

Toward an Open Science Ecosystem in Neuroimaging

Russ Poldrack
Stanford University

Conflicts of interest

- I have research grants from the US National Institutes of Health
- I am a paid consultant for the US National Institute of Mental Health

Transparency is essential for reproducibility

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

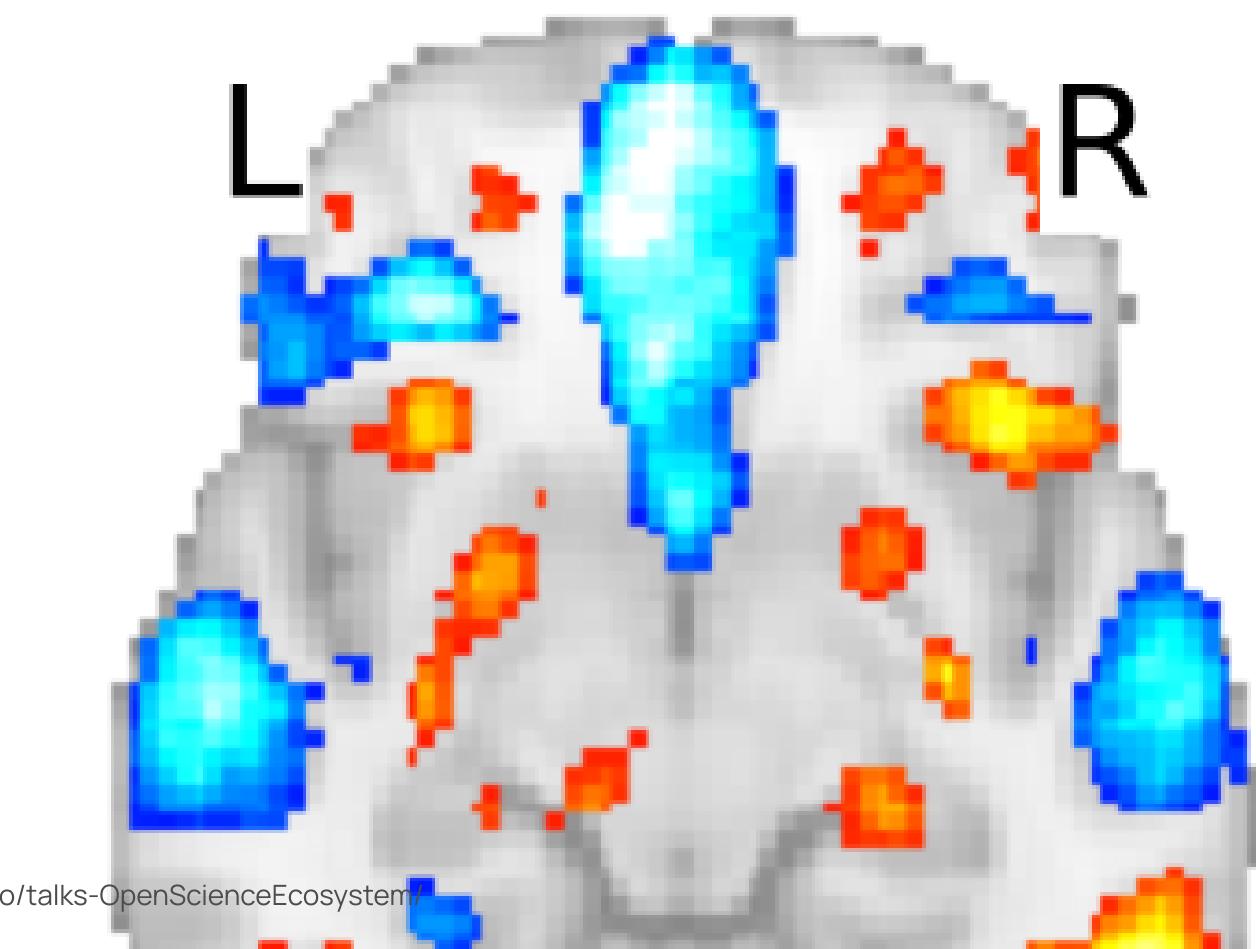
“we can distill Claerbout’s insight into a slogan:

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



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



A false start for fMRI data sharing

The fMRI Data Center
fMRIDC

SEARCH **fMRI**DC Database FOR SUBMIT

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

HOME

DATABASE

SUBMISSIONS

RESOURCES

DATA MANAGEMENT TOOL

HELP

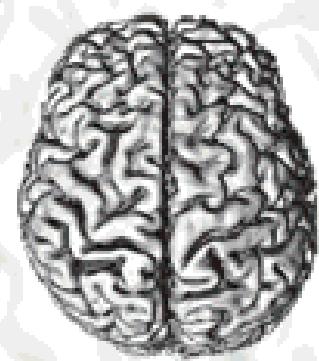
ABOUT US

[How do I get started?](#)
Answers to questions commonly posed by first-time visitors.

[Q&A about fMRI](#)
A comprehensive list of frequently asked questions about the fMRI DC.

[Available Datasets](#)
A list of datasets currently available.

[Information for Authors](#)

 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

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**

6 / 58

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?

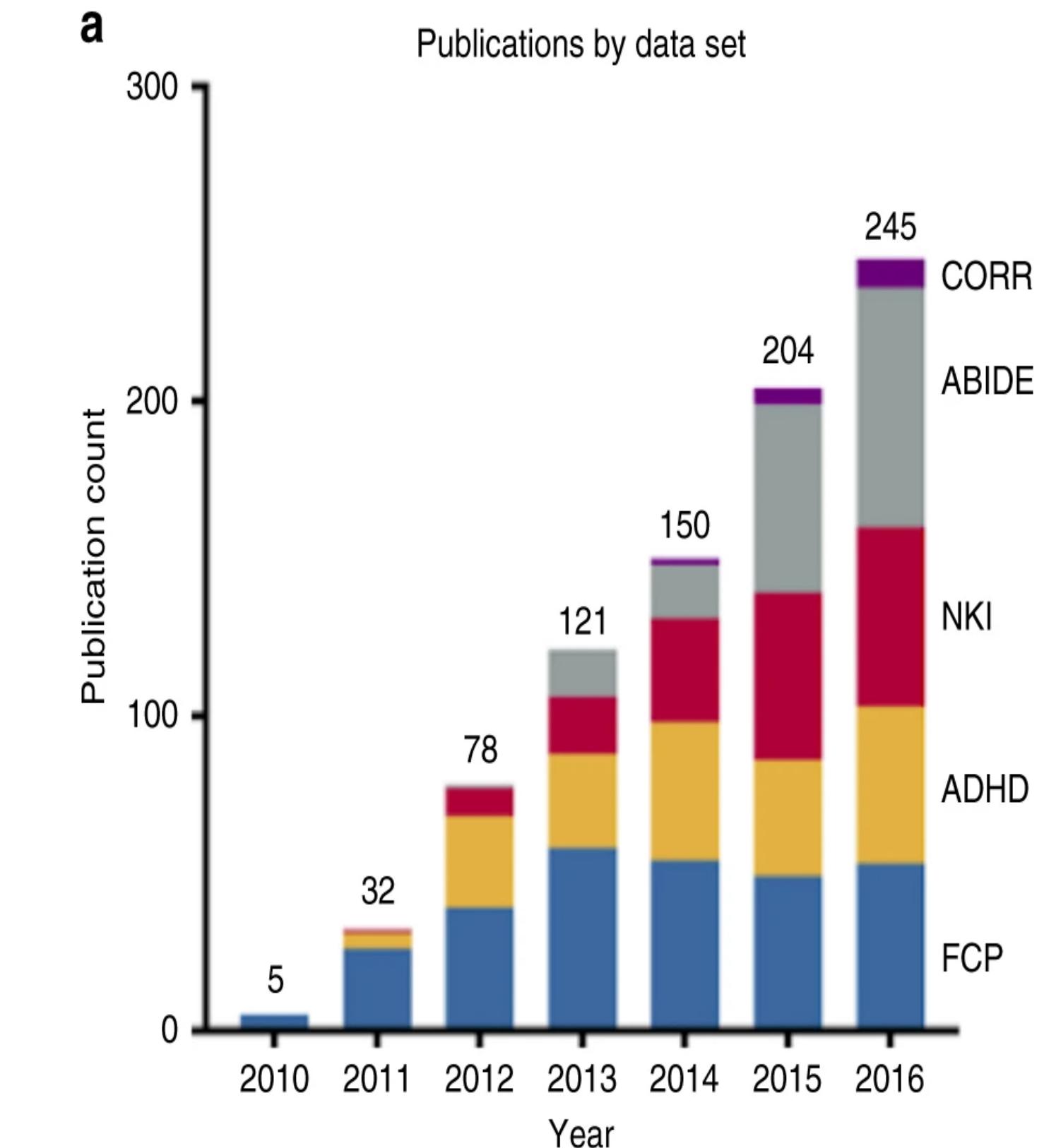
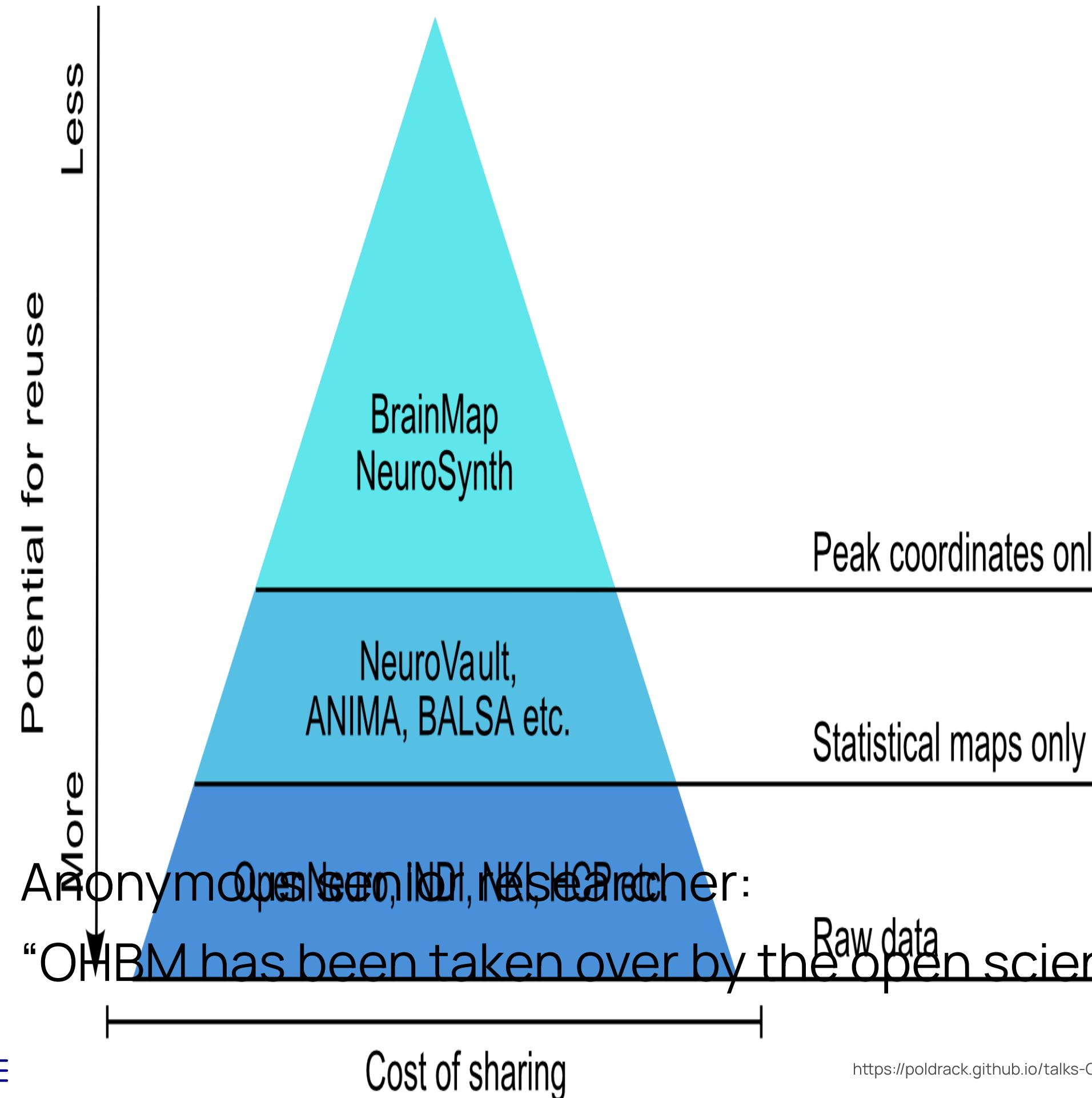
This letter comes from a group of scientists who are publishing papers using fMRI to understand the links between brain and behavior. We are writing in reaction to the recent announcement of the creation of the National fMRI Data Center (www.fmridc.org). In the letter announcing the creation of the center, 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. ...
<https://poldrack.github.io/talks-OpenScienceEcosystem/>
We are particularly concerned with any journal's

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}

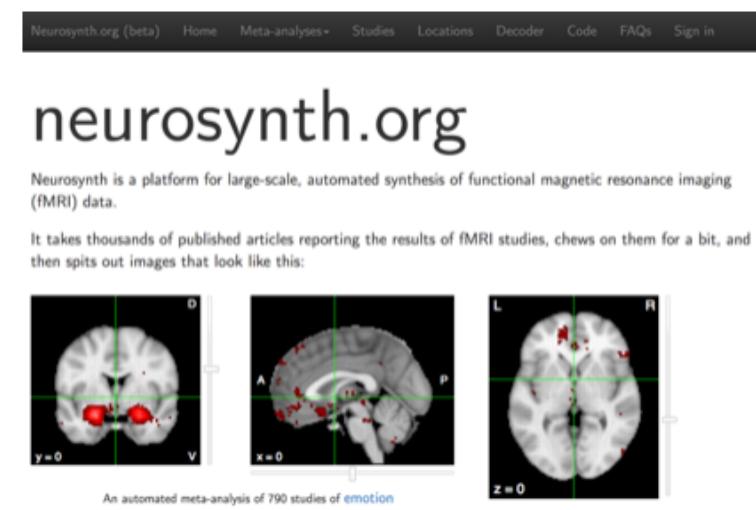
Data sharing is becoming the norm in neuroimaging



Milham et al., *Nature Communications*, 2018

An open ecosystem for retrospective data sharing

Breadth



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

- 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



- 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

Neurosynth: Sharing activation coordinates

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

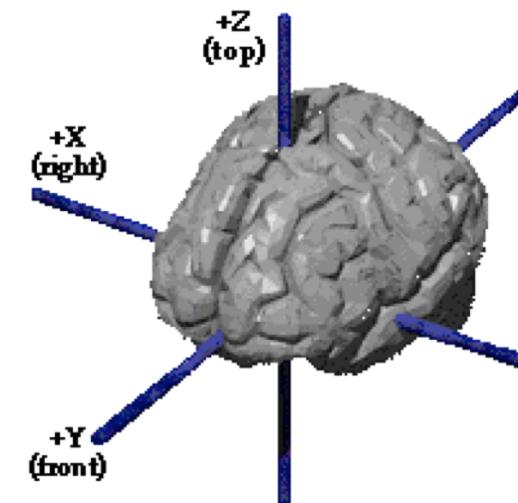


Table 1
Regions that showed a condition × time interaction in the ANOVA analysis

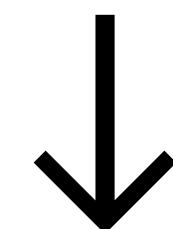
No.	Region	Hemisphere	BA	x	y	z	mm ³
1	Middle/superior temporal gyrus	L	21/22/37	-52	-54	9	13257
2	Inferior frontal gyrus	L	45/46/9	-49	26	6	2781
3	Posterior cerebellum	L		-19	-79	-38	2214
4	Dorsomedial PFC	L	9/8	-11	42	47	3051
5	Left anterior PFC	L	10	-37	49	15	2025
6	Inferior parietal cortex	L	40/7	-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		5	-53	29	405

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

Yarkoni et al. 2011 Nature Methods
against manual database (SumsDB)

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



working memory

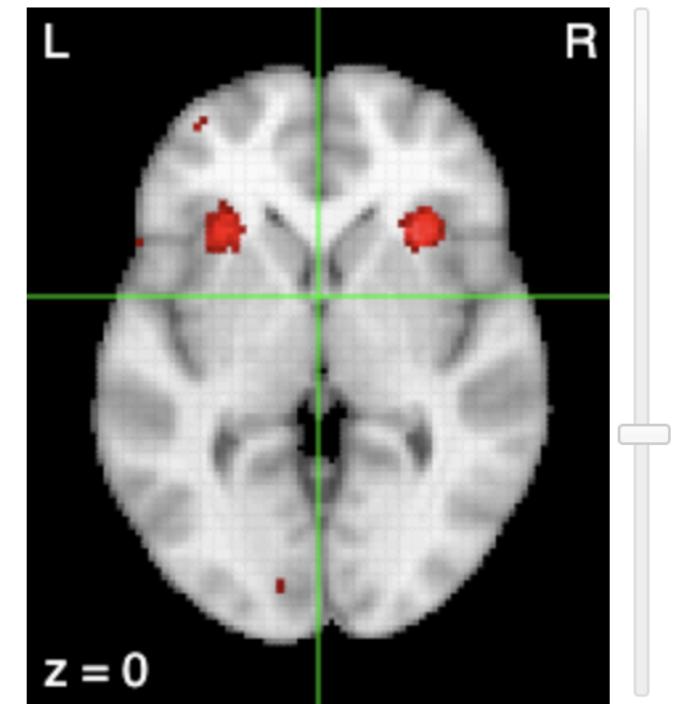
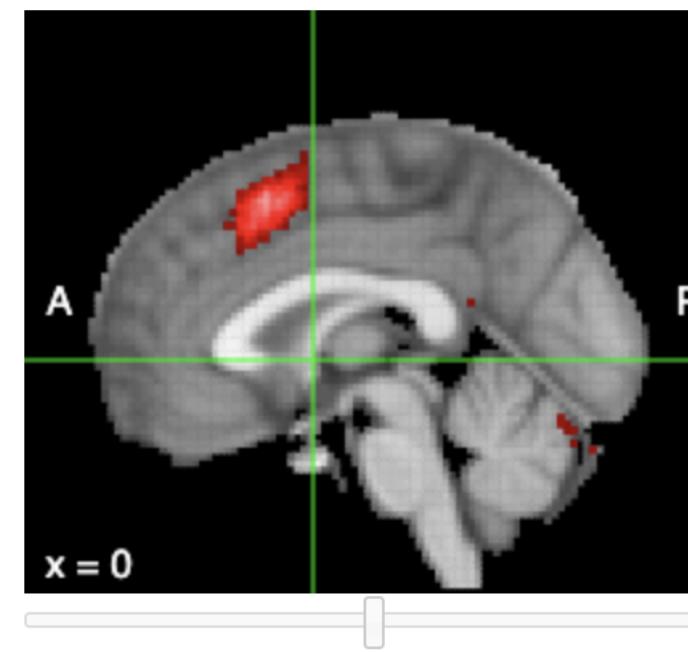
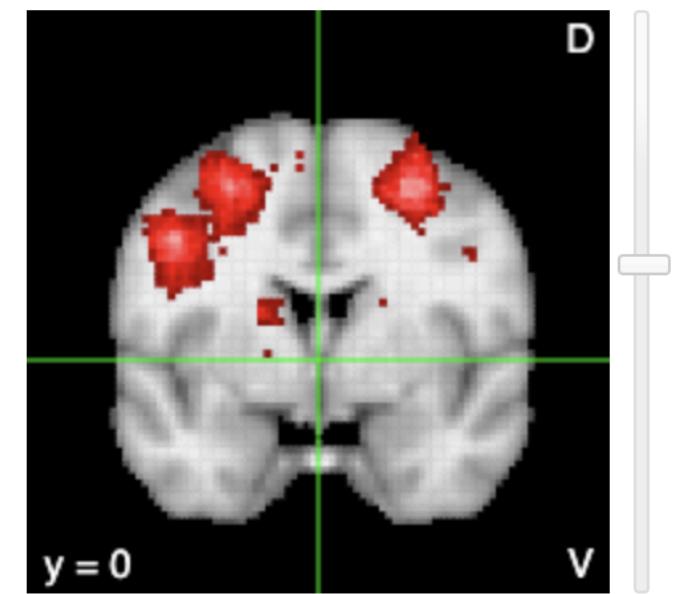
An automated meta-analysis of 1091 studies

Search for another term:

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 ▼

Crosshairs

Positive/Negative:

positive ▼

Pan/zoom

Labels

Thresholds:

0 0

Opacity:

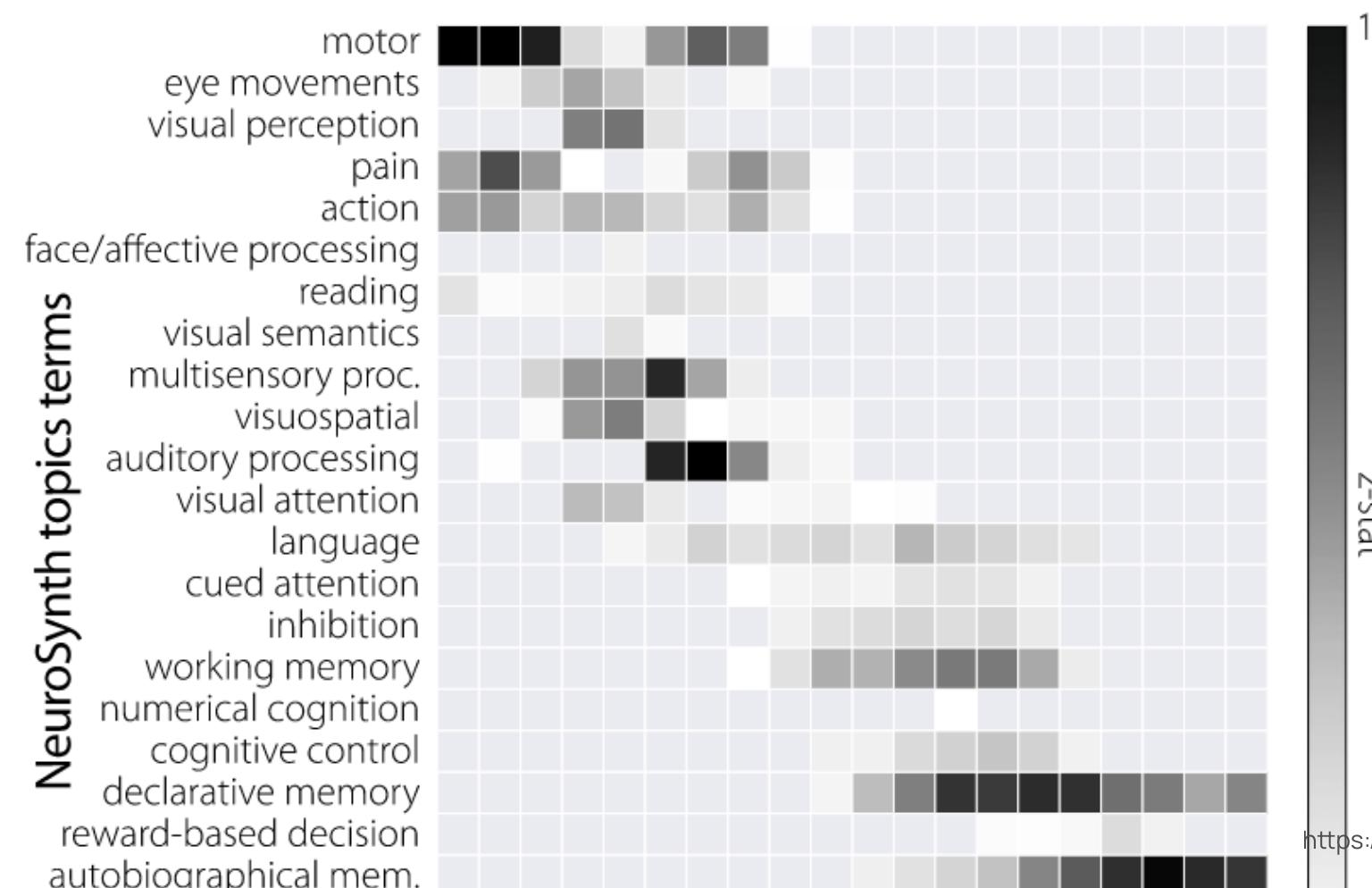
1

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}

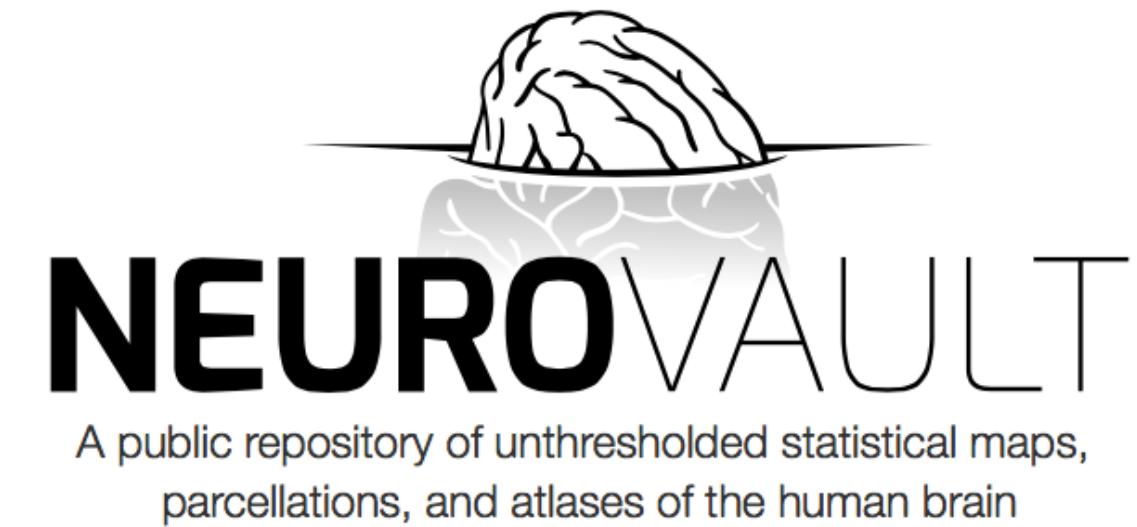
12574–12579 | PNAS | November 1, 2016 | vol. 113 | no. 44



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



- 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

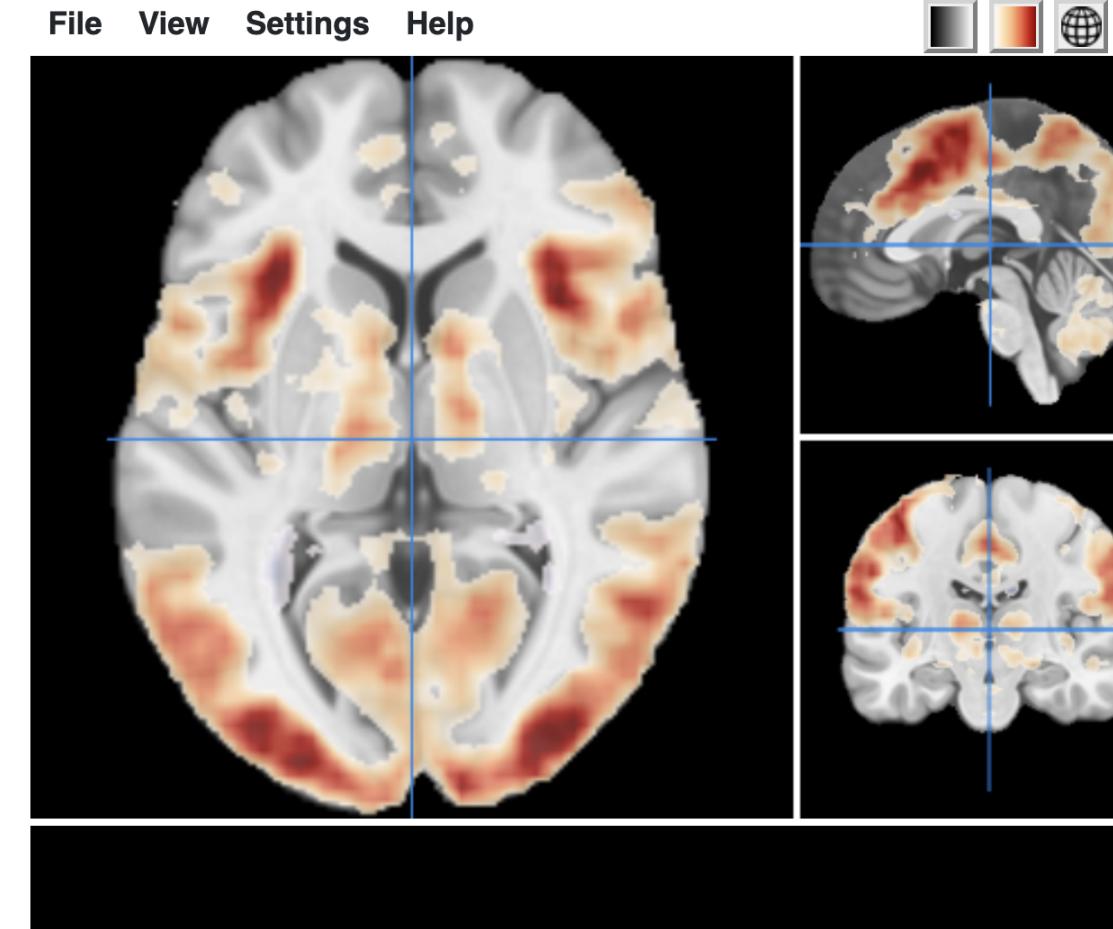
Preprocessed Consortium for Neuropsychiatric Phenomics dataset

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:

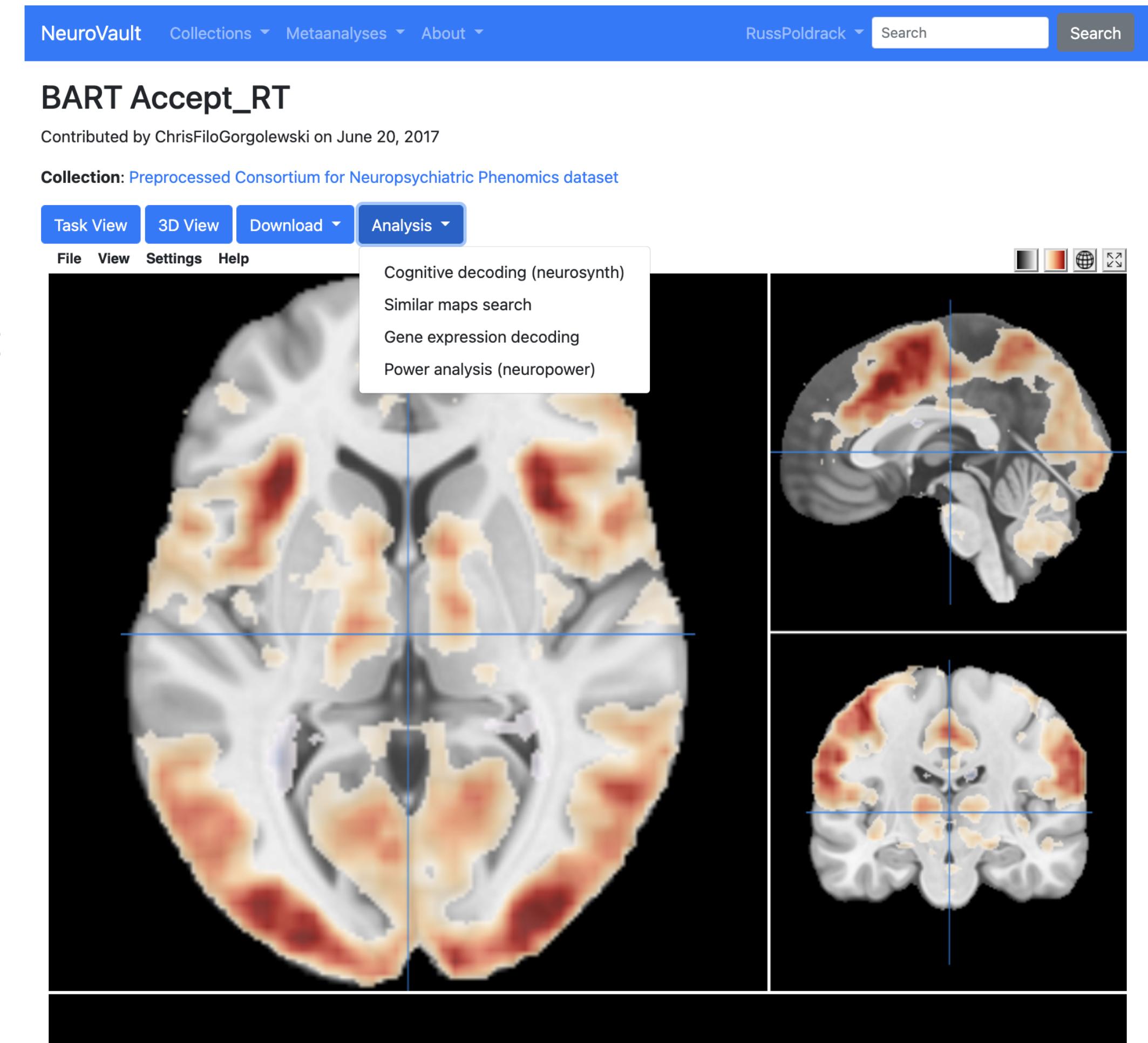
<https://identifiers.org/neurovault.collection:2606>

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

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

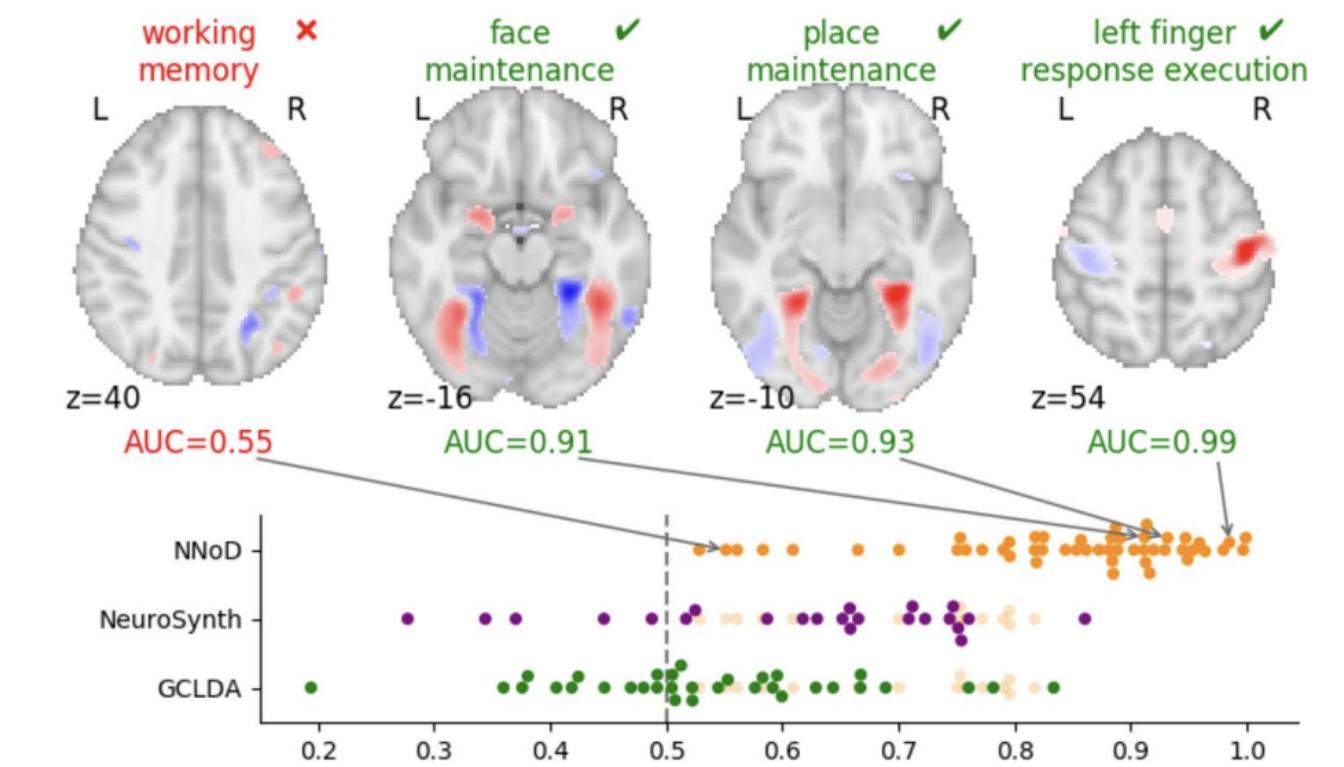
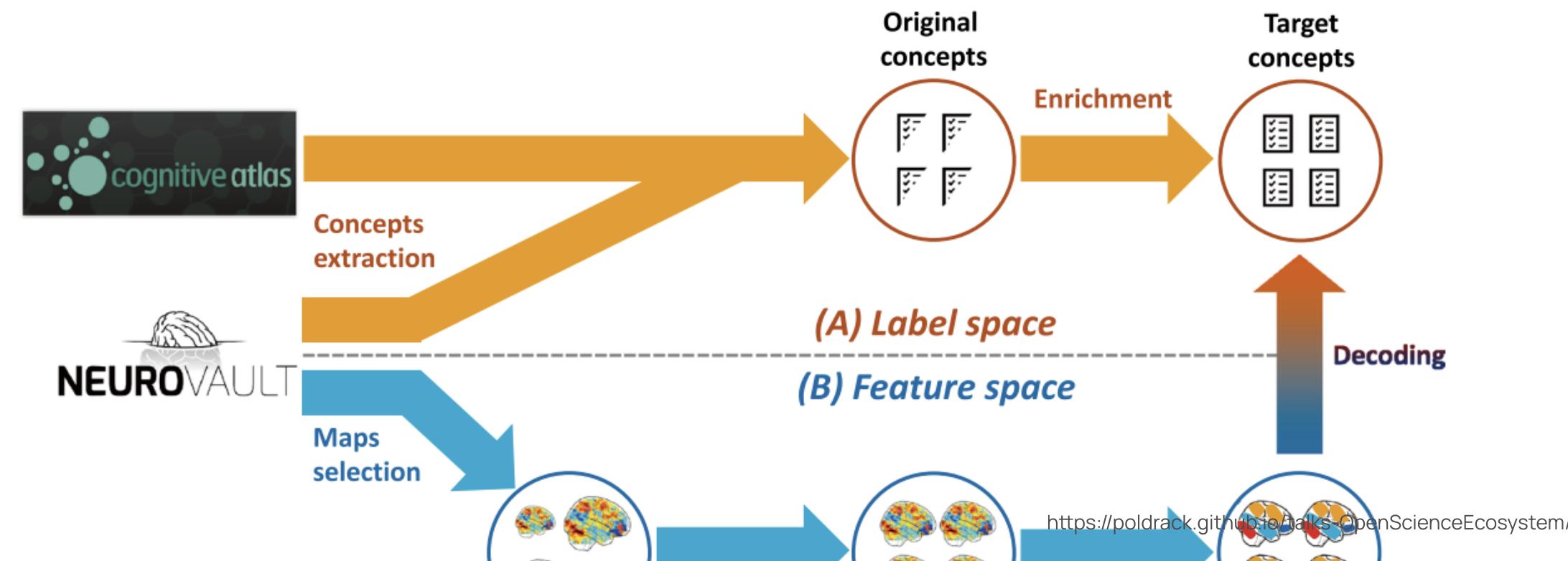


Example of Neurovault usage

Comprehensive decoding mental processes from Web repositories of functional brain images

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

Scientific Reports | (2022) 12:7050





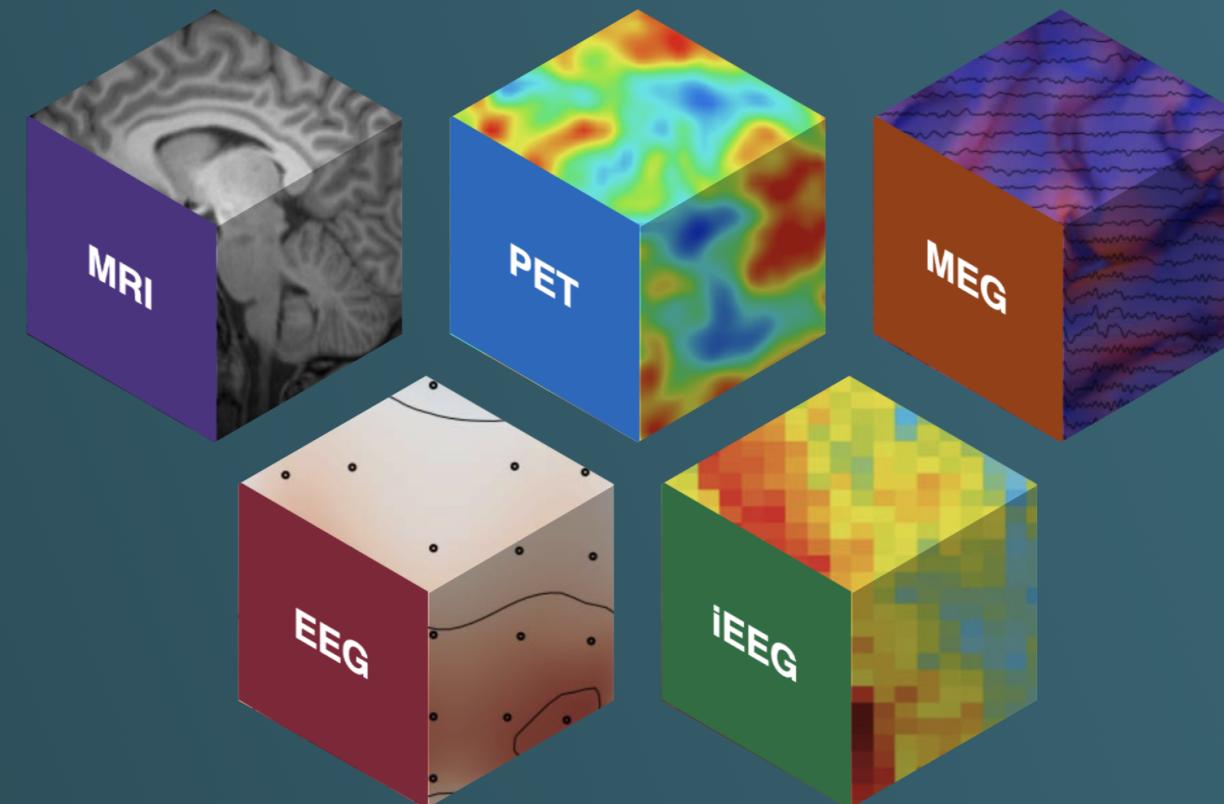
SEARCH SUPPORT FAQ

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

Google

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](#)

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?

Brain Imaging Data Structure (BIDS)

- A community-based open standard for neuroimaging data
 - A file organization standard
 - A metadata standard



The development of BIDS

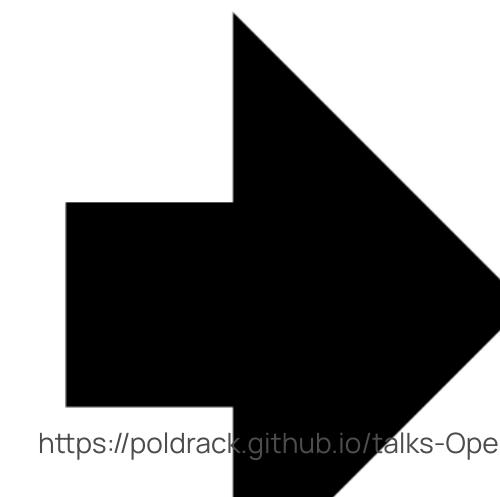
- January 2015
 - Initial stakeholder meeting at Stanford (funded by INCF)
 - Initiated development of a draft standard
- September 2015
 - Draft standard posted to BIDS web site with 22 example datasets
 - Solicited feedback from community
- June 2016

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*

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



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

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.

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)

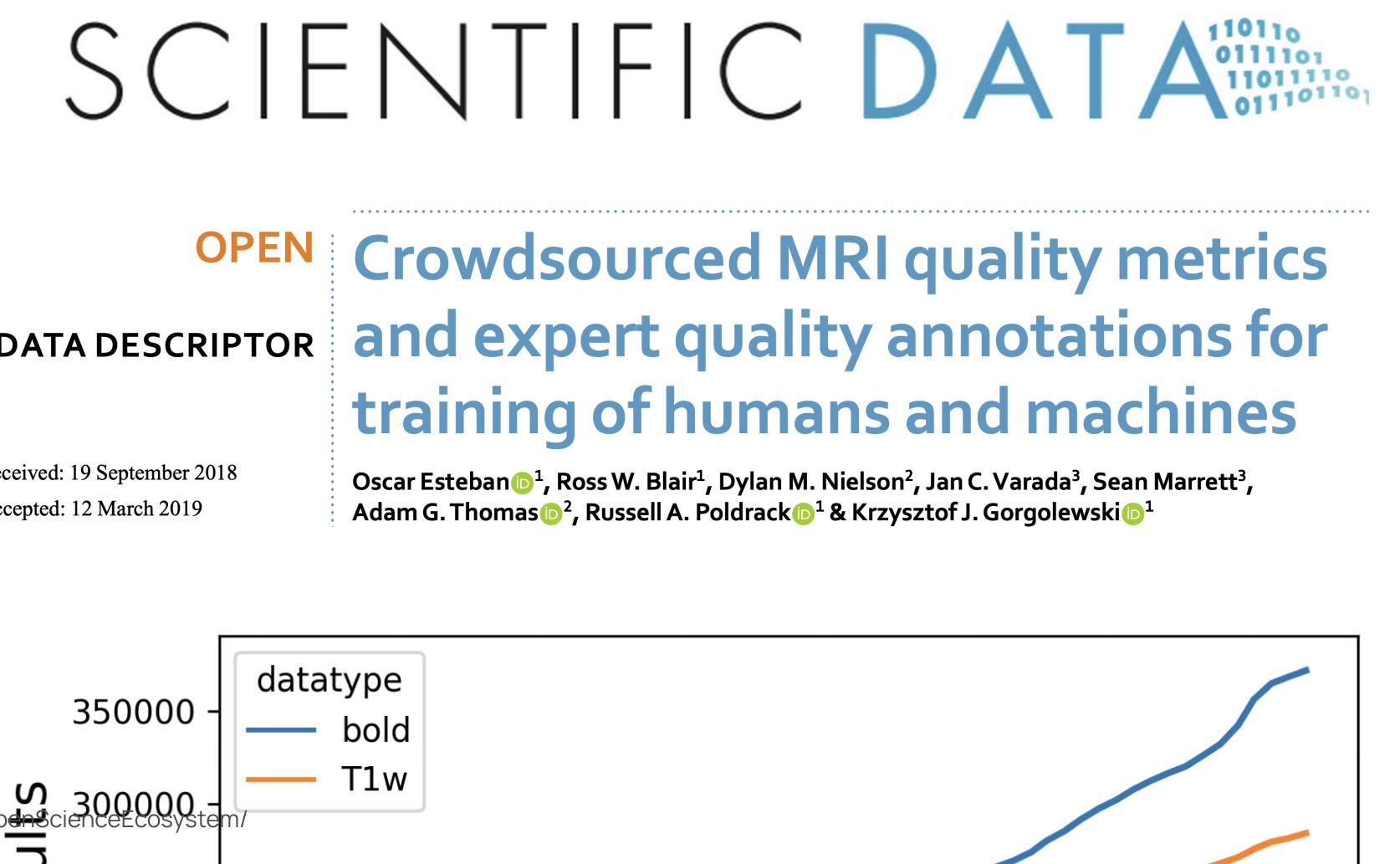
BEP006 Electroencephalography (EEG)

BEP007 Hierarchical Event Descriptor (HED)
Tags

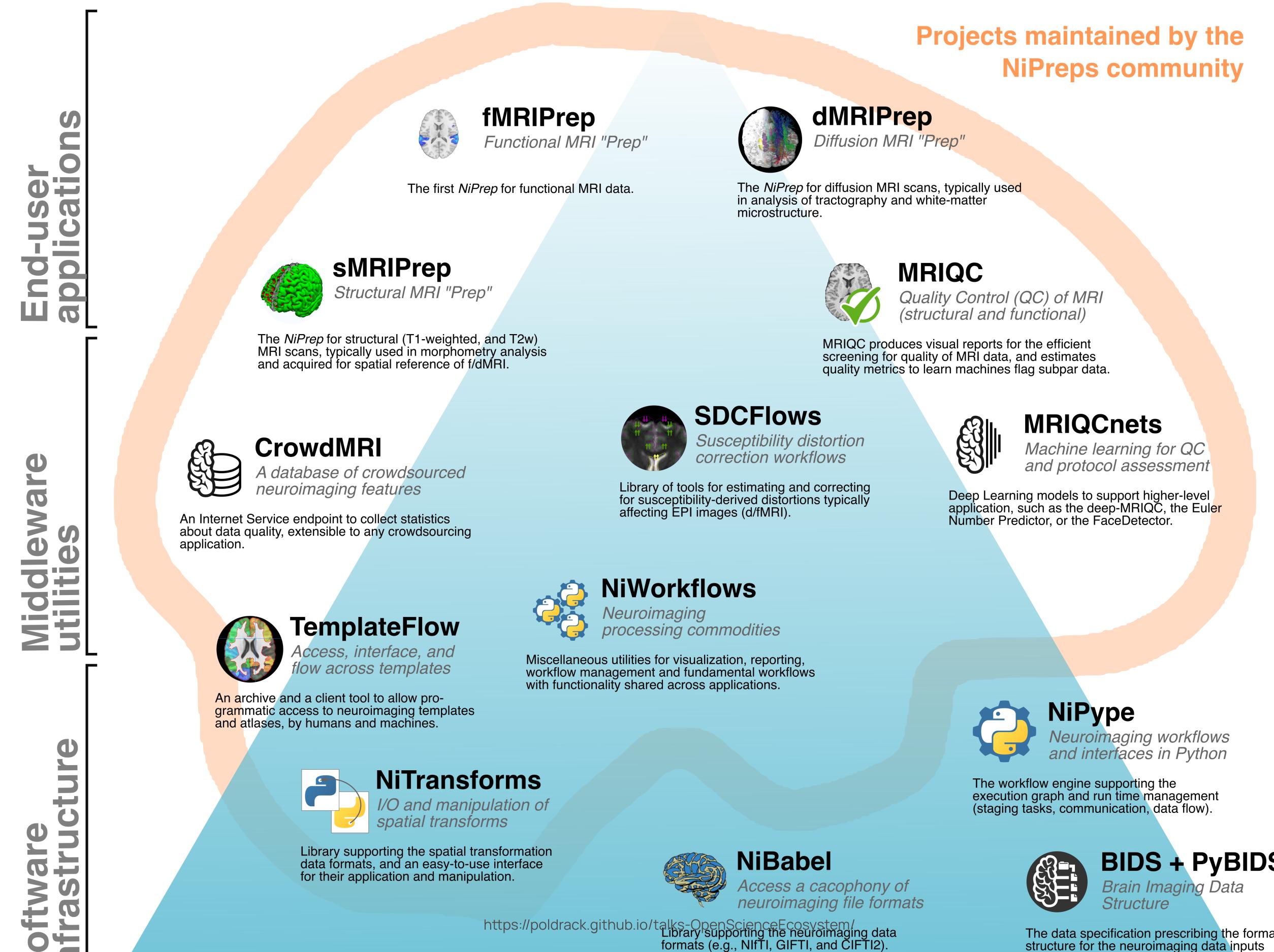
BEP008 Magnetoencephalography (MEG)

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/>



BIDS enables a growing open-source software ecosystem



The screenshot shows the homepage of the OpenNEURO platform. At the top left is the logo 'OpenNEURO' with a brain network icon. To its right are links for 'SEARCH', 'SUPPORT', and 'FAQ'. On the far right is a 'Sign in' button. The main content area features a dark teal background with several large, colorful 3D cubes representing different neuroimaging modalities: MRI (purple), PET (blue), MEG (orange), EEG (red), and iEEG (green). Below these cubes, there's a search bar containing the text 'reading' and a magnifying glass icon. To the left of the search bar is a dropdown menu labeled 'Browse by Modalities'. Above the search bar, the text 'A free and open platform for validating and sharing BIDS-compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#) data' is displayed. On the left side, there are two statistics: '29,064 Participants' and '761 Public Datasets'. At the bottom, there are links for 'SIGN IN', 'Google', and 'ORCID', along with a blue navigation bar.

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

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



Search All Datasets

Keywords ?



KEYWORD:
reading X

reading X

CLEAR ALL

SEARCH SUPPORT FAQ

Sign in

These filters return **194** datasets:

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



^A

MODALITIES:

MRI

Modalities

MRI

PET

EEG

iEEG

MEG

TASKS:

read task

rest

OPENNEURO ACCESSION NUMBER: [ds003974](#)

SESSIONS: 1

PARTICIPANTS: 52

PARTICIPANTS' AGES: N/A

SIZE: 46.67GB

FILES: 893

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

The Reading Brain Project L2 Adults



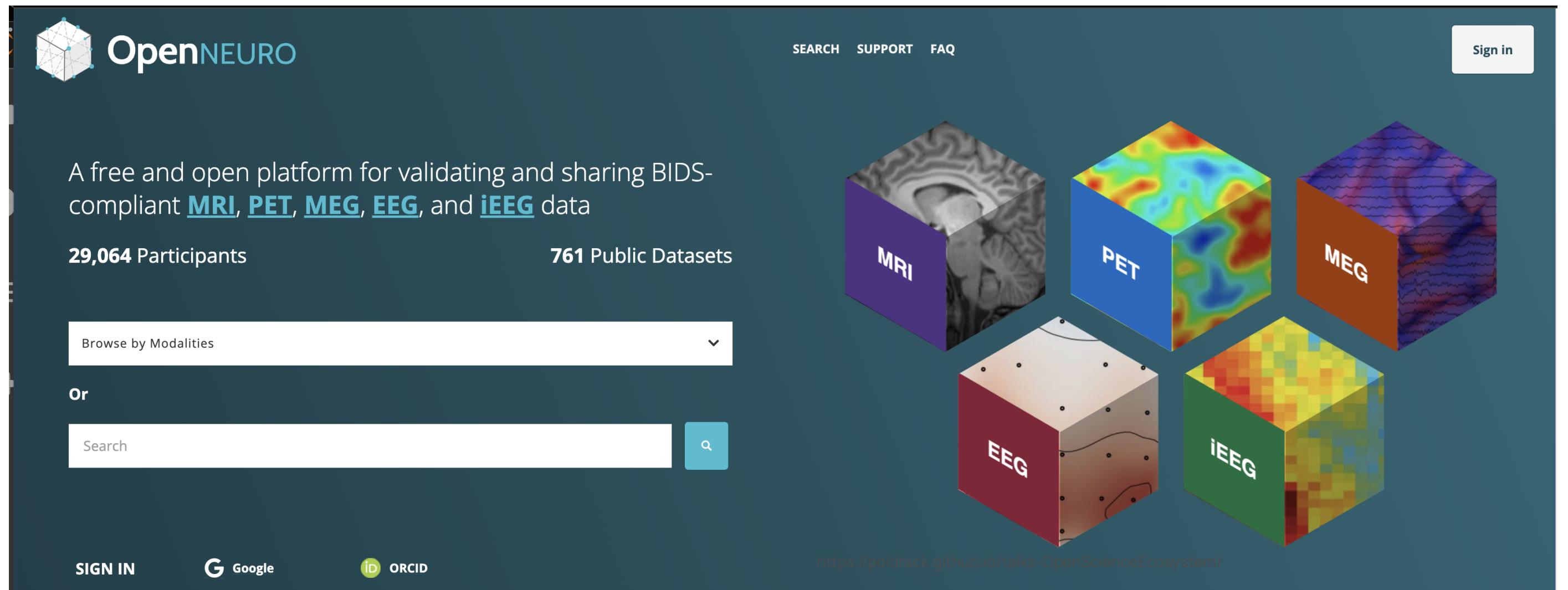
^A

Each shared dataset
is versioned and

receives a persistent
identifier (DOI)

The screenshot shows the OpenNeuro dataset page for 'The Reading Brain Project L2 Adults'. At the top, there's a navigation bar with 'SEARCH', 'SUPPORT', 'FAQ', and a 'Sign in' button. Below the header, the project title 'The Reading Brain Project L2 Adults' is displayed, along with a 'MRI' icon and social sharing buttons for 'Follow' (2) and 'Bookmark' (2). The main content area includes a 'BIDS Validation' section showing '4 WARNINGS' and a 'Valid' status, along with links for 'Files', 'Download', and 'Metadata'. A large text box contains the 'README' file, which provides a detailed description of the dataset, mentioning 56 participants, two sessions of testing (MRI and behavioral), and various tests like ANT, GSRT, LNS, PPVT-4, and Tower of Hanoi. It also notes that the data is from native speakers of Mandarin Chinese living in the United States and Beijing, China. Below the README is a note about an updated methodology document. To the right of the README, there are sections for '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), 'Tasks' (read task, rest), and 'Uploaded by' (https://poldrack.github.io/talks-OpenScienceEcosystem/).

Any valid BIDS dataset can be shared via OpenNeuro



The image shows the homepage of the OpenNEURO platform. The header features the "OpenNEURO" logo with a brain cube icon. Navigation links include "SEARCH", "SUPPORT", "FAQ", and a "Sign in" button. Key statistics are displayed: "29,064 Participants" and "761 Public Datasets". A "Browse by Modalities" dropdown menu is open, showing options like "MRI", "PET", "MEG", "EEG", and "iEEG". Below this is a search bar with a magnifying glass icon. The main visual area displays five 3D cubes, each representing a different modality: MRI (grey brain scan), PET (color-coded heatmap), MEG (color-coded waveforms), EEG (color-coded waveforms on a grid), and iEEG (color-coded waveforms on a grid). The URL at the bottom is <https://poldrack.github.io/talks-OpenScienceEcosystem/>.

OpenNEURO

SEARCH SUPPORT FAQ Sign in

29,064 Participants 761 Public Datasets

Browse by Modalities

Or

Search

SIGN IN G Google ORCID

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

The screenshot shows the OpenNEURO homepage with a dark background. In the center, a white sign-in modal is displayed. The modal features the OpenNEURO logo at the top left, followed by a "Sign in" button. Below the button are two dark blue rectangular buttons: one with a white "G" icon labeled "Google" and another with a white "ID" icon labeled "ORCID". To the right of these buttons is a small link "What is this?". At the top right of the modal is a small "x" icon. In the bottom right corner of the main page, there is a "Sign in" button.

OpenNEURO

A free and open platform for validating and sharing 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

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

The screenshot shows the homepage of the OpenNEURO platform. At the top, there is a navigation bar with links for SEARCH, SUPPORT, FAQ, UPLOAD (underlined), and My Account. Below the navigation bar, there is a large banner featuring five 3D cubes, each representing a different neuroimaging modality: MRI (purple), PET (blue), MEG (orange), EEG (red), and iEEG (green). To the left of the banner, there is a summary of the platform's statistics: 29,064 Participants and 761 Public Datasets. Below these statistics, there is a dropdown menu labeled "Browse by Modalities" and a search bar with a magnifying glass icon. The URL at the bottom of the page is <https://poldrack.github.io/talks-OpenScienceEcosystem/>.

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

SEARCH SUPPORT FAQ UPLOAD

My Account

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



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 Edit

This dataset is public.

BIDS Validation View

Files Share Versioning Admin

How to Download

Download with your browser

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](#)

<https://poldrack.github.io/talks-OpenScienceEcosystem/>
Not all subjects/sessions/runs have the same scanning parameters.

My Account

Follow 1 Bookmark 2

Them public.

uro Accession Number
03

Russell A. Poldrack

lit

e Modalities

39 / 58

The screenshot shows the OpenNEURO dataset upload interface. The main header says "OpenNEURO" with a logo. Below it, a card for "Rhyme judgment" (MRI study) has options to "Edit" and "This dataset is public". A "BIDS Validation" dropdown is open, showing "Files", "Share", "Versioning", and "Admin". A "How to Download" section includes "Download with your browser" and "Steps". The central part is a modal titled "Upload Dataset" with four tabs: "Step 1: Select Files", "Step 2: Validation", "Step 3: Metadata" (which is active), and "Step 4: Accept Terms". A note says incomplete fields will make search harder. Below are input fields for "DOI of papers from the source data lab" (with a "Papers that were published from the Lab that collected this dataset" link), "Species", "Study Type", "Domain Studied", and "Number of Trials (if applicable)". The right side of the modal shows a dark sidebar with "My Account", "Follow 1", "Bookmark 2", and other user stats.

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.

Steps

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)

My Account

Follow 1 Bookmark 2

Dataset public.

Open Access Number

03

Russell A. Poldrack

lit

Modalities

OpenNEURO

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 where you want to save the dataset. You can also choose to download specific files or entire sub-directories.

Download with curl

If you prefer to download the dataset using the command line, you can use the curl command:

```
curl -L https://openneuro.org/api/v1/datasets/03/zip
```

Replace `03` with the dataset ID of the dataset you want to download.

Download with Python

You can also download the dataset using Python:

```
import requests  
url = "https://openneuro.org/api/v1/datasets/03/zip"  
response = requests.get(url)  
with open("dataset.zip", "wb") as f:  
    f.write(response.content)
```

Replace `03` with the dataset ID.

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

My Account

Follow 1 Bookmark 2

Dataset Accession Number 03

Russell A. Poldrack

Data Modalities

 OpenNEURO

SEARCH SUPPORT FAQ UPLOAD [My Account](#)

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

OpenNeuro Accession Number
ds004338

Authors
Xue, G., Russell A. Poldrack [Edit](#)

Available Modalities
MRI

Version

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



OpenNEURO

SEARCH SUPPORT FAQ UPLOAD My Account

MRI Rhyme judgment

Follow 1 Bookmark 2

BIDS Validation ▾

1 ERROR

! Invalid

brainlife.io

Clone ▾

OpenNeuro Accession Number

ds000003

Authors

Xue, G., Russell A. Poldrack

Available Modalities

MRI

Versions

1.0.0

Created: 2020-05-14

Versions ▾

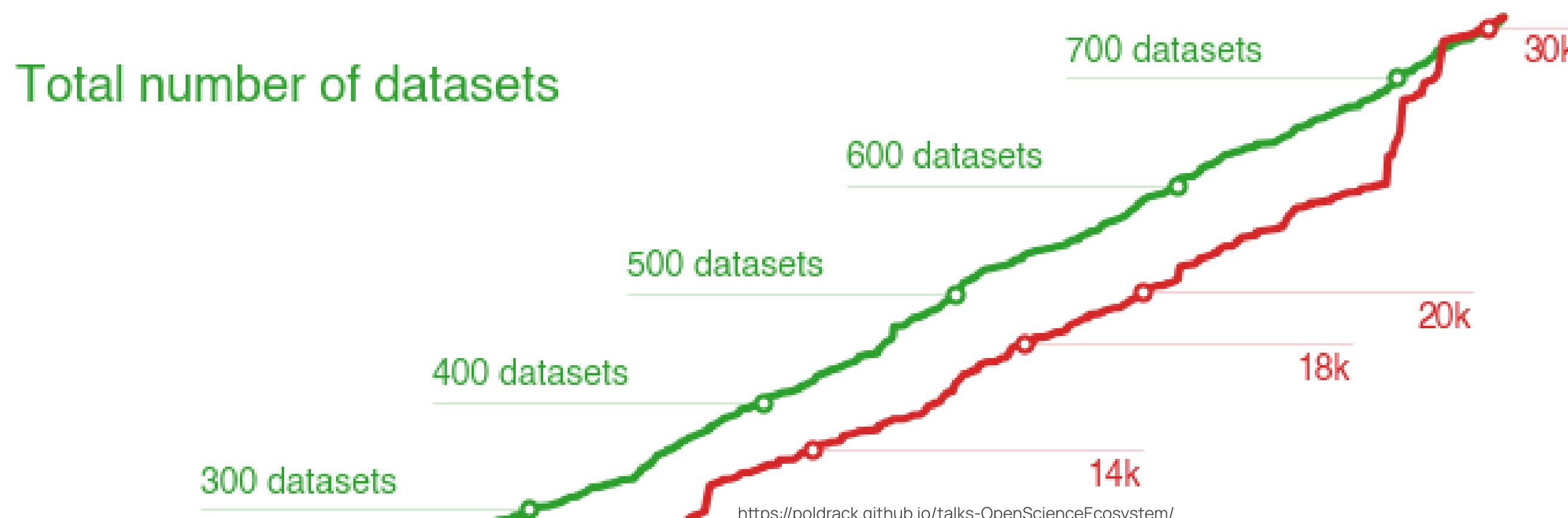
README

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

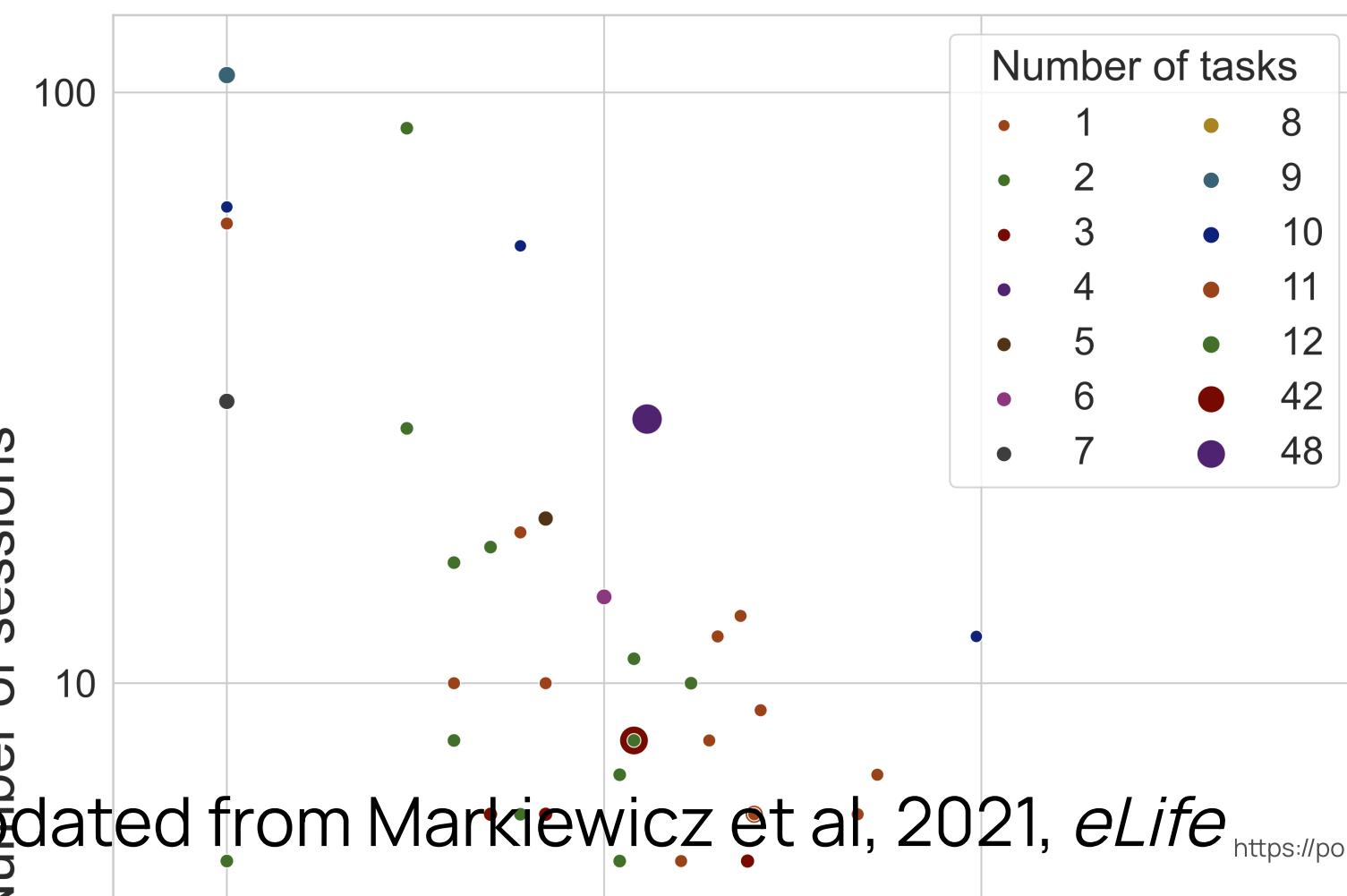
Release history: 10/06/2011: initial release 3/21/2013: Updated release with QA information 2/18/2016: Updated orientation information in nifti headers for improved left-right determination

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 users of the data will acknowledge the OpenfMRI project and NSF Grant OCI-1131441 (R. Poldrack, PI) in any publications.

The growth of OpenNeuro



The diversity of OpenNeuro datasets



Updated from Markiewicz et al, 2021, *eLife*

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

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

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

Scholarly reuse of OpenNeuro datasets

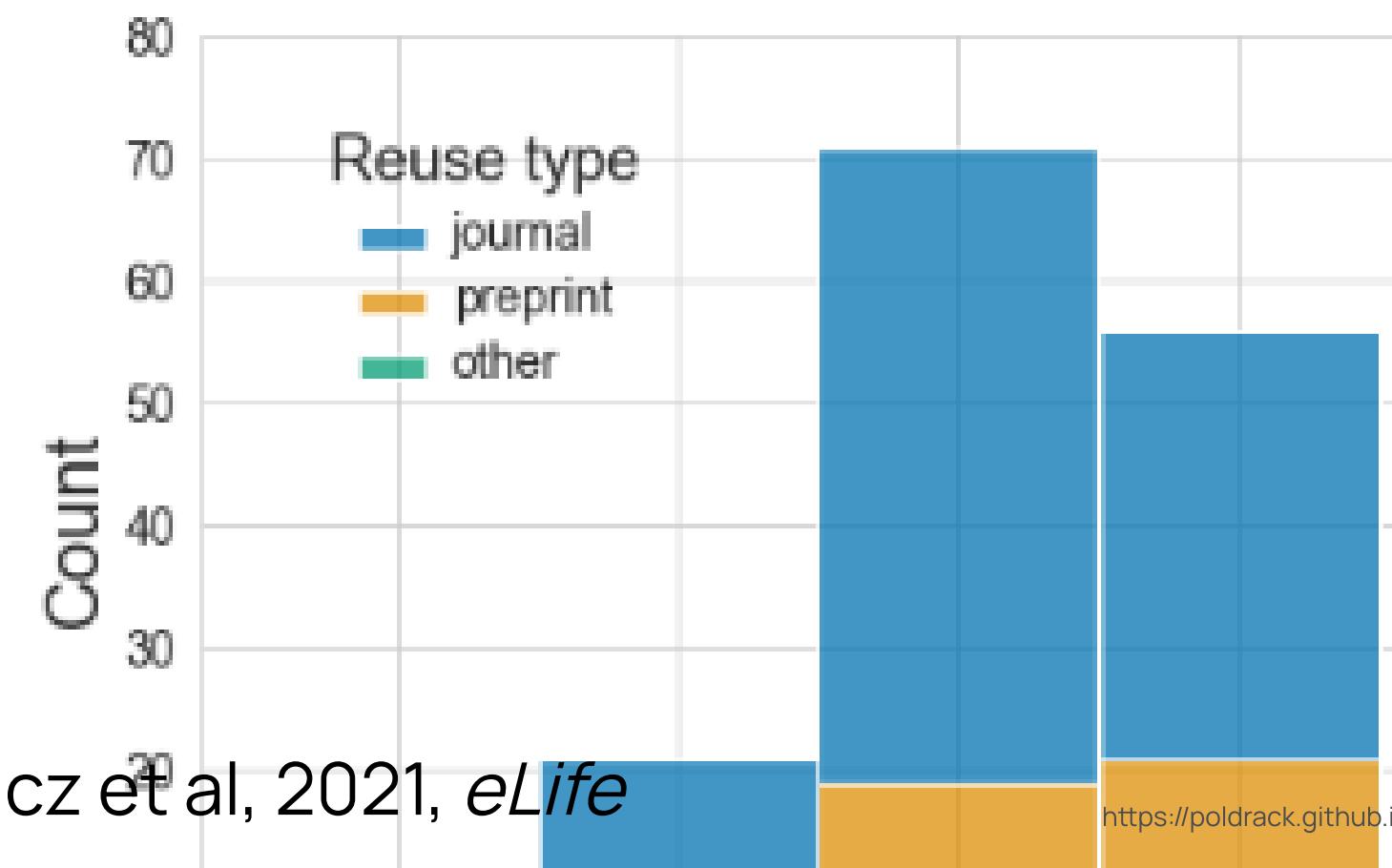
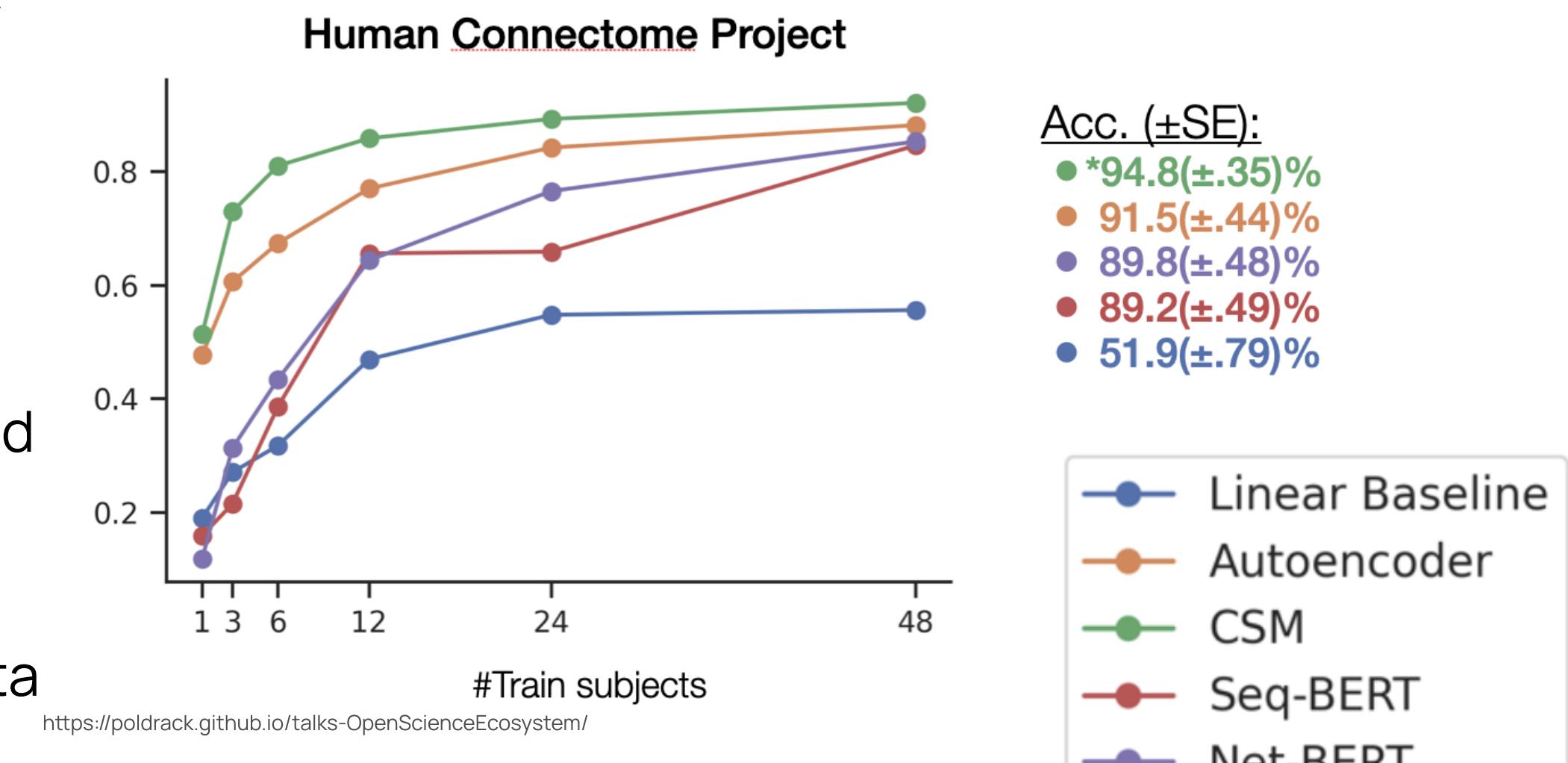


Figure 5. Published reuses of OpenNeuro datasets, split by the type of reuse. Note that the final bar includes only reuses identified through June 2021.

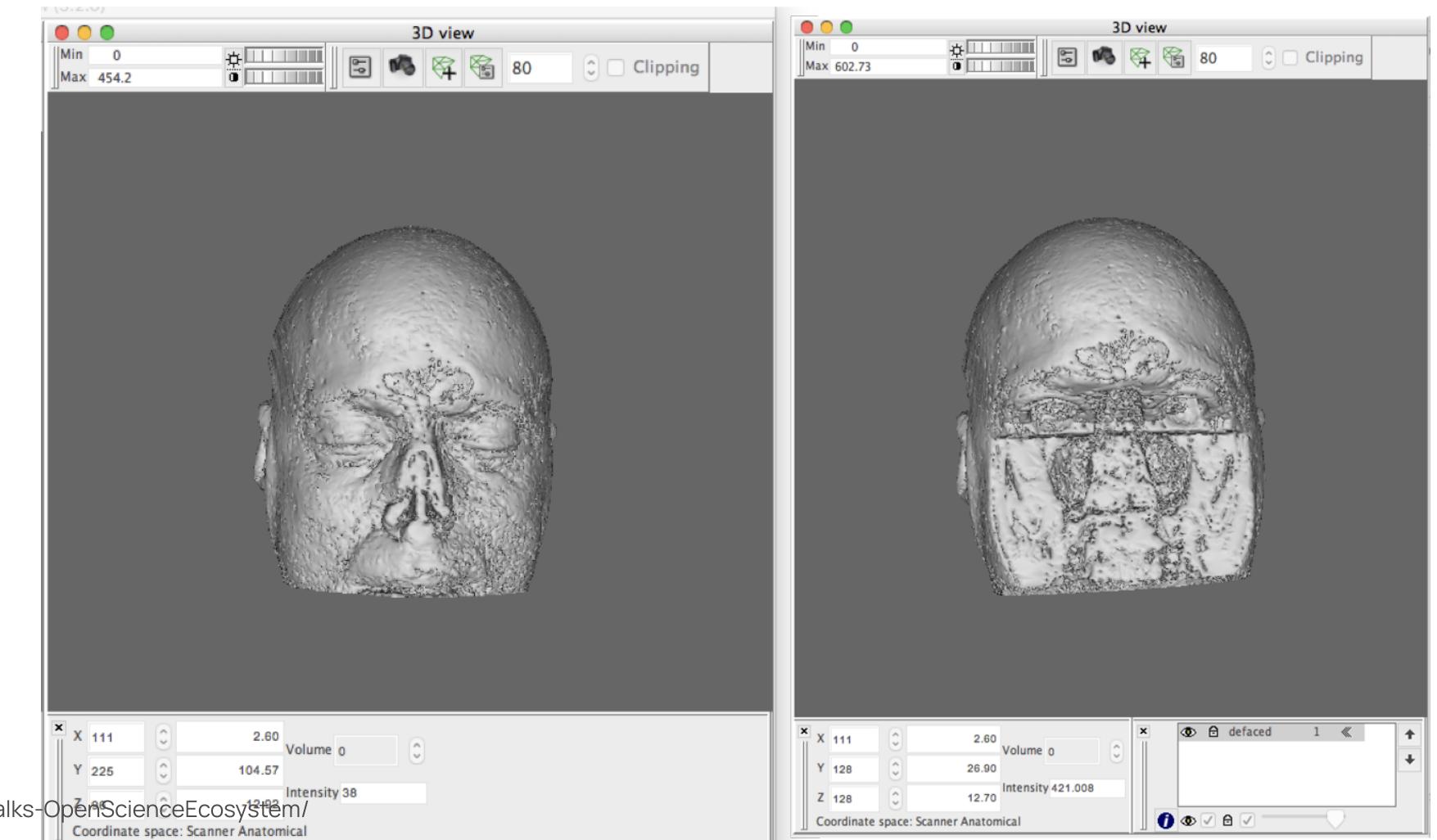
Example of OpenNeuro reuse

- 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



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



Keys to success in neuroimaging data sharing

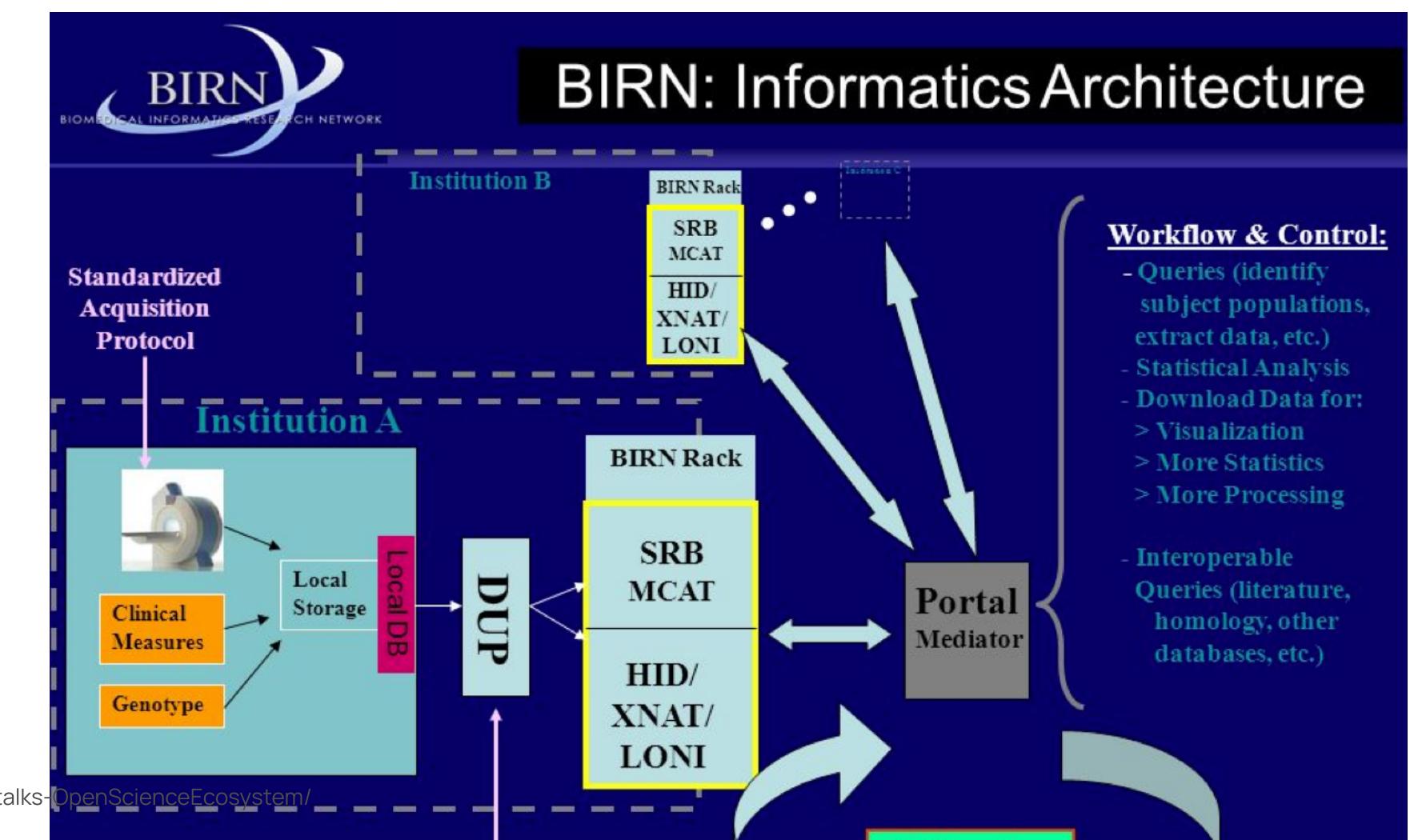
- *Data are digital end-to-end*
 - Minimizes manual steps in the process
- *Standardized file formats and data standards*
 - Makes data immediately usable by anyone
 - Reduces burden of curation and preparation
- *Demonstrated scientific utility*
- ≡ • *Numerous success stories*

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

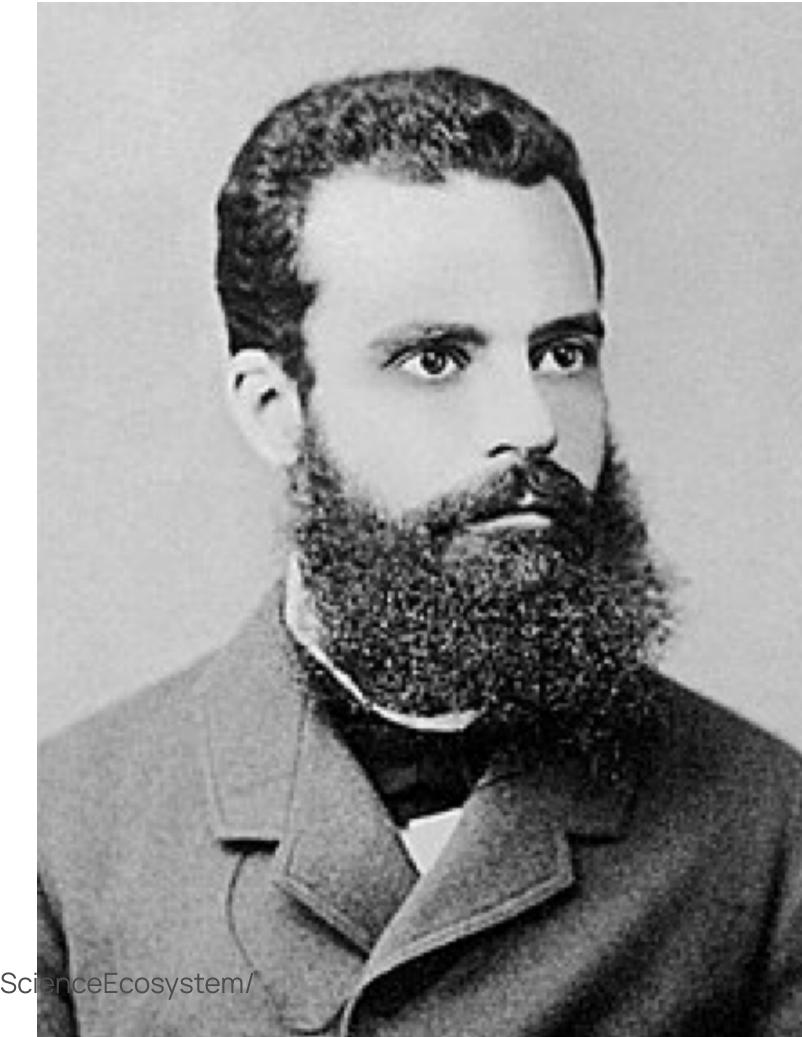
Lessons learned

- 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



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

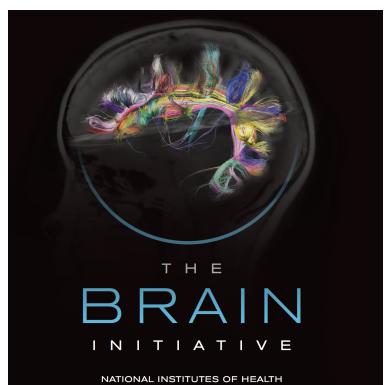
The Poldrack Lab



OpenNeuro Team



Funding

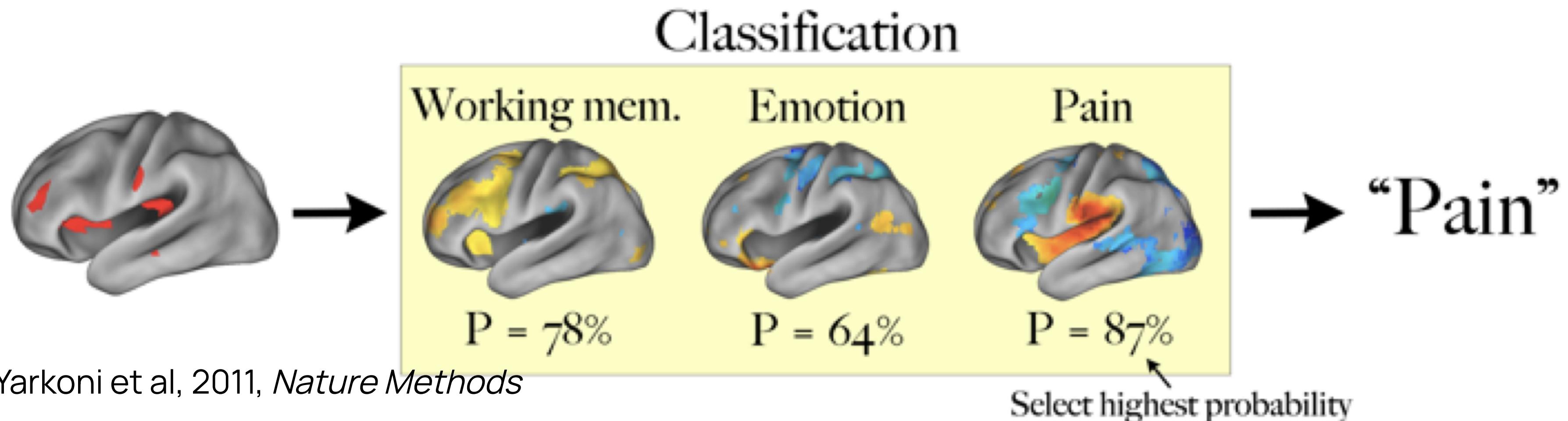


Collaborators



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?
- = Yarkoni et al. 2011, *Nature Methods*

