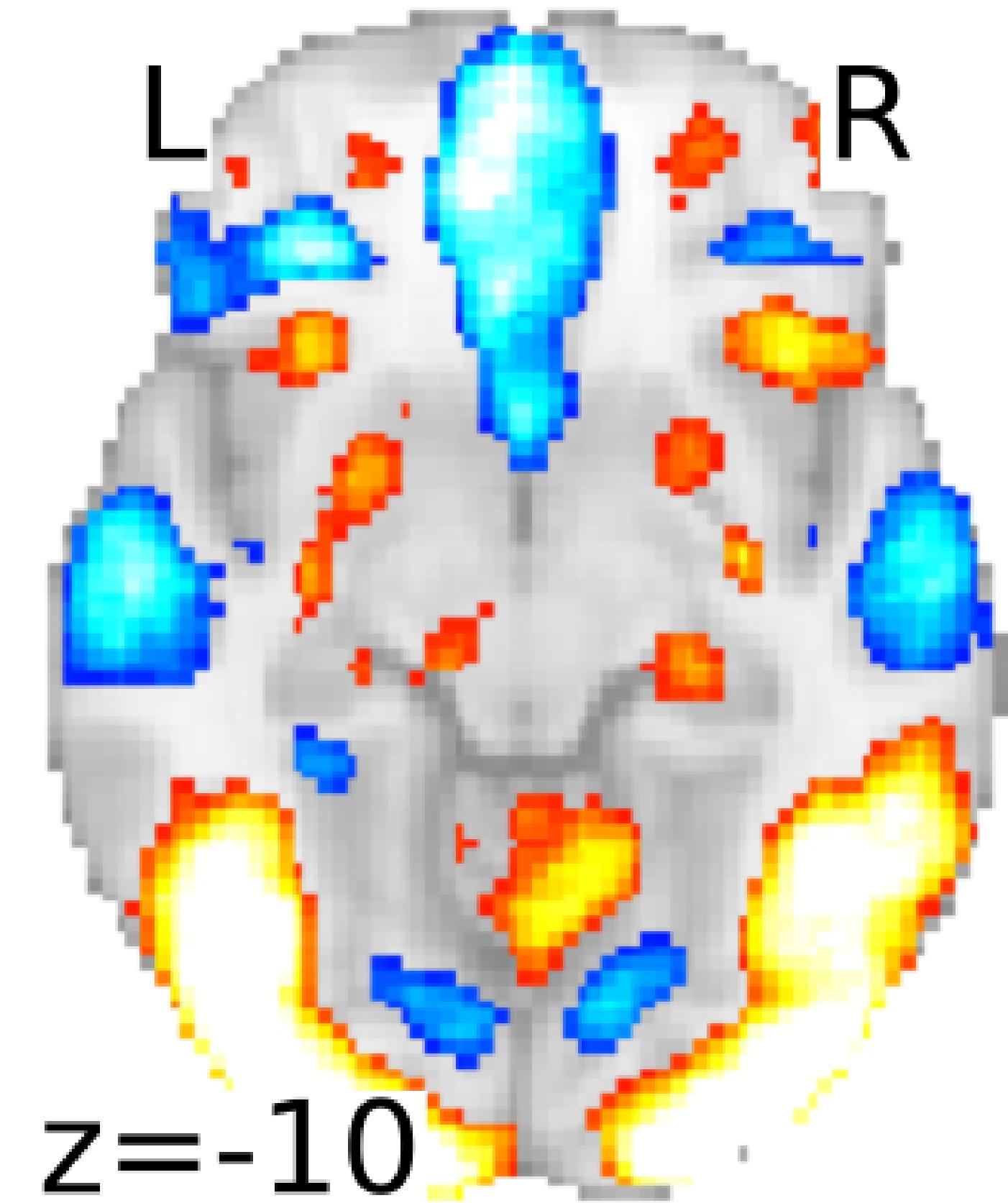
- Buckheit & Donoho, 1995



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

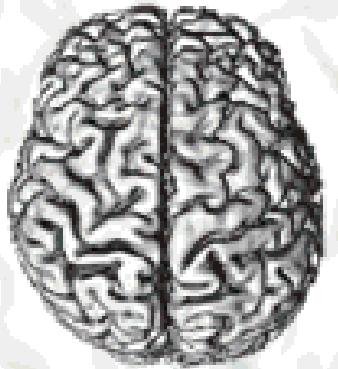
The fMRI Data Center
fMRI IDC

SEARCH fMRI IDC Database FOR SUBMIT

[My Account](#) [Request List \(Empty\)](#)

▶ HOME
DATABASE
SUBMISSIONS
RESOURCES
DATA MANAGEMENT TOOL
HELP
ABOUT US

[Sitemap](#) 
[Contact Us](#) 

 A public repository of peer-reviewed fMRI studies and their underlying data.
Funded By
The National Science Foundation
The W. M. Keck Foundation
The National Institutes of Mental Health
A Sun Center of Excellence for Neuroscience


INFORMATION
[How do I get started?](#)
Answers to questions commonly posed by first-time visitors.
[Q&A about fMRI IDC](#)
A comprehensive list of frequently asked questions about the fMRI IDC.
[Available Datasets](#)
A list of datasets currently available.
[Information for Authors](#)
How to submit your imaging data to the Data Center.
[fMRI Data Management Tool](#)
A tool to aid in the tracking, sharing, and searching of data from neuroimaging experiments.

fMRI IDC NEWS
[fMRI IDC Releases DCSearch \(beta\)](#)
November 10, 2005 - Now Search the fMRI Data Center archive by anatomical region, Brodmann area, talairach/MNI coordinates, and other fields.
[Michael Gazzaniga Elected to Institute of Medicine](#)
October 24, 2005 - fMRI IDC PI Receives Honor
[The PALS Brain Atlas from WUSTL](#)
October 17, 2005 - As seen in Science NetWatch, 7 Oct 2005
[Computer Generated Phantoms for Use in Evaluating fMRI Methods](#)
October 6, 2005 - Free test images for use in assessing data processing routines
[More news items...](#)

PROJECT STATISTICS
[Registered users:](#) 1912
[Datasets available:](#) 110
[Dataset requests:](#) 1789
[More database statistics...](#)

Updated November 18, 2005

 **Special Collections**
Data from special or rare populations of subjects.

 **Summer Workshops**

 **The New Perspectives in fMRI Research Award**

 **fMRI IDC Data Management Tool**
Written in Java.™

 **Our computing power. Your analysis.**
High-performance servers and databases available to researchers.

Maintained at [Dartmouth College](#), [Disclaimer](#) and [Privacy Statement](#).
©1999-2005, The fMRI Data Center. All rights reserved.

[Web site design and development by Paradigm Consulting Co.](#)

A false start for fMRI data sharing

nature
neuroscience

A debate over fMRI data sharing

nature

3 August 2000 Volume 406 Issue no 6795

Whose scans are they, anyway?

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 to 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

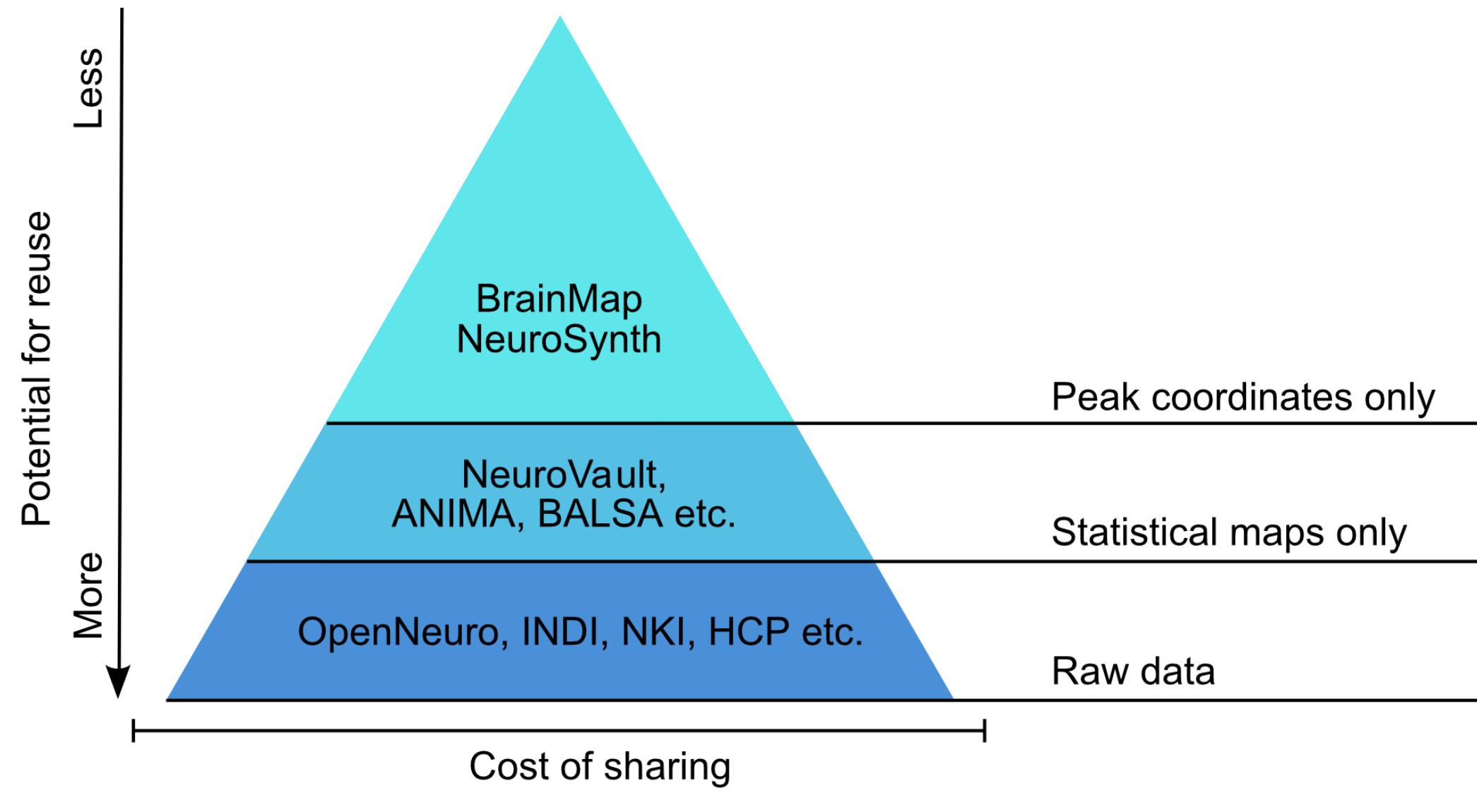
Toward discovery science of human brain function

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 Hampsonⁱ, Matthew J. Hoptman^j, James S. Hyde^k, Vesa J. Kiviniemi^l, Rolf Kötter^m, Shi-Jiang Liⁿ, 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 Tengⁱⁱ, Juha Veijola^{jj}, Arno Villringer^{ee,kk}, Martin Walter^{ll}, Lihong Wang^q, Xu-Chu Weng^{mm}, Susan Whitfield-Gabrieliⁿⁿ, Peter Williamson^{oo}, Christian Windischberger^{pp}, Yu-Feng Zang^{qq}, Hong-Ying Zhangⁱⁱ, F. Xavier Castellanos^{b,j}, and Michael P. Milham^{b,1}

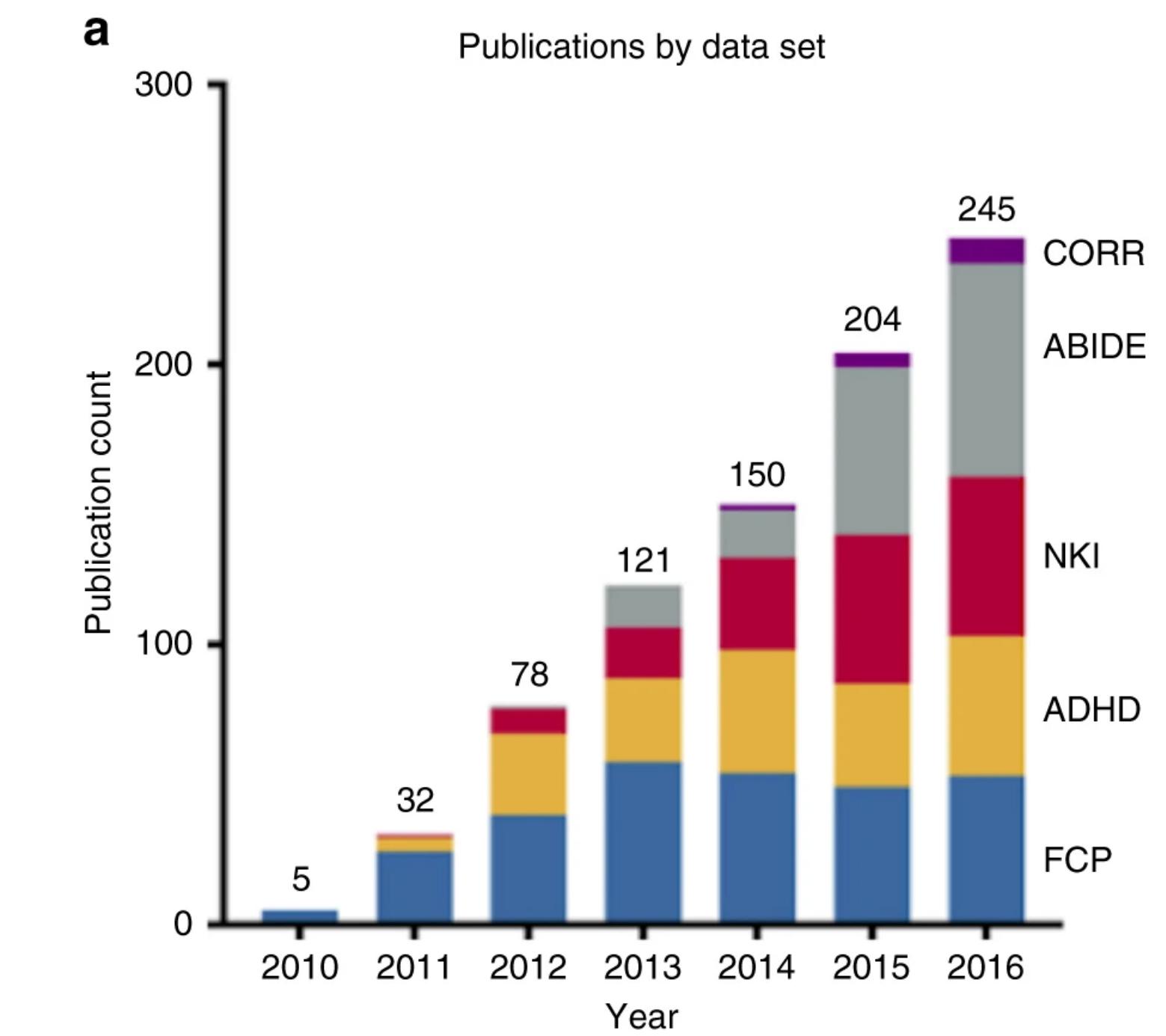
4734–4739 | PNAS | March 9, 2010 | vol. 107 | no. 10

- “Comprehensive mapping of the functional connectome, and its subsequent exploitation to discern genetic influences and brain-behavior relationships, will require multicenter collaborative datasets. Here we initiate this endeavor by gathering R-fMRI data from 1,414 volunteers collected independently at 35 international centers. We demonstrate a universal architecture of positive and negative functional connections, as well as consistent loci of inter-individual variability. ...”

Data sharing is becoming the norm in neuroimaging



Poldrack et al., *Annual Reviews in Biomedical Data Science*, 2019



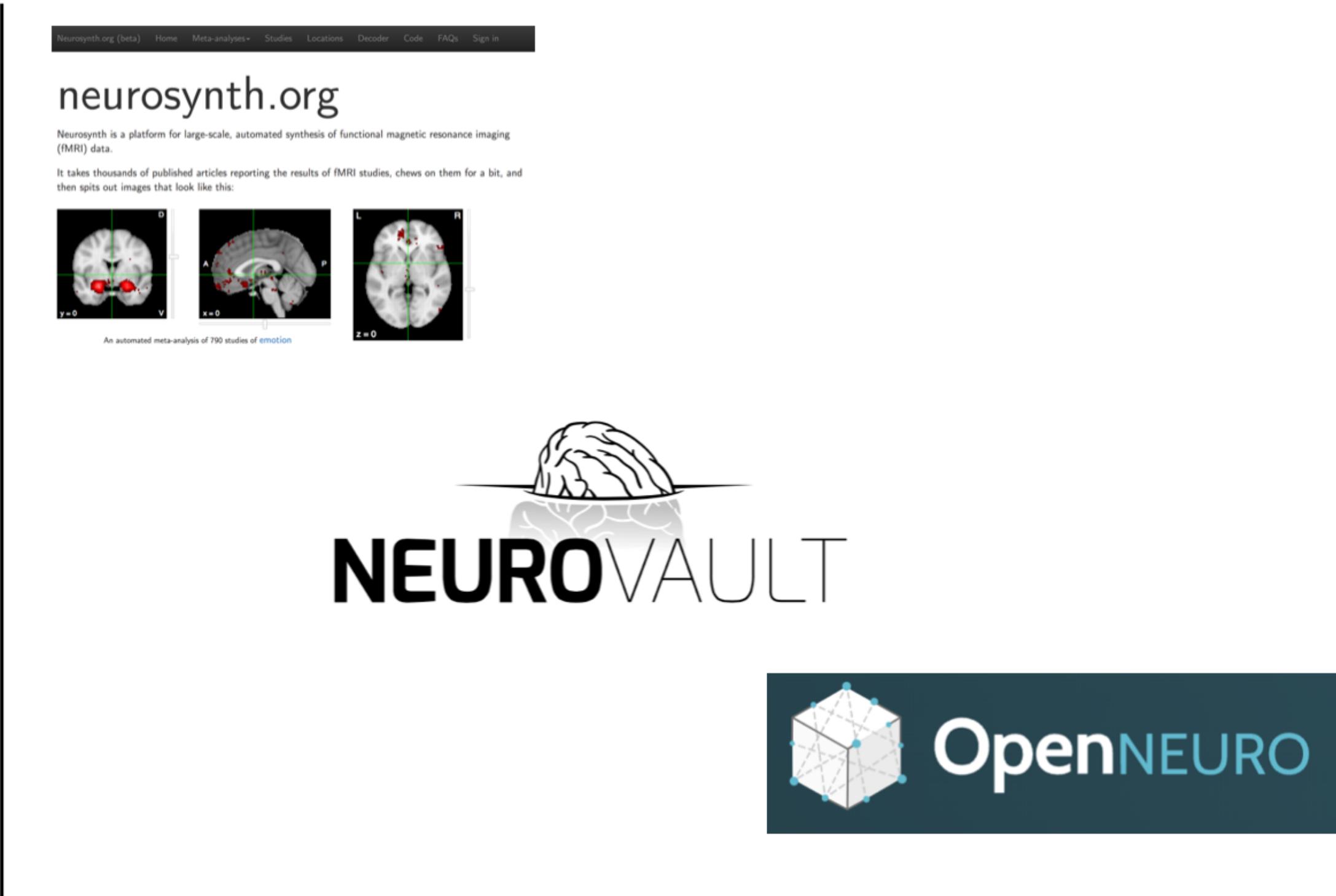
Milham et al., *Nature Communications*, 2018

Anonymous senior researcher:

“OHBM has been taken over by the open science zealots!”

An open ecosystem for retrospective data sharing

Breadth



Depth

- 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

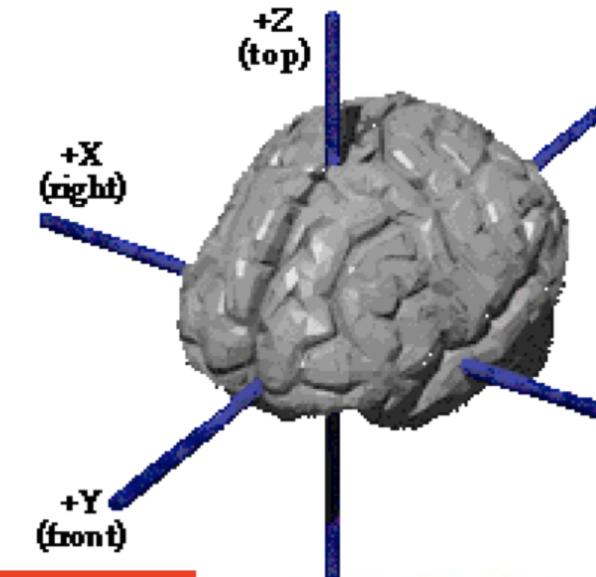


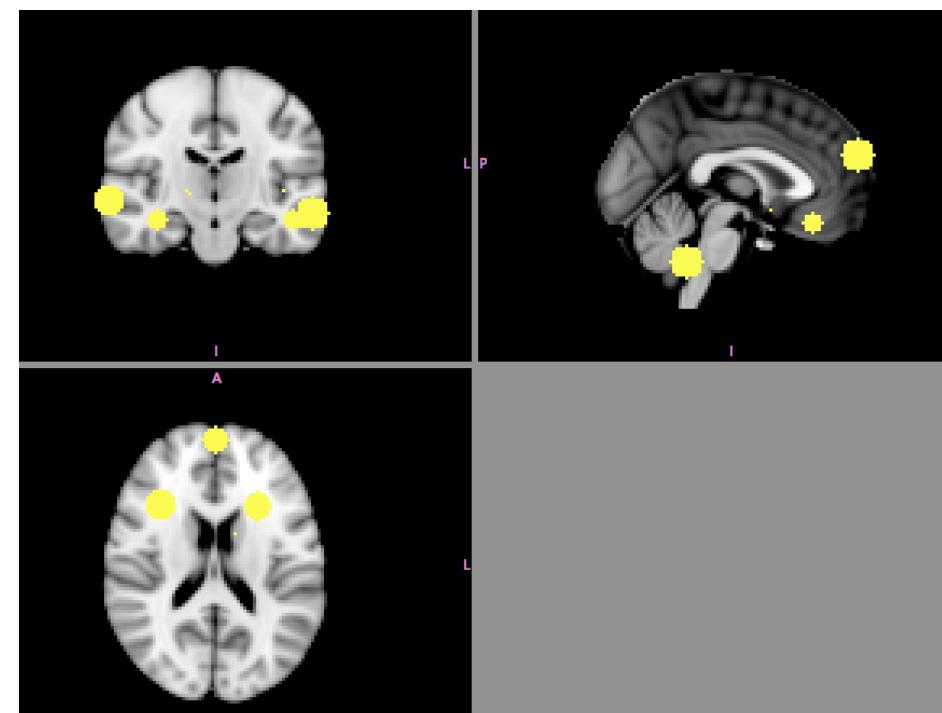
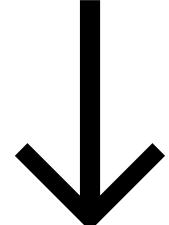
Table 1
Regions that showed a condition × time interaction in the ANOVA analysis

No.	Region	Hemisphere	BA	x	y	z	mm ³
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

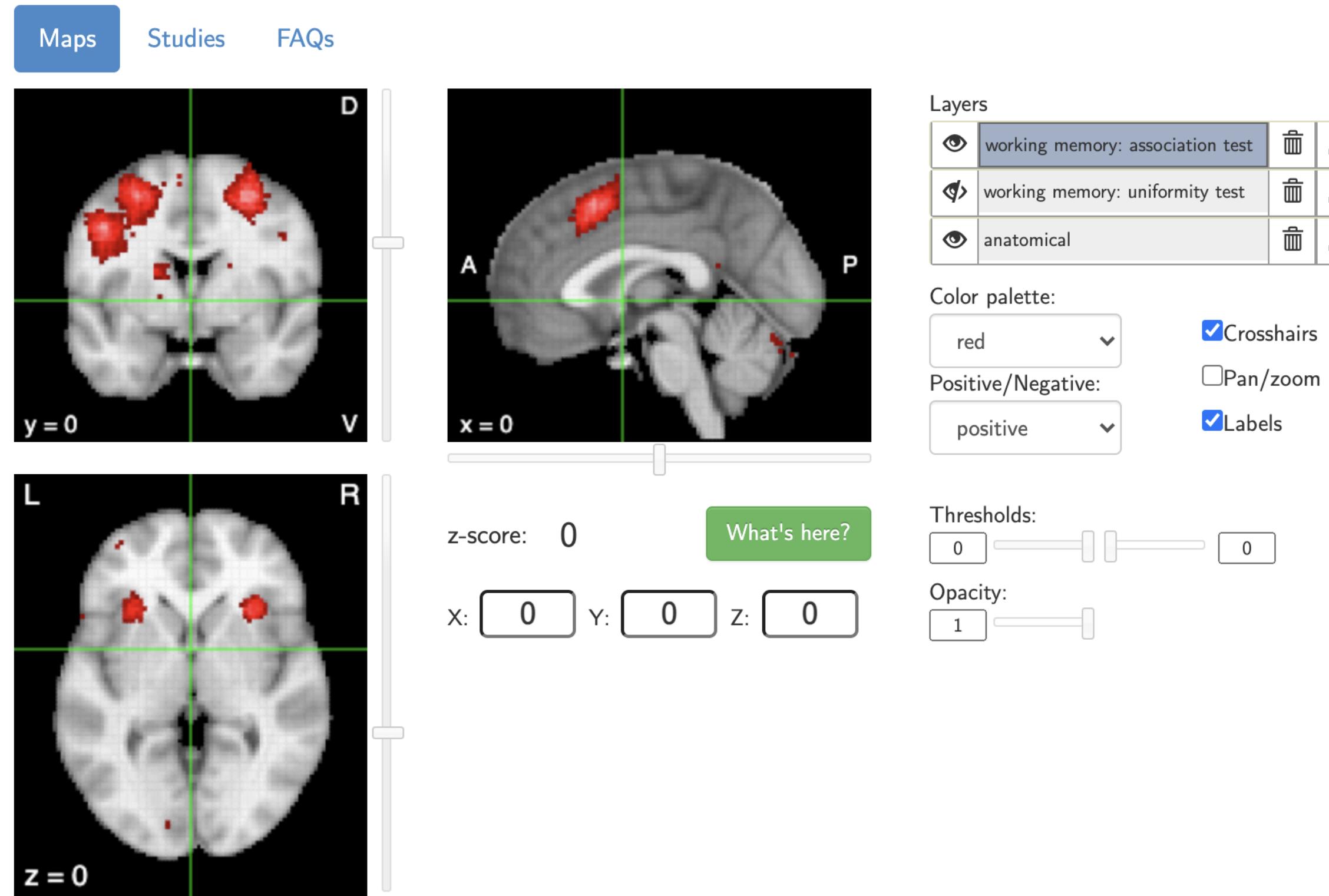
X	Y	Z
12	57	-6
33	21	15
24	-6	51
28	10	18



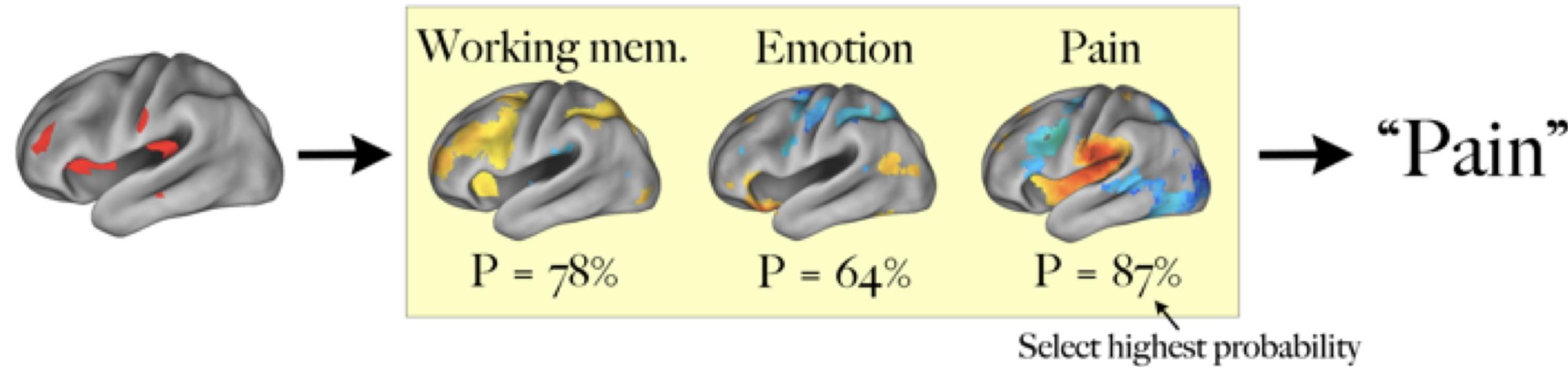
working memory

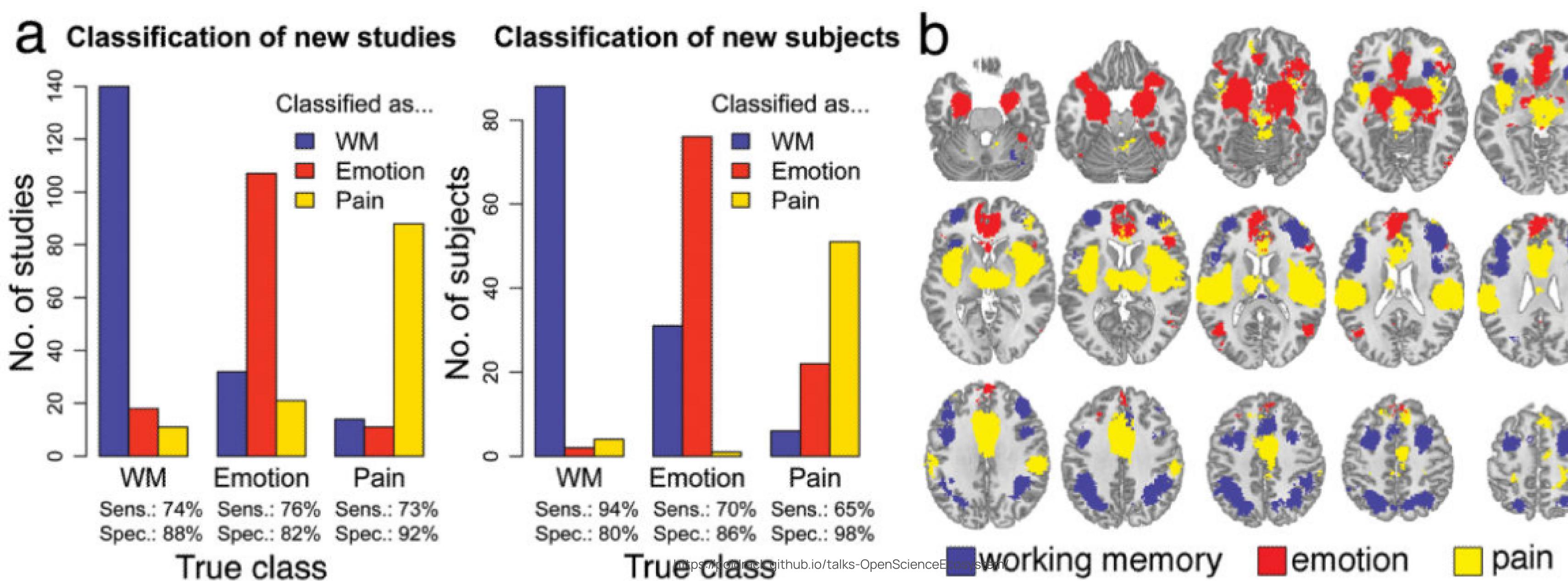
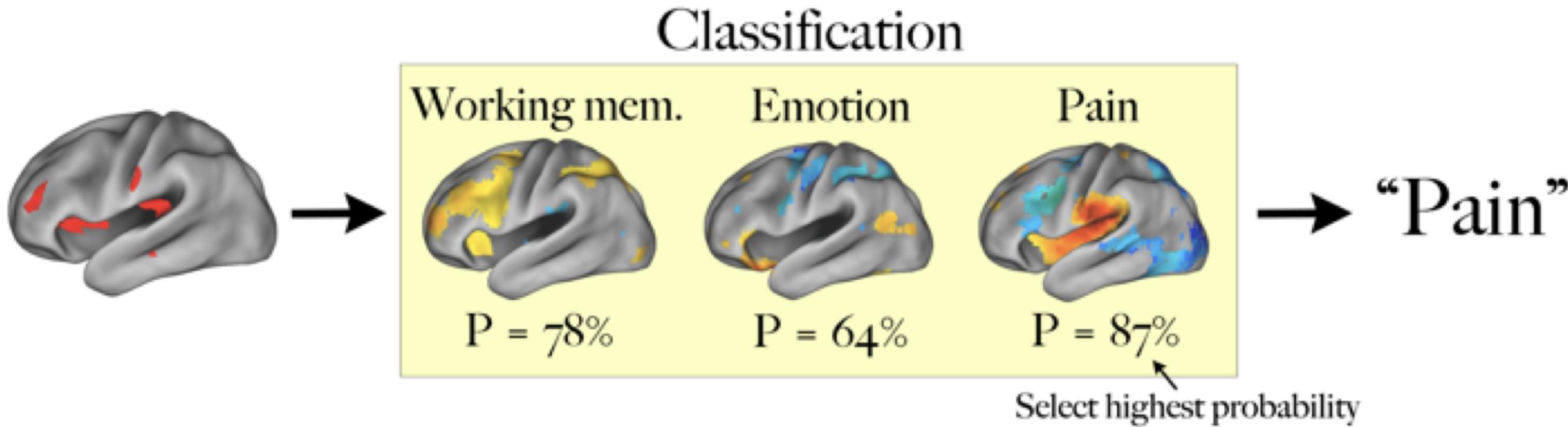
An automated meta-analysis of 1091 studies

Search for another term:

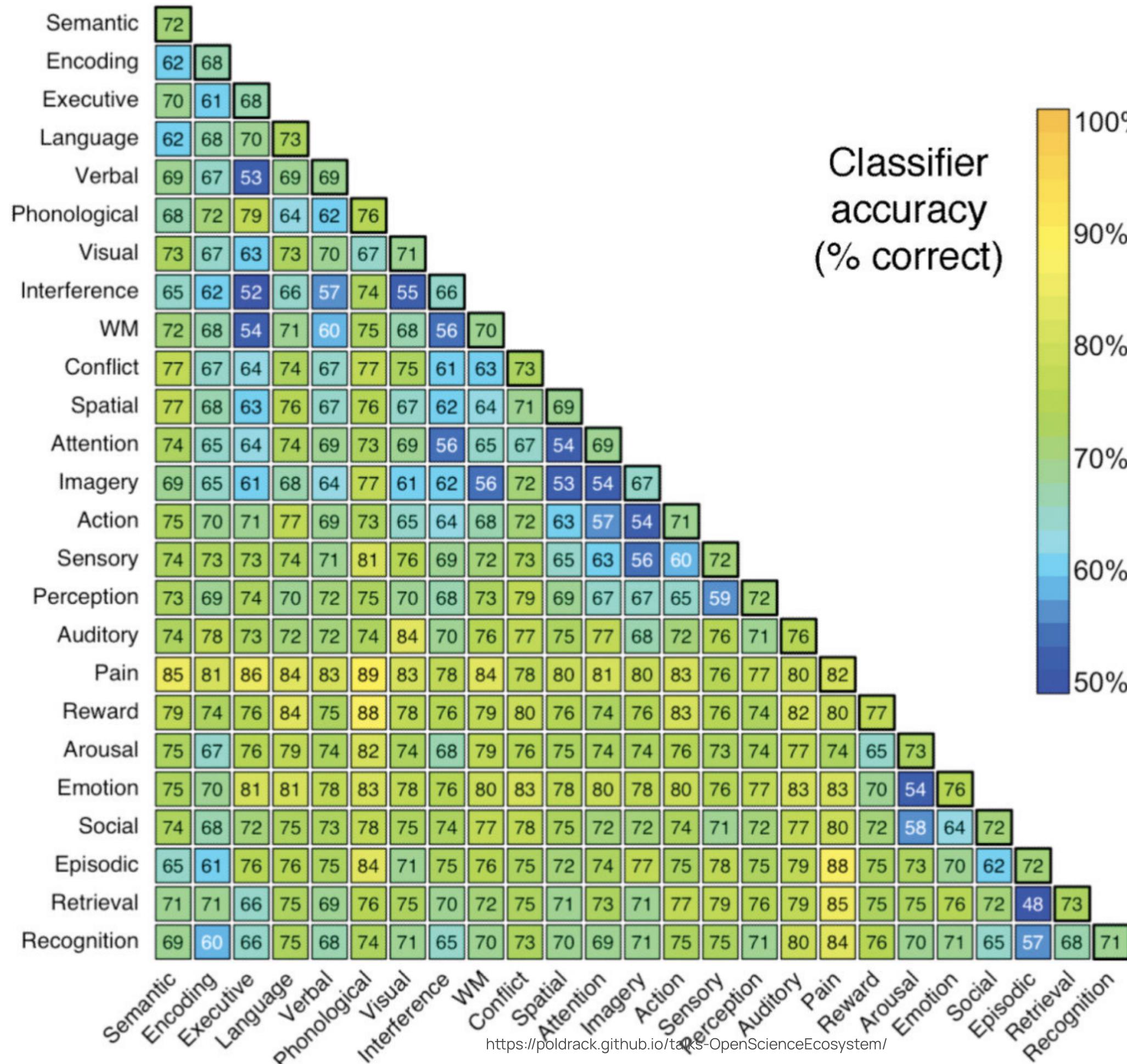


Classification



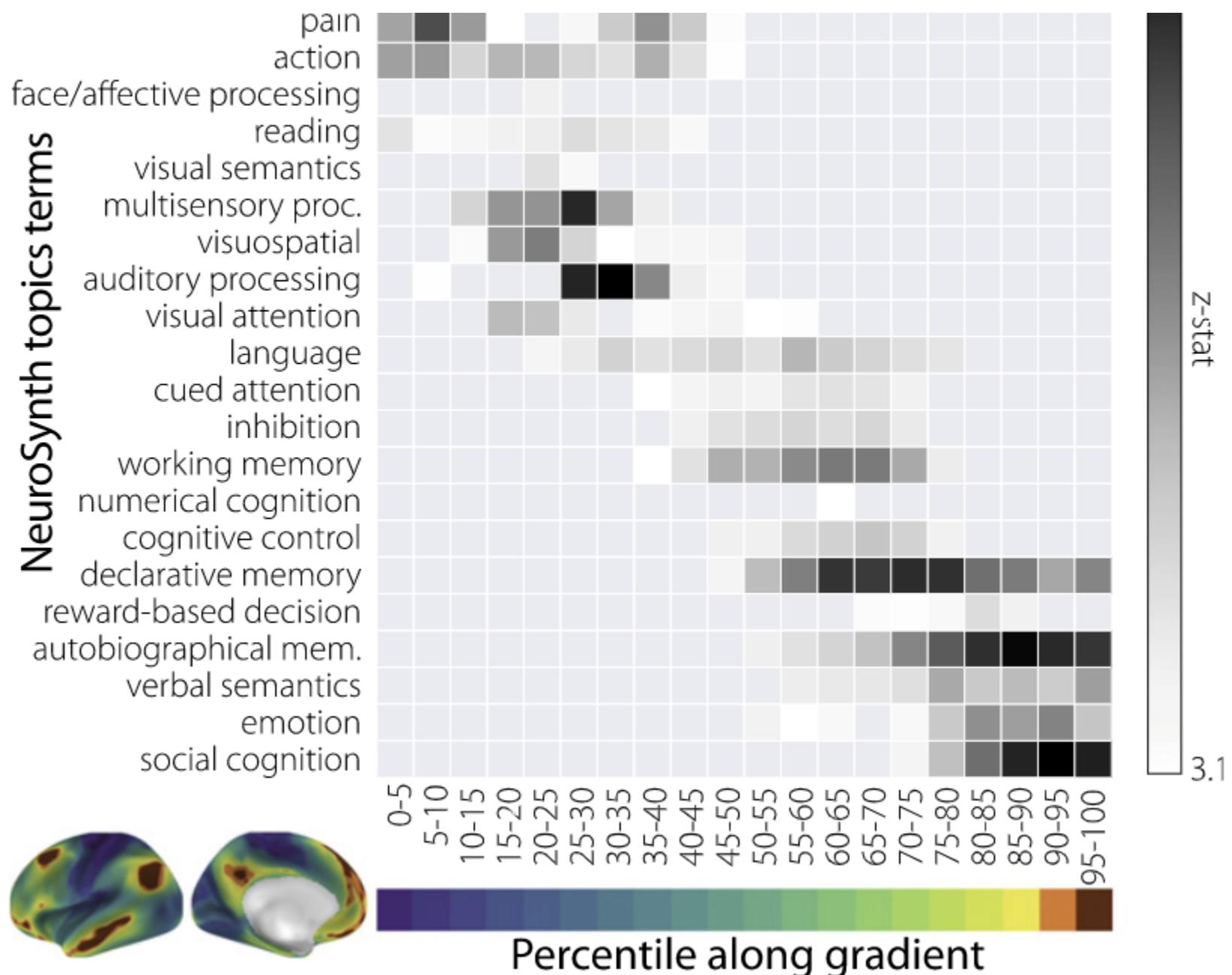


Decoding brain activity patterns using Neurosynth



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

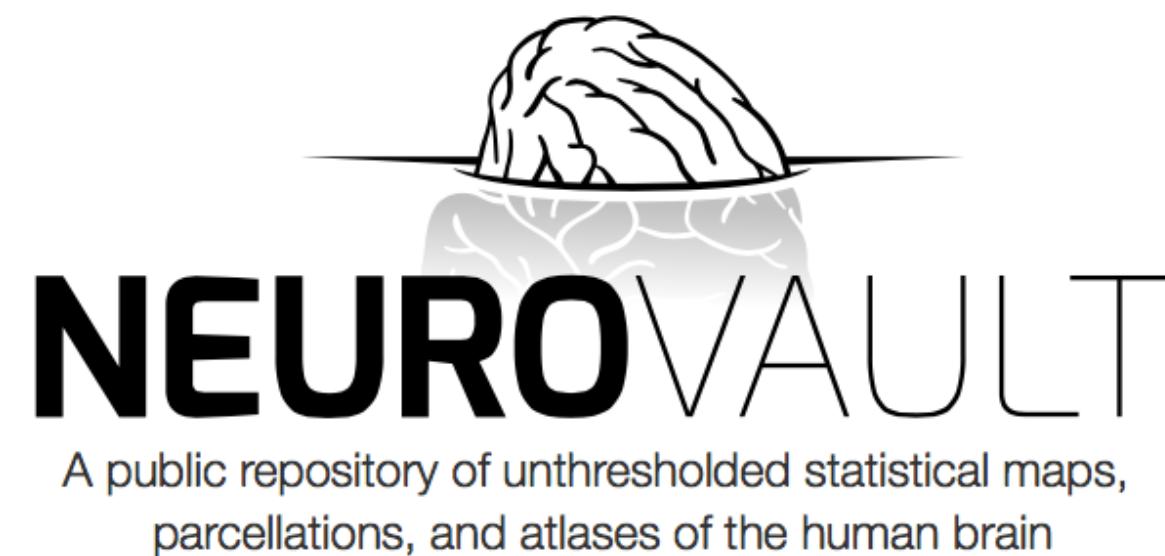
Daniel S. Margulies^{a,1}, Satrajit S. Ghosh^{b,c}, Alexandros Goulas^d, Marcel Falkiewicz^a, Julia M. Huntenburg^{a,e}, Georg Langs^{f,g}, Gleb Bezgin^h, Simon B. Eickhoff^{i,j}, F. Xavier Castellanos^{k,l}, Michael Petrides^m, Elizabeth Jefferies^{n,o}, and Jonathan Smallwood^{n,o}



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

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



What is it?

A place where researchers can publicly store and share unthresholded statistical maps, parcellations, and atlases produced by MRI and PET studies.

Why use it?

- Interactive visualization
- A permanent URL
- Publicly shareable
- Improves meta-analyses

Supported by



Get started and upload an image!

- 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

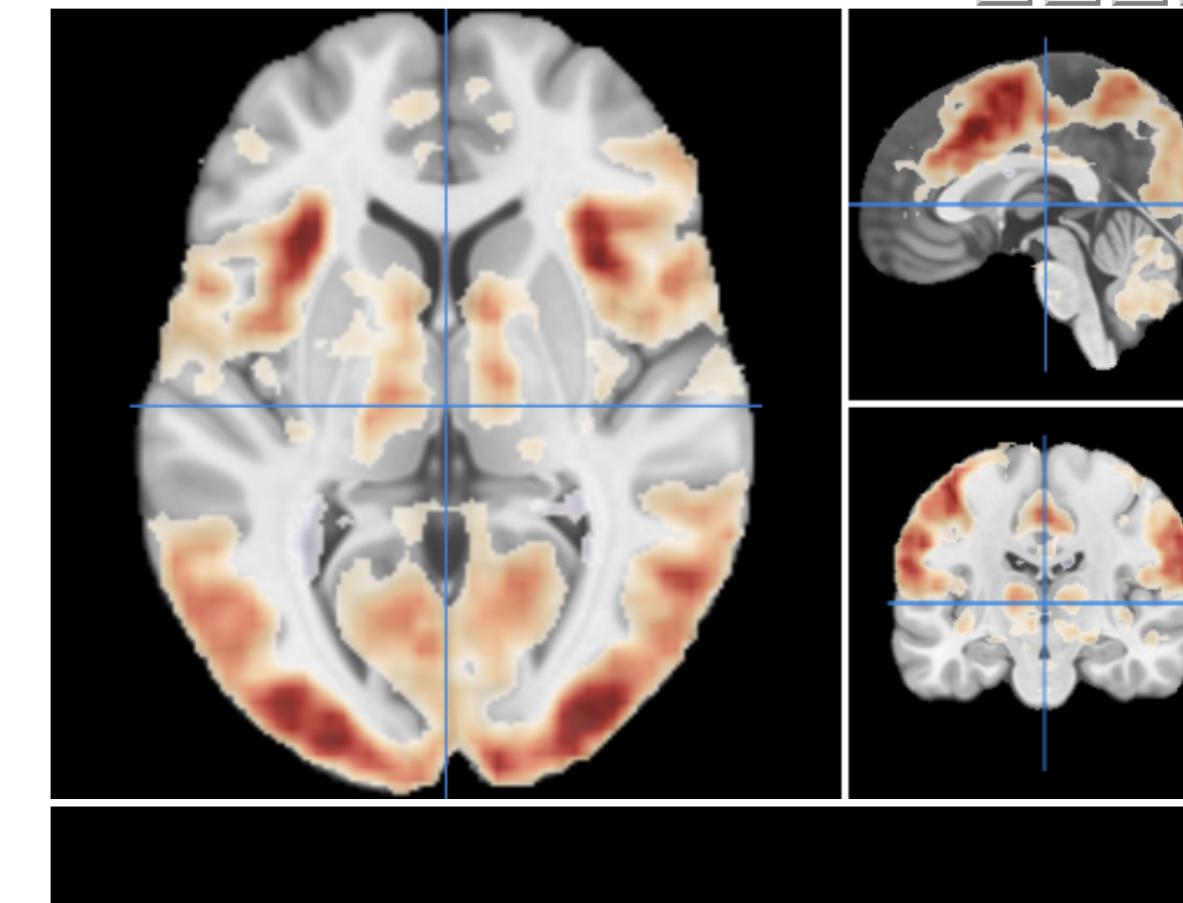
Preprocessed Consortium for Neuropsychiatric Phenomics dataset

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

Source data:

3D View Download

File View Settings Help



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:

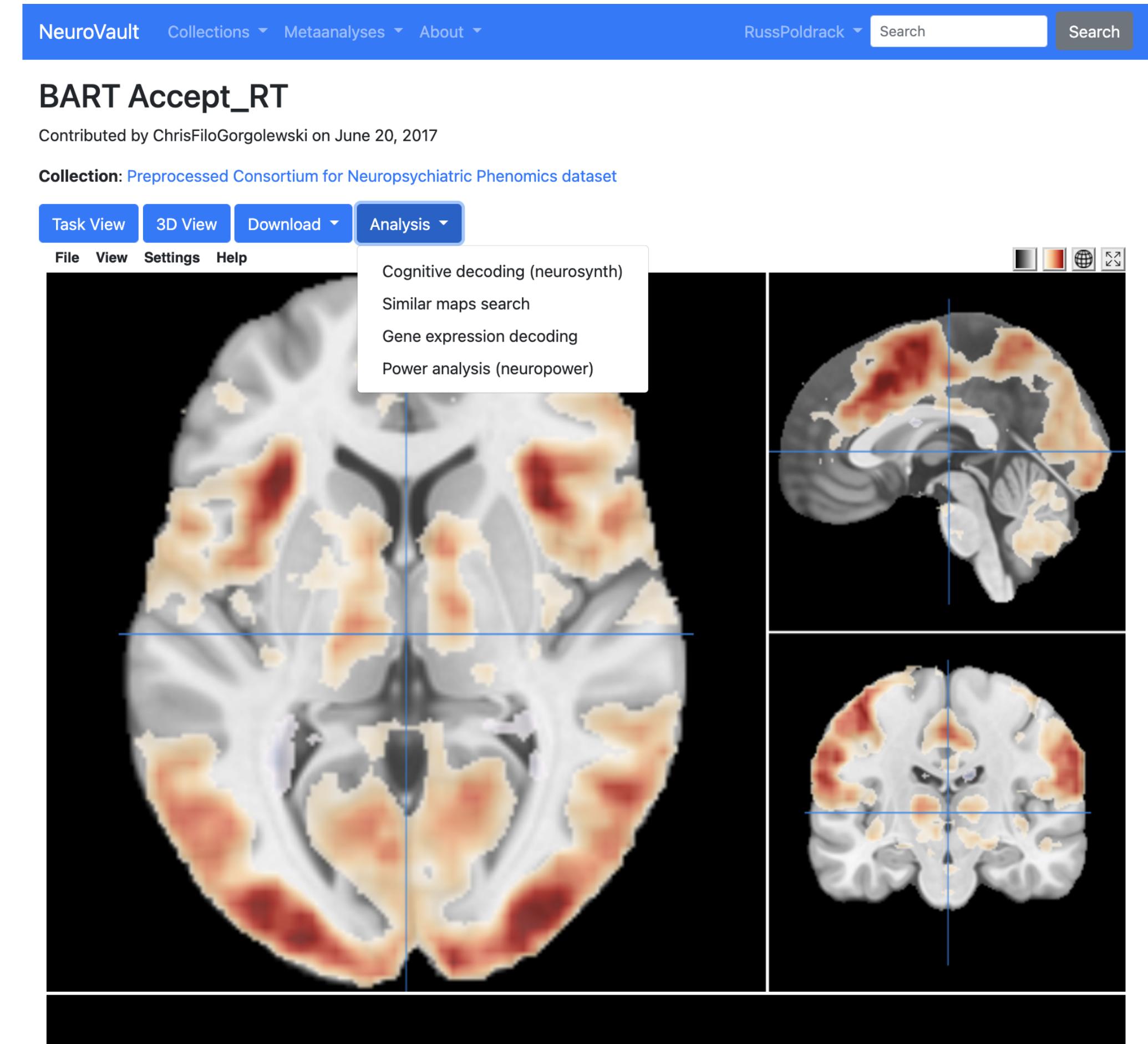
<https://identifiers.org/neurovault.collection:2606>

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

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

- **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

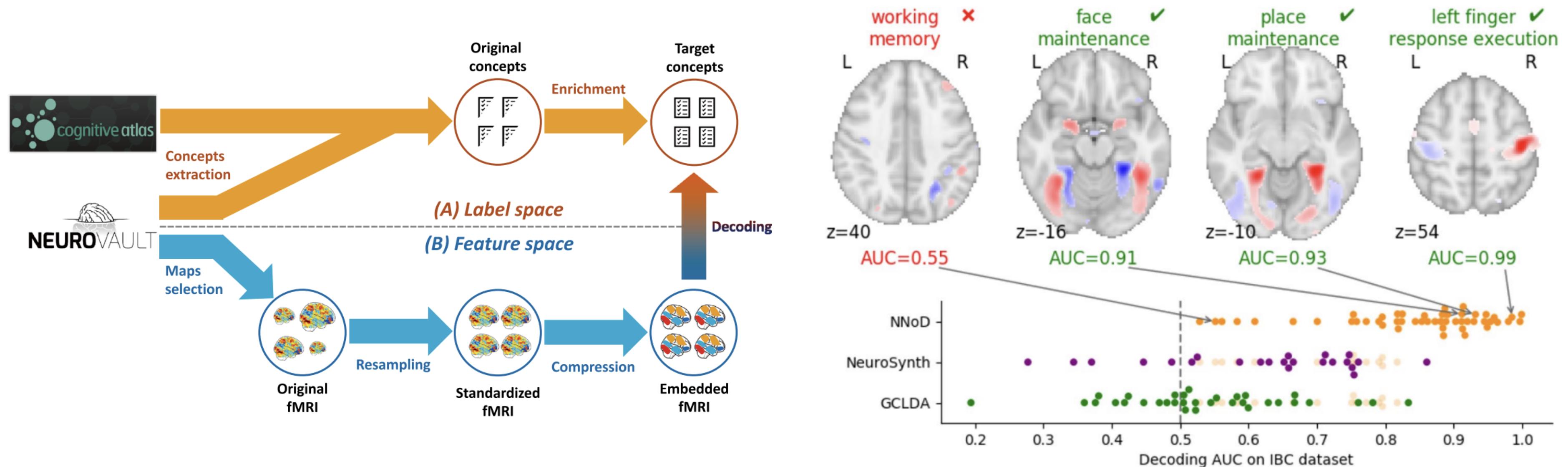


Example of Neurovault usage

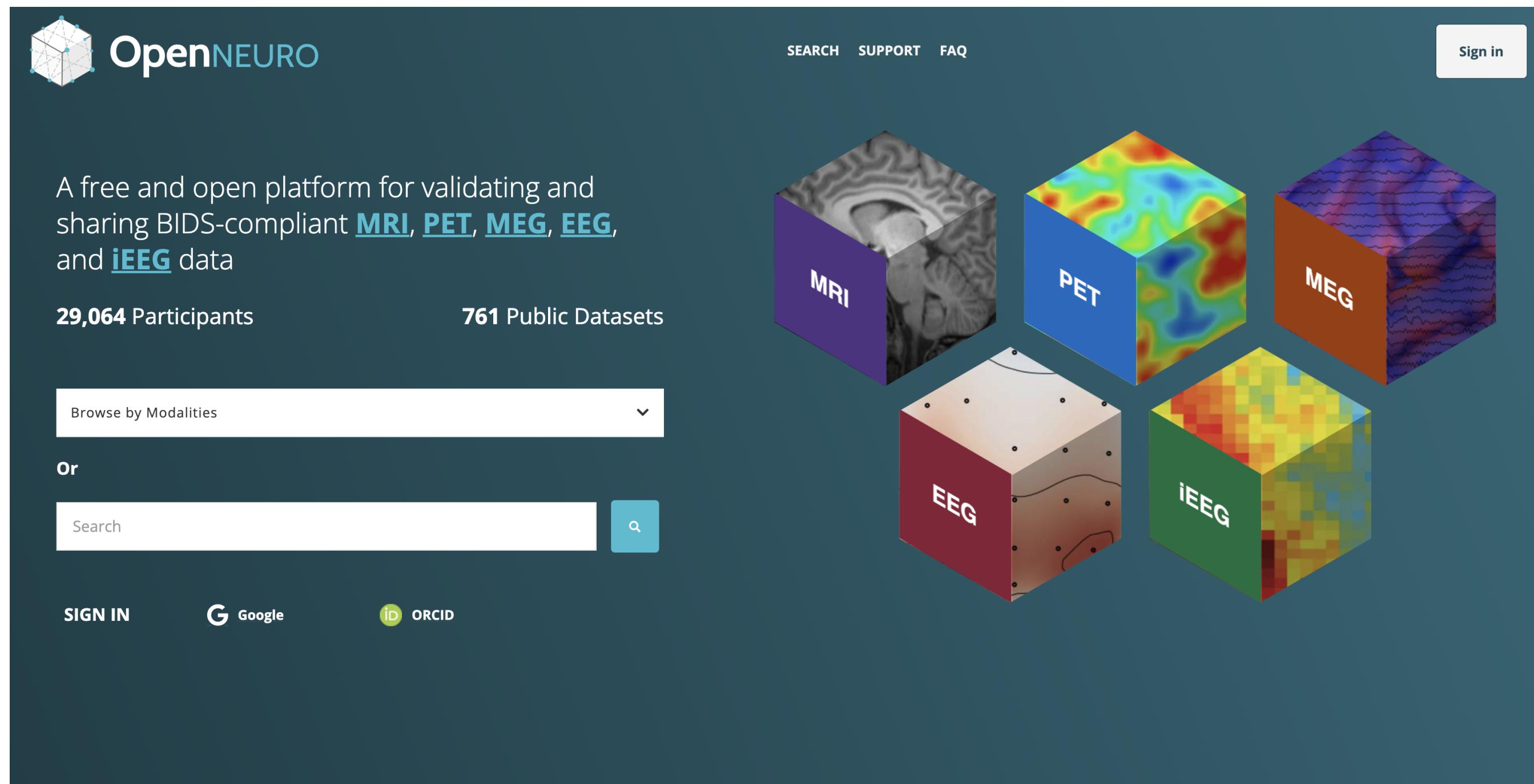
Comprehensive decoding mental processes from Web repositories of functional brain images

Romuald Menuet^{5,6}, Raphael Meudec^{1,2,3,6}, Jérôme Dockès⁴, Gael Varoquaux^{1,2,3} & Bertrand Thirion^{1,2,3}✉

Scientific Reports | (2022) 12:7050



OpenNeuro: Sharing raw and processed neuroimaging data



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-OpenScienceEcosystem/>



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
 - A file organization standard
 - A metadata standard



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

Received: 18 December 2015

Accepted: 19 May 2016

Published: 21 June 2016

Krzysztof J. Gorgolewski¹, Tibor Auer², Vince D. Calhoun^{3,4}, R. Cameron Craddock^{5,6}, Samir Das⁷, Eugene P. Duff⁸, Guillaume Flandin⁹, Satrajit S. Ghosh^{10,11}, Tristan Glatard^{7,12}, Yaroslav O. Halchenko¹³, Daniel A. Handwerker¹⁴, Michael Hanke^{15,16}, David Keator¹⁷, Xiangrui Li¹⁸, Zachary Michael¹⁹, Camille Maumet²⁰, B. Nolan Nichols^{21,22}, Thomas E. Nichols^{20,23}, John Pellman⁶, Jean-Baptiste Poline²⁴, Ariel Rokem²⁵, Gunnar Schaefer^{1,26}, Vanessa Sochat²⁷, William Triplett¹, Jessica A. Turner^{3,28}, Gaël Varoquaux²⁹ & Russell A. Poldrack¹

The development of BIDS

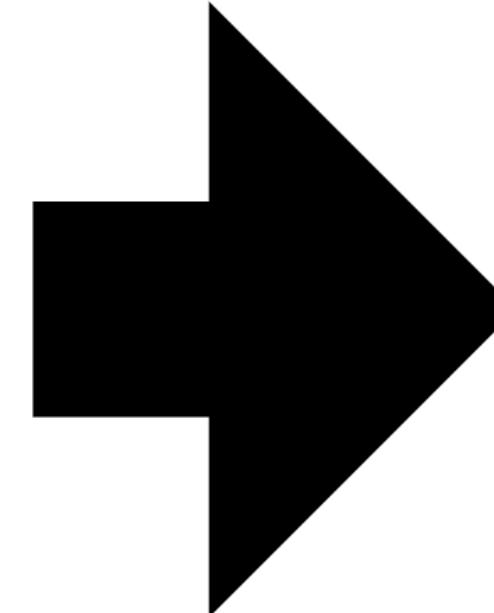
- January 2015
 - Initial stakeholder meeting at Stanford (funded by INCF)
 - Initiated development of a draft standard
- September 2015
 - Draft standard posted to BIDS web site with 22 example datasets
 - Solicited feedback from community
- June 2016
 - Published paper
- September 2018
 - BIDS-standard 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/
 - .nii.gz sub-01_T1w.nii.gz
 - 📁 func/
 - .nii.gz sub-01_task-rest_bold.nii.gz
 - /json sub-01_task-rest_bold.json
 - 📁 dwi/
 - .nii.gz sub-01_dwi.nii.gz
 - .json sub-01_dwi.json
 - /.bval sub-01_dwi.bval
 - /.bvec 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:

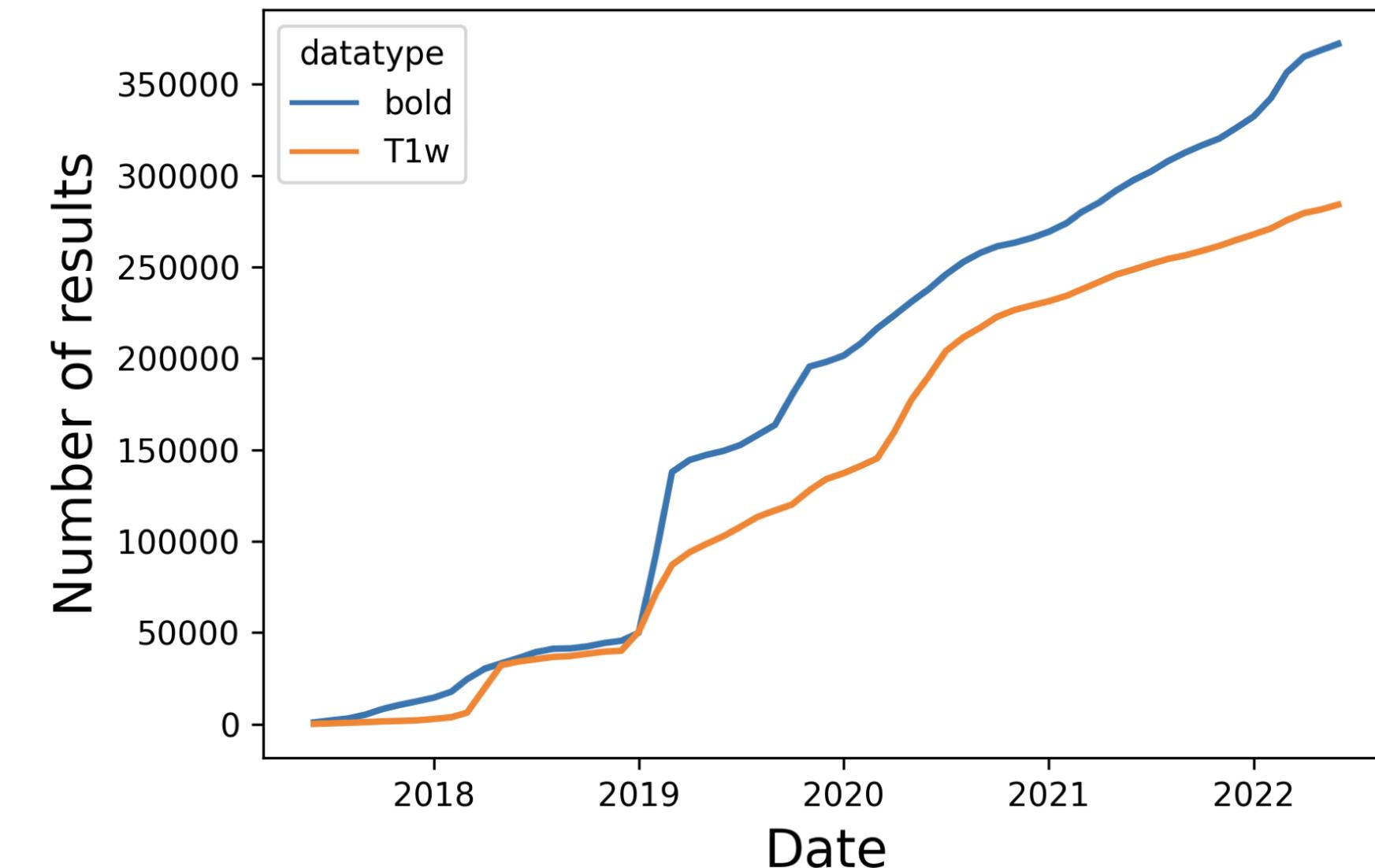
BEP #	Title
BEP001	Quantitative MRI (qMRI)
BEP003	Common Derivatives
BEP005	Arterial Spin Labeling (ASL)
BEP006	Electroencephalography (EEG)
BEP007	Hierarchical Event Descriptor (HED) Tags
BEP008	Magnetoencephalography (MEG)
BEP009	Positron Emission Tomography (PET)
BEP010	intracranial Electroencephalography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

The growing usage of BIDS: An example

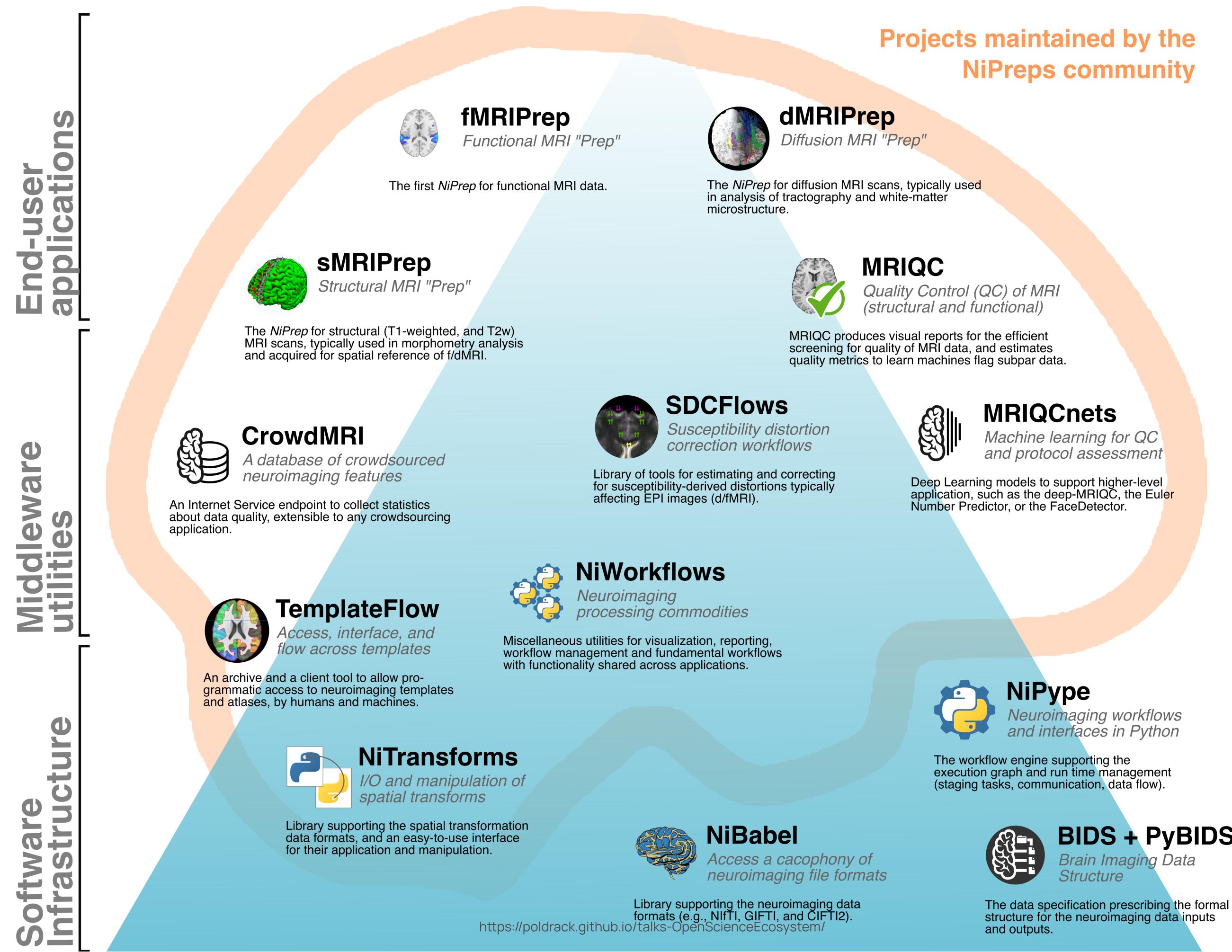
- MRIQC Web API
 - Crowdsourced database of MR QC metrics
 - QC metrics from ~375K unique BOLD scans and ~280K T1w scans as of June 2022
 - Publicly available:
<https://mriqc.nimh.nih.gov/>

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 Esteban¹, Ross W. Blair¹, Dylan M. Nielson², Jan C. Varada³, Sean Marrett³,
Adam G. Thomas¹, Russell A. Poldrack¹ & Krzysztof J. Gorgolewski¹



BIDS enables a growing open-source software ecosystem



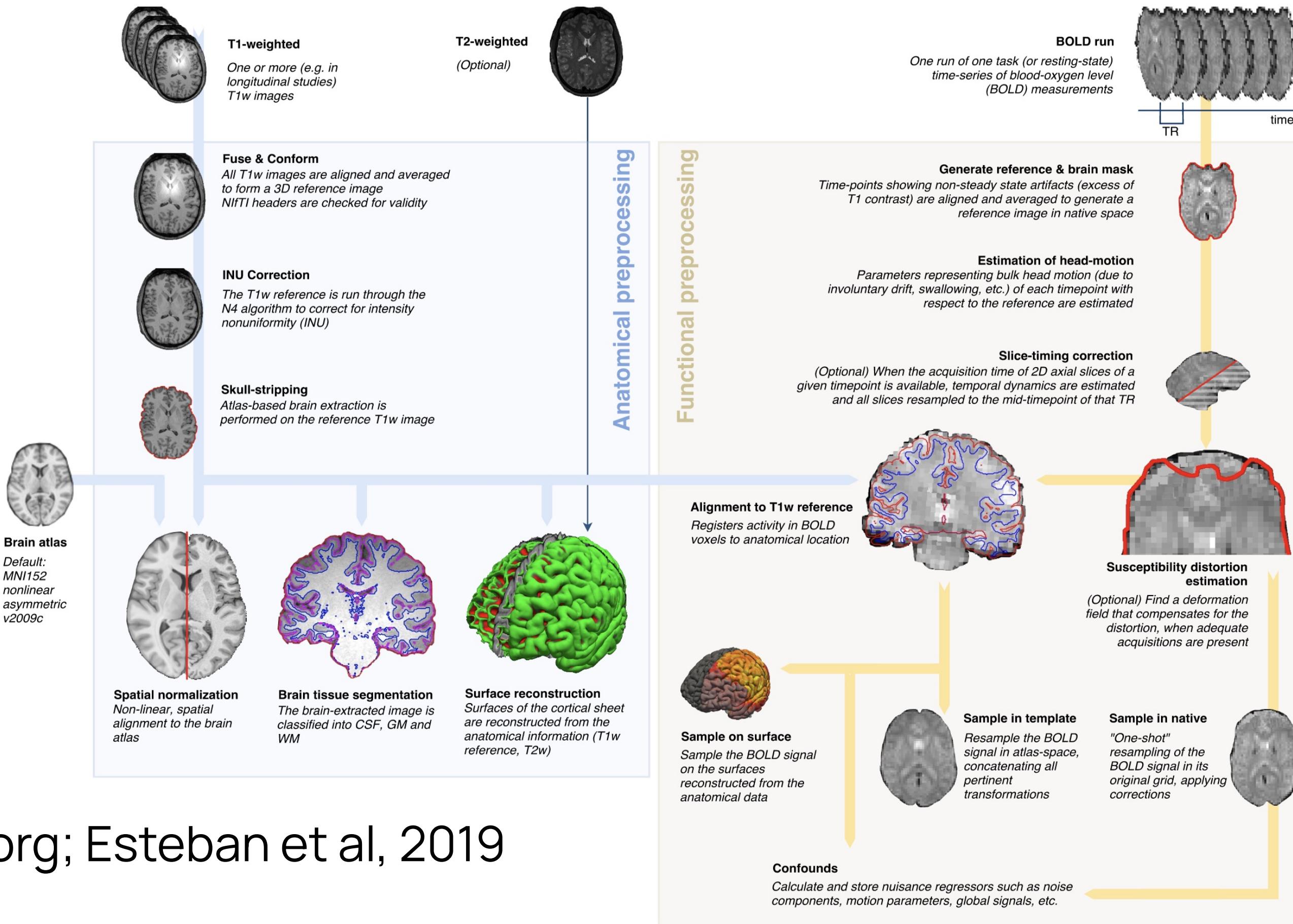
BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods

Krzysztof J. Gorgolewski^{1*}, Fidel Alfaro-Almagro², Tibor Auer³, Pierre Bellec^{4,5}, Mihai Capotă⁶, M. Mallar Chakravarty^{7,8}, Nathan W. Churchill⁹, Alexander Li Cohen¹⁰, R. Cameron Craddock^{11,12}, Gabriel A. Devenyi^{7,8}, Anders Eklund^{13,14,15}, Oscar Esteban¹, Guillaume Flandin¹⁶, Satrajit S. Ghosh^{17,18}, J. Swaroop Guntupalli¹⁹, Mark Jenkinson², Anisha Keshavan²⁰, Gregory Kiar^{21,22}, Franziskus Liem²³, Pradeep Reddy Raamana^{24,25}, David Raffelt²⁶, Christopher J. Steele^{7,8}, Pierre-Olivier Quirion¹⁵, Robert E. Smith²⁶, Stephen C. Strother^{24,25}, Gaël Varoquaux²⁷, Yida Wang⁶, Tal Yarkoni²⁸, Russell A. Poldrack¹

PLOS Computational Biology | <https://doi.org/10.1371/journal.pcbi.1005209> March 9, 2017

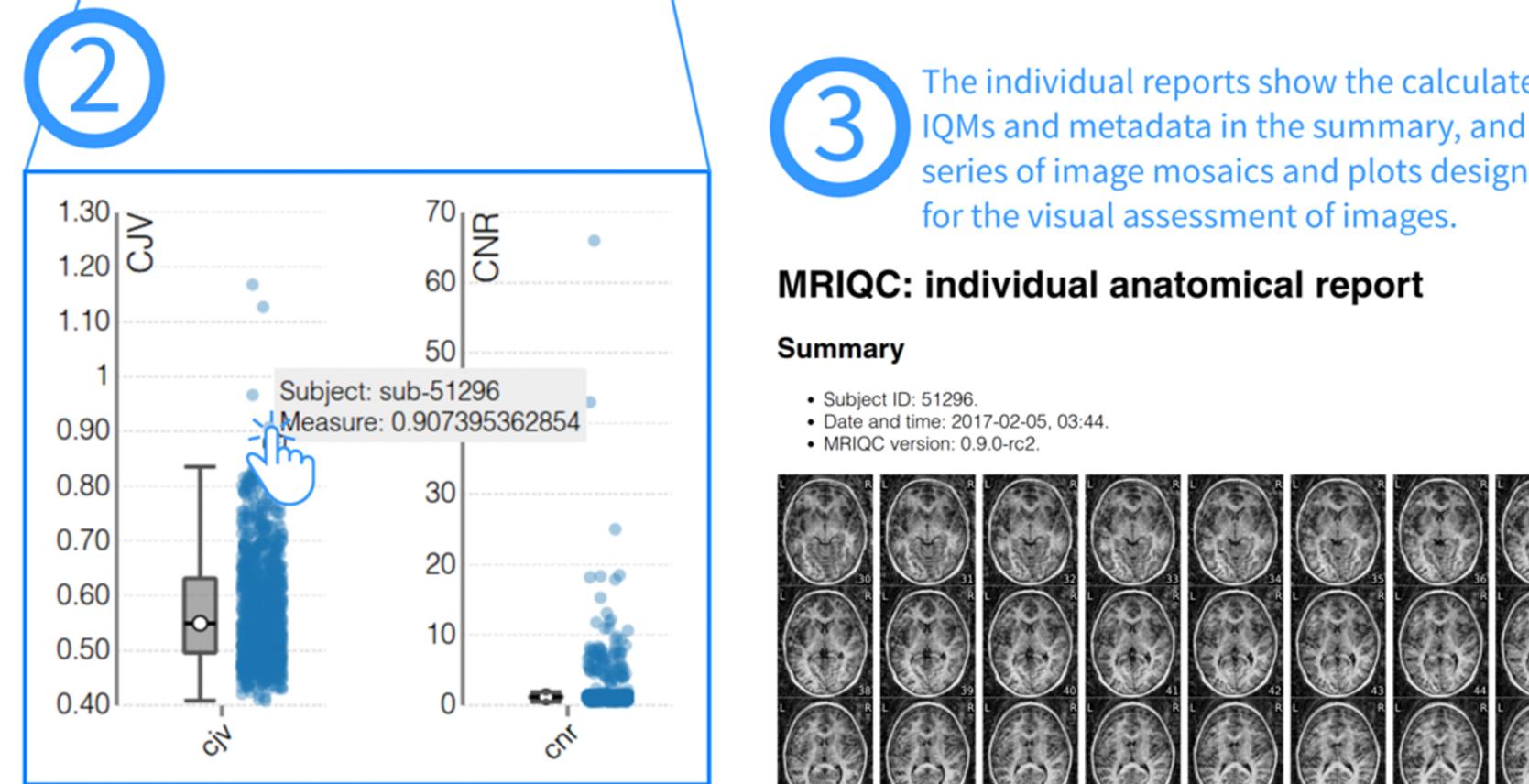
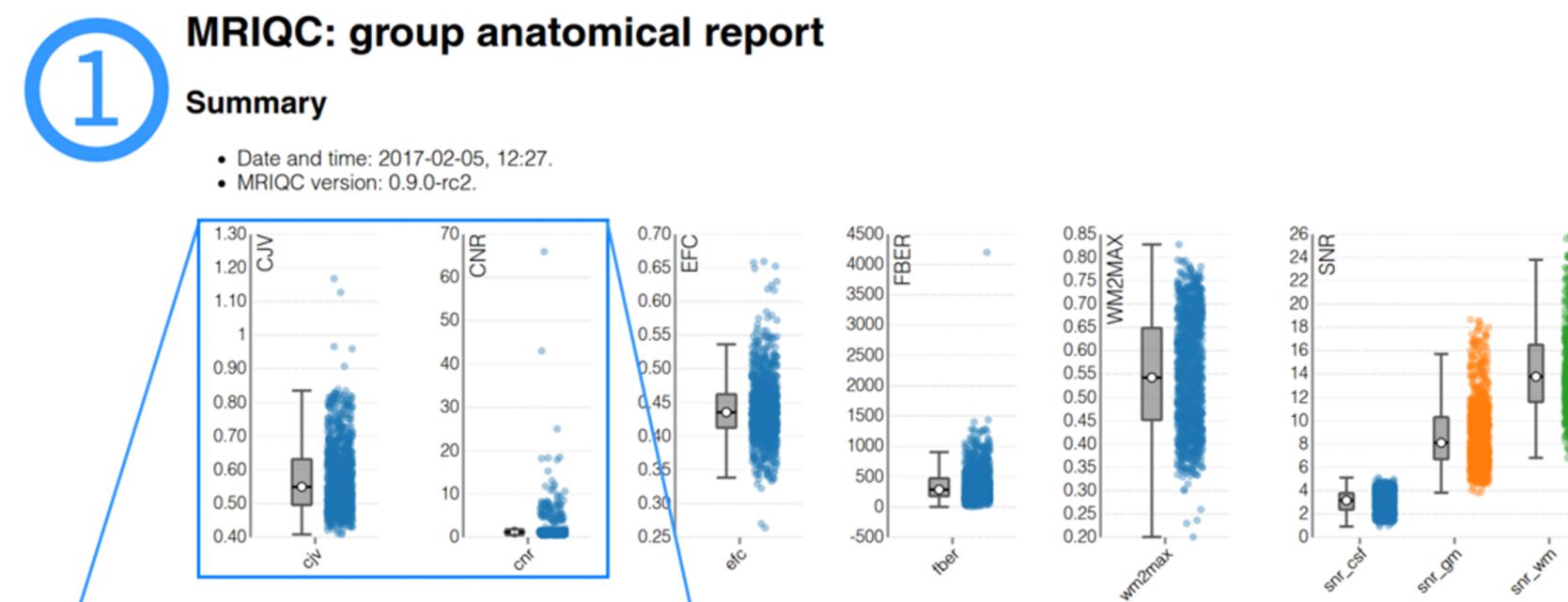
- Containerized applications that can be run on a BIDS dataset
 - Containers provide ease of use as well as better reproducibility

fMRIprep: Robust preprocessing of fMRI data



fmriprep.org; Esteban et al, 2019

MRIQC: MRI quality control for BIDS data



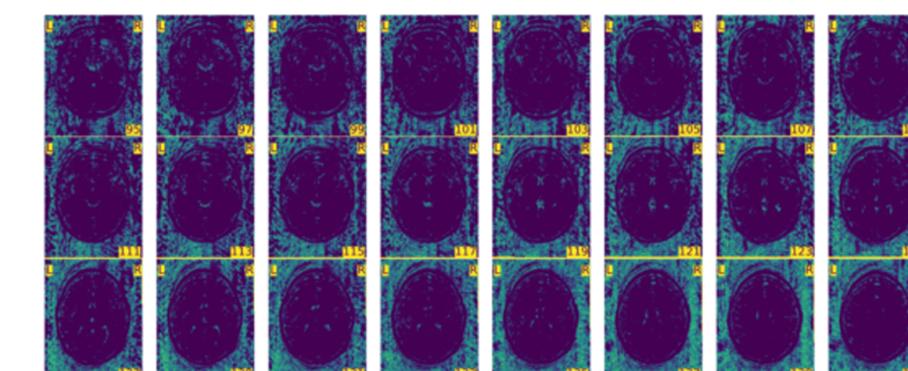
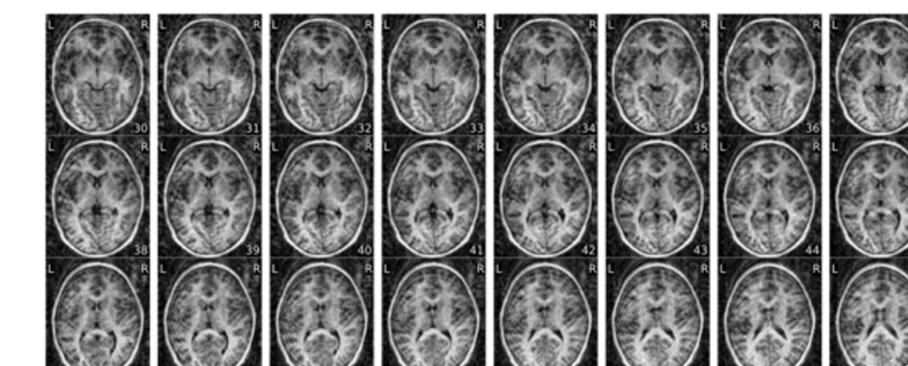
Data points in the scatter plots of the group report can be clicked to open the corresponding individual report. This feature is particularly useful to identify low-quality datasets visually.

③ The individual reports show the calculated IQMs and metadata in the summary, and a series of image mosaics and plots designed for the visual assessment of images.

MRIQC: individual anatomical report

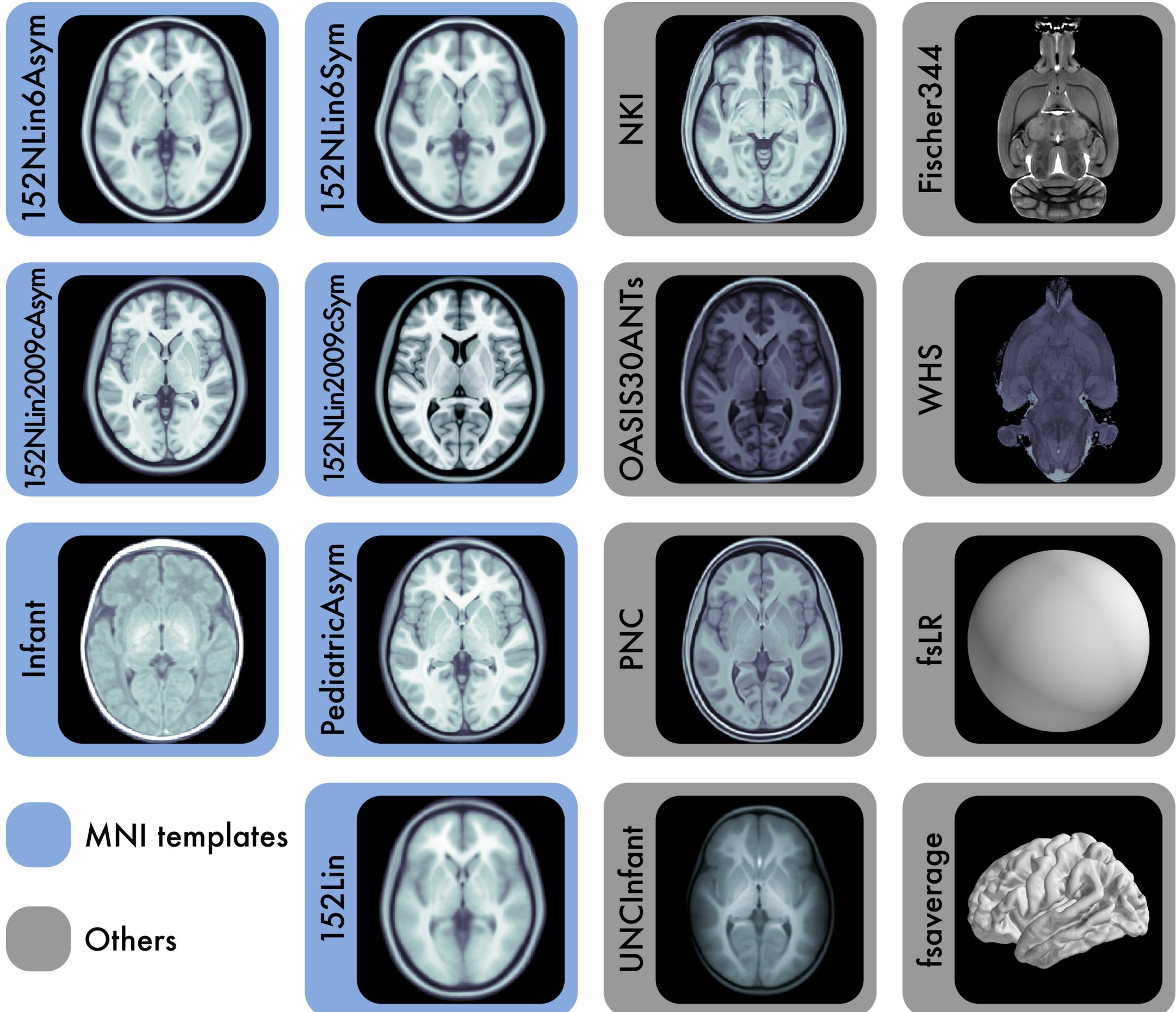
Summary

- Subject ID: 51296.
- Date and time: 2017-02-05, 03:44.
- MRIQC version: 0.9.0-rc2.

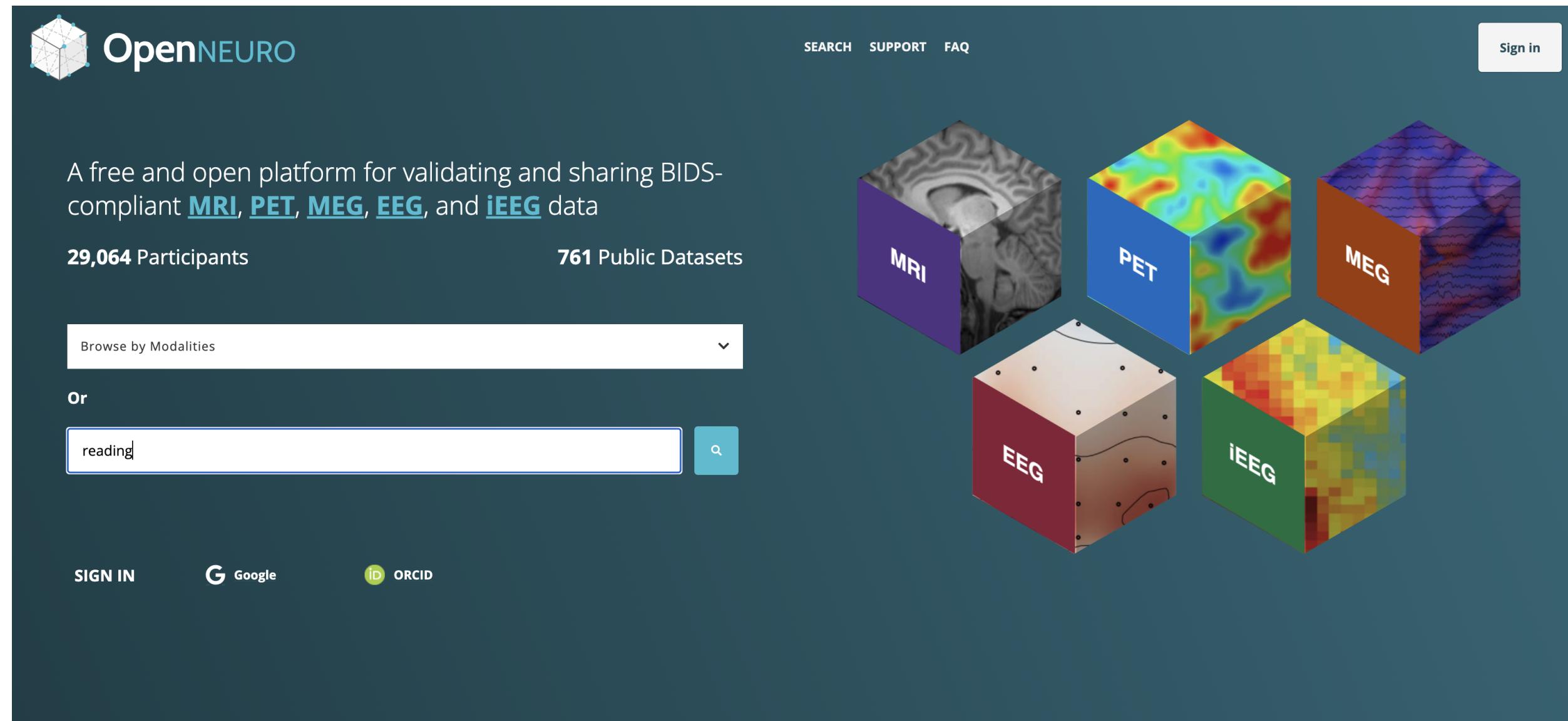


Tentemplateflow: FAIR Sharing of Neuroimaging Templates

- Templates and atlases are commonly used in neuroimaging
- There is a significant lack of clarity in the use of these templates
 - There are numerous versions of the widely used “MNI template”
- Templateflow provides programmatic access to templates and mappings between them in a BIDS-like format



OpenNeuro: A BRAIN Initiative archive for BIDS data



- Supports sharing of any validated BIDS dataset



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](#)

Search All Datasets

Keywords ?

Enter Keyword(s) to Search



reading X

These filters return **194** datasets:

CLEAR ALL

KEYWORD:

reading X

SORT BY: Relevance ↗

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

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

 OpenNEURO

SEARCH SUPPORT FAQ Sign in

MRI The Reading Brain Project L2 Adults

Follow 2 Bookmark 2

BIDS Validation ▾ 4 WARNINGS Valid brainlife.io Clone ▾

Files Download Metadata

README

OpenNeuro curator note: This dataset was previously accessible at ds002317. The dataset was reuploaded due to privacy considerations.

This dataset contains the bilingual (L2) adult subset of the Reading Brain Project (RBP) data, focusing on 56 participants who underwent two sessions of testing: MRI scanning and behavioral tests tests; collected 2-3 days apart. During the first session, following the scanning of structural (T1-weighted) and resting-state data, participants performed a reading task with simultaneous eye-tracking and fMRI scanning, and the session ended with a diffusion tensor imaging (DTI) scan. The second session consisted of only behavioral tests, including five standardized tests: the Attention Network Test (ANT), Gray's Silent Reading Test (GSRT), Letter-Number Sequencing (LNS), Peabody Picture Vocabulary Test (PPVT-4), and Tower of Hanoi (ToH), followed by a survey: Reading Background Questionnaire (RBQ), which includes familiarity rating for the topics of our five reading texts. The raw data are all provided here. The first 28 subjects (sub-01 to sub-28) were native speakers of Mandarin Chinese living in the United States. Their data was collected in Hershey, PA. The second 28 subjects (sub-29 to sub-56) were native speakers of Mandarin Chinese who lived in China. Their data was collected in Beijing, China.

An updated version of the methodology document will be available at http://blclab.org/reading_brain under L2 Adult dataset/

The Reading Brain Project L2 Adults ▾

Files: 960 Size: 63.82GB

- CHANGES
- README
- dataset_description.json
- .bidsignore
- L2_Adult_eyetracking.xlsx
- L2_Adults_LHQ.tsv
- L2_Adults_RBQ.tsv
- L2_adults_demographics.tsv
- T1w.json
- dir-AP_epi.json
- dir-PA_epi.json
- dwi.json
- task-read_bold.json

OpenNeuro Accession Number
ds003988

Authors
Ping Li, Chun-Ting Hsu, Ben Schloss, Anya Yu, Lindsey Ma, Marissa Scotto, Friederike Seyfried, Chanyuan Gu

Available Modalities
MRI

Versions

1.0.0 Created: 2022-02-01 Versions ▾

Tasks
read task, rest

Uploaded by
Chanyuan Gu on 2022-01-11 - 10 months ago

Last Updated
2022-02-01 - 10 months ago

Sessions
1

Participants
56

Dataset DOI
[doi:10.18112/openneuro.ds003988.v1.0.0](https://doi.org/10.18112/openneuro.ds003988.v1.0.0)

License
CC0

How To Cite

Text BibTeX Copy

Ping Li and Chun-Ting Hsu and Ben Schloss and Anya Yu and Lindsey Ma and Marissa Scotto and Friederike Seyfried and Chanyuan Gu (2022). The Reading Brain Project L2 Adults. OpenNeuro. [Dataset] doi: [doi:10.18112/openneuro.ds003988.v1.0.0](https://doi.org/10.18112/openneuro.ds003988.v1.0.0)

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

Any valid BIDS dataset can be shared via
OpenNeuro



SEARCH SUPPORT FAQ

Sign in

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

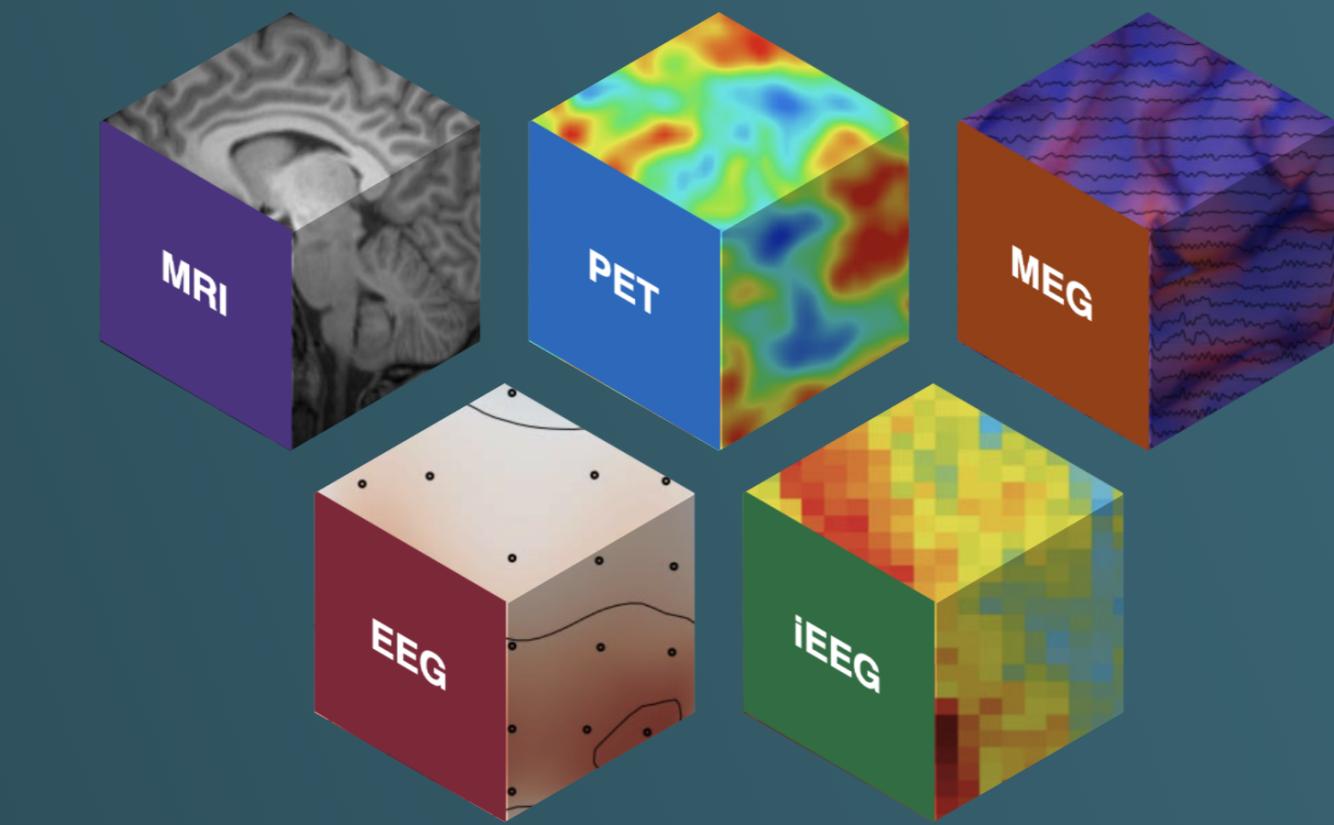
29,064 Participants

761 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.



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](#)

The screenshot shows the OpenNEURO homepage with a dark background. At the top left is the OpenNEURO logo with a brain cube icon. Below it is a callout: "A free and open platform for validating and sharing compliant **MRI**, **PET**, **MEG**, **EEG**, and **iEEG** datasets". A large number of participants is listed as "29,064 Participants". The total number of public datasets is "761 Public Datasets". A "Browse by Modalities" dropdown menu is open, showing "MRI", "PET", "MEG", "EEG", and "iEEG". Below it is a search bar with the placeholder "Search" and a magnifying glass icon. Navigation links include "SIGN IN", "Google", and "ORCID". A modal window titled "Sign in" is displayed, featuring "Google" and "ORCID" buttons, a "Sign in" button in the top right, and a "What is this?" link.

BIDS

Validation Using BIDS

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

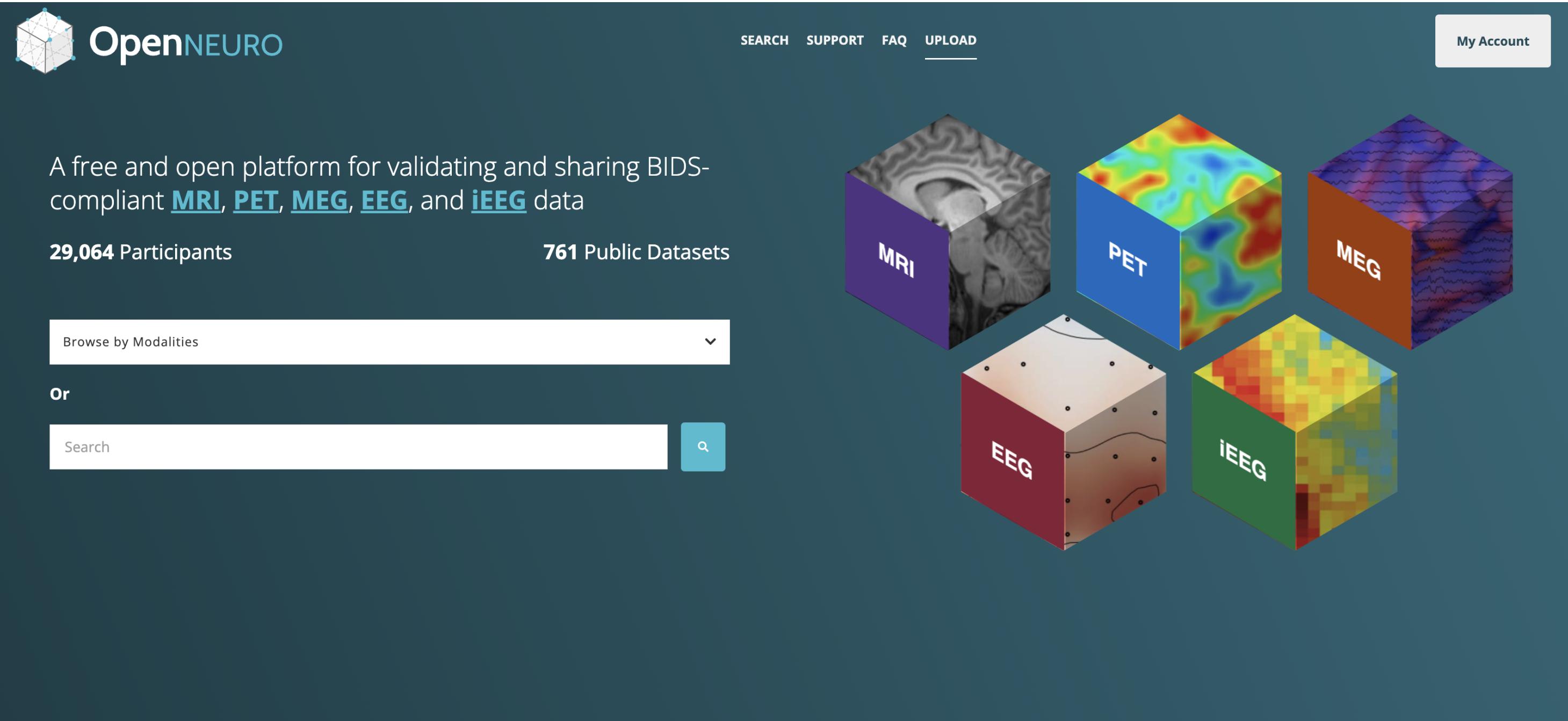
Want to contribute to BIDS?
Visit the [Google discussion group](#) to contribute.

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](#)



The screenshot shows the OpenNEURO homepage. At the top left is the logo with a brain cube icon and the text "OpenNEURO". To the right are navigation links: SEARCH, SUPPORT, FAQ, UPLOAD (underlined), and My Account. Below the header, a main message reads: "A free and open platform for validating and sharing BIDS-compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#) data". To the left, there are two statistics: "29,064 Participants" and "761 Public Datasets". Below these are search and browse options: "Browse by Modalities" with a dropdown arrow, "Or", and a search bar with a magnifying glass icon. To the right, five 3D cube icons represent different data modalities: MRI (grey brain scan), PET (color-coded heatmap), MEG (wavy lines), EEG (red grid with black dots), and iEEG (color-coded heatmap). The background is dark teal.



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

A free and open platform for validating compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and

29,064 Participants

Browse by Modalities

Or

Search

Upload Dataset

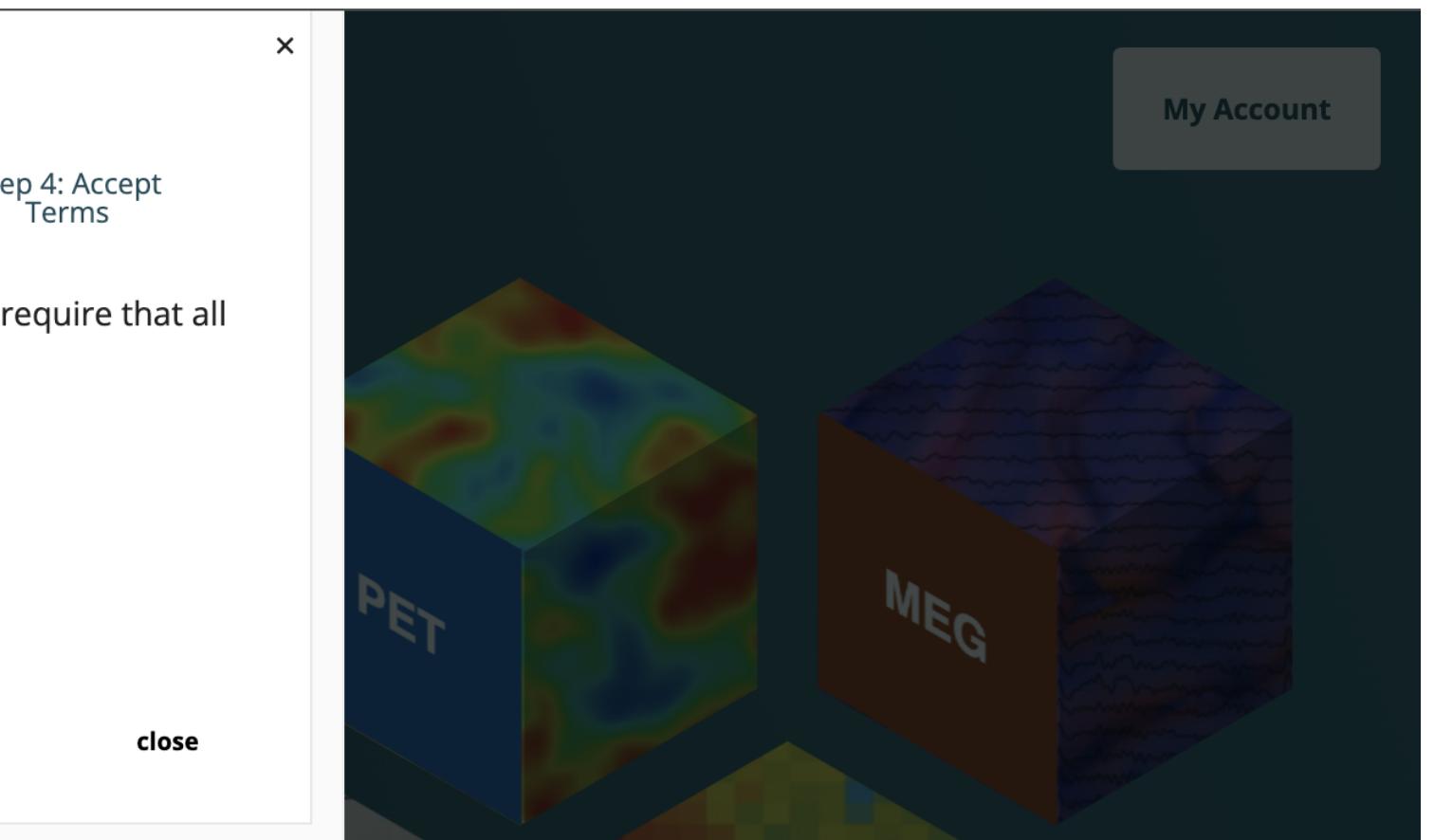
Step 1: Select Files Step 2: Validation Step 3: Metadata Step 4: Accept Terms

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



 OpenNEURO

MRI Rhyme judgment 

This dataset is public.

BIDS Validation ▾

Files Share Versioning Admin

How to Download

Download with your browser

This method is convenient and allows you to select a local directory to save the dataset.

Steps

1. Select a local directory to save the dataset and grant permission to download.
2. Download will run in the background, please leave the page open.
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  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  Not all subjects/sessions/runs have the same scanning parameters.

Warning: 3  Tabular file contains custom columns not described in a data dictionary

Click to view details on [BIDS specification](#)

close

My Account

Follow 1 Bookmark 2

Dataset Accession Number: 03

Russell A. Poldrack

Dataset Modality: fMRI

Created: 2020-05-14

Versions ▾

Uploaded by: Chris Gorgolewski on 2016-10-13 - about 6 years ago

OpenNEURO

MRI Rhyme judgment [Edit](#)

This dataset is public.

BIDS Validation

- Files
- Share
- Versioning
- Admin

How to Download

Download with your browser

This method is convenient and allows you to select a local directory to save the dataset.

Steps

1. Select a local directory to save the dataset and grant permission to OpenNeuro to access it.
2. Download will run in the background, please leave the page open until it completes.
3. A notification will appear when complete.

[Download](#)

Download from S3

The most recently published snapshot can be downloaded from S3. For example, you can use the AWS CLI:

```
aws s3 sync --no-sign-request s3://openneuro.ds000003.d  
To download unpublished datasets or older snapshots, see the Node.js documentation.
```

Download with Node.js

Using [@openneuro/cli](#) you can download this dataset from the command line. This package has no dependencies and is stable across all platforms, though it may have unstable connections, but has known issues on Windows.

```
openneuro download --draft ds000003 ds000003-d  
DOI: https://doi.org/10.11112/openneuro.ds000003.v1.0.0
```

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

MRI Rhyme judgment Edit

This dataset is public

BIDS Validation ▼

Files Share Versioning Admin

How to Download

Download with your browser

This method is convenient and allows you to select a local directory to save the dataset.

Steps

1. Select a local directory to save the dataset and grant permission to OpenNeuro to access it.
2. Download will run in the background, please leave the page open until it completes.
3. A notification will appear when complete.

Download

Download from S3

The most recently published snapshot can be downloaded from S3. For example, you can use the AWS CLI:

```
aws s3 sync --no-sign-request s3://openneuro.org/datasets/rhyme-judgment/03/ /path/to/local/directory
```

To download unpublished datasets or older snapshots, see the [Download section](#).

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

 Rhyme judgment  Following 1  Bookmark 0This 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



You must add at least one change message to create a new 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](https://doi.org/10.18112/openneuro.ds000003.v1.0.0)

Google Calendar - week of Jul 19, 2015
<https://www.google.com/calendar/render?pli=1&>



OpenNEURO

SEARCH SUPPORT FAQ UPLOAD My Account

Rhyme judgment

Follow 1 Bookmark 2

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

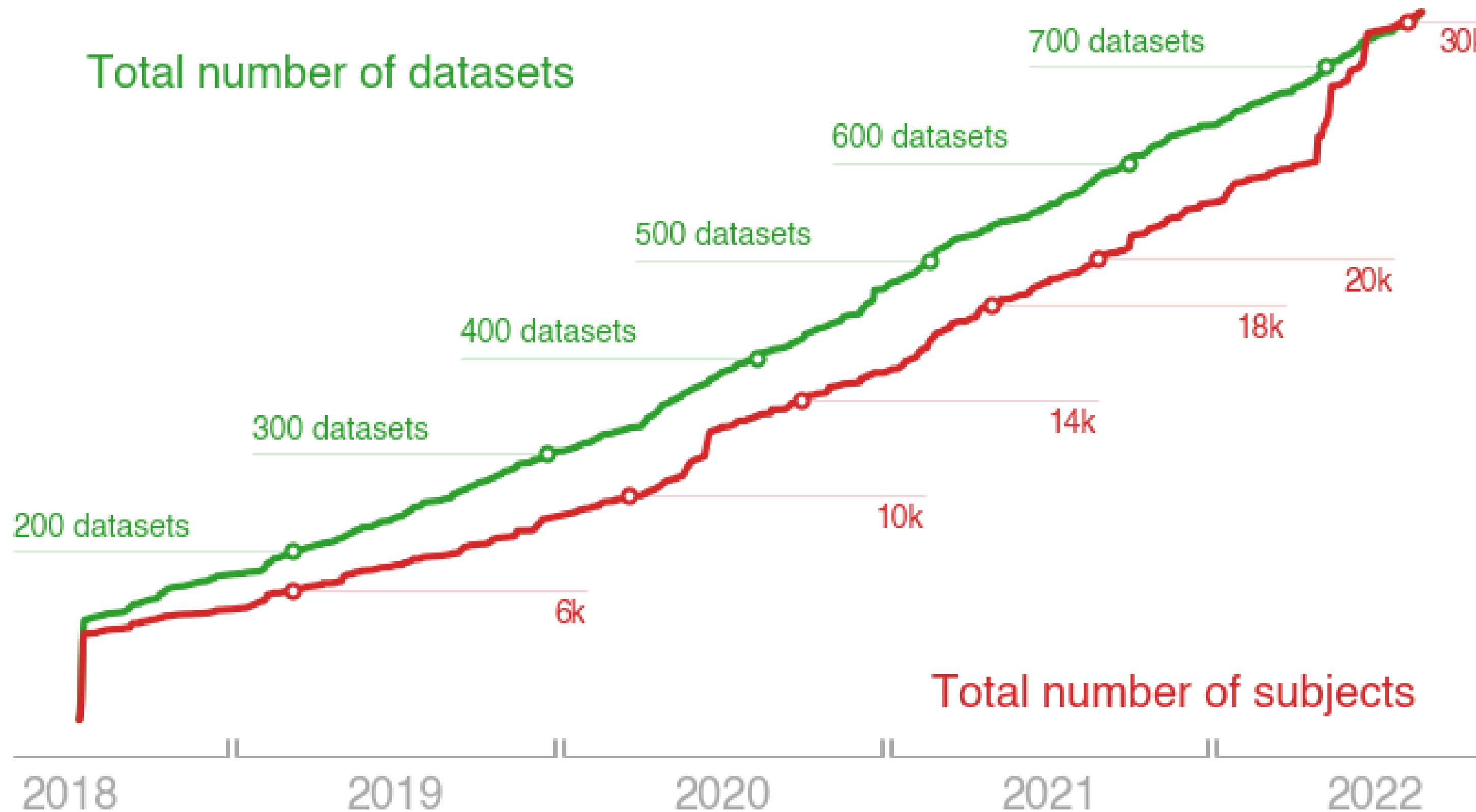
Sessions
1

Participants
13

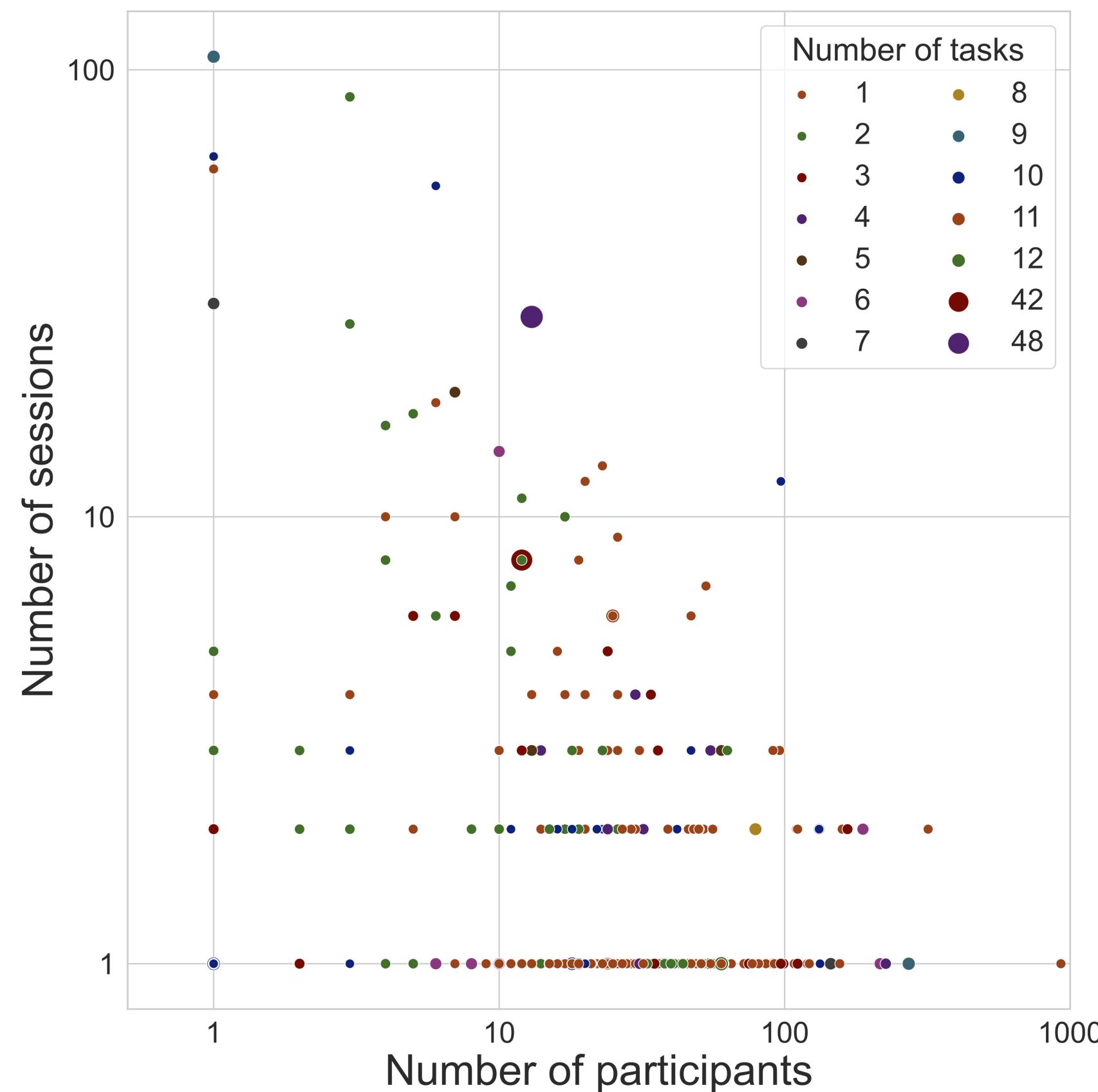
Dataset DOI
<doi:10.18112/openneuro.ds000003.v1.0.0>

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

Scholarly reuse of OpenNeuro datasets

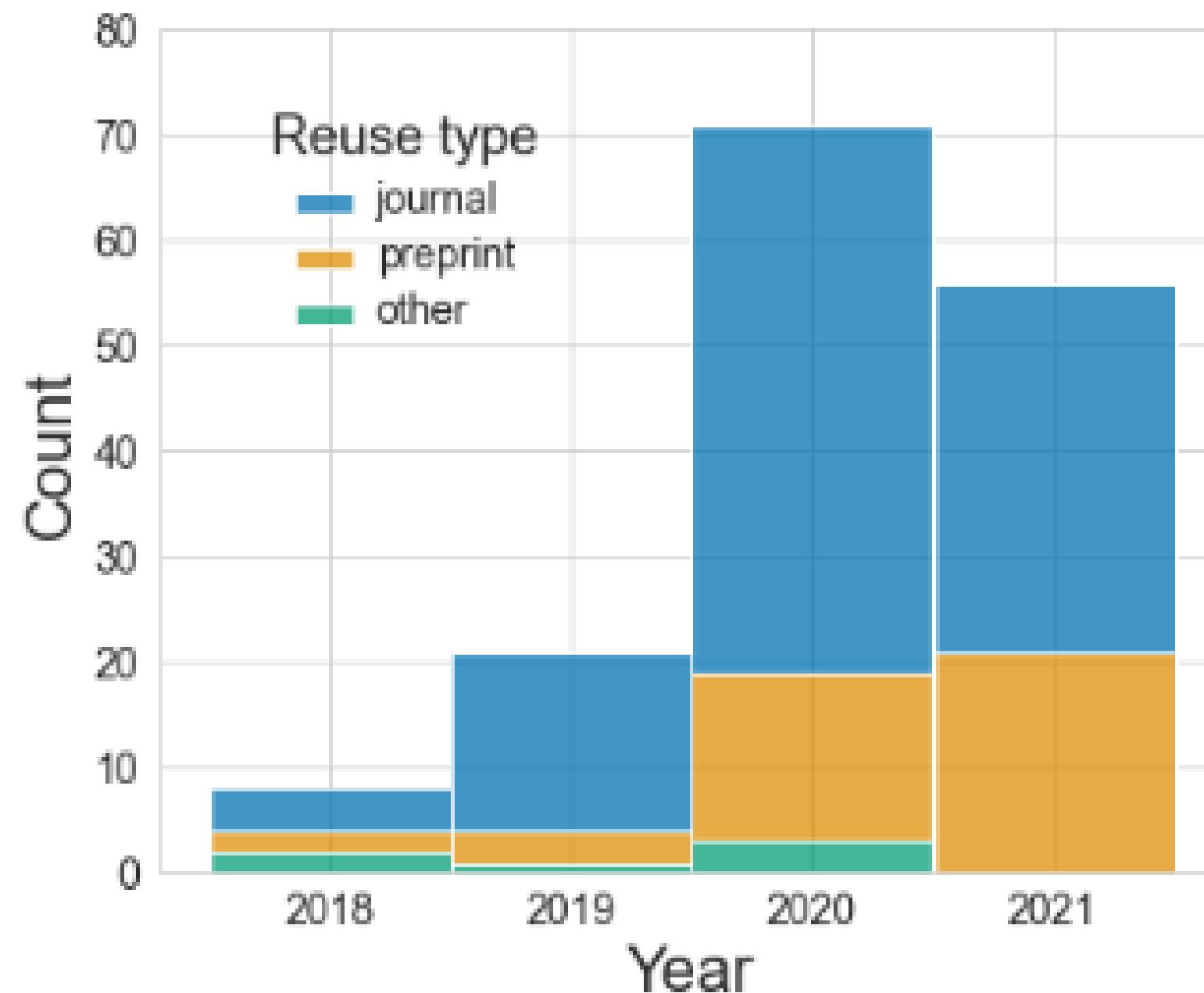


Figure 5. Published reuses of OpenNeuro datasets, split by the type of reuse. Note that the final bar includes only reuses identified through June 2021.

Processing of OpenNeuro data

brainlife.io: processing of MRI data

brainlife < Search Projects

DATASETS

PROJECTS

APPS

PUBLICATIONS

DATATYPES

PUBLIC/PROTECTED PROJECTS

HCP 3T / Diffusion
anat/t1w transform/nifti anat/t2w
hcp/freesurferpost raw dwi freesurfer
Human Connectome Project Datasets - Diffusion MRI 3T (1200-subjects-data-
1112 sub | 6880 objs (4.51 TB)

HCP 7T / Diffusion
anat/t1w dwi
Human Connectome Project Datasets - Diffusion MRI 3T (184 out of 1200-
150 sub | 300 objs (22.56 MB)

HCP 3T Retest / Diffusion
transform/nifti dwi hcp/freesurferpost
anat/t2w freesurfer anat/t1w
45 HCP 3T subjects retested
45 sub | 315 objs (101.23 GB)

O3D
networkneuro freesurfer anat/t1w wmc-deprecated
dwi dtiinit recon track/tck LiFE track/trk
O3D (Open Diffusion Data and Derivative) is a reference repository for precision
12 sub | 2708 objs (2.28 TB)

IXI
anat/t2w dwi anat/t1w
Information eXtraction from Images (EPSRC GR/S21533/02)
583 sub | 1553 objs (13.01 GB)

Human Olfactory Atlases
anat/t1w reho
This repository contains: -Probabilistic atlas of the lateral olfactory tracts based on
1 sub | 20 objs (8.72 MB)

NEMAR.org: processing of EEG/MEG data

Data Summary

Subject: sub-001

Session(s): 1, **Run(s):** 1

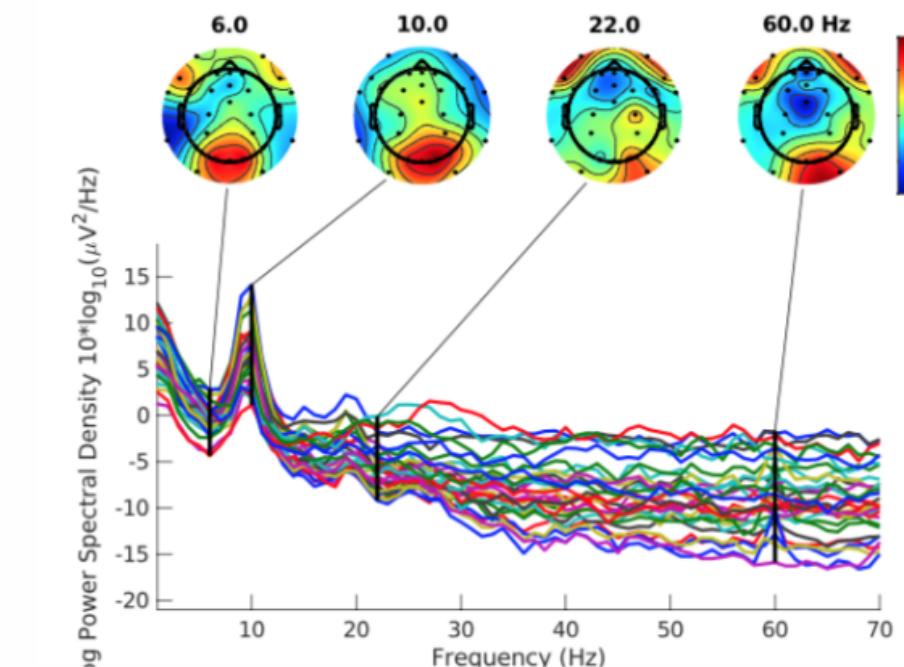
Data size: 32 channels, 298k frames

Acceptable scalp channels: 100.0% (32 of 32) i

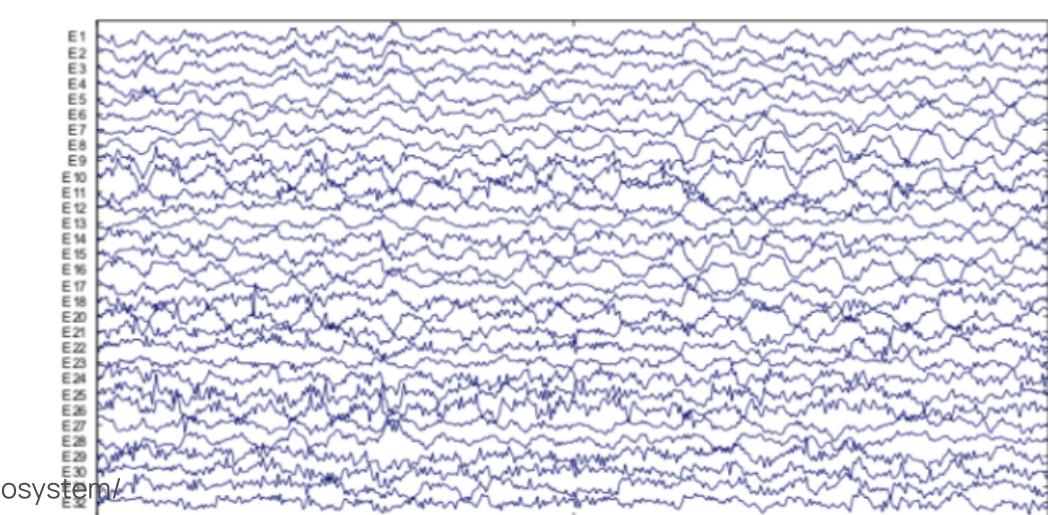
Acceptable data points channels: 90.9% (271k of 298k) i

Source quality metric based on independent component: 48.4% i

Scalp channel log spectra

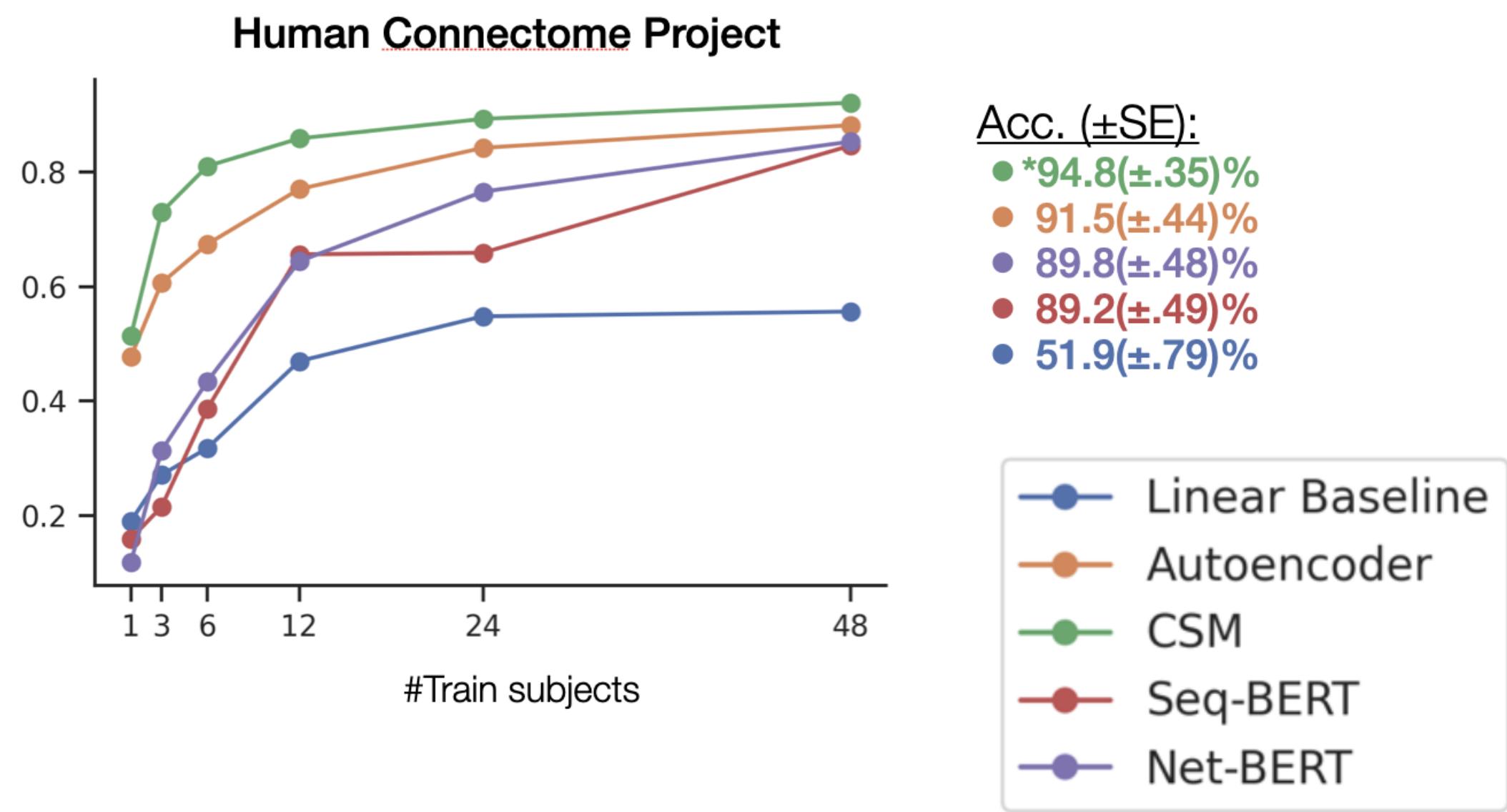


Sample scalp channel data (mid 2 seconds)



Example of OpenNeuro reuse

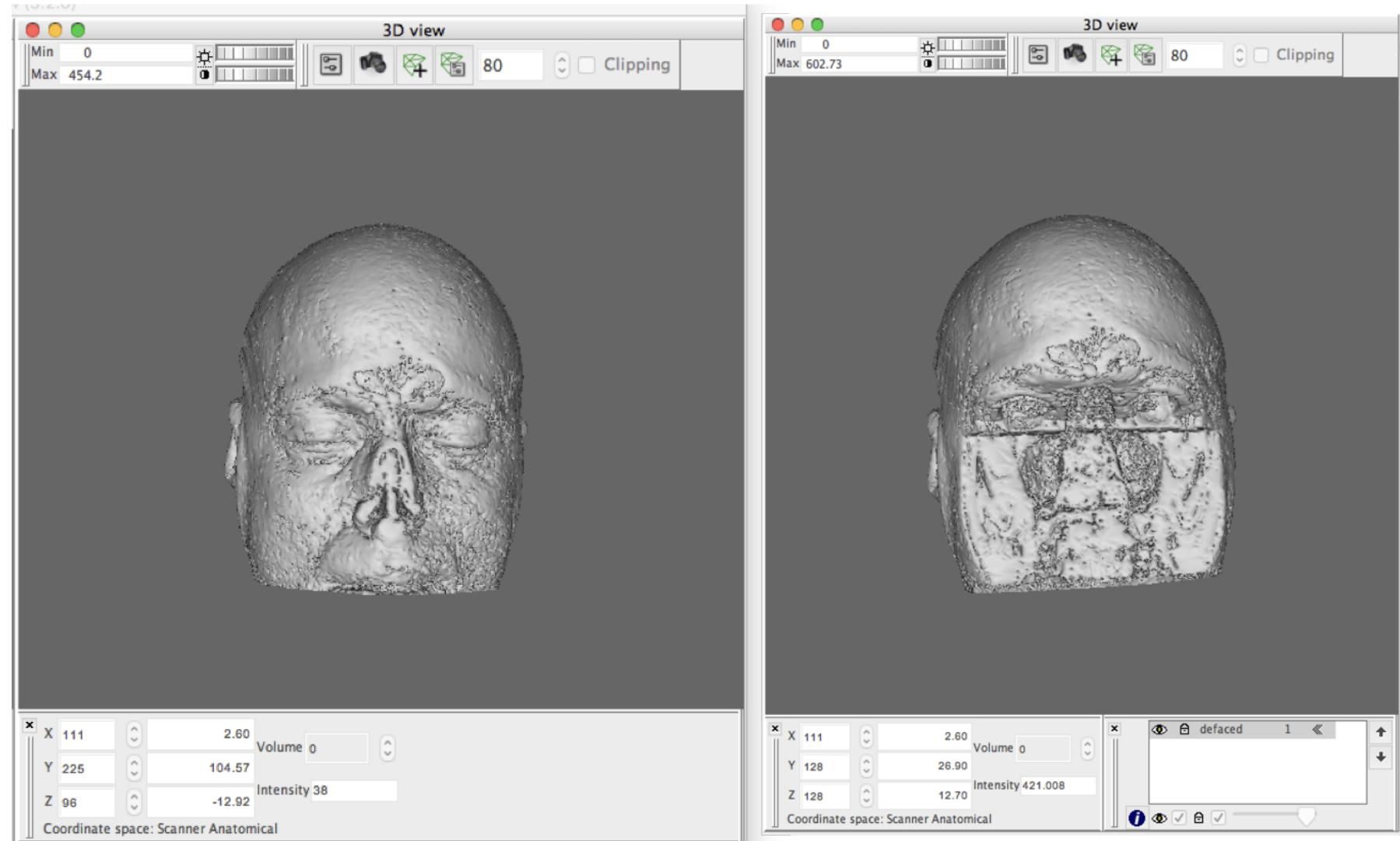
- A challenge for decoding brain activity from fMRI data is that most datasets are very small
- We used OpenNeuro to train a “foundation model”
 - A pre-trained model that can be used as a starting point for decoding models on smaller datasets
- We pre-train models on broad fMRI data from OpenNeuro: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



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

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

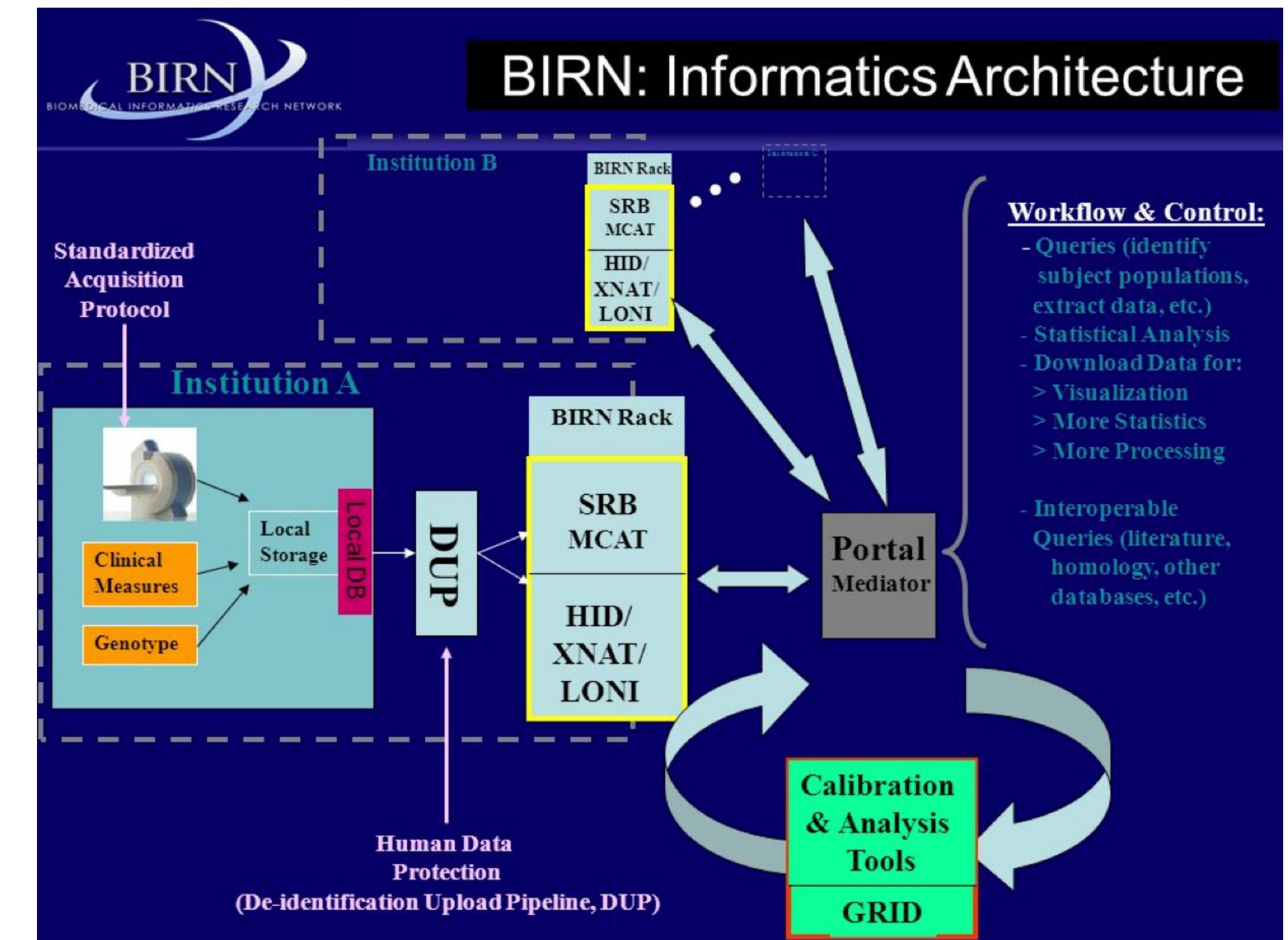
- *Data are digital end-to-end*
 - Minimizes manual steps in the process
- *Standardized file formats and data standards*
 - Makes data immediately usable by anyone
 - Reduces burden of curation and preparation
- *Demonstrated scientific utility*
- *Numerous success stories*

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

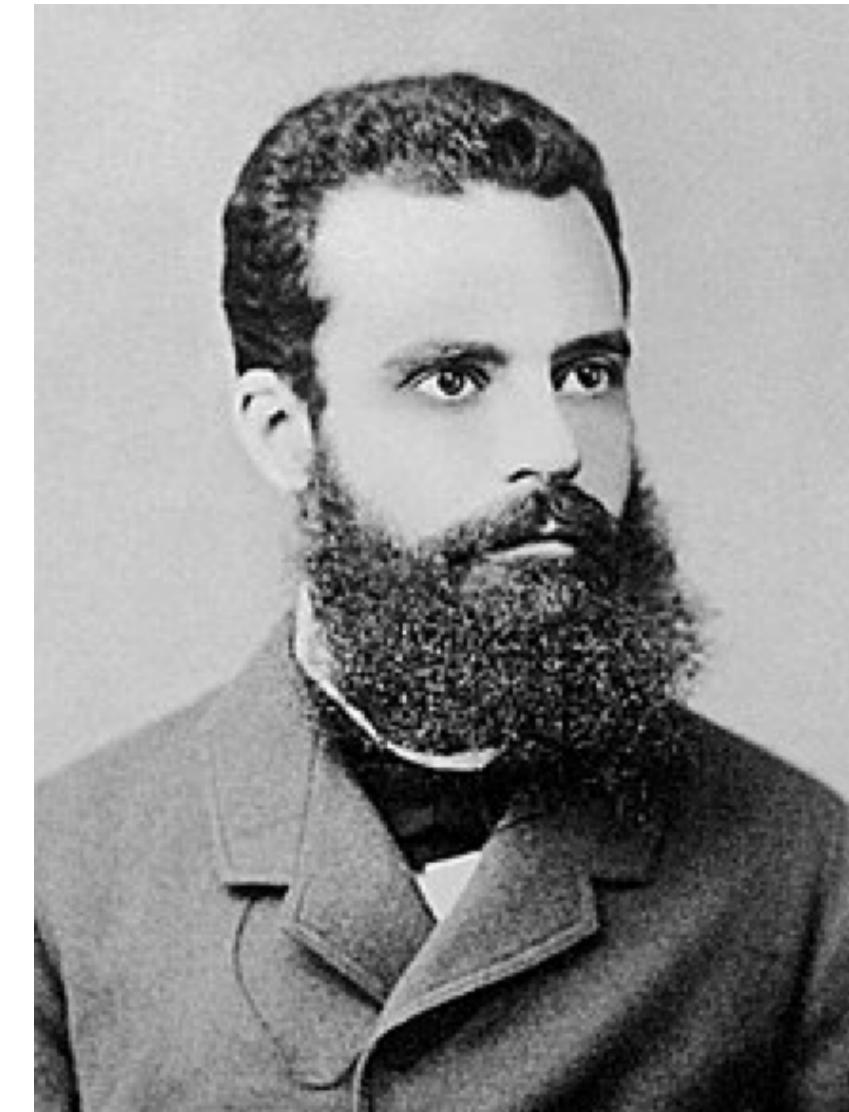
Lessons learned

- 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 model ecosystem for open science and data sharing
- 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

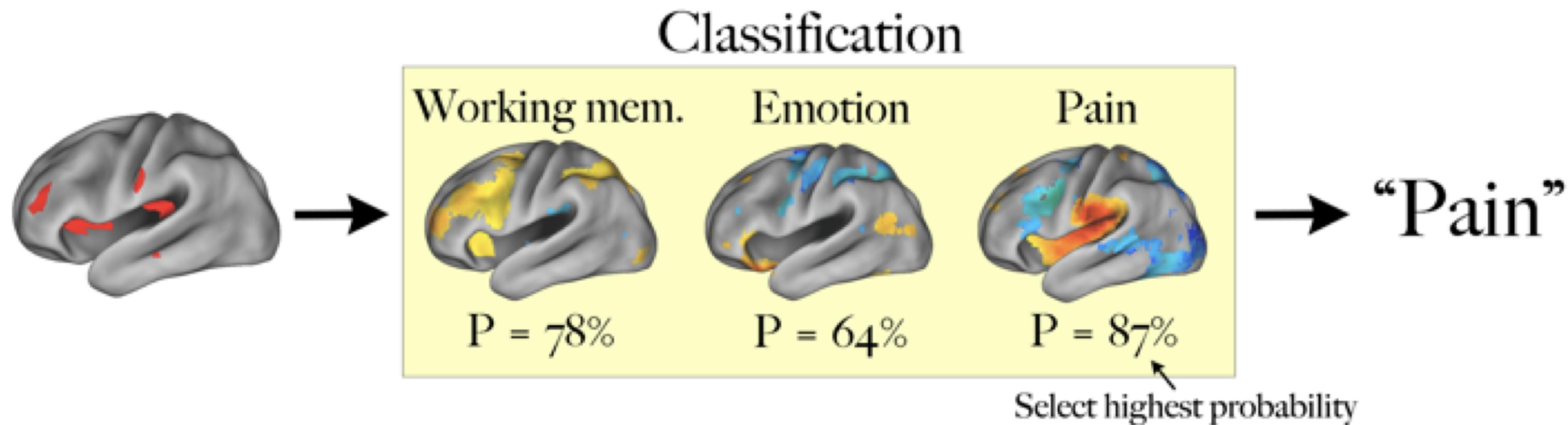


Collaborators



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.



- 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?

