

# 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

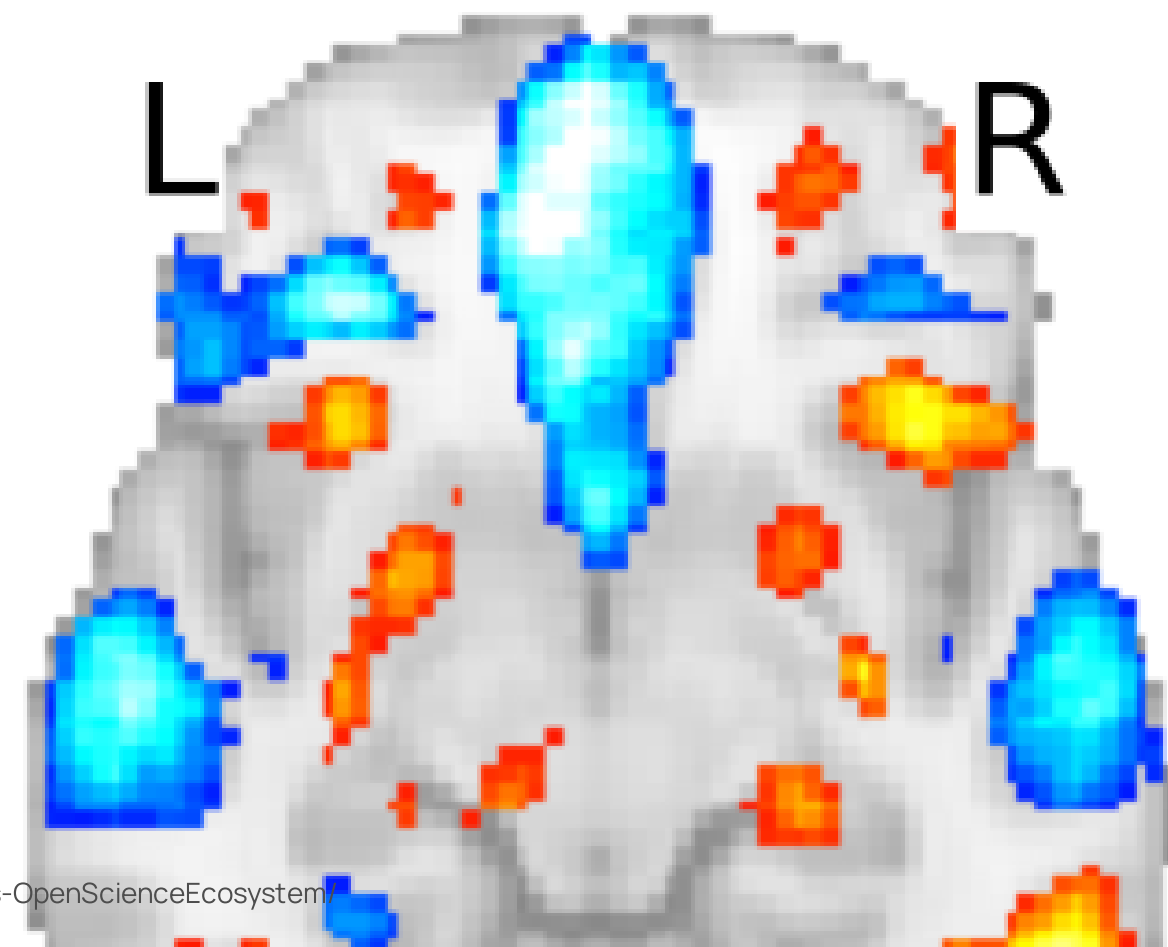
environment and the complete set of

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



# 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

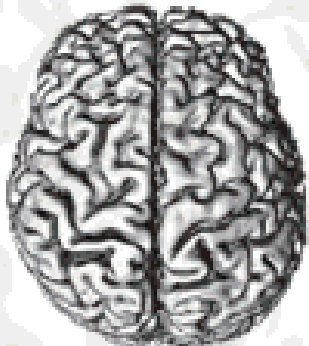
The screenshot shows the homepage of the fMRIDC (The fMRI Data Center). The page features a navigation menu on the left with links for HOME, DATABASE, SUBMISSIONS, RESOURCES, DATA MANAGEMENT TOOL, HELP, and ABOUT US. A search bar at the top allows users to search the fMRIDC Database. The main content area includes a central banner with a brain image and text describing the repository as a public repository of peer-reviewed fMRI studies and their underlying data. It lists funding sources: The National Science Foundation, The W. M. Keck Foundation, The National Institutes of Mental Health, and A Sun Center of Excellence for Neuroscience. A 'PROJECT STATISTICS' section provides data: Registered users: 1912, Datasets available: 110, and Dataset requests: 1789. There are also sections for 'fMRIDC NEWS' and 'Special Collections'.

The fMRI Data Center  
**fMRIDC**


SEARCH  FOR

My Account Request List (Empty)

► HOME  
DATABASE  
SUBMISSIONS  
RESOURCES  
DATA MANAGEMENT TOOL  
HELP  
ABOUT 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



**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**

**INFORMATION**  
[How do I get started?](#)  
Answers to questions commonly posed by first-time visitors.  
[Q&A about fMRIDC](#)  
A comprehensive list of frequently asked questions about the fMRIDC.  
[Available Datasets](#)  
A list of datasets currently available.  
[Information for Authors](#)

**fMRIDC NEWS**  
[fMRIDC Releases DCSearch \(beta\)](#)  
November 10, 2005 - Now Search the fMRI Data Center archive by anatomical region, Brodman area, talairach/MNI coordinates, and other fields.  
[Michael Gazzaniga Elected to Institute of Medicine](#)  
October 24, 2005 - [fMRIDC/POL Pack](#) [pub.io/talks-OpenScienceEcosystem/](#)  
Honor

[Sitemap](#)

5 / 65

# 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](http://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. ...

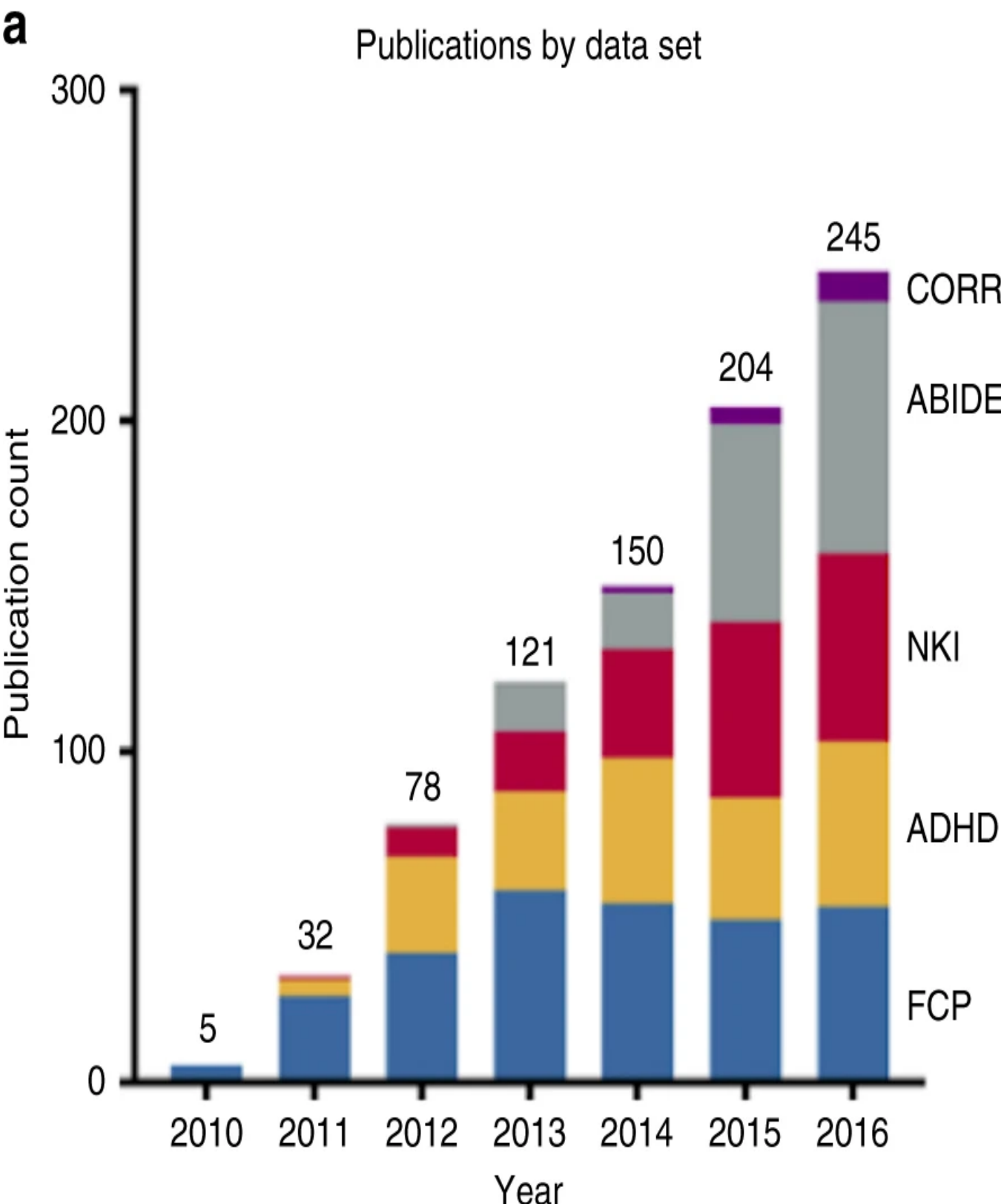
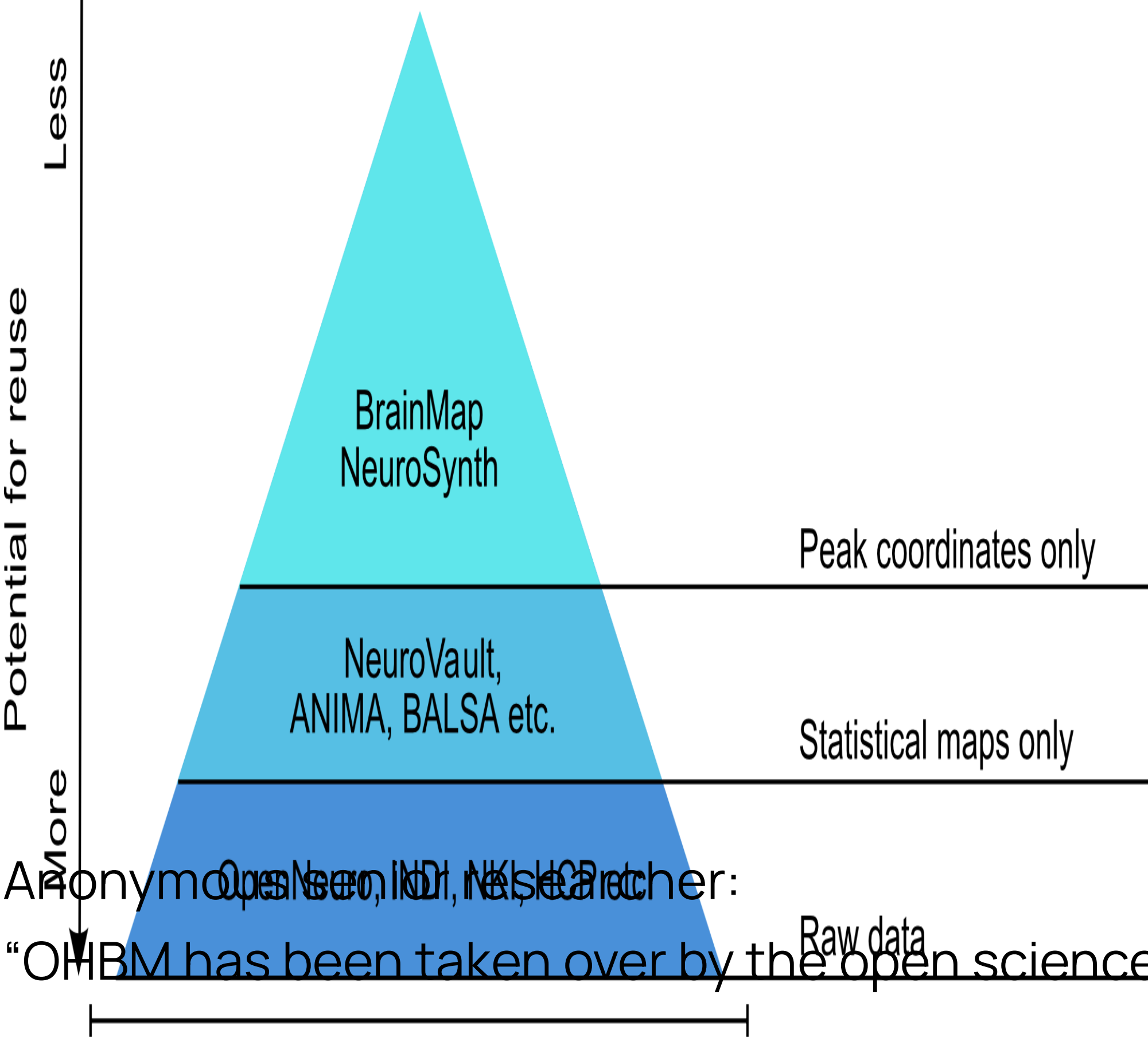
We are particularly concerned with any journal's

# 2010: The year data sharing broke in neuroimaging

## Toward discovery science of human brain function

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

# Data sharing is becoming the norm in neuroimaging

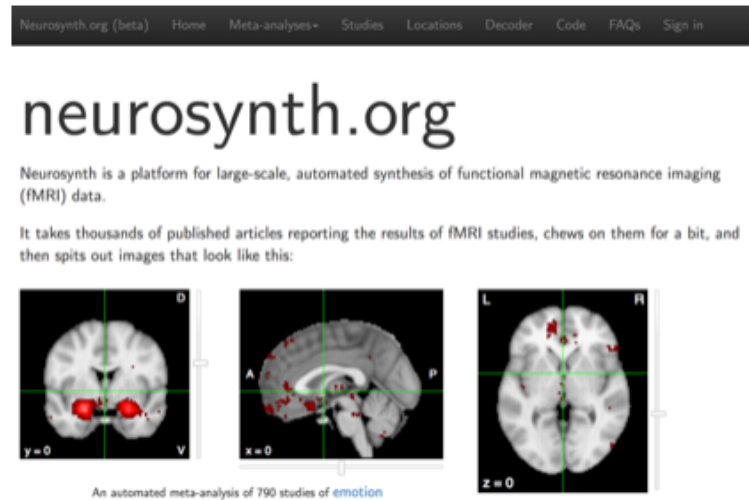


Milham et al., *Nature Communications*, 2018



# 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

# 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

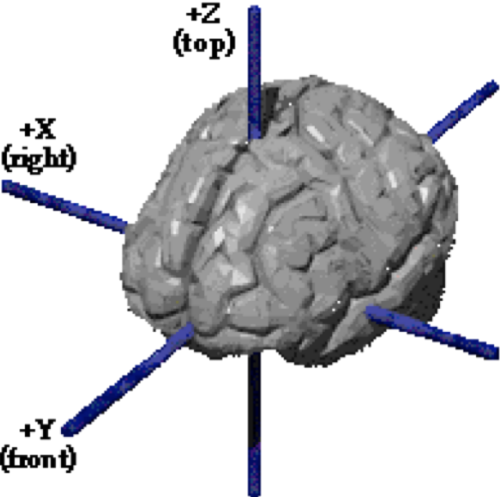


Table 1  
Regions that showed a condition × time interaction in the ANOVA analysis

No.	Region	Hemisphere	BA	x	y	z	mm <sup>3</sup>
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

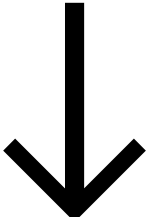
<https://poldrack.github.io/talks-OpenScienceEcosystem/>

# 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



- Meta-analytic maps created for each

# working memory

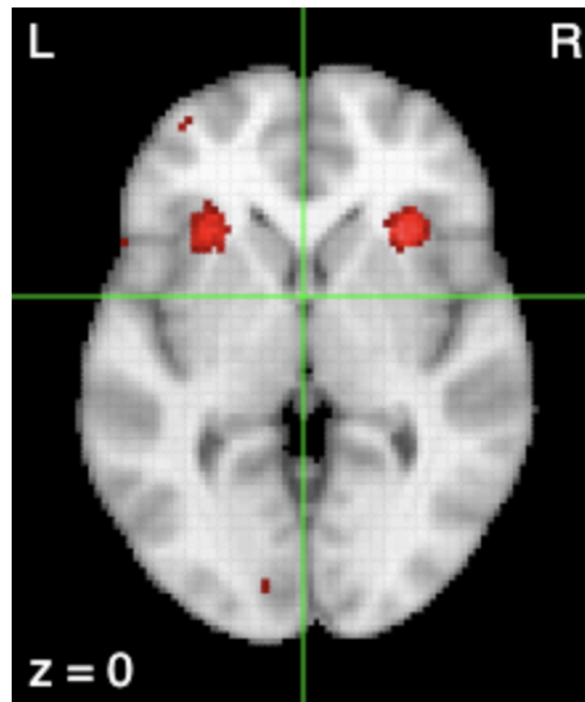
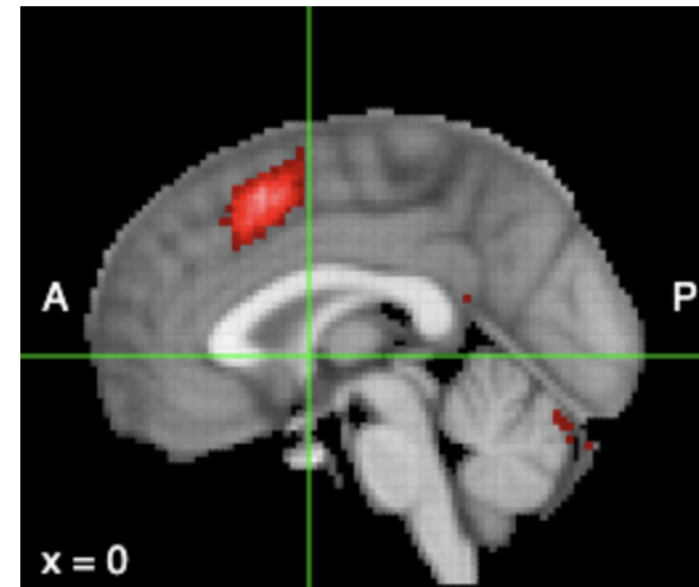
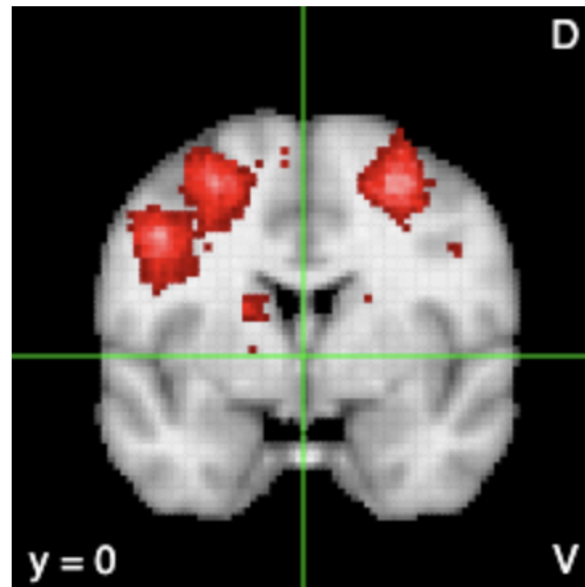
An automated meta-analysis of 1091 studies

Search for another term:

Maps

Studies

FAQs



z-score: 0

What's here?

X:  Y:  Z:

Layers

<input checked="" type="checkbox"/>	working memory: association test	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	working memory: uniformity test	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	anatomical	<input type="checkbox"/>	<input type="checkbox"/>

Color palette:

red

Crosshairs

Positive/Negative:

positive

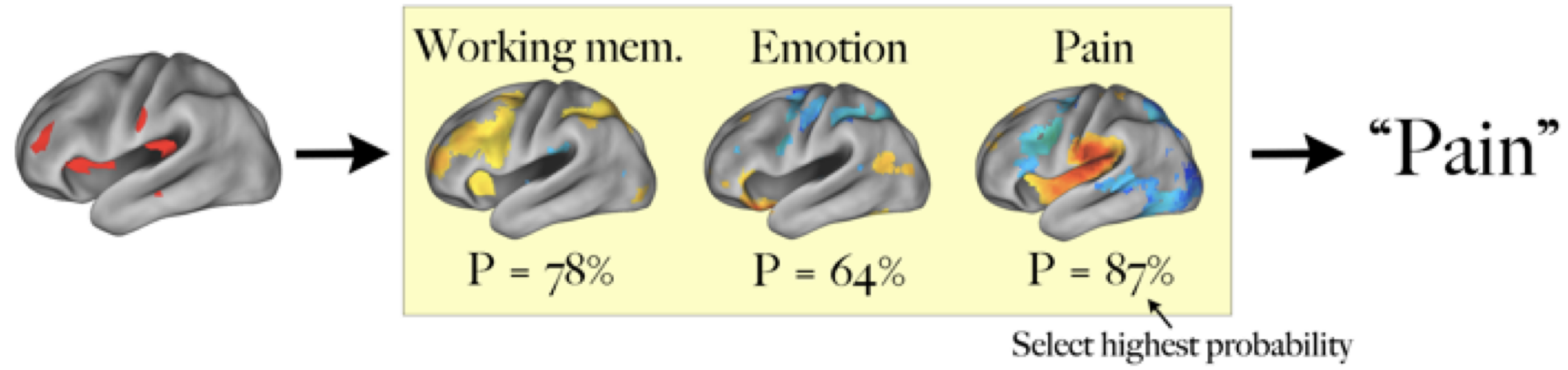
Pan/zoom

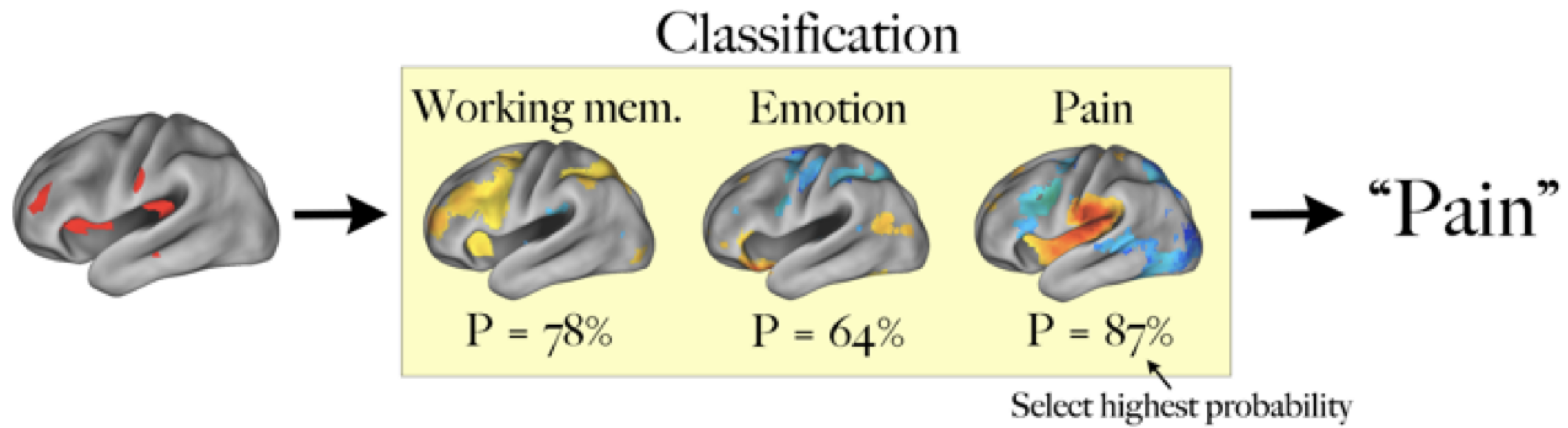
Labels

Thresholds:

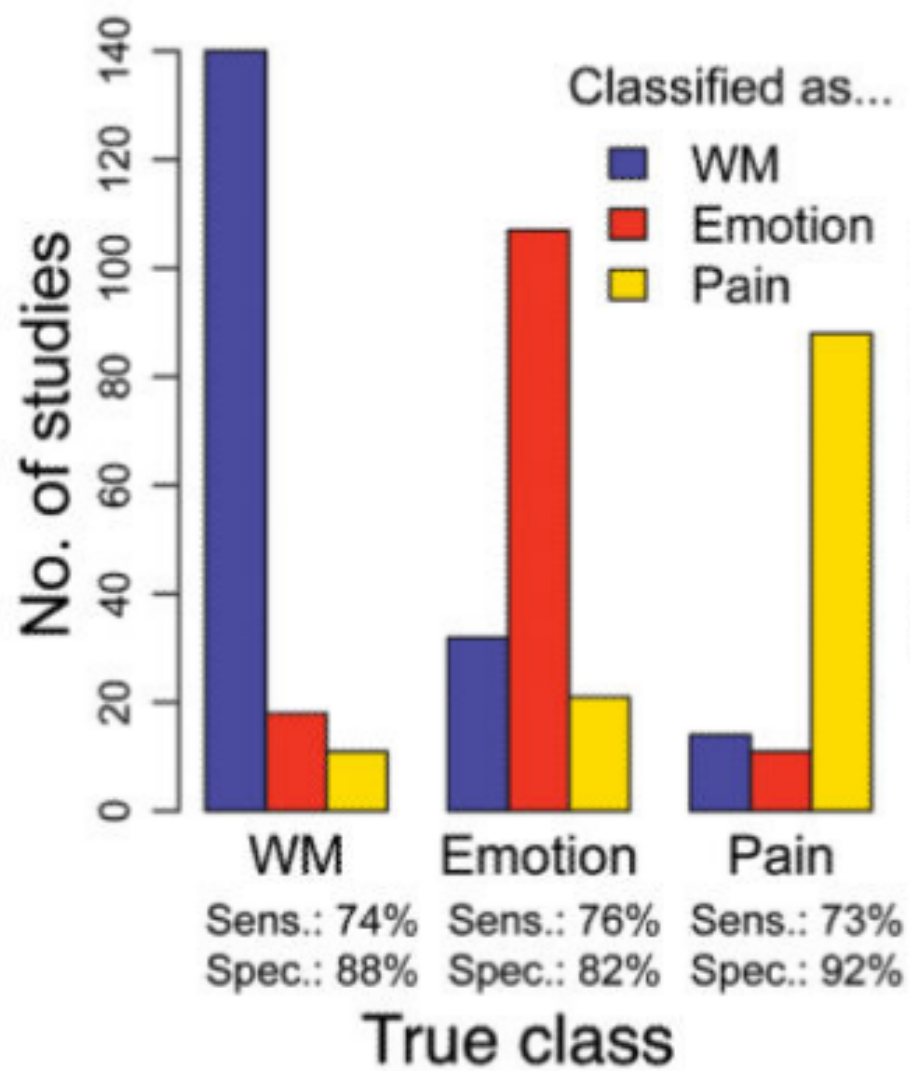
Opacity:

# Classification

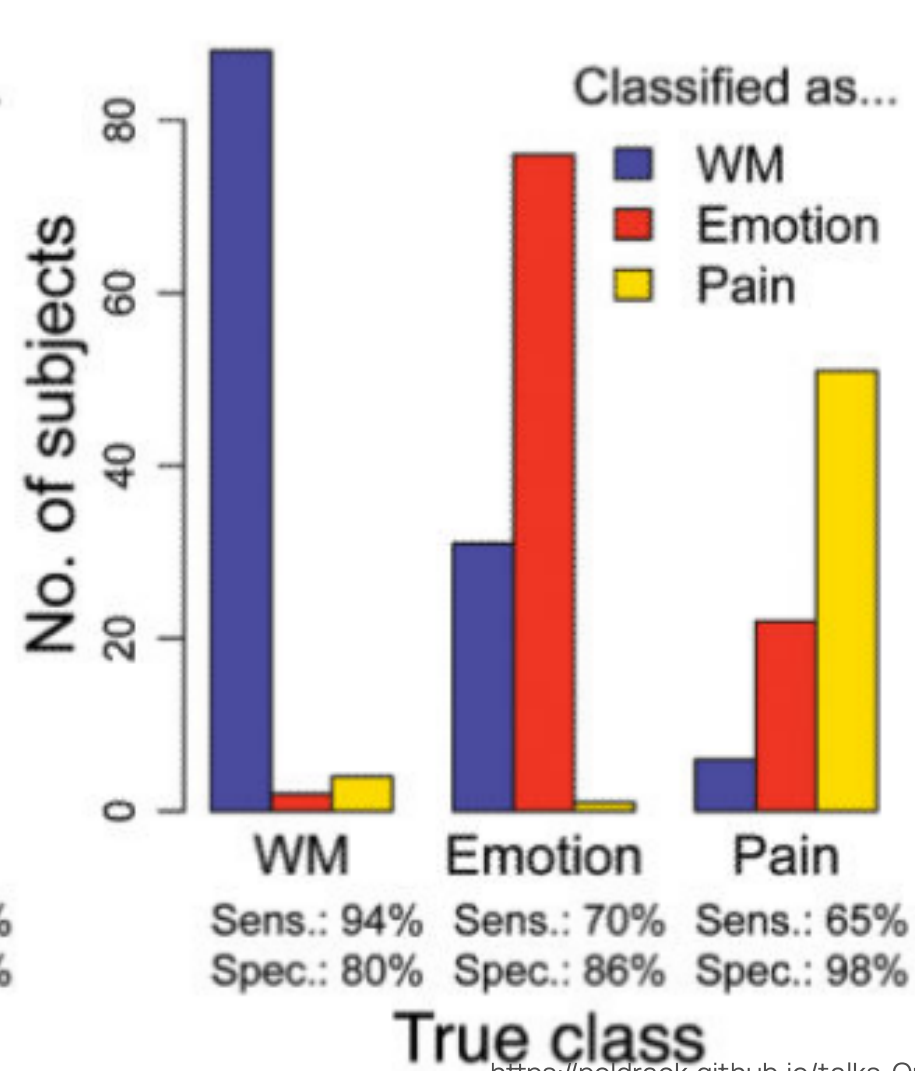




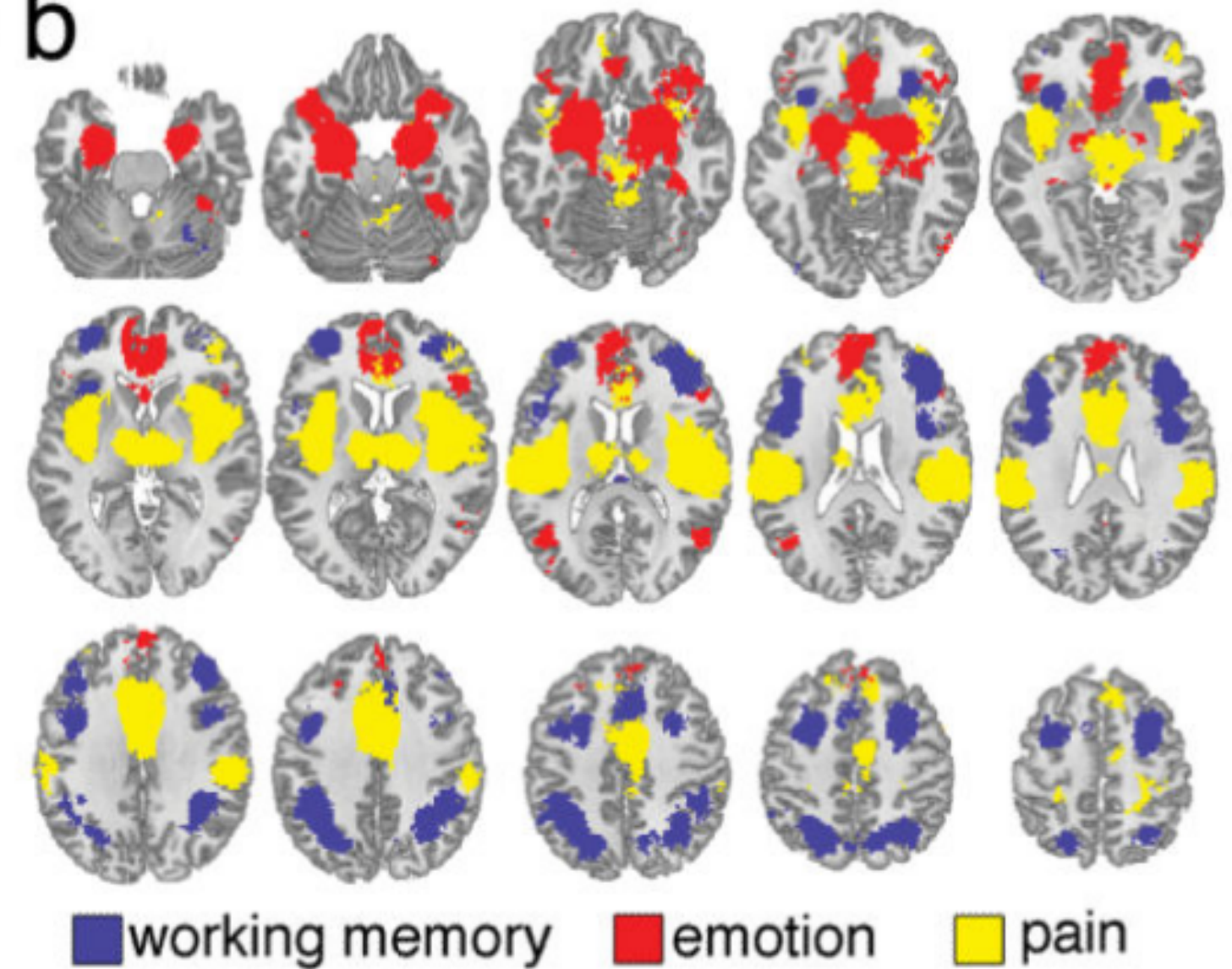
**a** Classification of new studies



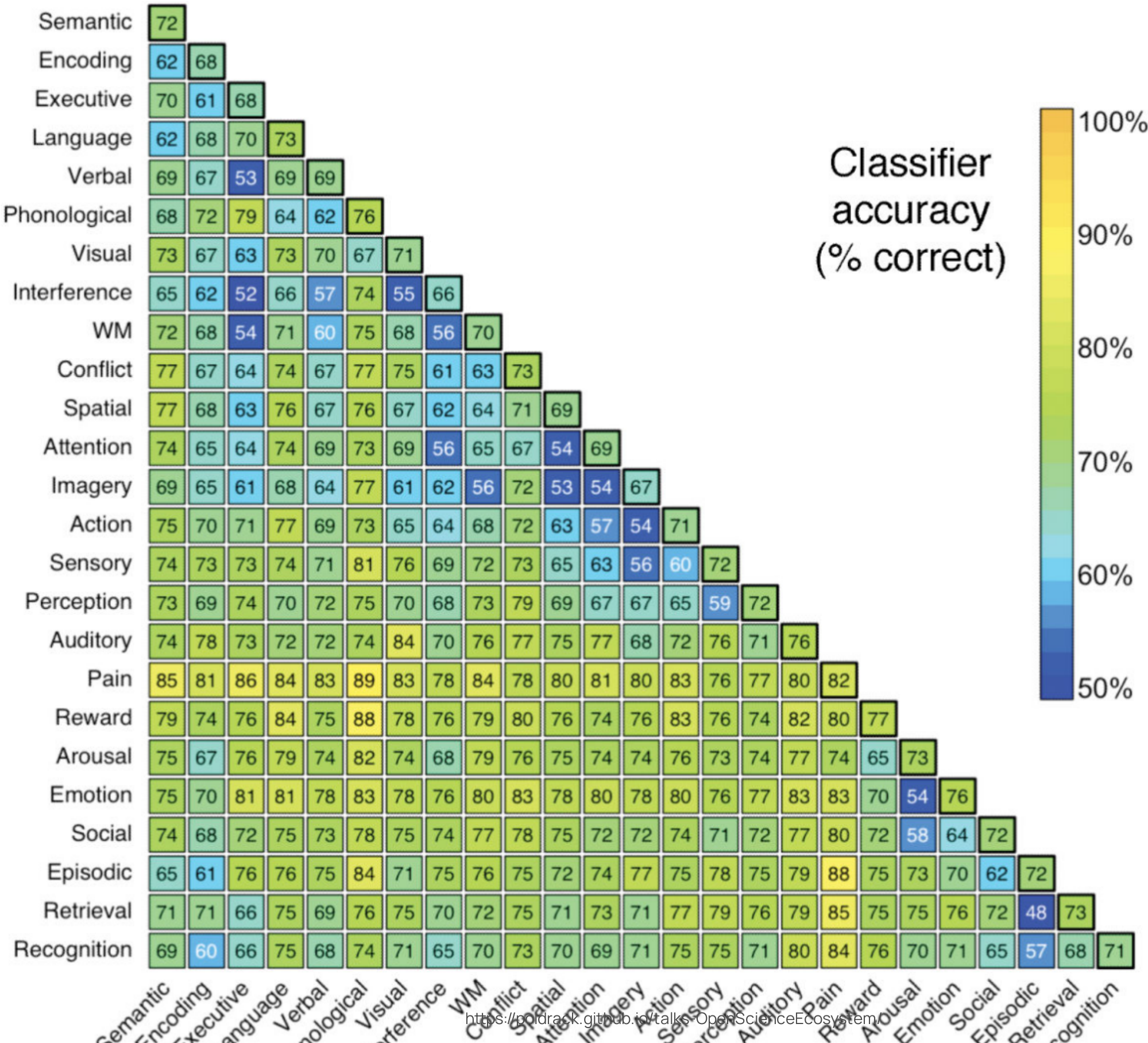
Classification of new subjects



**b**



# Decoding brain activity patterns using Neurosynth



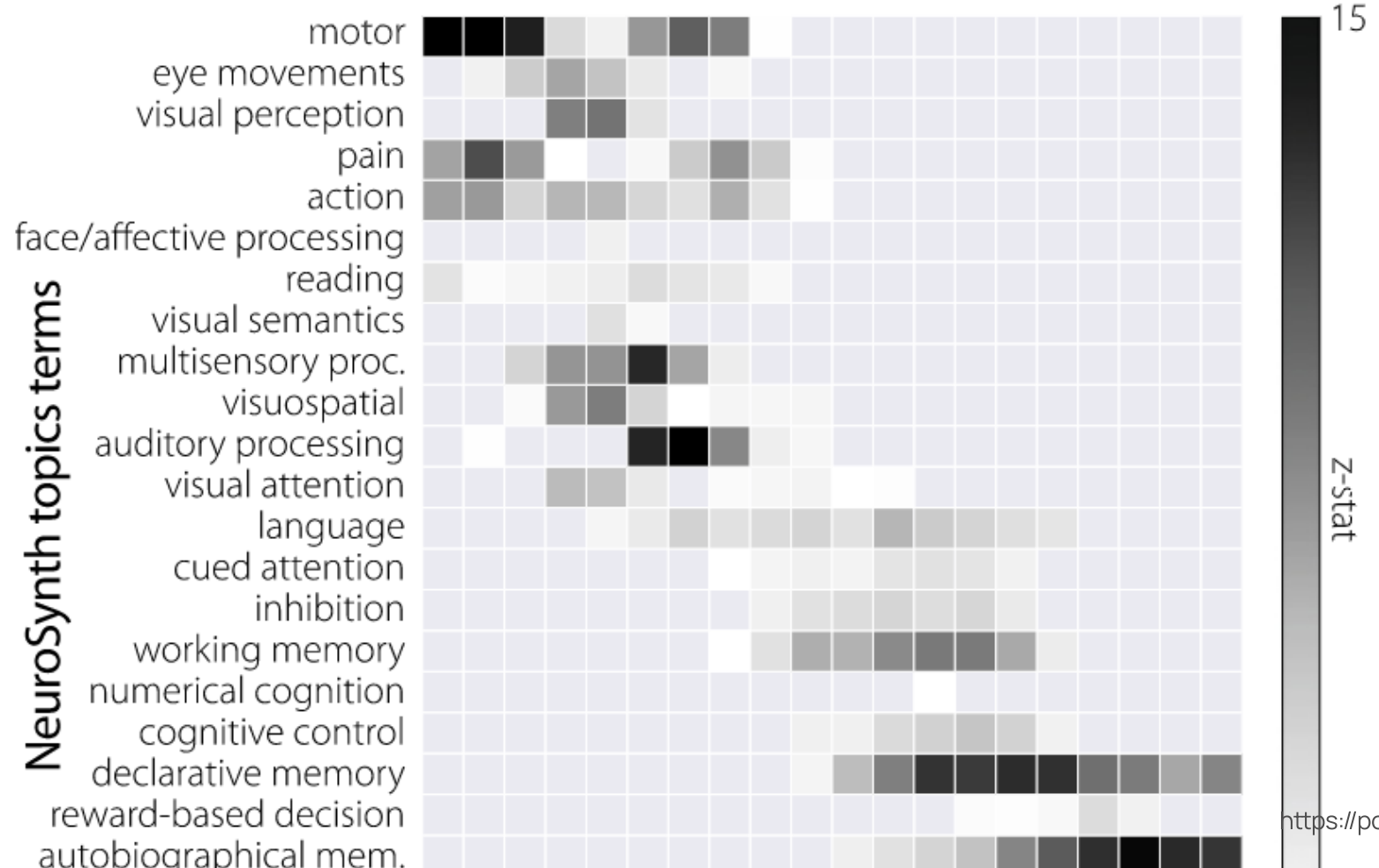


# Example of Neurosynth usage

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

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

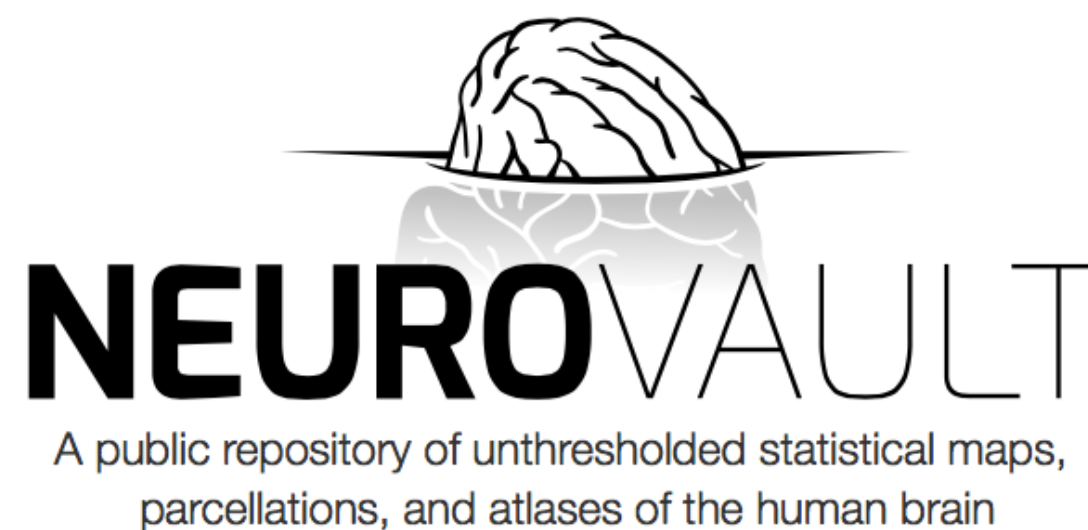
12574–12579 | PNAS | November 1, 2016 | vol. 113 | no. 44



- 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

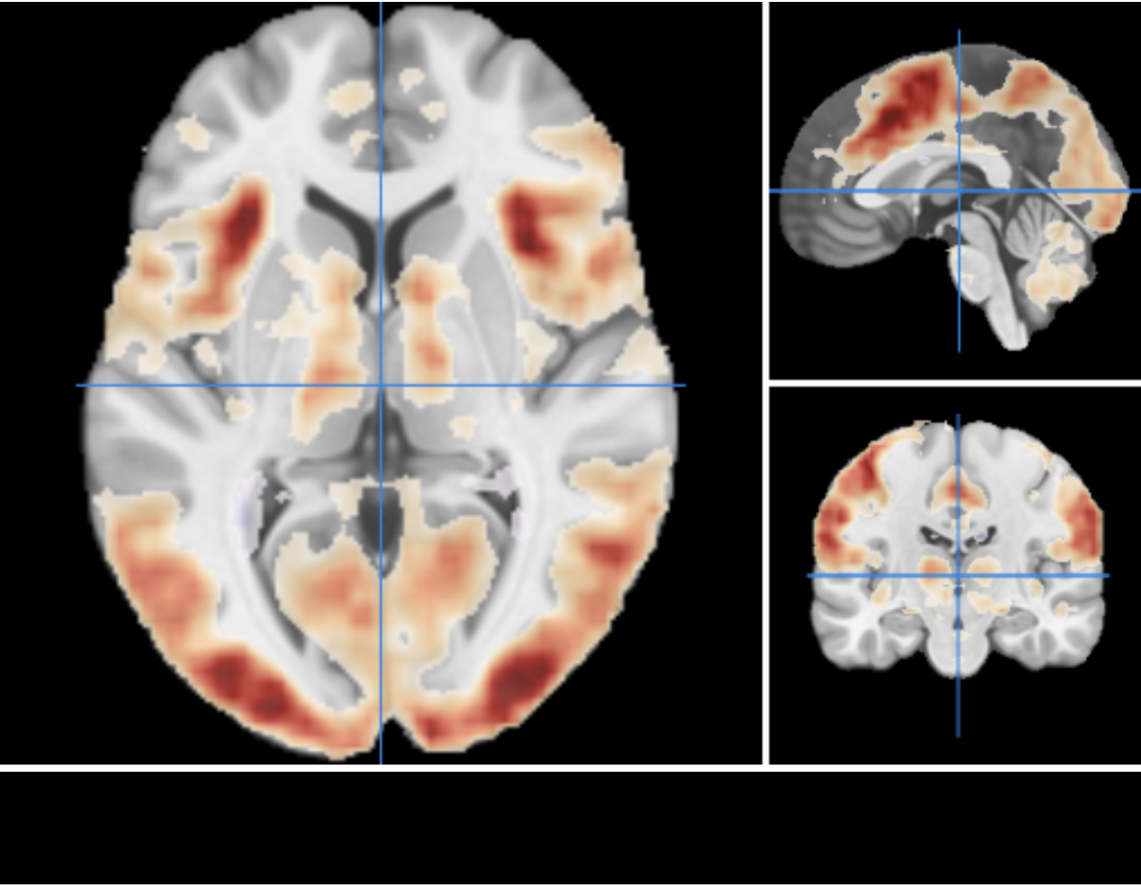
### Preprocessed Consortium for Neuropsychiatric Phenomics dataset

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
<input type="checkbox"/>	49974	BART Accept	T map
<input type="checkbox"/>	49975	BART AcceptParam - ExplodeParam	T map
<input type="checkbox"/>	49976	BART AcceptParam - RejectParam	T map
<input type="checkbox"/>	49977	BART AcceptParametric	T map
<input checked="" type="checkbox"/>	49978	BART Accept_RT	T map
<input type="checkbox"/>	49979	BART Control	T map
<input type="checkbox"/>	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**

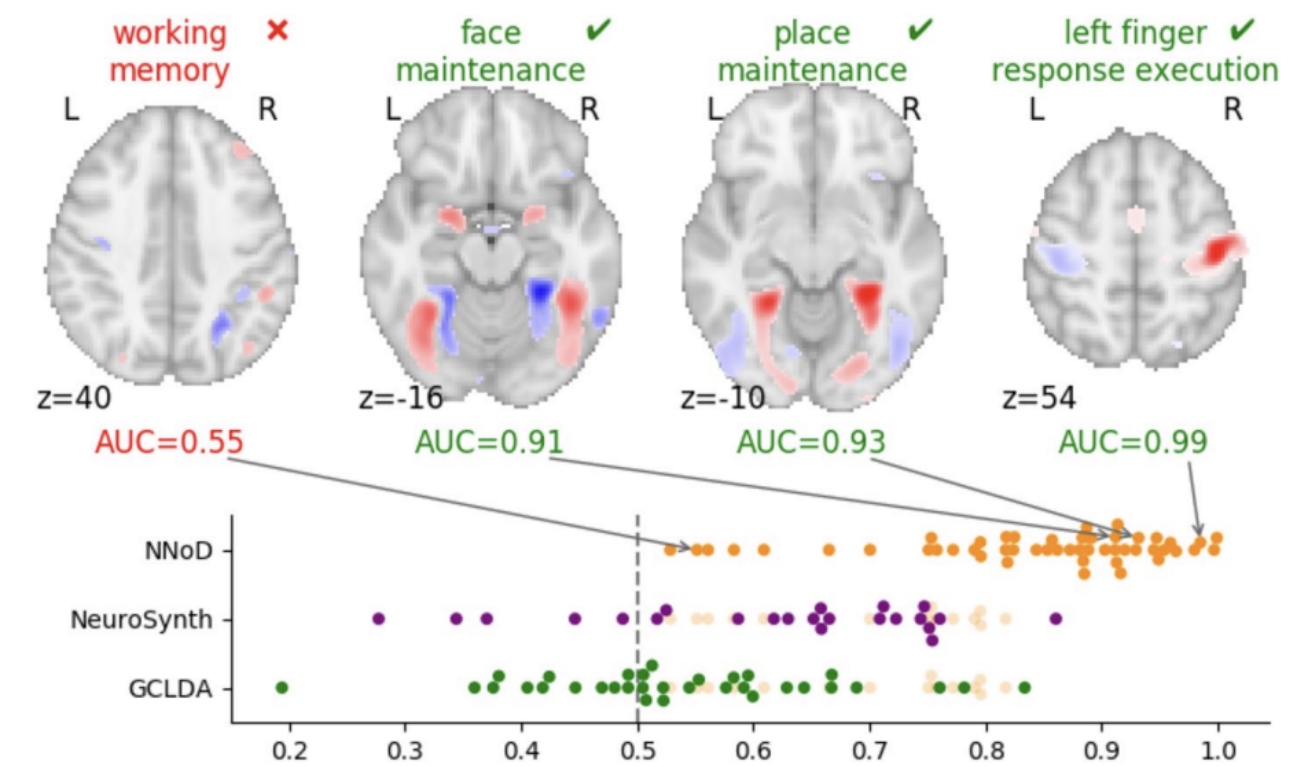
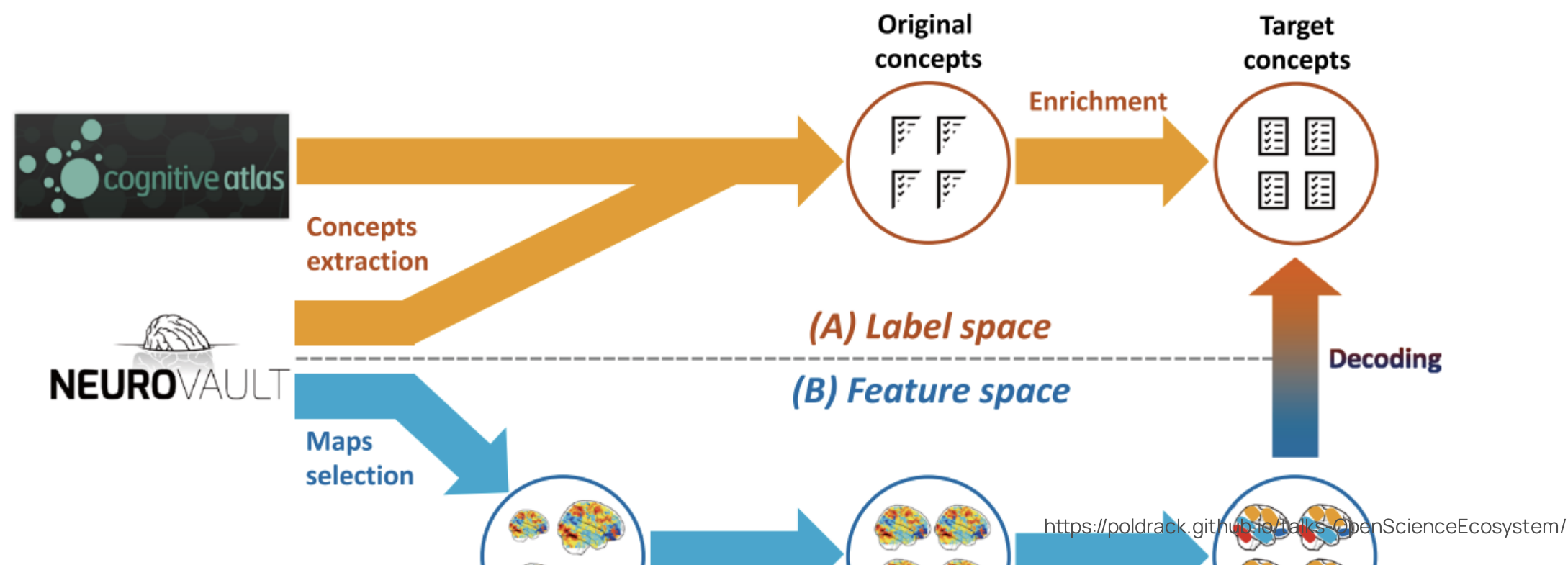
- Individual images can be browsed and downloaded
- A number of analysis tools can also be applied
- Each image also receives a persistent identifier

# Example of Neurovault usage

## Comprehensive decoding mental processes from Web repositories of functional brain images

Romuald Menuet<sup>5,6</sup>, Raphael Meudec<sup>1,2,3,6</sup>, Jérôme Dockès<sup>4</sup>, Gael Varoquaux<sup>1,2,3</sup> & Bertrand Thirion<sup>1,2,3</sup>✉

Scientific Reports | (2022) 12:7050



# OpenNeuro: Sharing raw and processed neuroimaging data

**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 Google ORCID

**MRI** **PET** **MEG** **EEG** **iEEG**



### 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-

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

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



- 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.



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

# 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

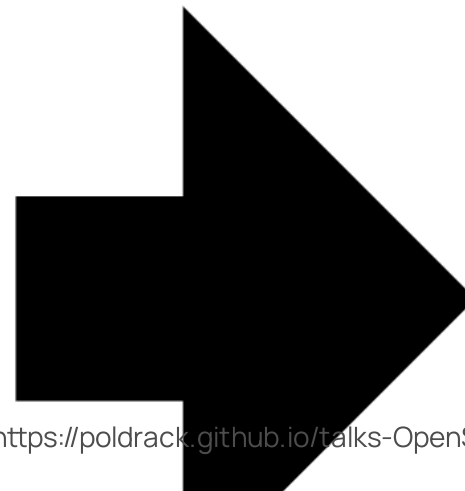
# BIDS Principles

- *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/>

SCIENTIFIC DATA 

OPEN

DATA DESCRIPTOR

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

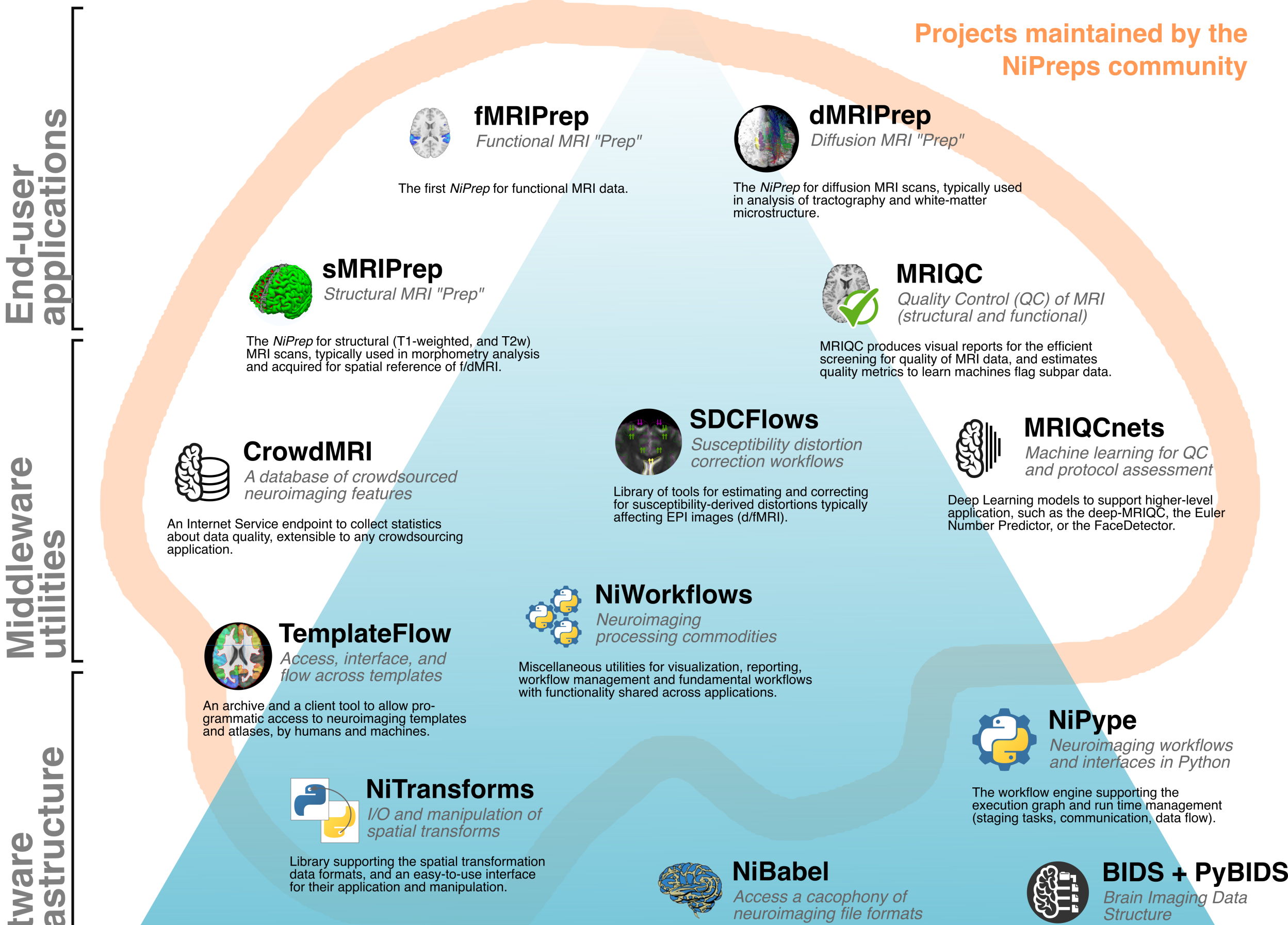
Received: 19 September 2018  
Accepted: 12 March 2019

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



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

# BIDS enables a growing open-source software ecosystem



<https://poldrack.github.io/talks-OpenScienceEcosystem/>  
Library supporting the neuroimaging data formats (e.g., NIfTI, GIFTI, and CIFTI2).



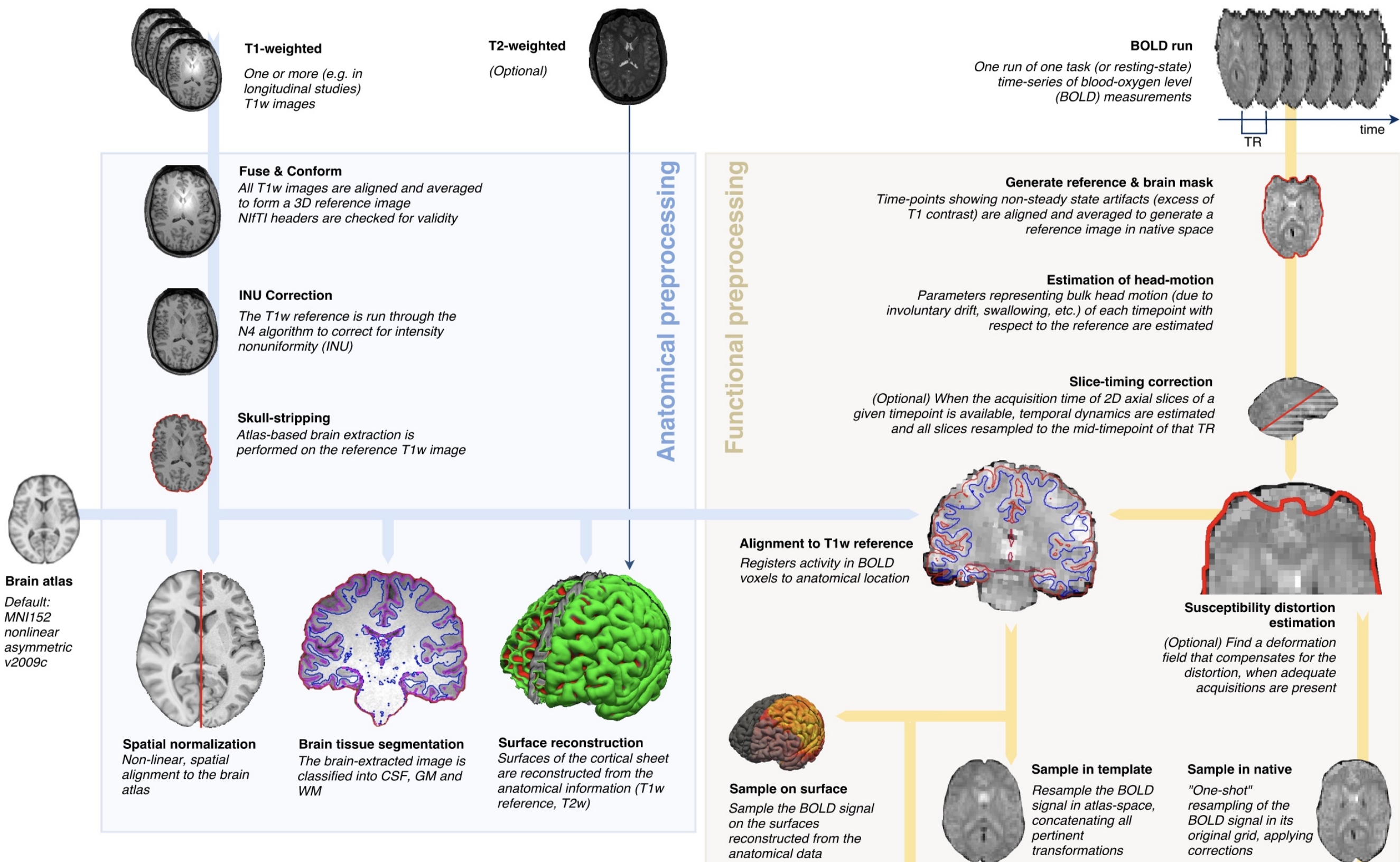
# BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods

Krzysztof J. Gorgolewski<sup>1\*</sup>, Fidel Alfaro-Almagro<sup>2</sup>, Tibor Auer<sup>3</sup>, Pierre Bellec<sup>4,5</sup>, Mihai Capotă<sup>6</sup>, M. Mallar Chakravarty<sup>7,8</sup>, Nathan W. Churchill<sup>9</sup>, Alexander Li Cohen<sup>10</sup>, R. Cameron Craddock<sup>11,12</sup>, Gabriel A. Devenyi<sup>7,8</sup>, Anders Eklund<sup>13,14,15</sup>, Oscar Esteban<sup>1</sup>, Guillaume Flandin<sup>16</sup>, Satrajit S. Ghosh<sup>17,18</sup>, J. Swaroop Guntupalli<sup>19</sup>, Mark Jenkinson<sup>2</sup>, Anisha Keshavan<sup>20</sup>, Gregory Kiar<sup>21,22</sup>, Franziskus Liem<sup>23</sup>, Pradeep Reddy Raamana<sup>24,25</sup>, David Raffelt<sup>26</sup>, Christopher J. Steele<sup>7,8</sup>, Pierre-Olivier Quirion<sup>15</sup>, Robert E. Smith<sup>26</sup>, Stephen C. Strother<sup>24,25</sup>, Gaël Varoquaux<sup>27</sup>, Yida Wang<sup>6</sup>, Tal Yarkoni<sup>28</sup>, Russell A. Poldrack<sup>1</sup>

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



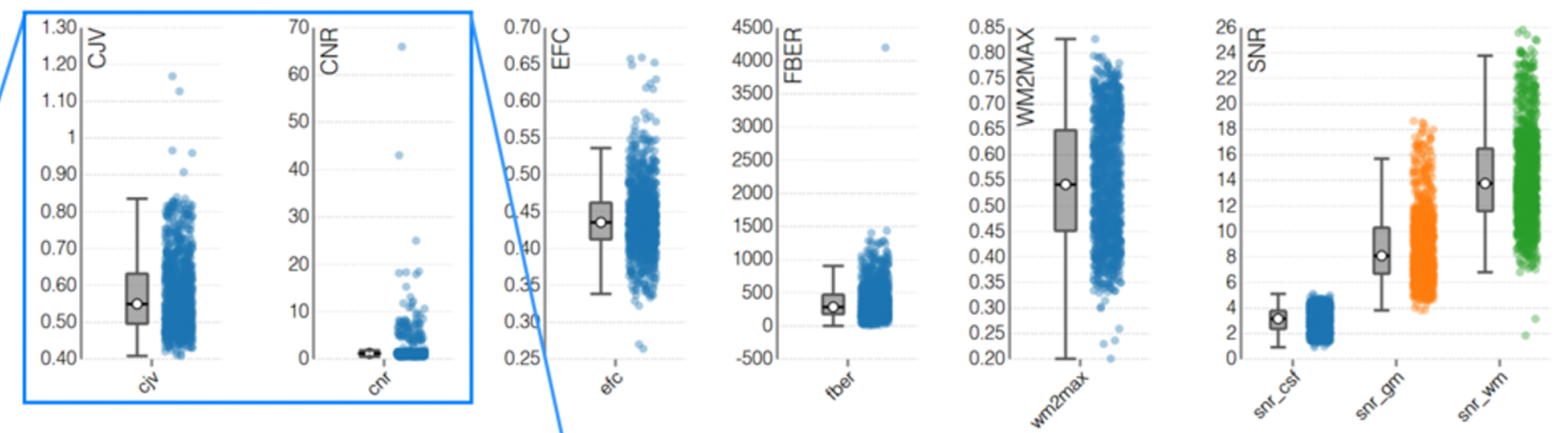
# MRIQC: MRI quality control for BIDS data

## 1 MRIQC: group anatomical report

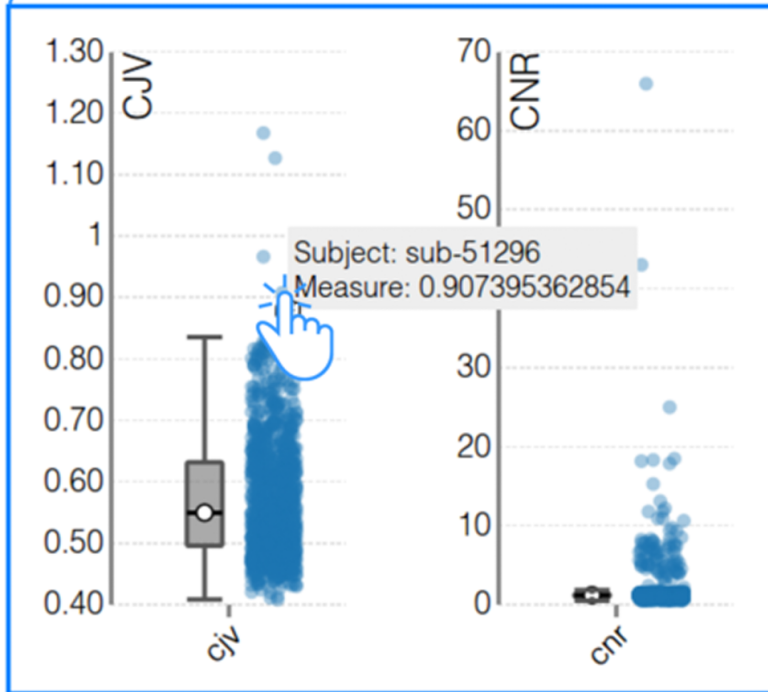
1

### Summary

- Date and time: 2017-02-05, 12:27.
- MRIQC version: 0.9.0-rc2.



2



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.

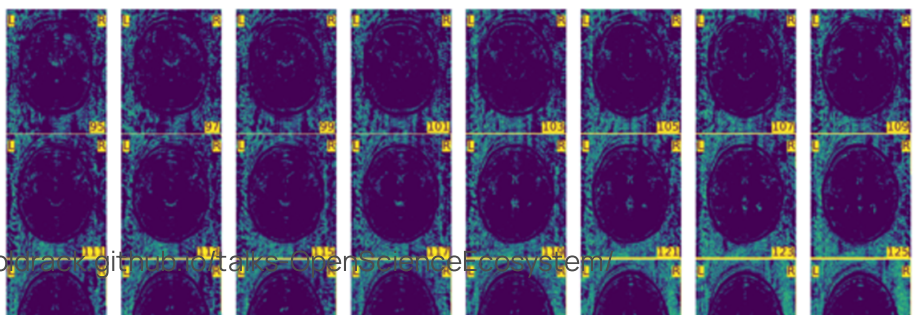
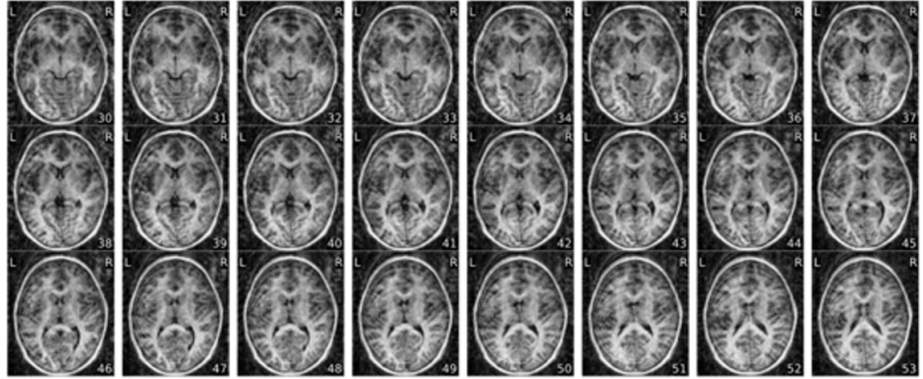
3

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

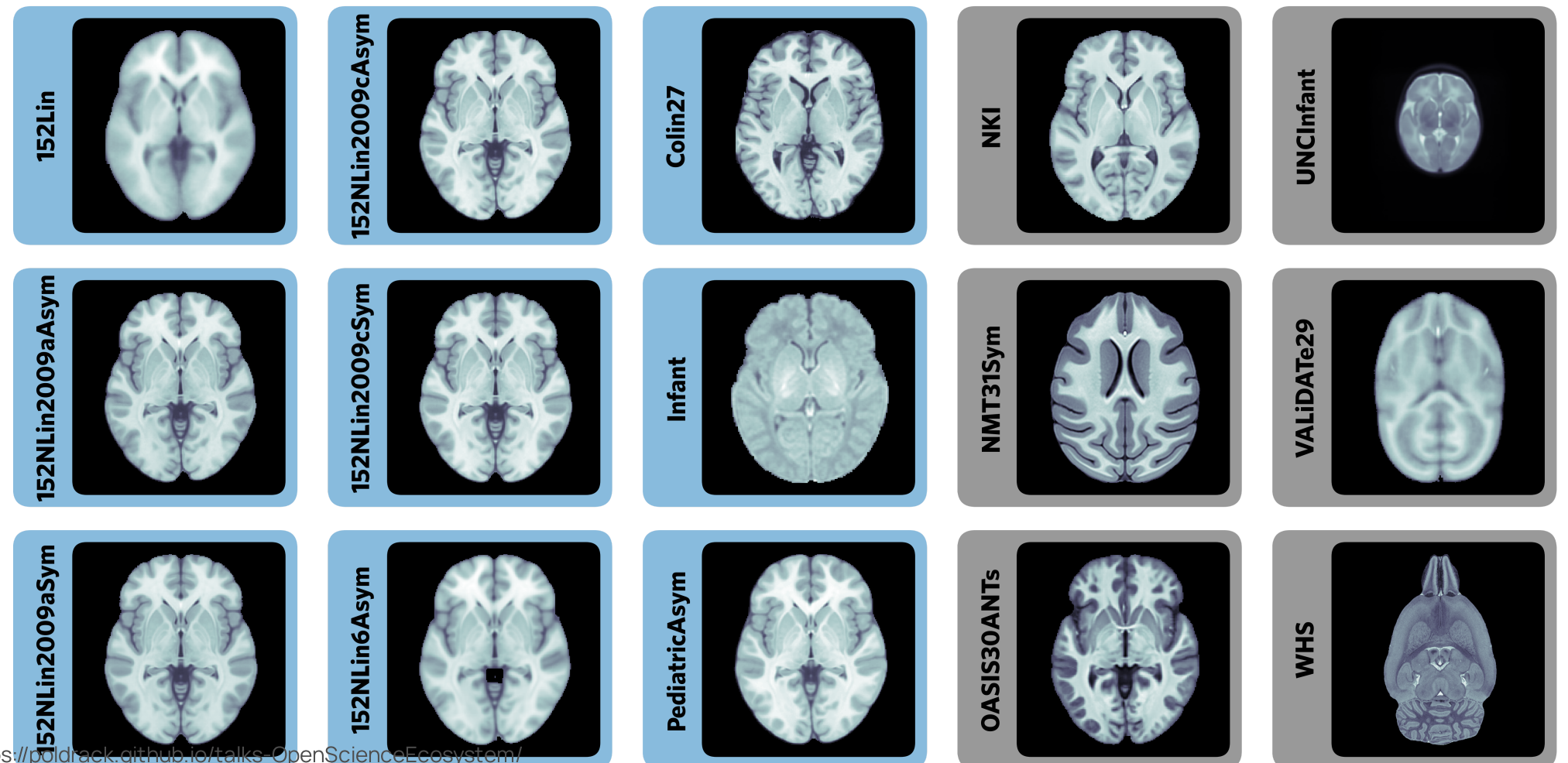
### Summary

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



# Tenplateflow: 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



# OpenNeuro: A BRAIN Initiative archive for BIDS data

The screenshot shows the OpenNeuro website interface. At the top left is the OpenNEURO logo. To the right are links for SEARCH, SUPPORT, and FAQ, and a Sign in button. Below the logo is a description: "A free and open platform for validating and sharing BIDS-compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#) data". It also displays "29,064 Participants" and "761 Public Datasets". There is a "Browse by Modalities" dropdown menu and a search bar with the text "reading" and a search icon. At the bottom left are "SIGN IN" and social login options for Google and ORCID. At the bottom center is a URL: <https://poldrack.github.io/talks-OpenScienceEcosystem/>. In the center of the page, there are five 3D cubes representing different data modalities: MRI (purple), PET (blue), MEG (orange), EEG (red), and iEEG (green). Each cube shows a representative image of its respective modality.

- Supports sharing of any validated BIDS dataset

## Search All Datasets

Keywords ?



reading ×

These filters return **194** datasets:

[CLEAR ALL](#)

KEYWORD:

reading ×

**SORT BY:** Relevance 

### Modalities

- MRI
- PET
- EEG
- iEEG
- MEG

**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**

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

**The Reading Brain Project L2 Adults**  

Each shared dataset is versioned and

receives a persistent identifier (DOI)

The screenshot shows the OpenNeuro website interface. At the top, there is a navigation bar with the OpenNeuro logo, search, support, and FAQ links, and a sign-in button. Below this is a dark purple header for the dataset 'The Reading Brain Project L2 Adults', which includes 'Follow' and 'Bookmark' buttons, both showing a count of 2. The main content area features a 'BIDS Validation' section with a dropdown menu, a '4 WARNINGS' indicator, a 'Valid' status with a green checkmark, and buttons for 'brainlife.io' and 'Clone'. Below the validation section are tabs for 'Files', 'Download', and 'Metadata'. The 'Files' tab is active, displaying a 'README' section. The README text states: '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](http://blclab.org/reading_brain) under L2 Adult dataset/ <https://poldrack.github.io/talks-OpenScienceEcosystem/>

On the right side of the dataset page, there are several sections: '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'.

Any valid BIDS dataset can be shared via  
OpenNeuro



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

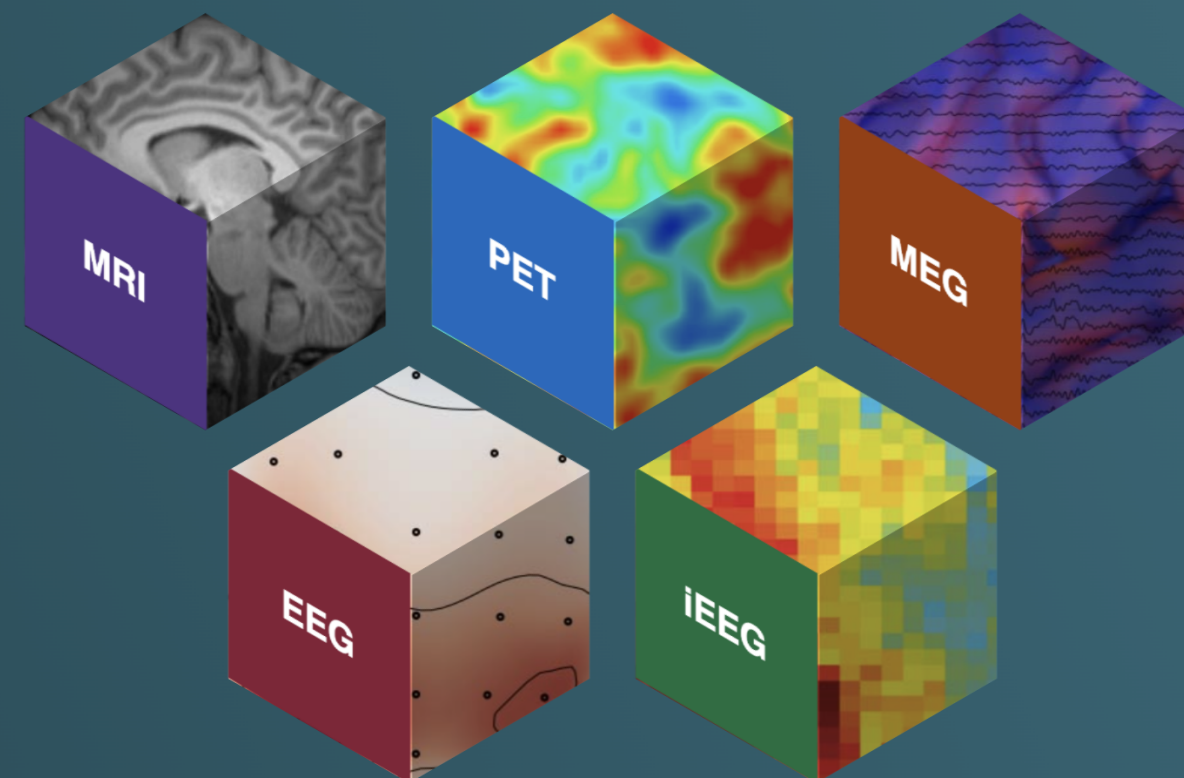
**29,064** Participants

**761** Public Datasets

Browse by Modalities ▼

Or

Search





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

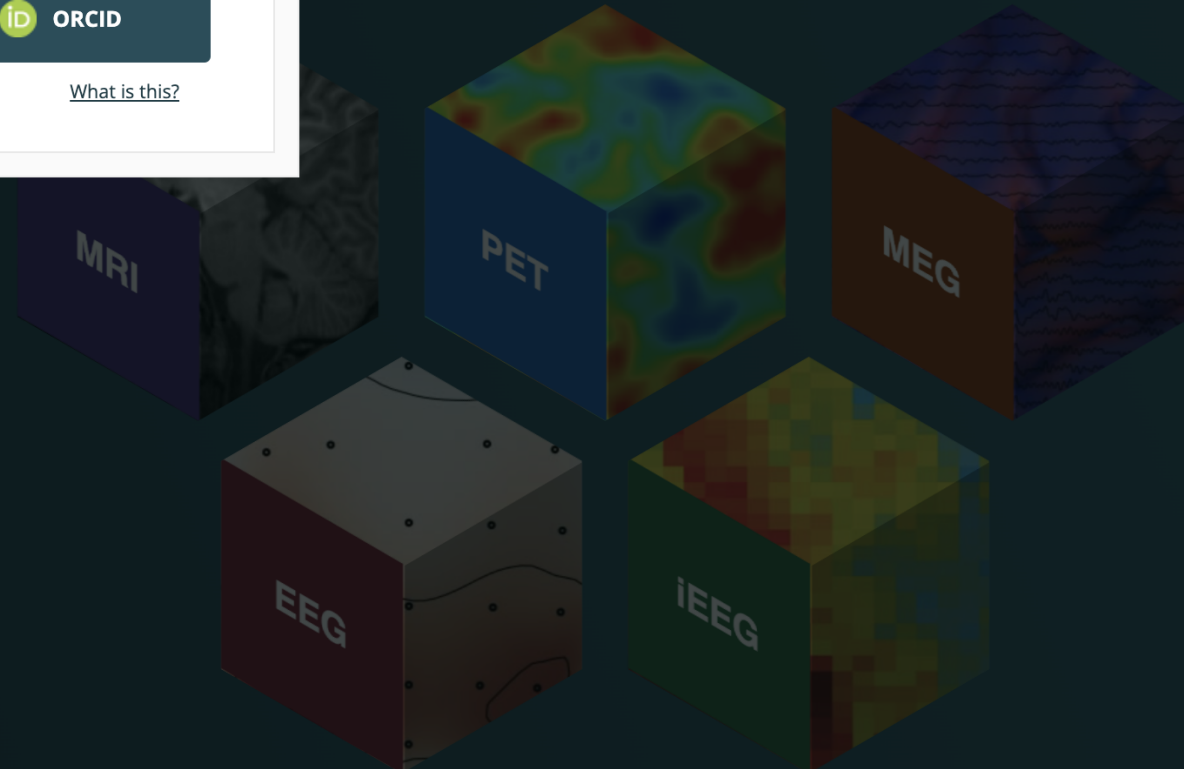
29,064 Participants

761 Public Datasets

Browse by Modalities

Or

Search



OpenNEURO

Sign in



Google



ORCID

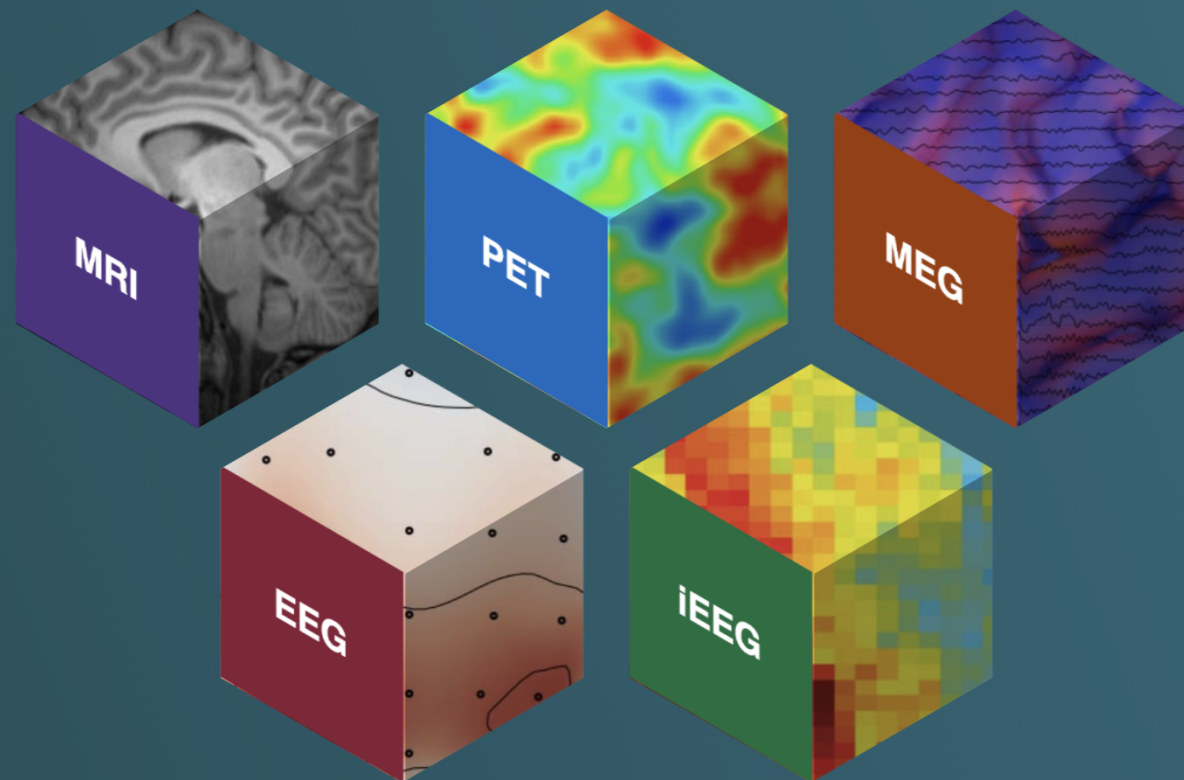
[What is this?](#)

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





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

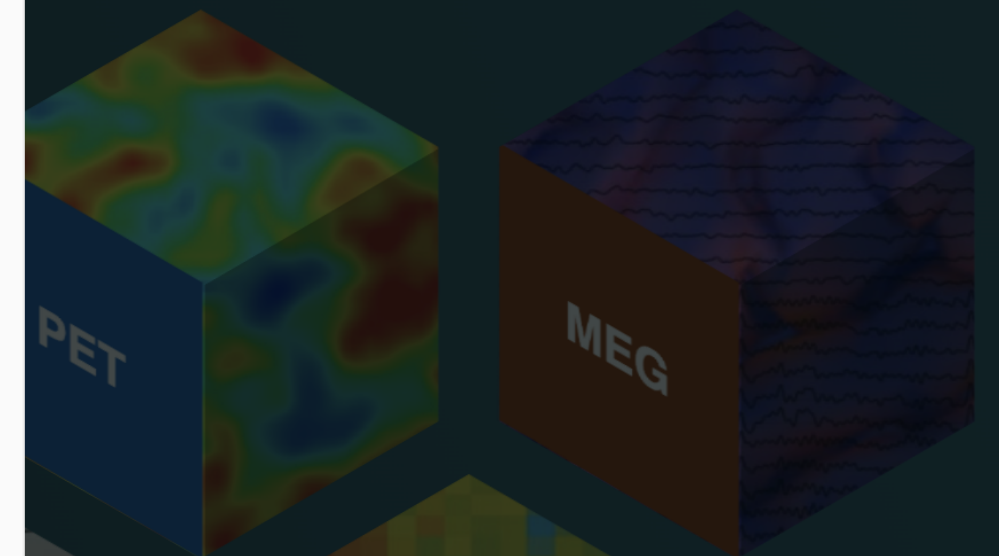
To protect the privacy of the individuals who have been scanned, we require that all scan data be defaced before publishing a dataset.

Select a [BIDS dataset](#) to upload

Select folder

close

My Account



This dataset

BIDS Validation

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 https://poldrack.github.io/talks-OpenScienceEcosystem/ VIEW 1 FILE

Not all subjects/sessions/runs have the same scanning parameters.

My Account

Follow 1 Bookmark 2

Make them public.


Accession Number


03

Russell A. Poldrack

lit

Modalities

 **OpenNEURO**

 **Rhyme judgment** [Edit](#)

This dataset is **BIDS Validated**

[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 data.

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

[My Account](#)

[Follow](#) 1 [Bookmark](#) 2

Make them public.


Neuro Accession Number

03

Russell A. Poldrack

lit

Modalities



**Rhyme judgment** Edit

This dataset is available under a [Creative Commons CC0 license](#).

**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 for the data.

### 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

Make them public.

**OpenNeuro Accession Number**  
03

Russell A. Poldrack

**Modalities**

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



MRI 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.

OpenNeuro Accession Number

ds000003

Authors

Xue, G., Russell A. Poldrack

Available Modalities

MRI

Versions

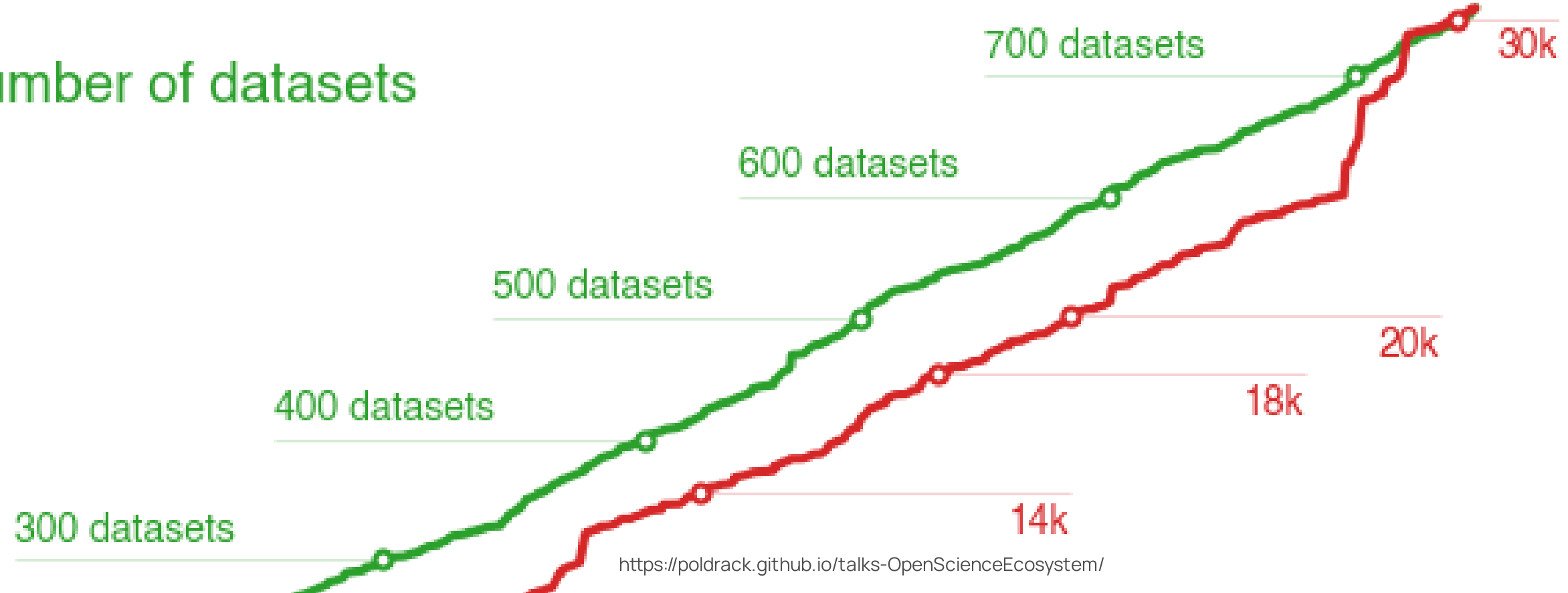
1.0.0

Created: 2020-05-14

Versions ▾

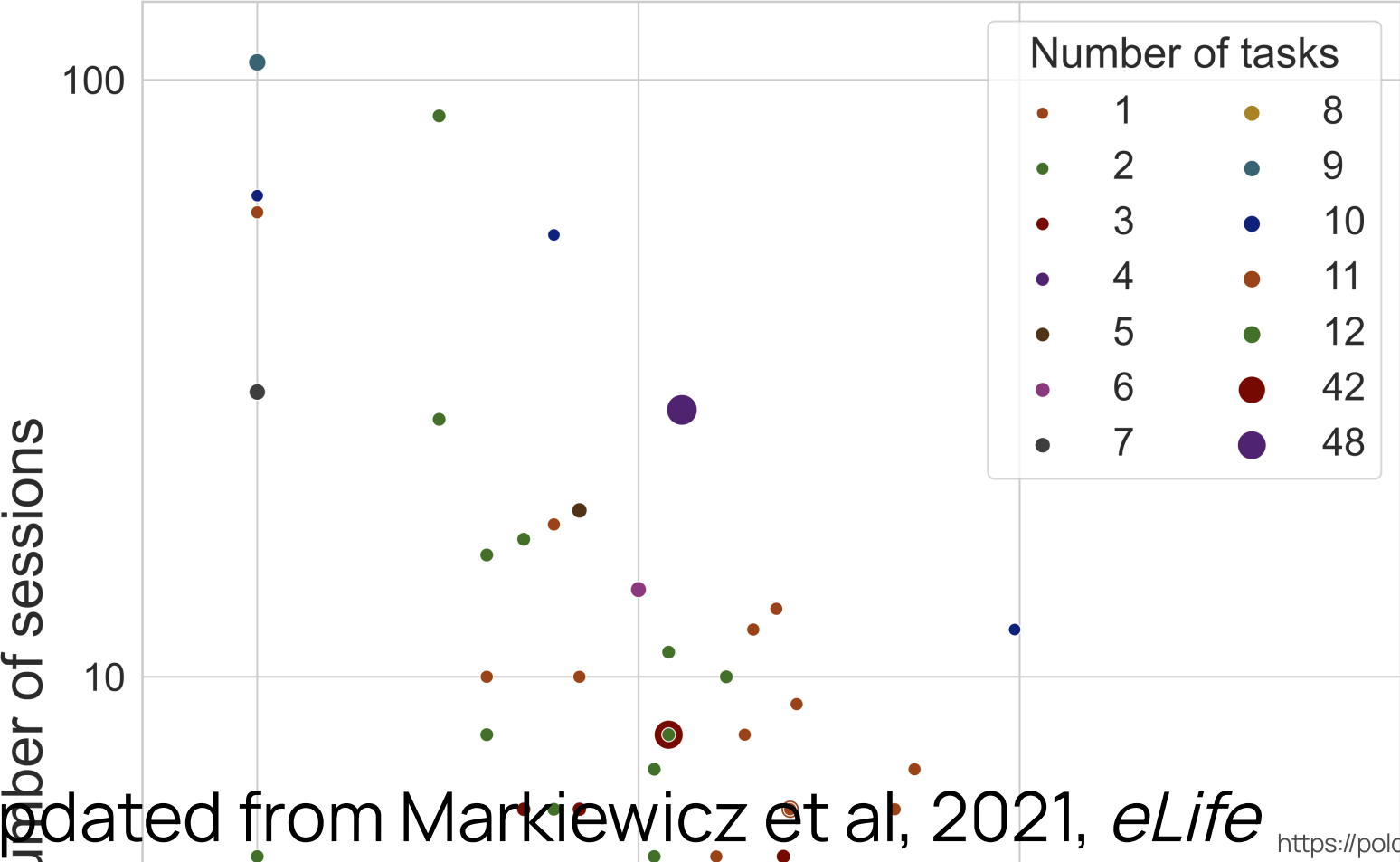
# The growth of OpenNeuro

Total number of datasets



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

# The diversity of OpenNeuro datasets



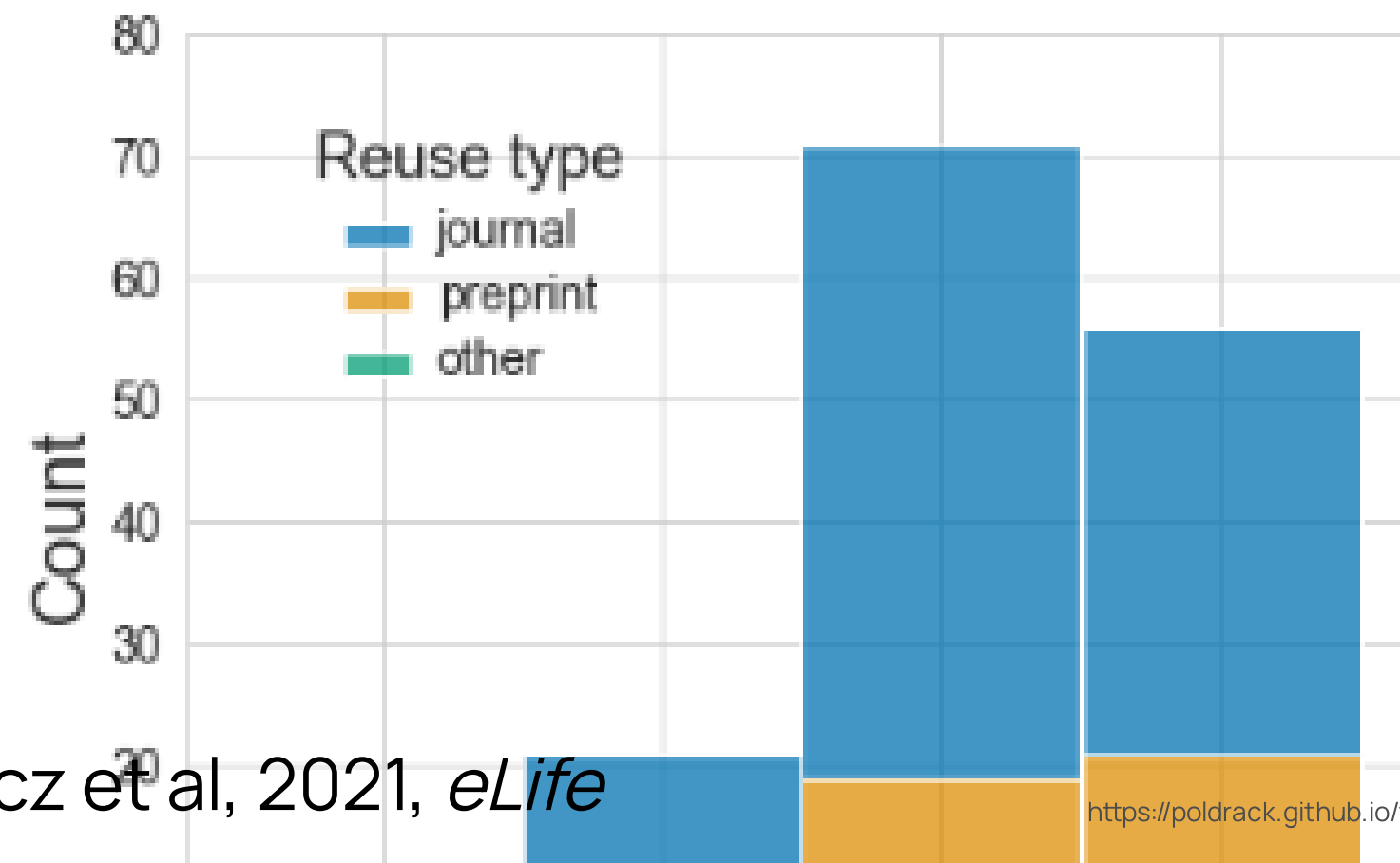
Datatype	#
mri - anat	597
mri - func	521
eeg	120
mri - dwi	67
meg	30
ieeg	17
beh	13

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

updated from Markiewicz et al, 2021, *eLife*

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

# 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

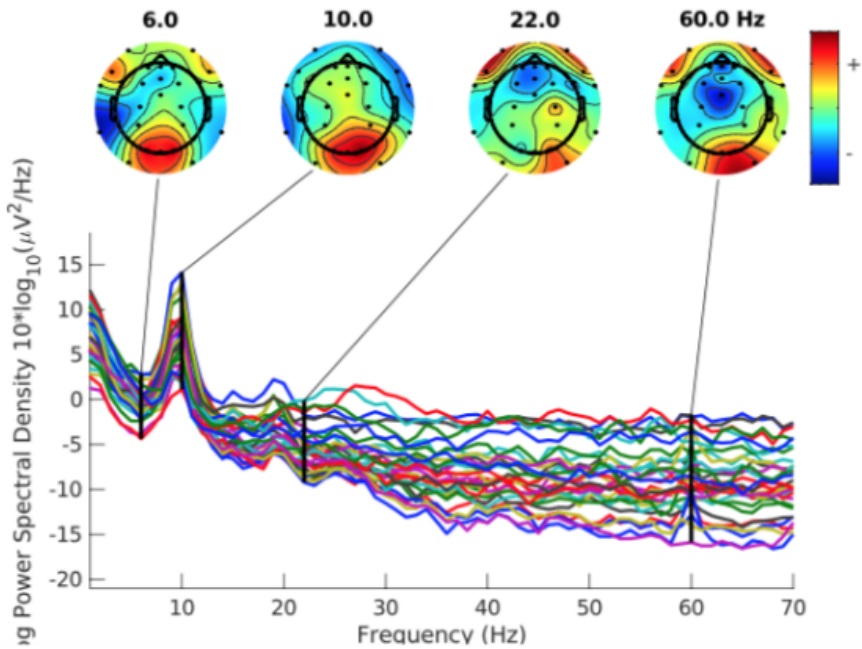
The screenshot shows the brainlife.io interface with a sidebar on the left containing navigation options: DATASETS, PROJECTS, APPS, PUBLICATIONS, and DATATYPES. The main content area is titled 'PUBLIC/PROTECTED PROJECTS' and features a search bar. Below the search bar, there are four project cards:

- HCP 3T / Diffusion:** 1112 sub | 6880 objs (4.51 TB). Includes data types: anat/t1w, transform/nifti, anat/t2w, hcp/freesurferpost, raw, dwi, freesurfer. Description: Human Connectome Project Datasets - Diffusion MRI 3T (1200-subjects-data-).
- HCP 7T / Diffusion:** 150 sub | 300 objs (22.56 MB). Includes data types: anat/t1w, dwi. Description: Human Connectome Project Datasets - Diffusion MRI 3T (184 out of 1200-).
- HCP 3T Retest / Diffusion:** 45 HCP 3T subjects retested. Includes data types: transform/nifti, dwi, hcp/freesurferpost, anat/t2w, freesurfer, anat/t1w.
- O3D:** networkneuro, freesurfer, anat/t1w, wmc-deprecated, dwi, dtiinit, recon, track/tck, LIFE, track/trk. Description: O3D (Open Diffusion Data and Derivative) is a reference repository for precision.

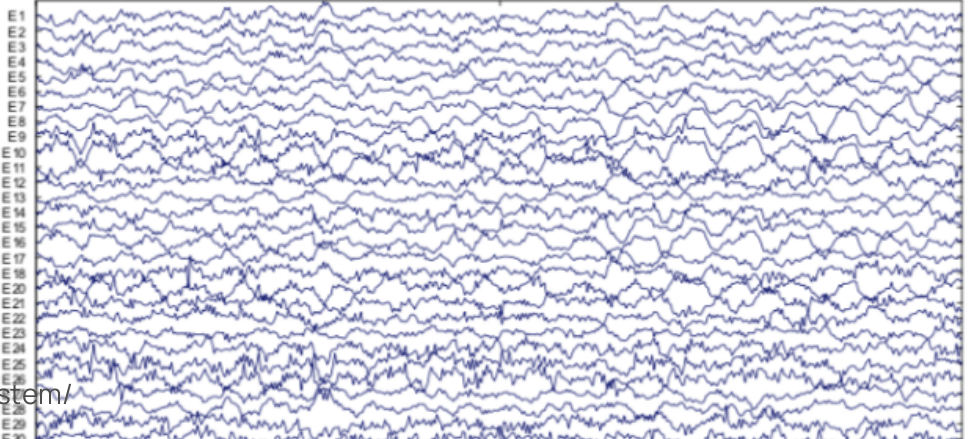
## Data Summary

**Subject:** sub-001  
**Session(s):** 1, **Run(s):** 1  
**Data size:** 32 channels, 298k frames  
**Acceptable scalp channels:** 100.0% (32 of 32) ⓘ  
**Acceptable data points channels:** 90.9% (271k of 298k) ⓘ  
**Source quality metric based on independent component:** 48.4% ⓘ

## Scalp channel log spectra



## Sample scalp channel data (mid 2 seconds)



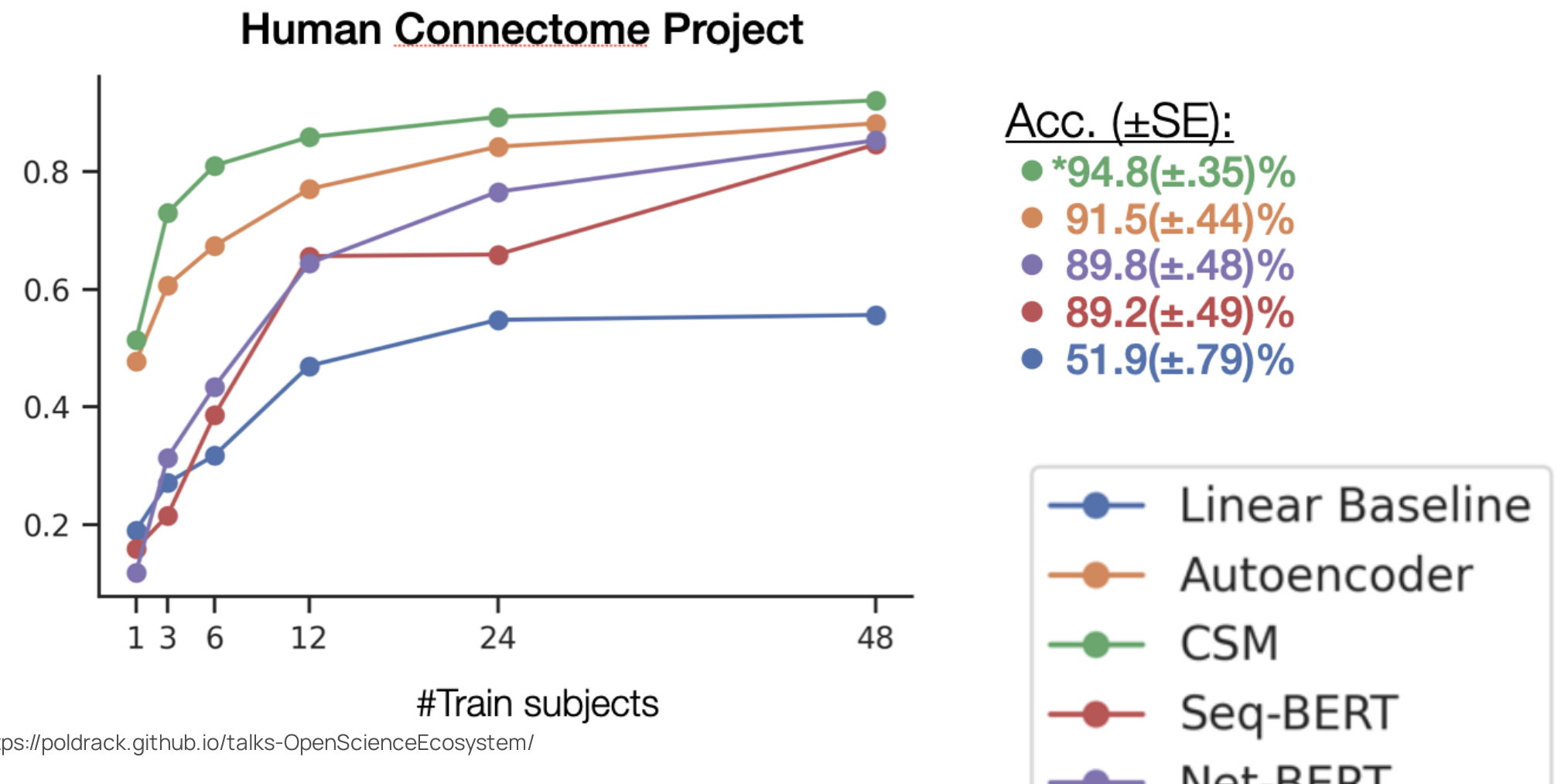
NEM

ta

# Example of OpenNeuro reuse

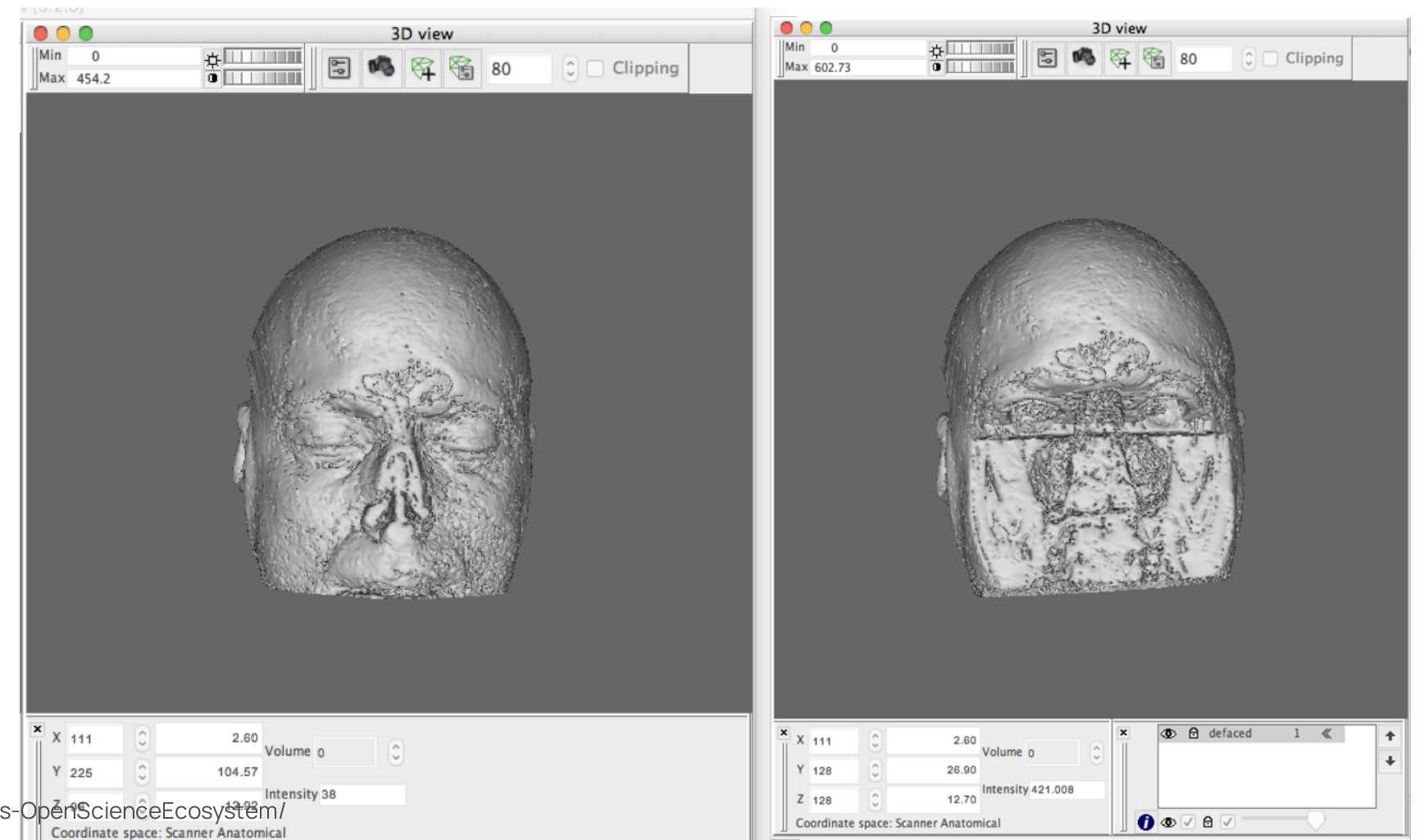
- A challenge for decoding brain activity from fMRI data is that most datasets are very small
- We used OpenNeuro to train a “foundation model”
  - A pre-trained model that can be used as a starting point for decoding models on smaller datasets

• 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*

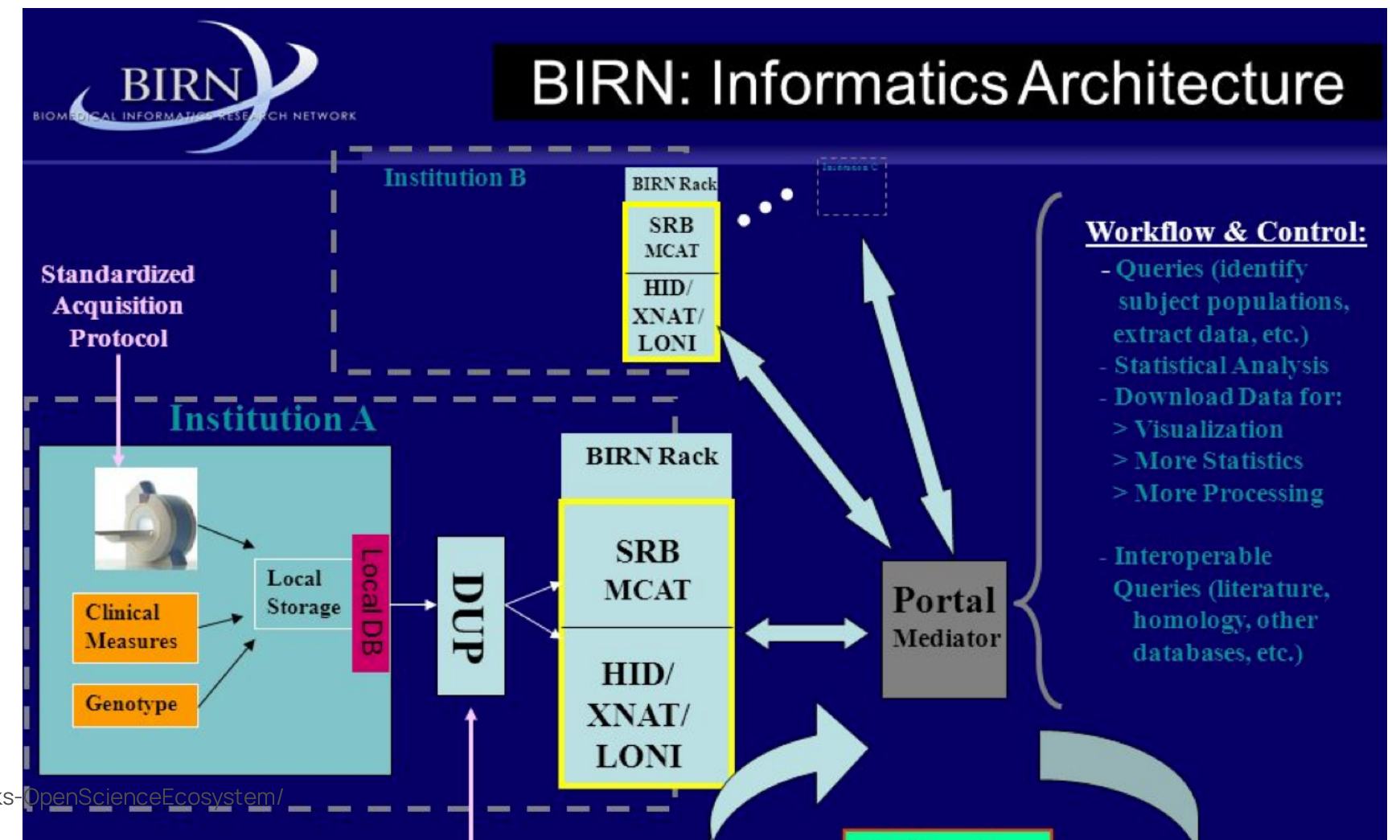


# Lessons learned

- 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

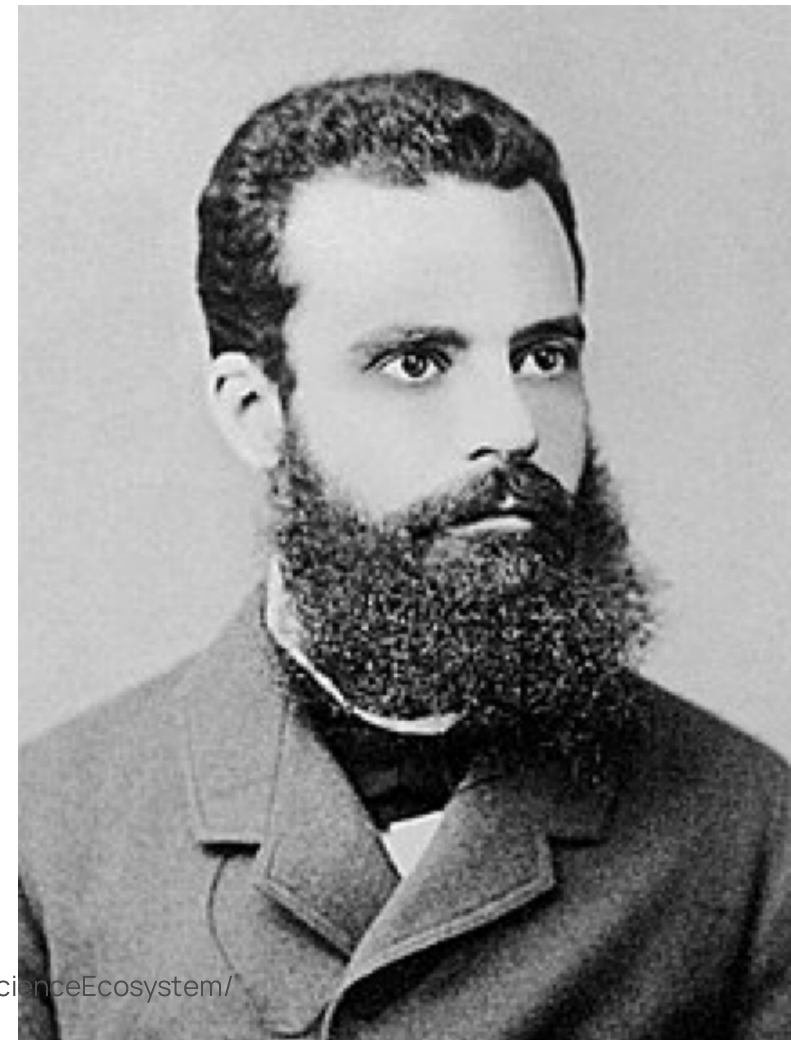
# Lessons learned

- 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

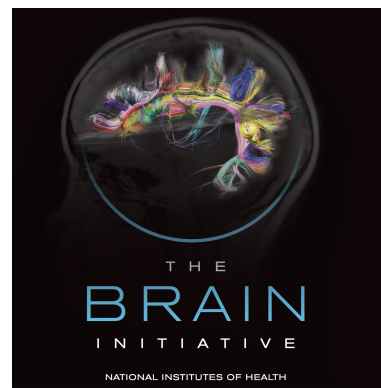
# The Poldrack Lab



# OpenNeuro Team



# Funding



# Collaborators



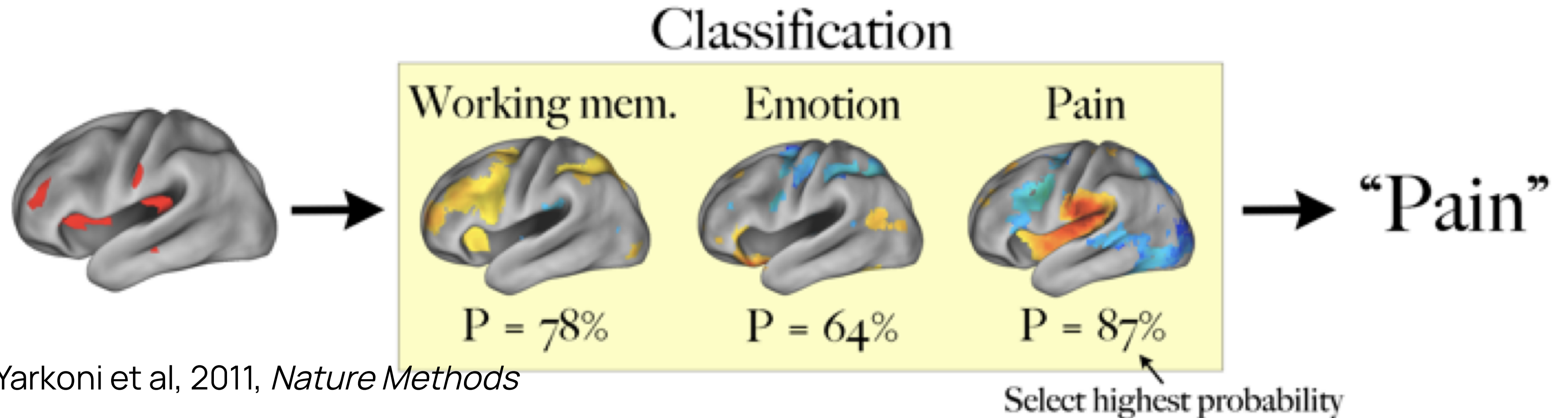
National Institute  
of Mental Health





# 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

