

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

<https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html>

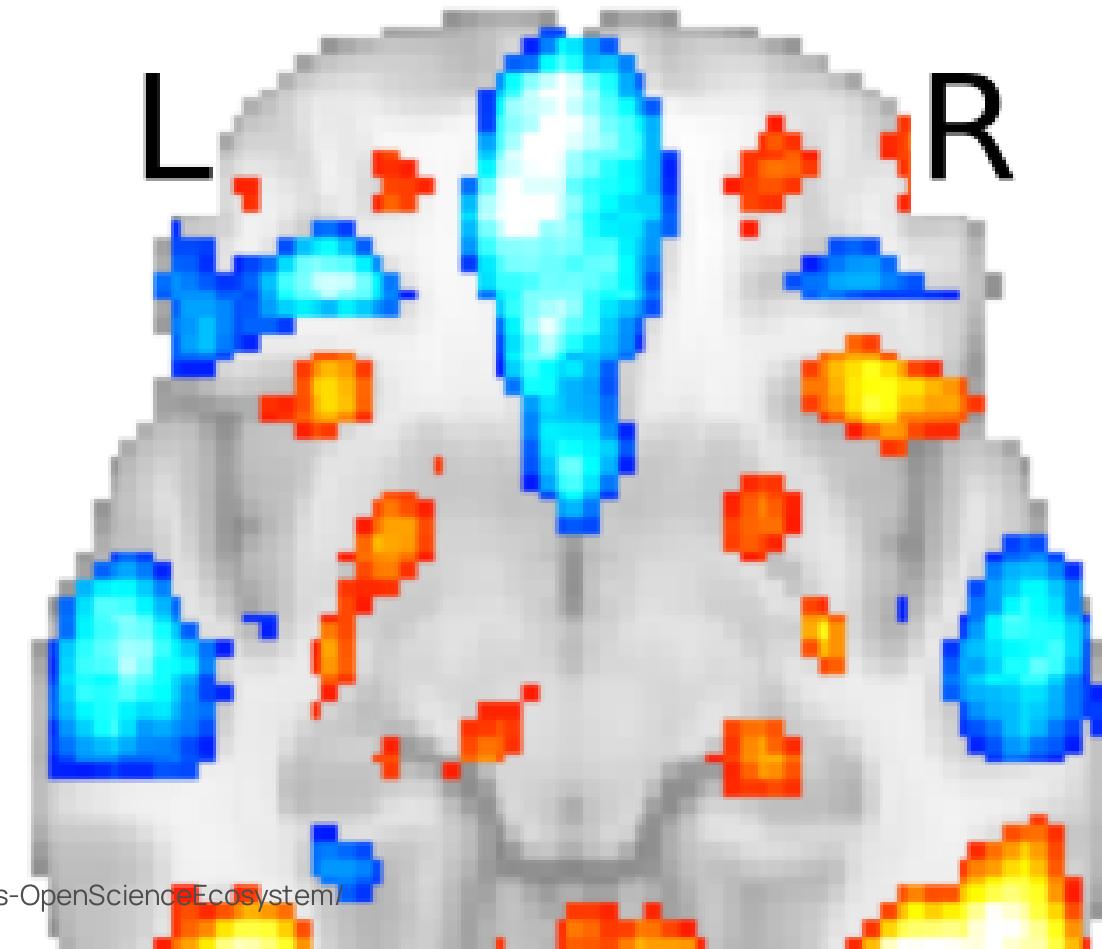
“we can distill Claerbout’s insight into a slogan:

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



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

fMRI DC

SEARCH fMRI DC Database FOR SUBMIT

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

▶ HOME

DATABASE

SUBMISSIONS

RESOURCES

DATA MANAGEMENT TOOL

HELP

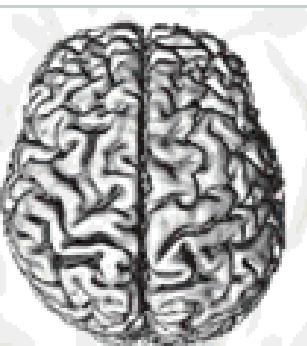
ABOUT US

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

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

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

[Information for Authors](#)
How to submit your imaging data to the Data



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**

[fMRI DC NEWS](#)
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 DC PI Receives Honor
[https://poldrack.github.io/talks/OpenScienceEcosystem/The_PALS_Brain_Atlas_from_WUSTL](#)
October 17, 2005 - A new fMRI DC

Sitemap 

Contact Us

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?

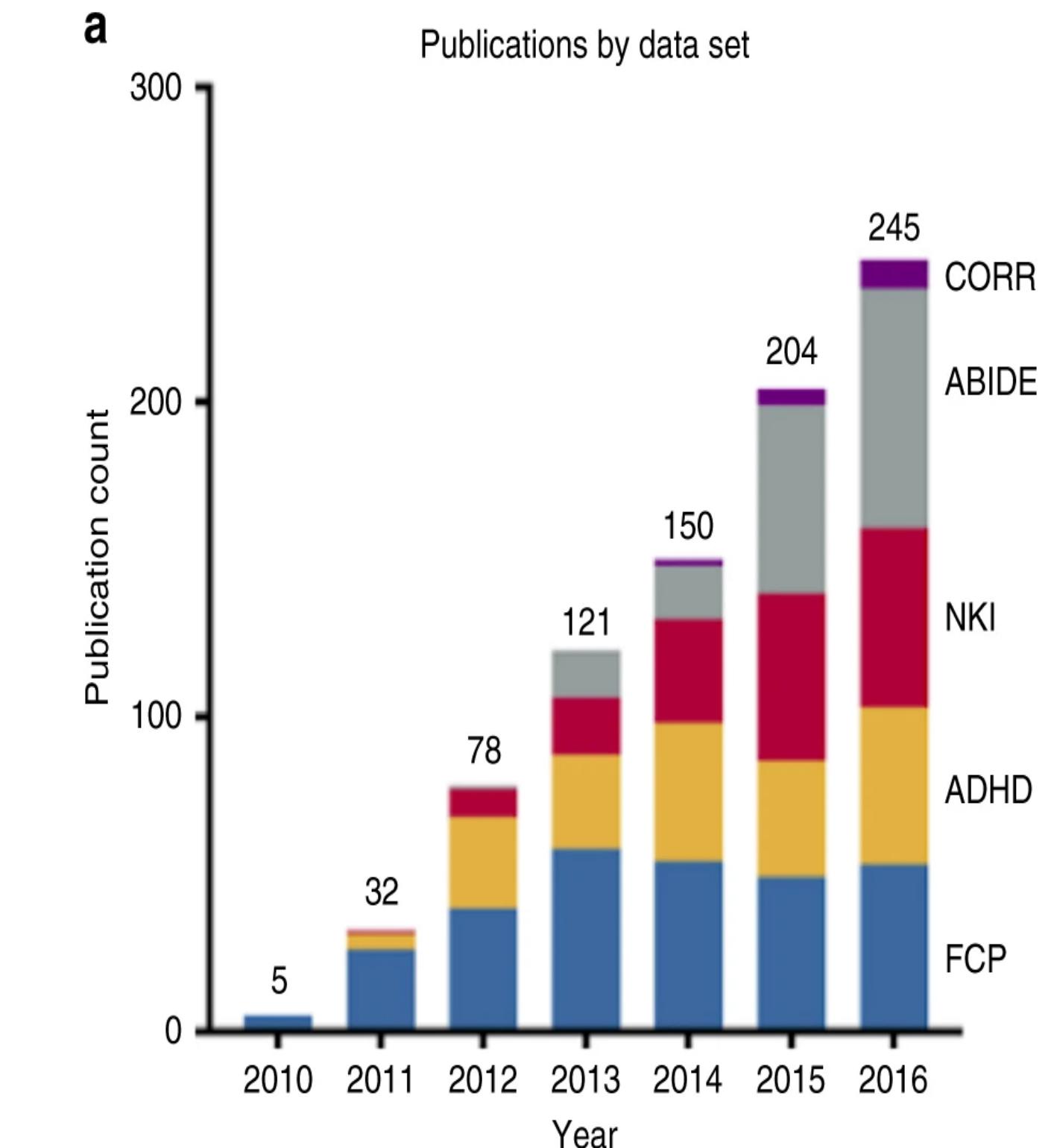
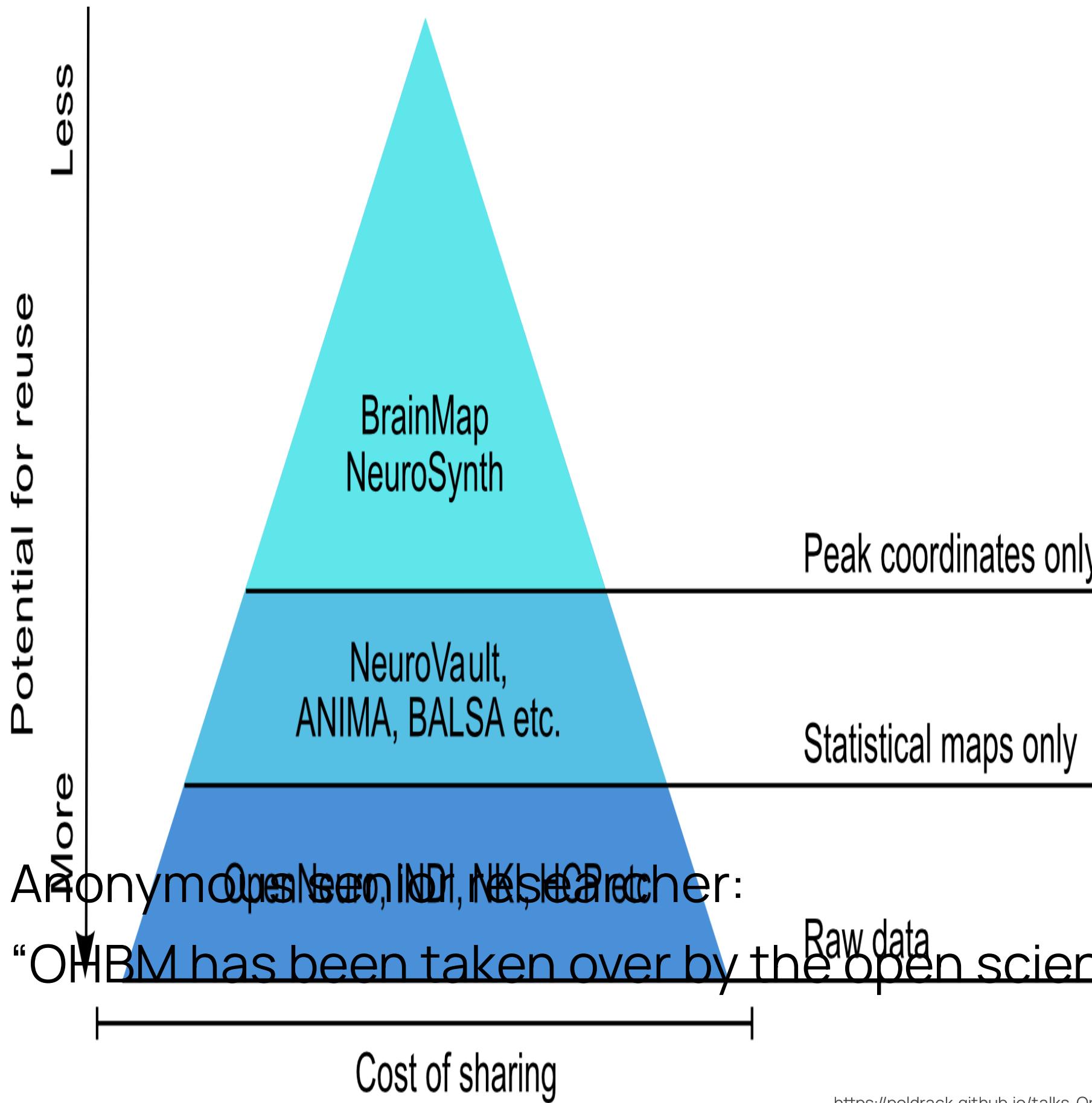
This letter comes from a group of scientists who are publishing papers using fMRI to understand the links between brain and behavior. We are writing in reaction the recent announcement of the creation of the National fMRI Data Center (www.fmridc.org). In the letter announcing the creation of the center, 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 authors to submit fMRI data to

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 Goel^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}

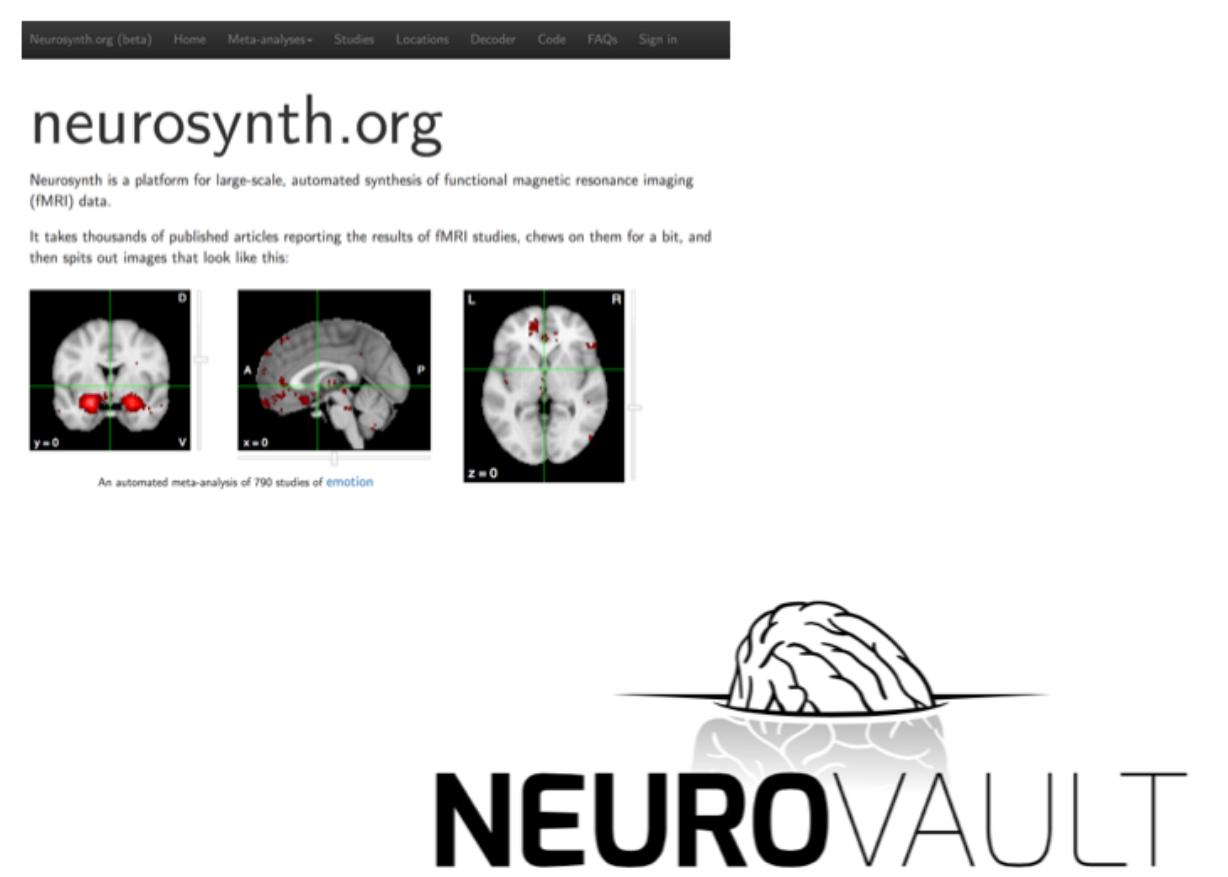
Data sharing is becoming the norm in neuroimaging



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

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 part of the public domain.

Neurosynth: Sharing activation coordinates

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

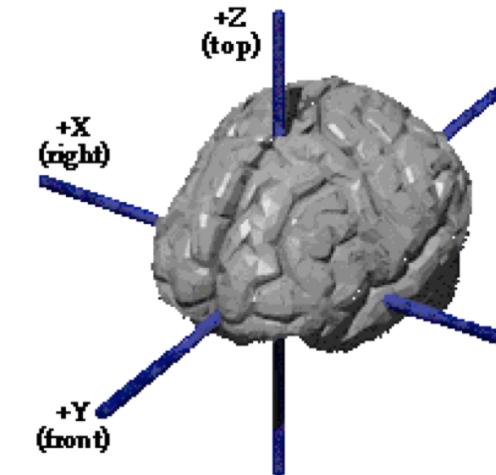


Table 1
Regions that showed a condition × time interaction in the ANOVA analysis

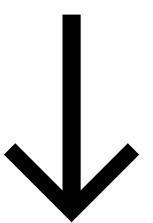
No.	Region	Hemisphere	BA	x	y	z	mm ³
1	Middle/superior temporal gyrus	L	21/22/37	-52	-54	9	13257
2	Inferior frontal gyrus	L	45/46/9	-49	26	6	2781
3	Posterior cerebellum	L		-19	-79	-38	2214
4	Dorsomedial PFC	L	9/8	-11	42	47	3051
5	Left anterior PFC	L	10	-37	49	15	2025
6	Inferior parietal cortex	L	40/7	-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
13	Right anterior PFC	R		10	38	42	5022

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* (SumsDB)
against manual database

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



- Meta-analytic maps created for each

working memory

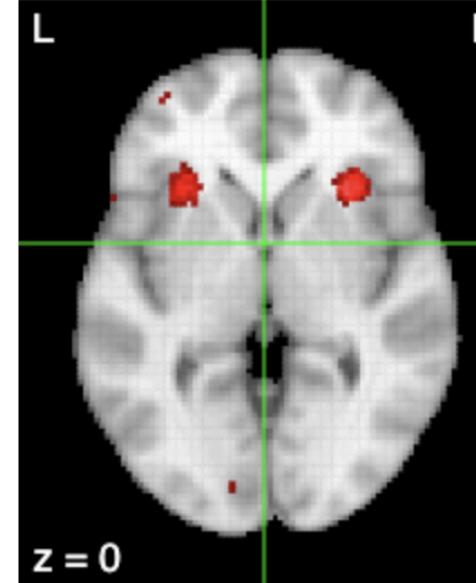
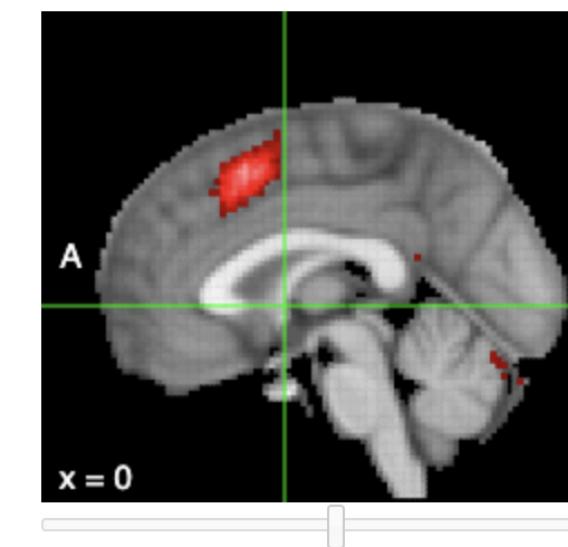
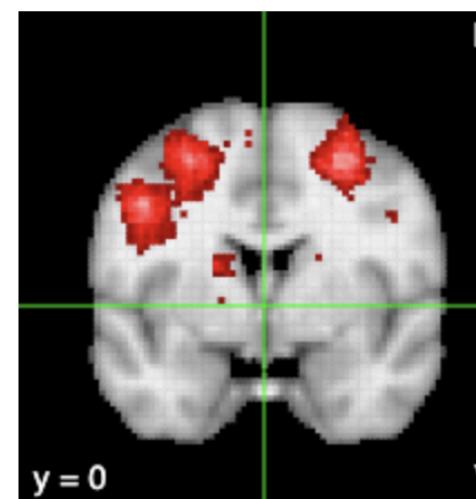
An automated meta-analysis of 1091 studies

Search for another term:

Maps

Studies

FAQs



z-score: 0

What's here?

X: 0 Y: 0 Z: 0

Layers

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

Color palette:

red

Crosshairs

Positive/Negative:

positive

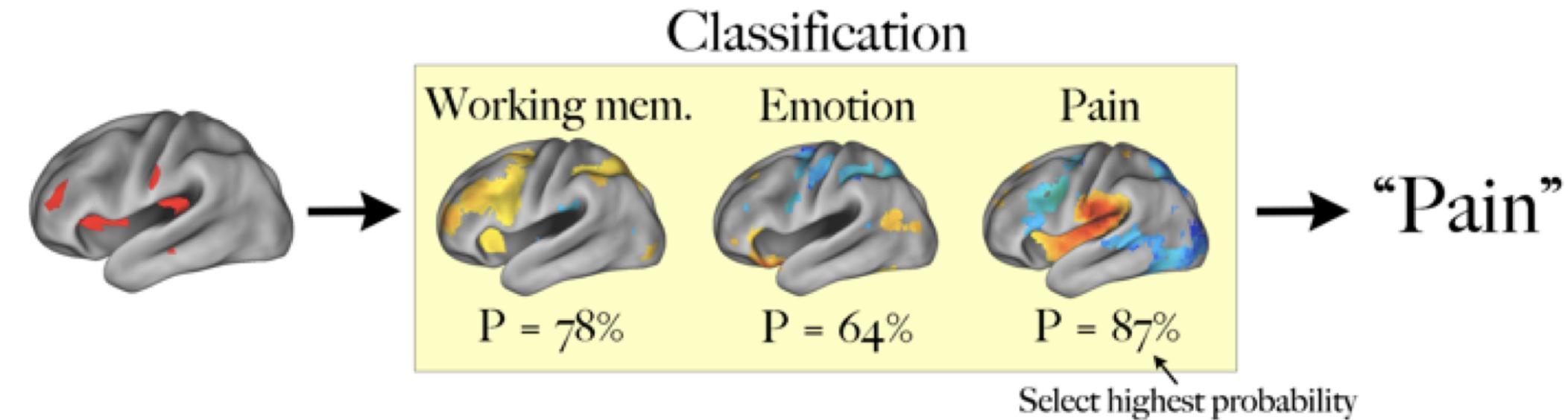
Labels

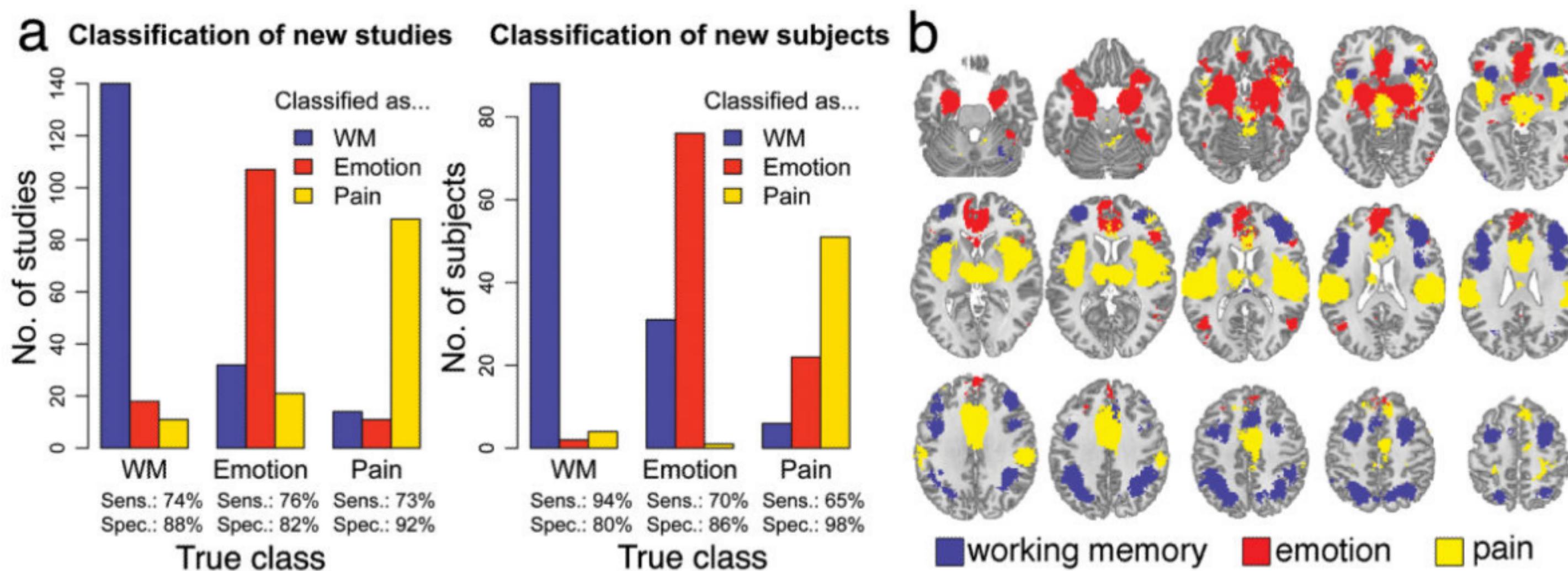
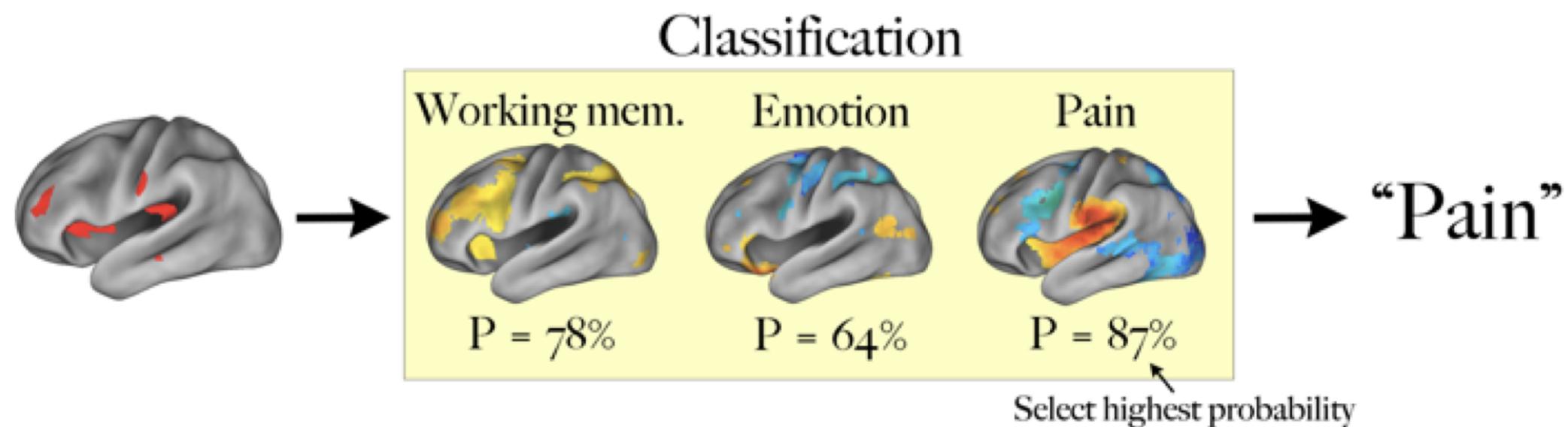
Thresholds:

0 0

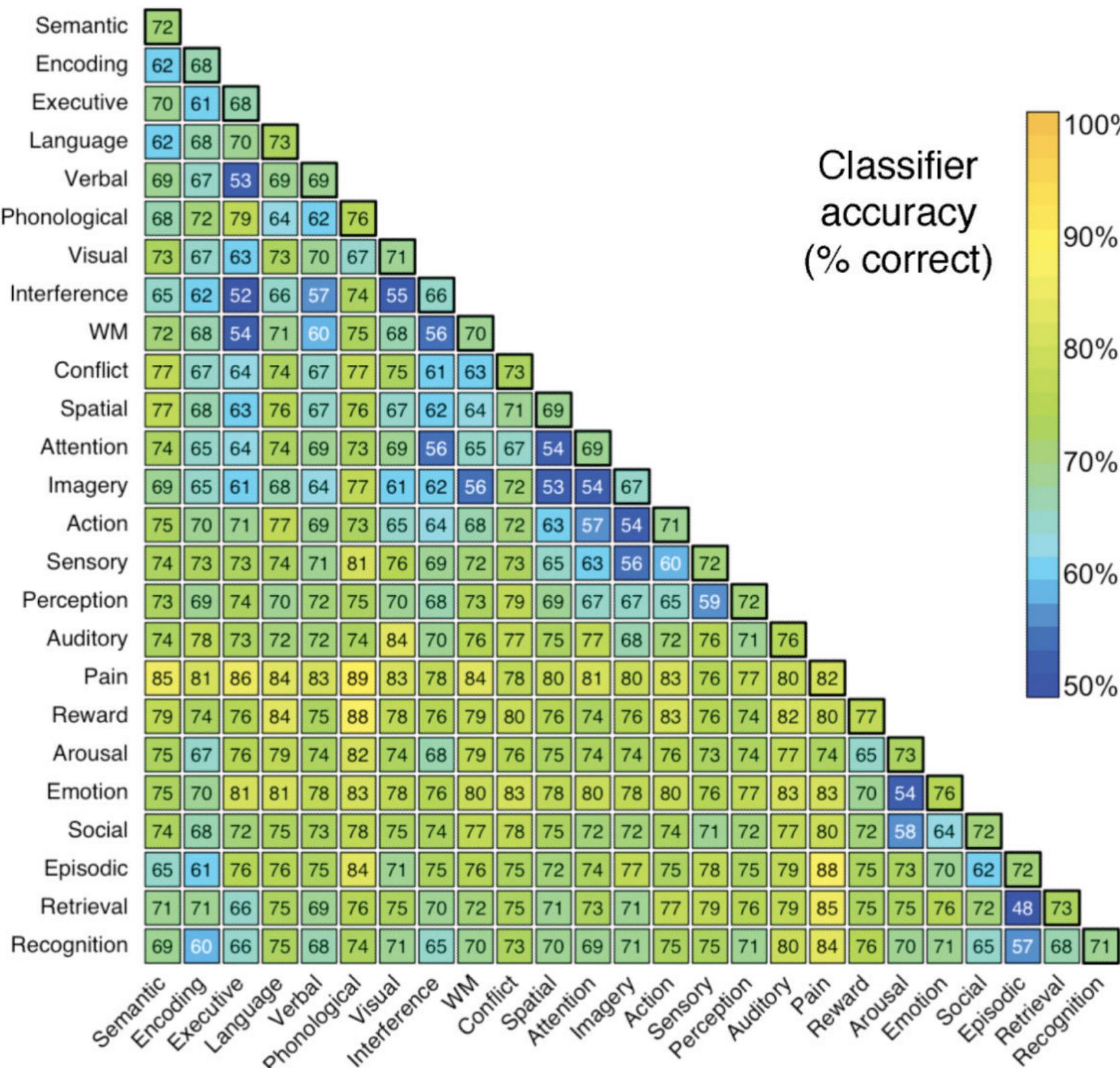
Opacity:

1



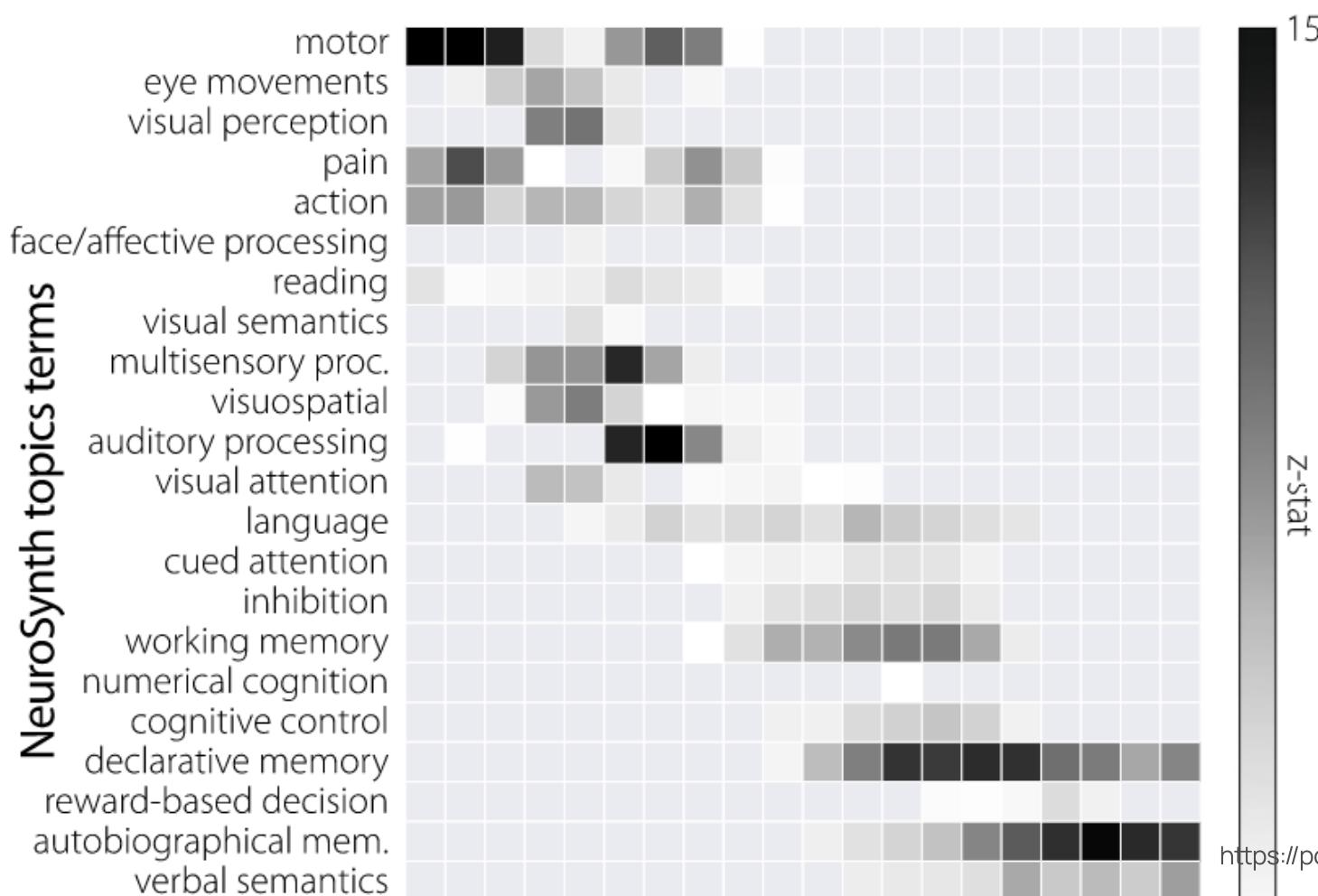


Decoding brain activity patterns using Neurosynth



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}

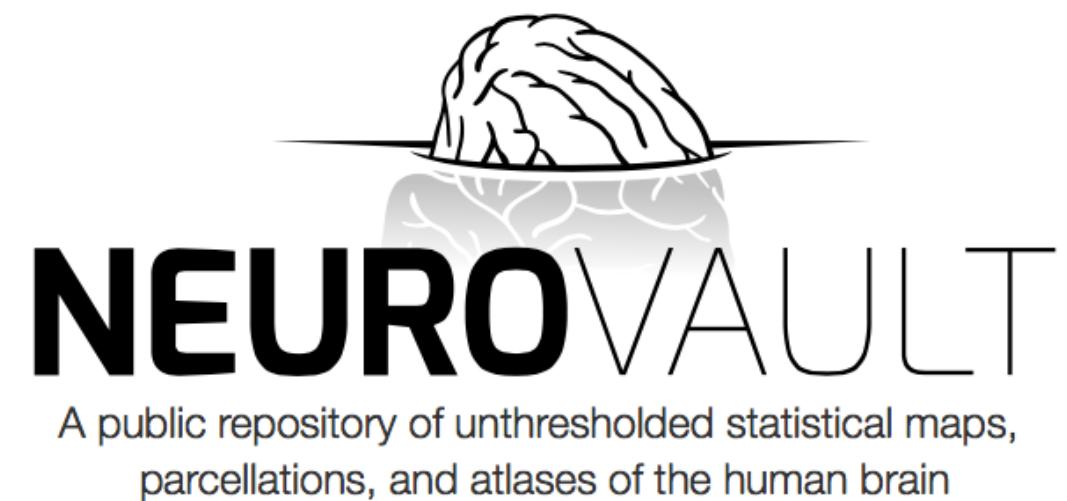


- Identified gradients of functional organization across the cortex
- Used Neurosynth to identify the most common terms associated with each gradient

Neurovault: Sharing neuroimaging results

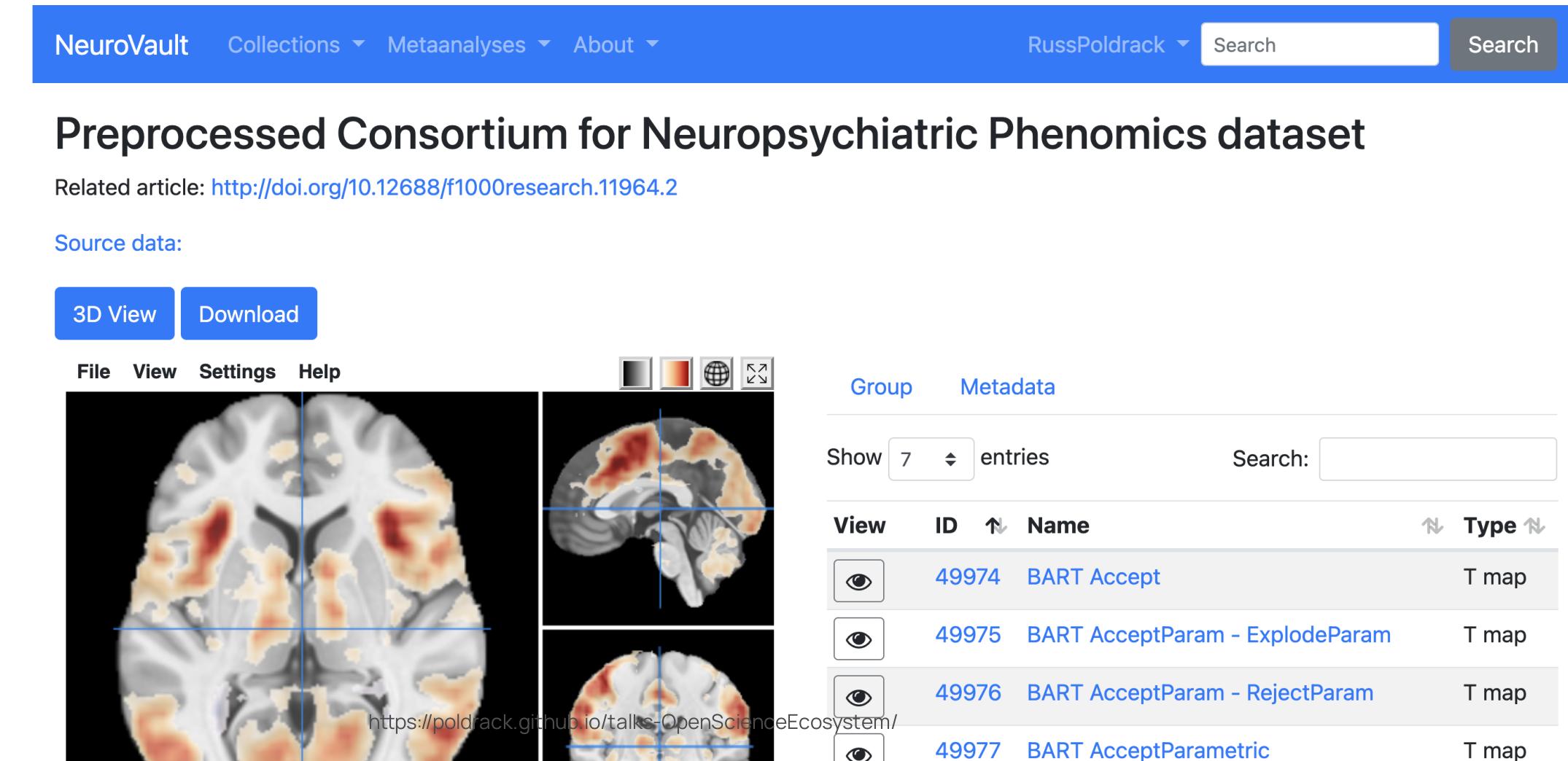
- The results of most neuroimaging studies are images with statistical estimates at each voxel
- Neurovault.org is an open archive for these results

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



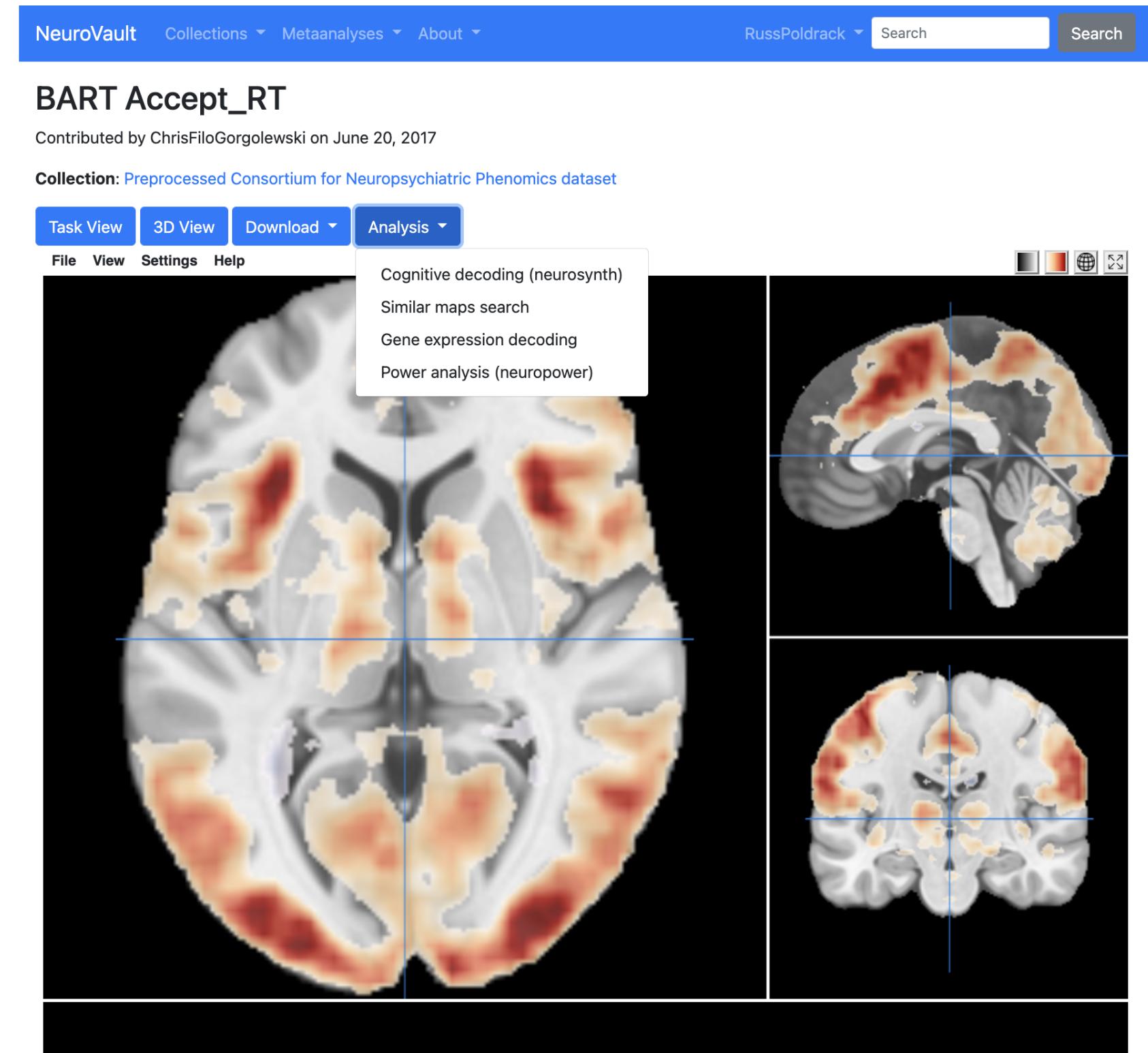
- **Collections**

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



- **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

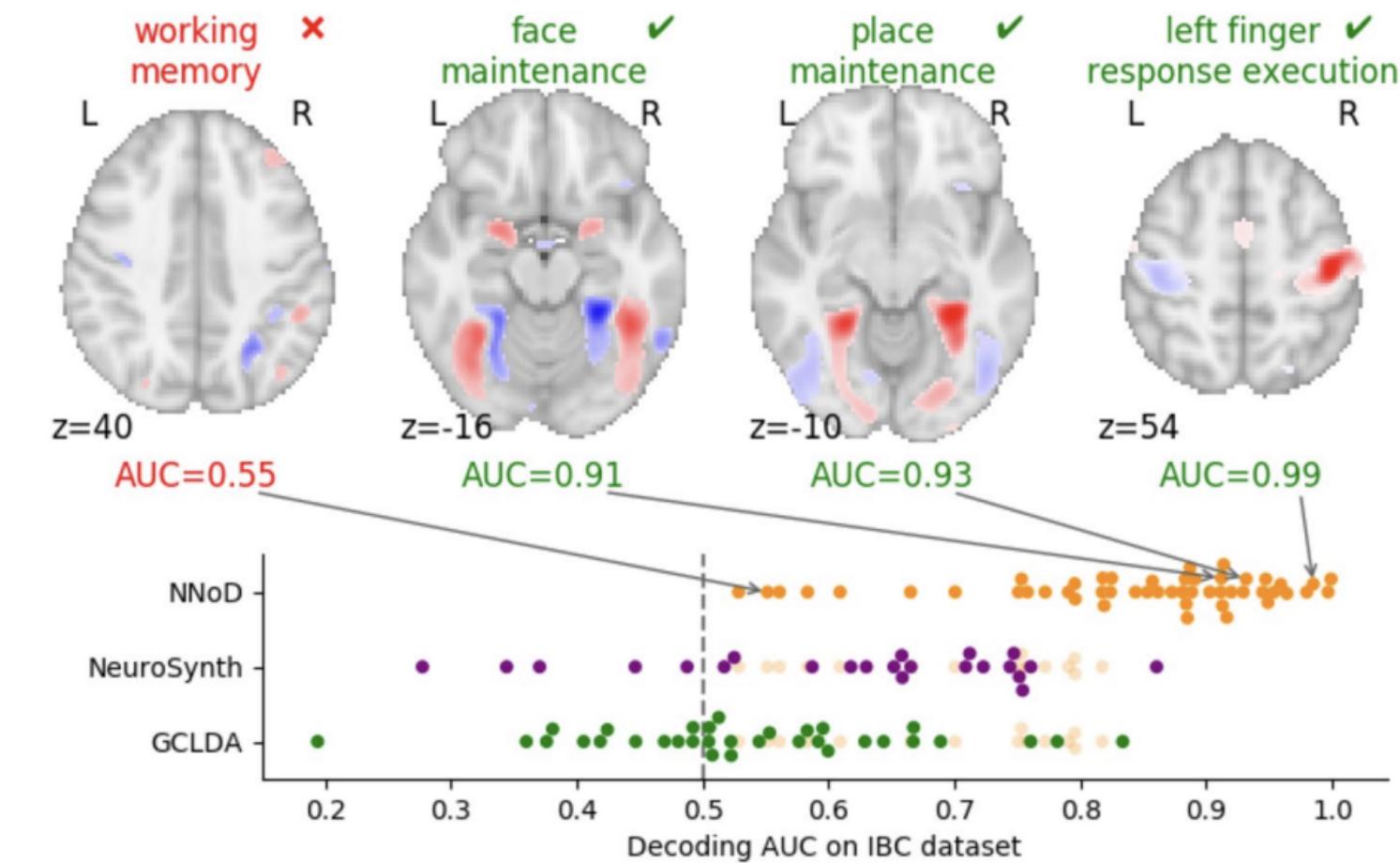
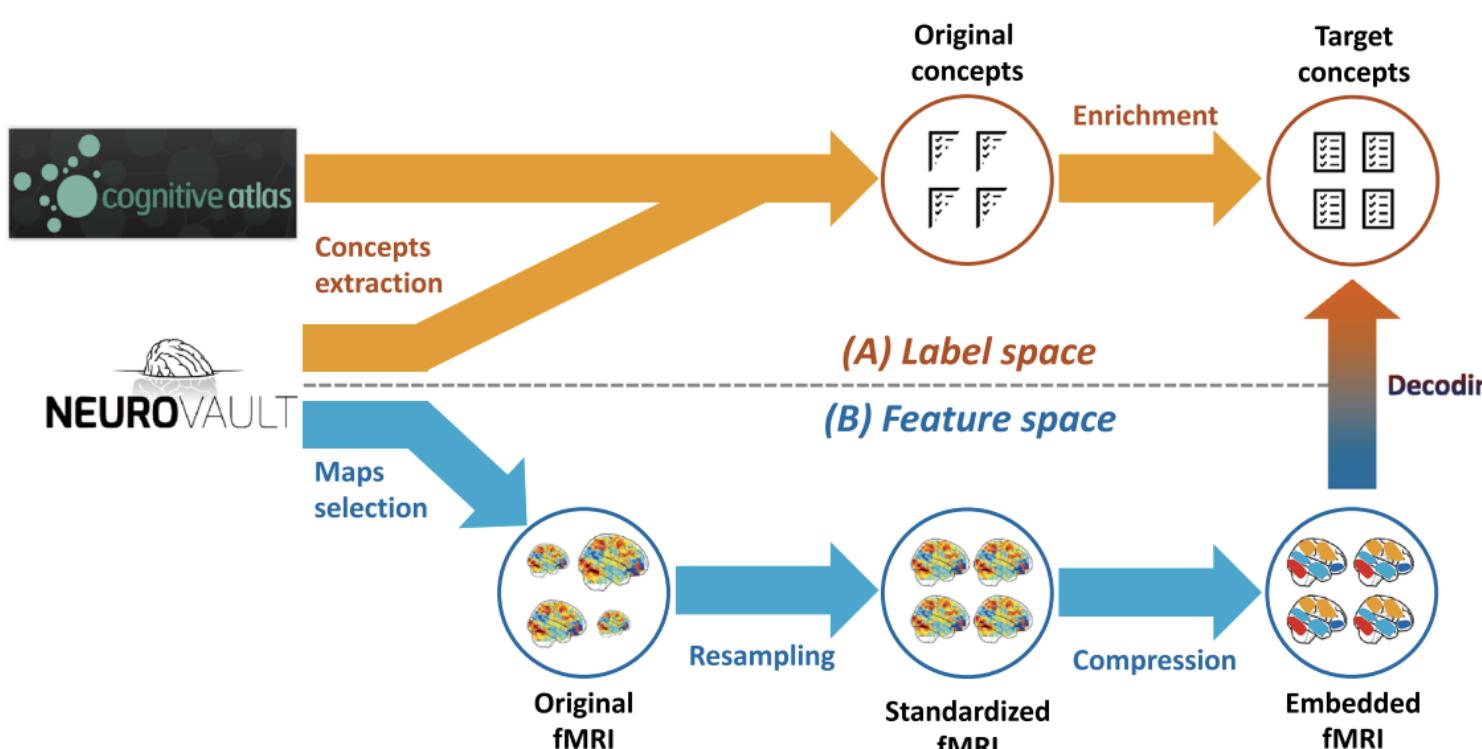


Example of Neurovault usage

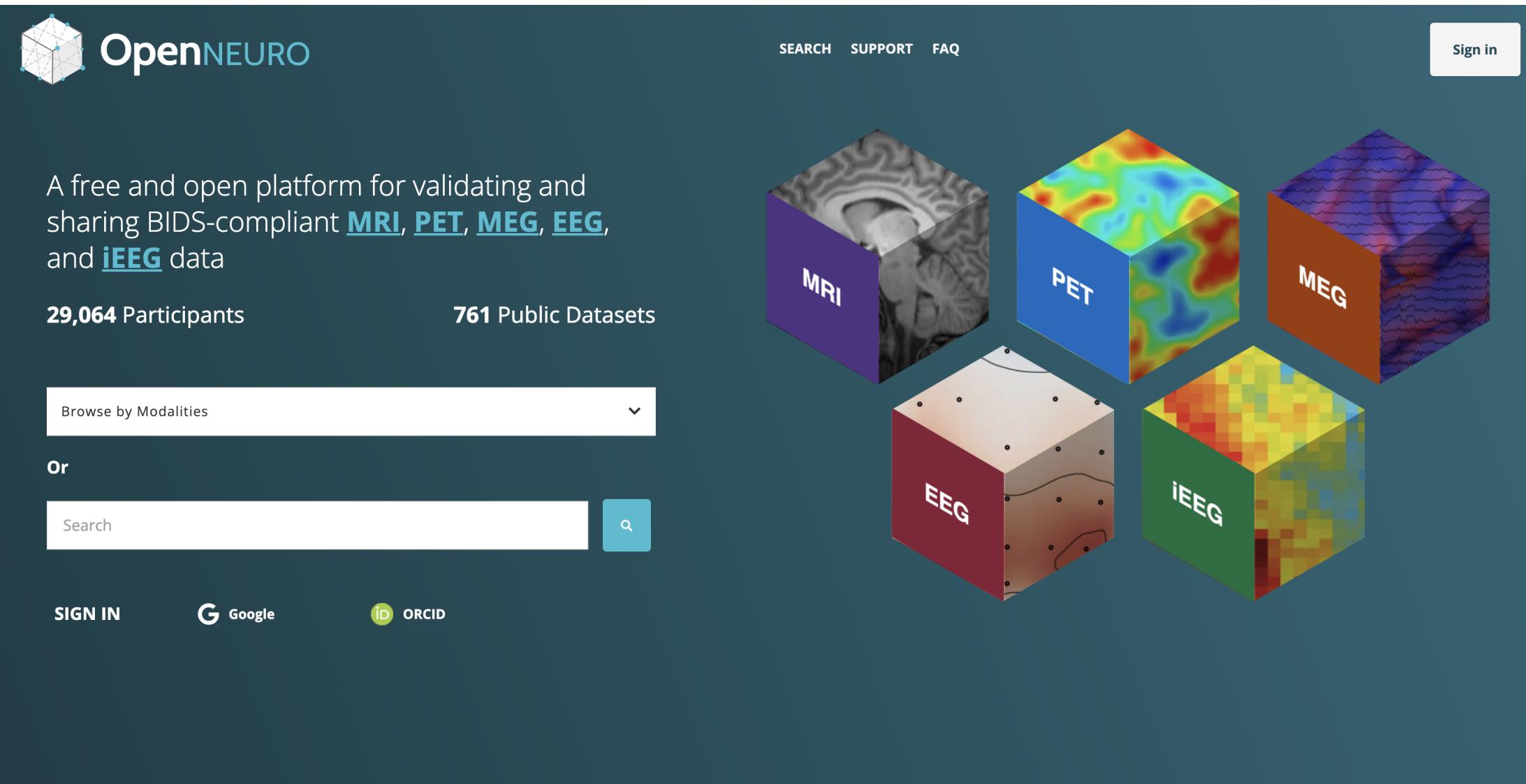
Comprehensive decoding mental processes from Web repositories of functional brain images

Romuald Menet^{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.



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

- 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

Received: 18 December 2015

Accepted: 19 May 2016

Published: 21 June 2016

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

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

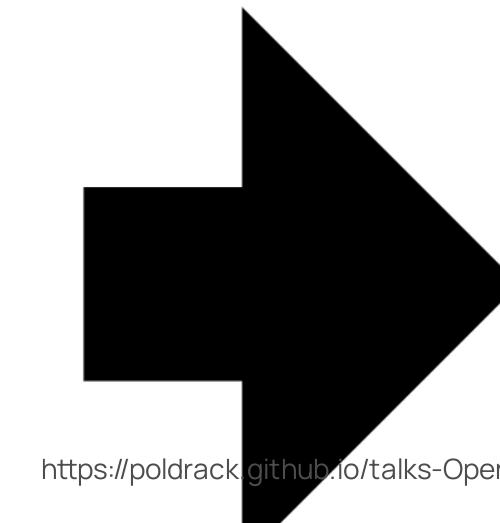
- January 2015
 - Initial stakeholder meeting at Stanford (funded by INCF)
 - Initiated development of a draft standard
- September 2015
 - Draft standard posted to BIDS web site with 22 example datasets
 - Solicited feedback from community
- June 2016

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
 sub-01_dwi.json

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

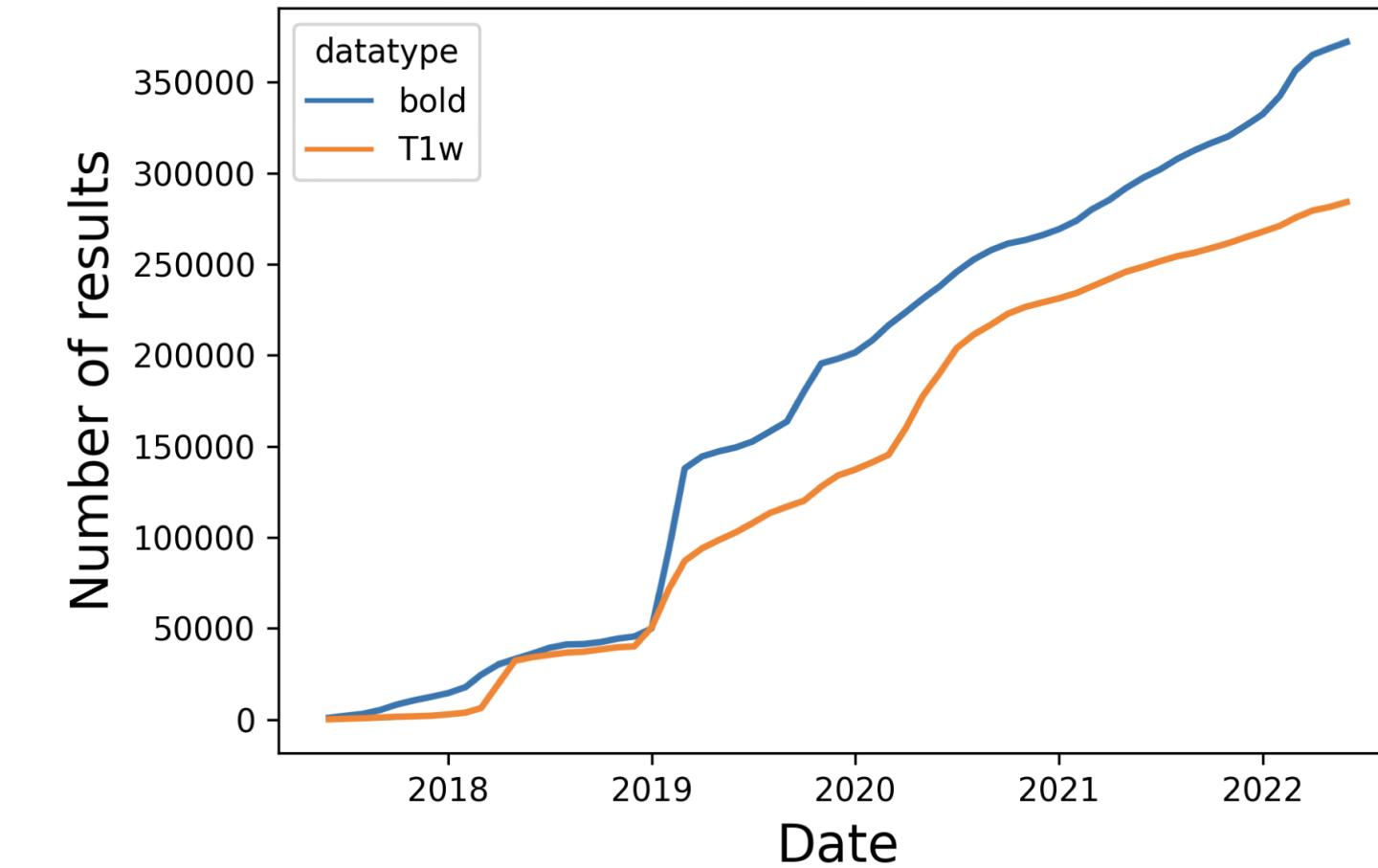
BEP008 Magnetoencephalography (MEG)

The growing usage of BIDS: An example

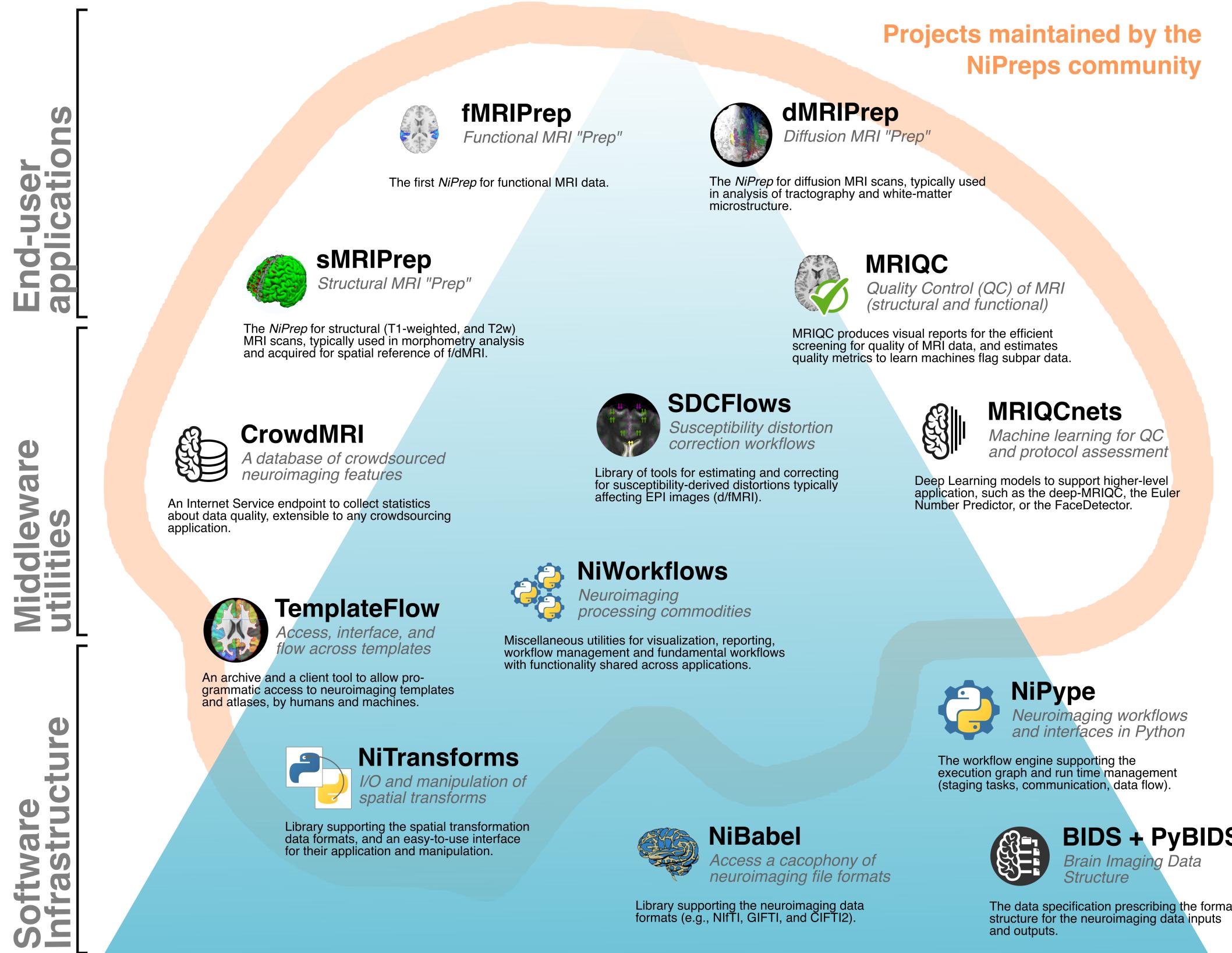
SCIENTIFIC DATA 

OPEN
DATA DESCRIPTOR
Received: 19 September 2018
Accepted: 12 March 2019
Crowdsourced MRI quality metrics and expert quality annotations for training of humans and machines
Oscar Esteban  ¹, Ross W. Blair  ¹, Dylan M. Nielson  ², Jan C. Varada  ³, Sean Marrett  ³, Adam G. Thomas  ², Russell A. Poldrack  ¹ & Krzysztof J. Gorgolewski  ¹

- 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



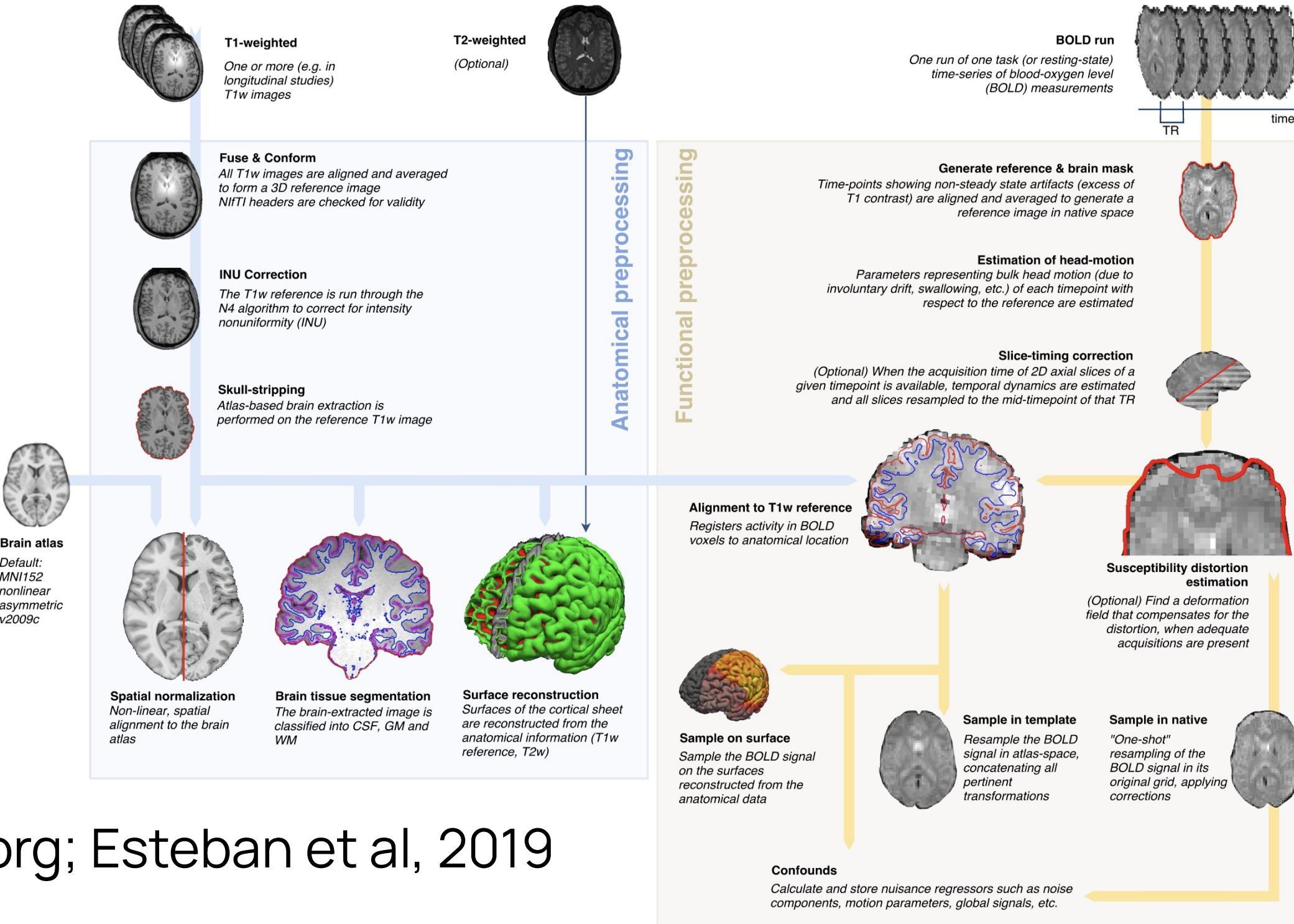
BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods

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

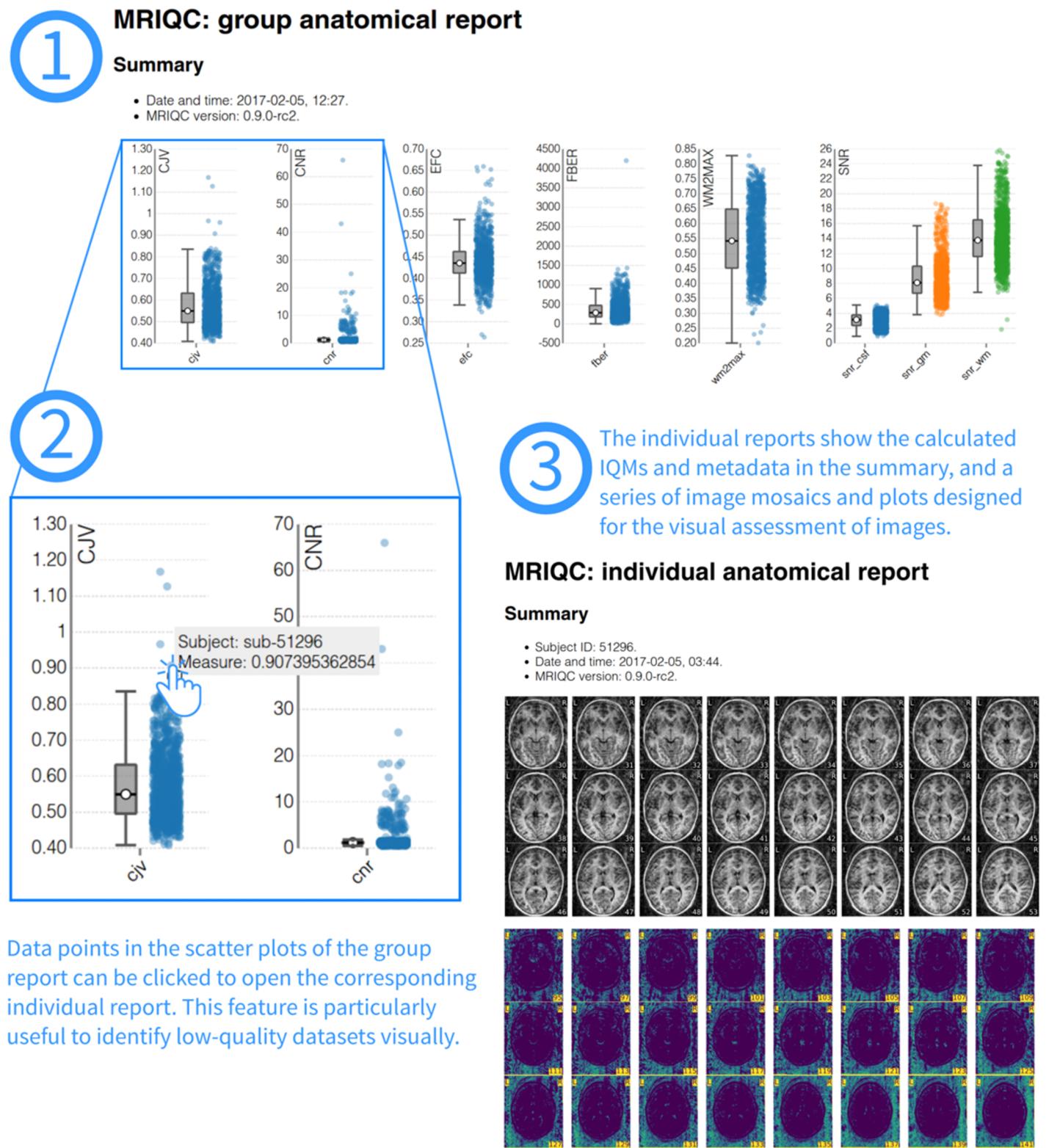
- Containerized applications that can be run on a BIDS dataset
 - Containers provide ease of use as well as better reproducibility

fMRIprep: Robust preprocessing of fMRI data



fmriprep.org; Esteban et al, 2019

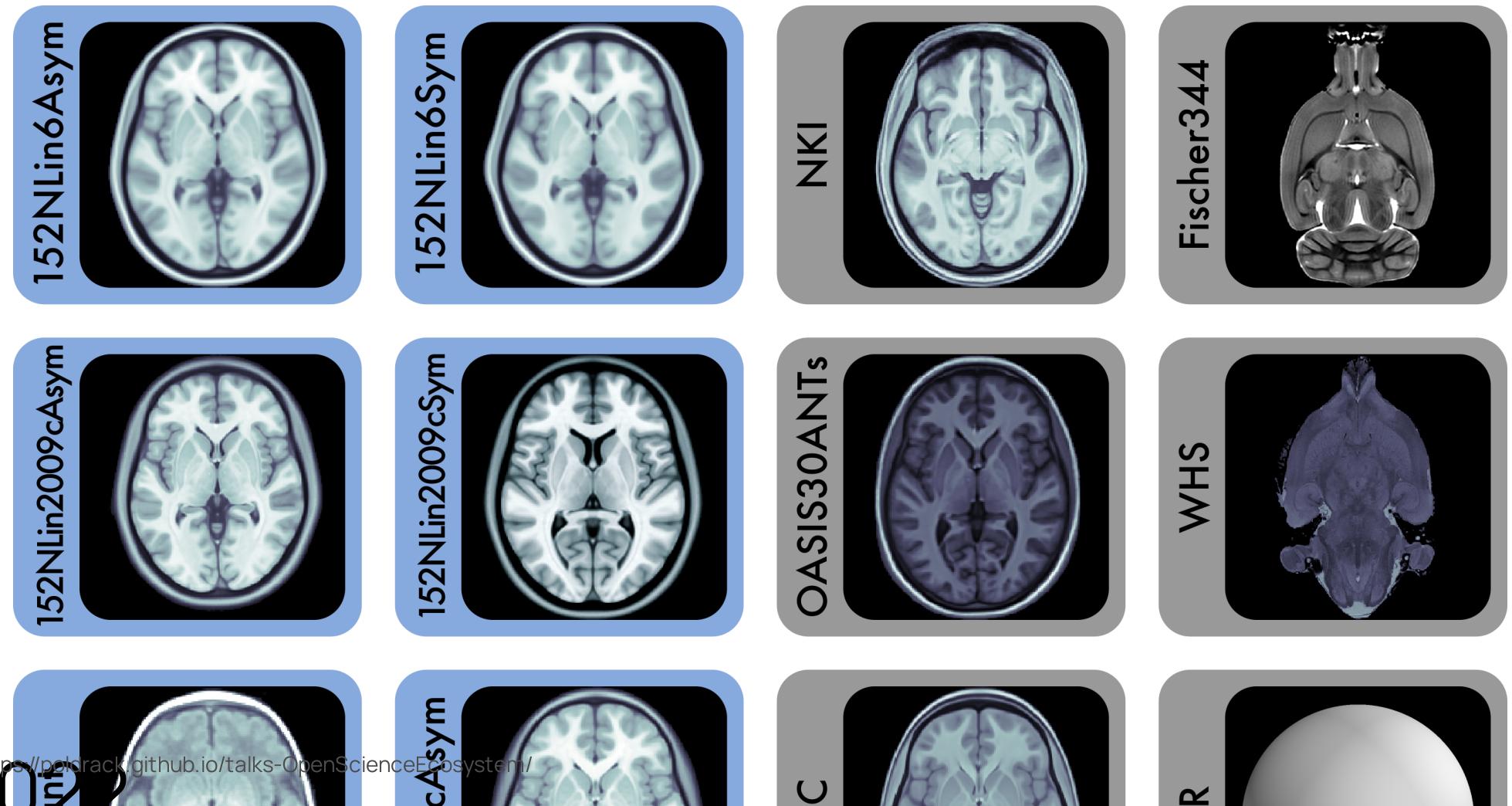
MRIQC: MRI quality control for BIDS data



mriqc.org; Esteban et al, 2017

Tentemplateflow: FAIR Sharing of Neuroimaging Templates

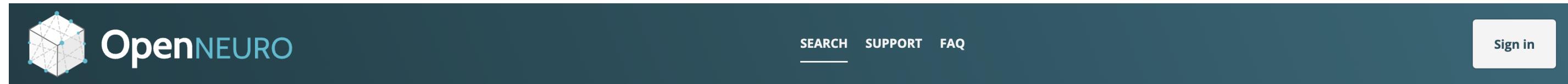
- Templates and atlases are commonly used in neuroimaging
- There is a significant lack of clarity in the use of these templates
 - There are numerous versions of the widely used “MNI template”



OpenNeuro: A BRAIN Initiative archive for BIDS data

The screenshot shows the OpenNeuro website homepage with a dark teal background. At the top left is the OpenNeuro logo (a white wireframe cube icon next to the text "OpenNEURO"). At the top right are links for "SEARCH", "SUPPORT", "FAQ", and "Sign in". Below the header, there's a main banner with the text "A free and open platform for validating and sharing BIDS-compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#) data". To the left of the banner are two statistics: "29,064 Participants" and "761 Public Datasets". Below these are two search/filter sections: "Browse by Modalities" (with a dropdown menu) and a search bar containing the text "reading" with a magnifying glass icon. At the bottom of the page are links for "SIGN IN", "Google", and "ORCID". In the center-right area, there are five 3D cube icons representing different data modalities: MRI (purple), PET (blue), MEG (orange), EEG (red), and iEEG (green). The MRI and PET cubes show brain scan images, while the MEG, EEG, and iEEG cubes show waveforms or heatmaps.

- Supports sharing of any validated BIDS dataset



Search All Datasets

Keywords ?



KEYWORD:
reading X

CLEAR ALL

Modalities

MRI

PET

EEG

iEEG

MEG

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



MODALITY:



TASKS:



OPENNEURO ACCESSION NUMBER: [ds003974](#)

SESSIONS: 1

PARTICIPANTS: 52

PARTICIPANTS' AGES: N/A

SIZE: 46.67GB

FILES: 893

The Reading Brain Project L2 Adults

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

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



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'. The top navigation bar includes links for SEARCH, SUPPORT, FAQ, and SIGN IN. Below the header, the dataset title 'The Reading Brain Project L2 Adults' is displayed, along with a small MRI icon. The page features a dark purple header bar with 'MRI' and the dataset name. Below this, there's a summary section with 'BIDS Validation' (Valid), '4 WARNINGS', and a 'brainlife.io' button. The main content area includes a 'Files' tab (selected), a 'Download' button, and a 'Metadata' tab. A 'README' box contains a note about the dataset being republished due to privacy concerns and a detailed description of the study's methodology and participants. To the right, detailed dataset metadata is listed: 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 an 'Uploaded by' section. At the bottom, a URL (https://poldrack.github.io/talks-OpenScienceEcosystem/) and file statistics (Files: 960, Size: 63.82GB) are shown.

Any valid BIDS dataset can be shared via OpenNeuro

The screenshot shows the homepage of the OpenNEURO platform. At the top left is the logo, which is a white wireframe cube with blue dots representing a brain. To its right is the text "OpenNEURO". In the top right corner are three small navigation links: "SEARCH", "SUPPORT", and "FAQ". Further to the right is a "Sign in" button.

Below the header, there is a main descriptive text: "A free and open platform for validating and sharing BIDS-compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#) data".

Two large numerical statistics are displayed: "29,064 Participants" and "761 Public Datasets".

Below these stats are two search/filter sections. The first section is a dropdown menu labeled "Browse by Modalities" with a downward arrow icon. The second section is a search bar labeled "Search" with a magnifying glass icon.

On the right side of the page, there is a graphic featuring five 3D cubes, each representing a different neuroimaging modality:

- A purple cube labeled "MRI" showing a grayscale brain scan.
- A blue cube labeled "PET" showing a color-coded heatmap.
- An orange cube labeled "MEG" showing a purple wavy pattern.
- A red cube labeled "EEG" showing a grid of black dots on a white background.
- A green cube labeled "iEEG" showing a colorful pixelated heatmap.

At the bottom of the page, there are several footer links: "SIGN IN", "Google", "ORCID", and a "≡" icon. The URL "https://poldrack.github.io/talks-OpenScienceEcosystem/" is also visible at the bottom.

The screenshot shows the OpenNEURO website interface. At the top left is the OpenNEURO logo with a brain network icon. Below it is a dark banner with the text "A free and open platform for validating and sharing compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#) data". To the right of the banner are statistics: "29,064 Participants" and "761 Public Datasets". Below these are buttons for "Browse by Modalities" and "Search" with a magnifying glass icon. At the bottom are links for "SIGN IN", "Google", and "ORCID". A central feature is a 3D cube visualization composed of five smaller cubes labeled "MRI", "PET", "MEG", "EEG", and "iEEG", each showing a different type of brain scan or signal. A "Sign in" button is located in the top right corner of the main page area.

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

Google

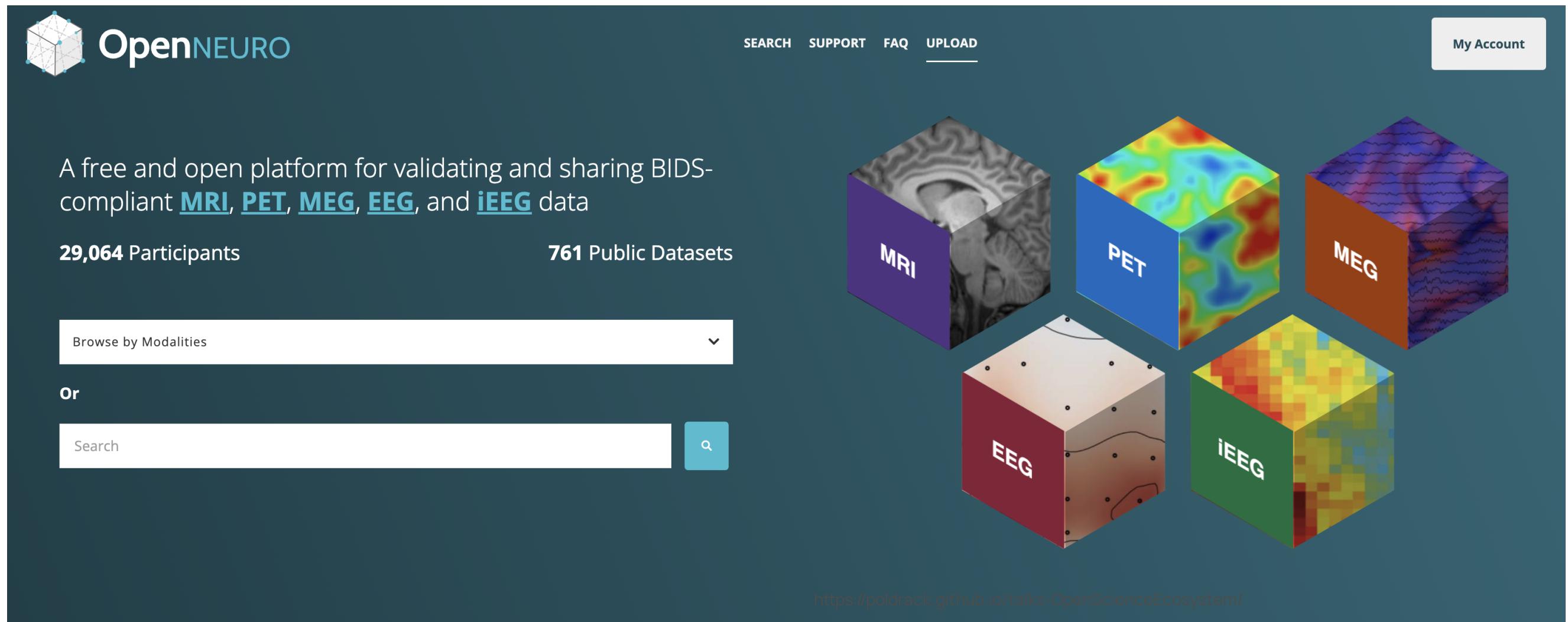
ORCID

Sign in

What is this?

MRI PET MEG EEG iEEG

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



The image shows the homepage of the OpenNEURO platform. At the top left is the OpenNEURO logo, which consists of a 3D cube with blue dots representing a brain network. To the right of the logo is the text "OpenNEURO". Above the main content area are navigation links: "SEARCH", "SUPPORT", "FAQ", "UPLOAD" (underlined), and "My Account". Below these links is a large text block: "A free and open platform for validating and sharing BIDS-compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#) data". To the right of this text are five 3D cubes, each representing a different data modality: MRI (purple), PET (blue), MEG (orange), EEG (red), and iEEG (green). Below the MRI cube is a "Browse by Modalities" dropdown menu. Below the PET cube is the text "761 Public Datasets". Below the MEG cube is the text "29,064 Participants". Below the EEG cube is the text "Search" with a magnifying glass icon. Below the iEEG cube is the text "Or". At the bottom of the page is a footer bar with the URL "https://poldrack.github.io/talks-OpenScienceEcosystem/" and a page number "43 / 65".

The screenshot shows the OpenNEURO website's dataset upload interface. The main navigation bar includes 'OpenNEURO', 'My Account', and a search bar. Below the navigation, a dark banner features the text 'A free and open platform for validating compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#)' and '29,064 Participants'. A 'Browse by Modalities' button is also present. The central area is titled 'Upload Dataset' and is divided into four steps: Step 1: Select Files (highlighted in dark blue), Step 2: Validation, Step 3: Metadata, and Step 4: Accept Terms. A note states: 'To protect the privacy of the individuals who have been scanned, we require that all scan data be defaced before publishing a dataset.' Below this is a placeholder text 'Select a [BIDS dataset](#) to upload' and a 'Select folder' button. In the bottom right corner, there are four large, semi-transparent cubes representing different modalities: PET (blue), MEG (orange), EEG (red), and iEEG (green). The URL <https://poldrack.github.io/talks-OpenScienceEcosystem/> is visible at the bottom.

 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 to download the dataset to.

Upload Dataset

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

We found 3 warnings in your dataset. You are not required to fix warnings, but doing so will make your dataset more BIDS compliant. Continue or fix the issues and select folder again.

Continue

VIEW 3 WARNINGS IN 15 FILES ▲

Warning: 1 **VIEW 13 FILES**

You should define 'SliceTiming' for this file. If you don't provide this information slice time correction will not be possible. 'Slice Timing' is the time at which each slice was acquired within each volume (frame) of the acquisition. Slice timing is not slice order -- rather, it is a list of times containing the time (in seconds) of each slice acquisition in relation to the beginning of volume acquisition.

Warning: 2 **VIEW 1 FILE**

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

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

My Account

Follow 1 Bookmark 2

Make them public.

Neuro Accession Number 03

Russell A. Poldrack

Public

Scanning Modalities

The screenshot shows a modal window titled "Upload Dataset" with four tabs: "Step 1: Select Files", "Step 2: Validation", "Step 3: Metadata" (which is active), and "Step 4: Accept Terms". Below the tabs, a note says: "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." The "Step 3: Metadata" section contains the following fields:

- 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

On the right side of the modal, there is a dark sidebar with "My Account", "Follow 1", "Bookmark 2", and other user information.

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 to save the dataset.

Steps

1. Select a local directory to save the dataset and grant permission to OpenNEURO.
2. Download will run in the background, please leave the page open.

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

Created: 2020-05-14

Versions

46 / 65

OpenNEURO

MRI Rhyme judgment

This dataset

BIDS Validation

Files Share Versioning Admin

How to Download

Download with your browser

This method is convenient and allows you to select a local di

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.

My Account

Follow 1 Bookmark 2

e them public.

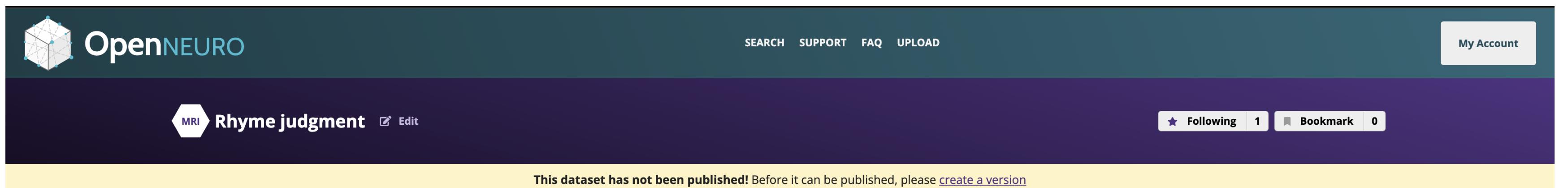
uro Accession Number

03

Russell A. Poldrack

lit

Modality



This screenshot shows the OpenNEURO platform interface for a dataset titled "Rhyme judgment".

Header:

- Logo: OpenNEURO (with a brain cube icon)
- Navigation: SEARCH, SUPPORT, FAQ, UPLOAD
- User Account: My Account

Middle Section:

- Dataset Type: MRI
- Title: Rhyme judgment
- Actions: Edit, Following (1), Bookmark (0)

Message Bar:

This dataset has not been published! Before it can be published, please [create a version](#)

BIDS Validation:

- Status: 2 WARNINGS, Valid
- Actions: Clone

File Actions:

- Files, Publish, Share, Versioning (selected), Admin, Download, Metadata, Delete

Versioning:

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

Available Modalities:

- MRI

Version:

Draft Updated: 2022-11-16 [Create Version](#)

Page Footer:

48 / 65

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 ▾

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.

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

OpenNeuro Accession Number

ds000003

Authors

Xue, G., Russell A. Poldrack

Available Modalities

MRI

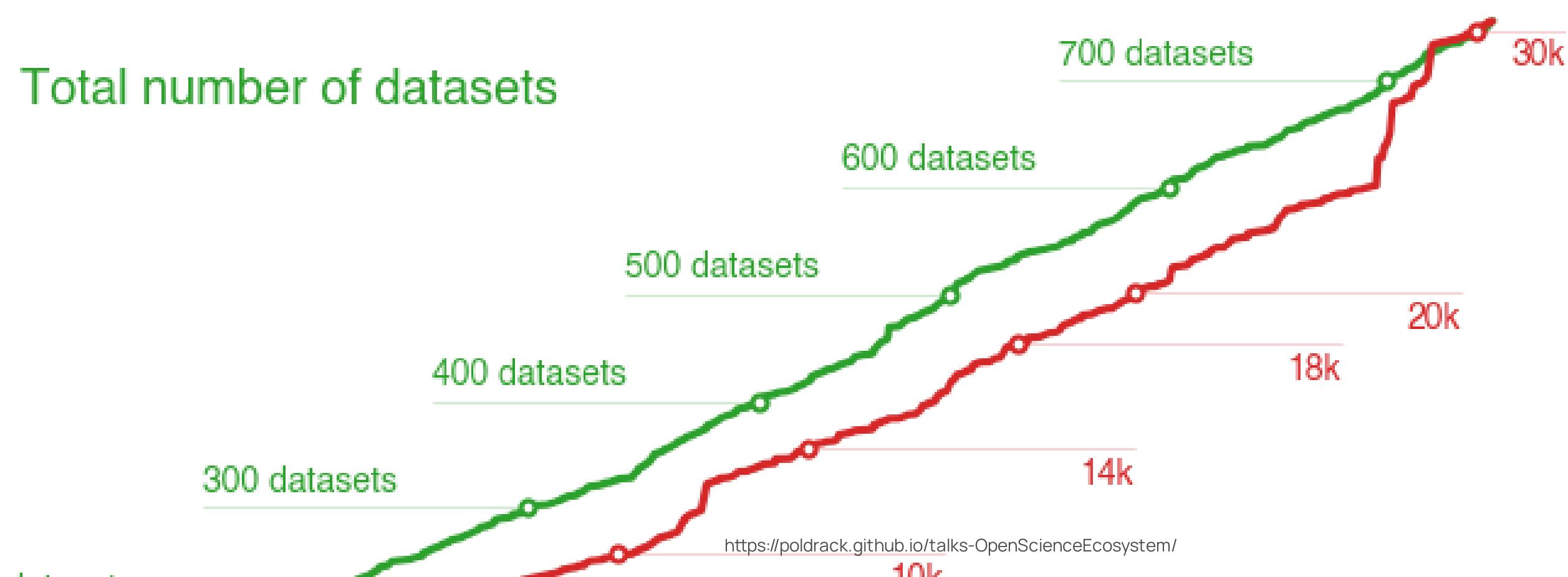
Versions

1.0.0

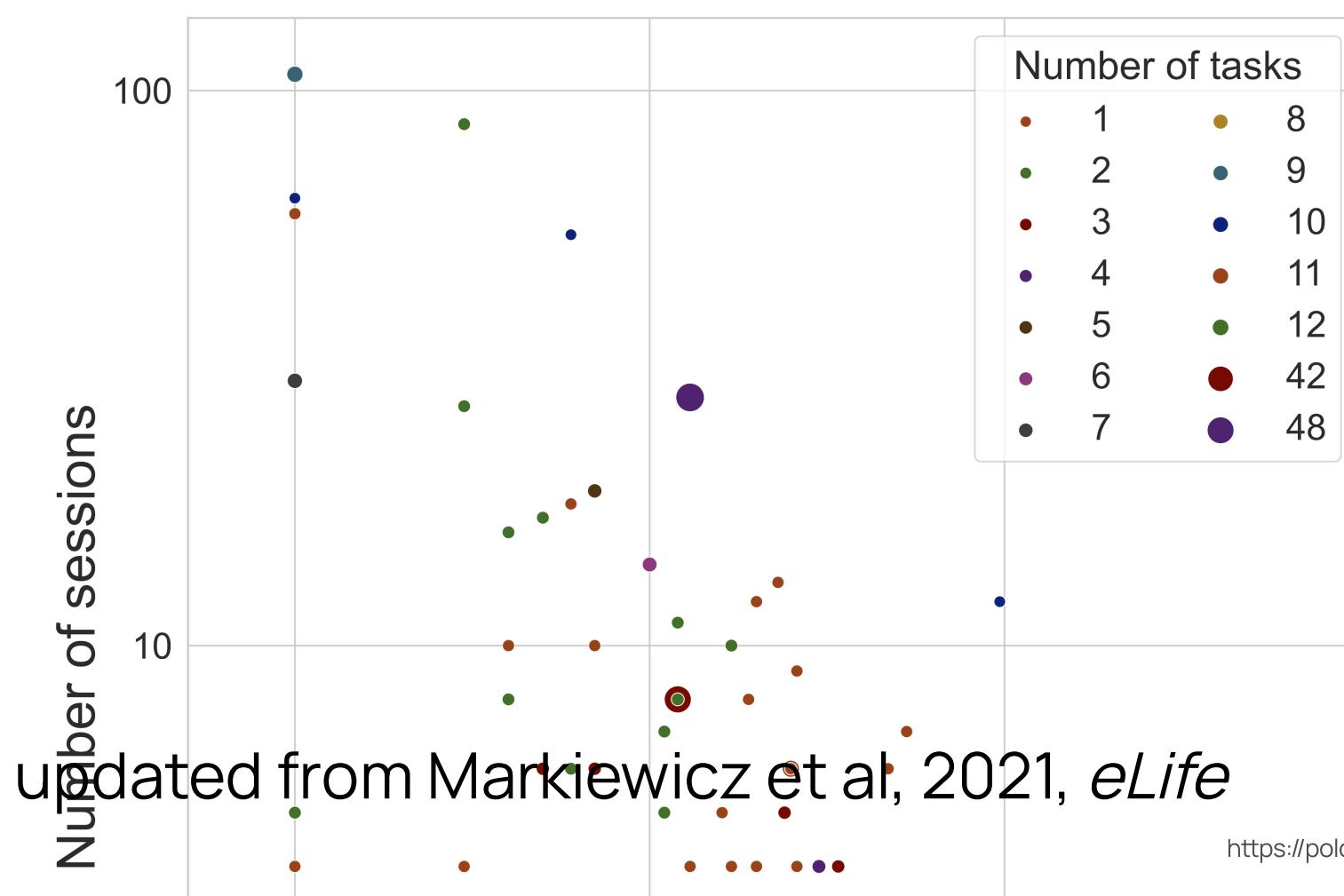
Created: 2020-05-14

Versions ▾

The growth of OpenNeuro



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
Doa	1

Scholarly reuse of OpenNeuro datasets

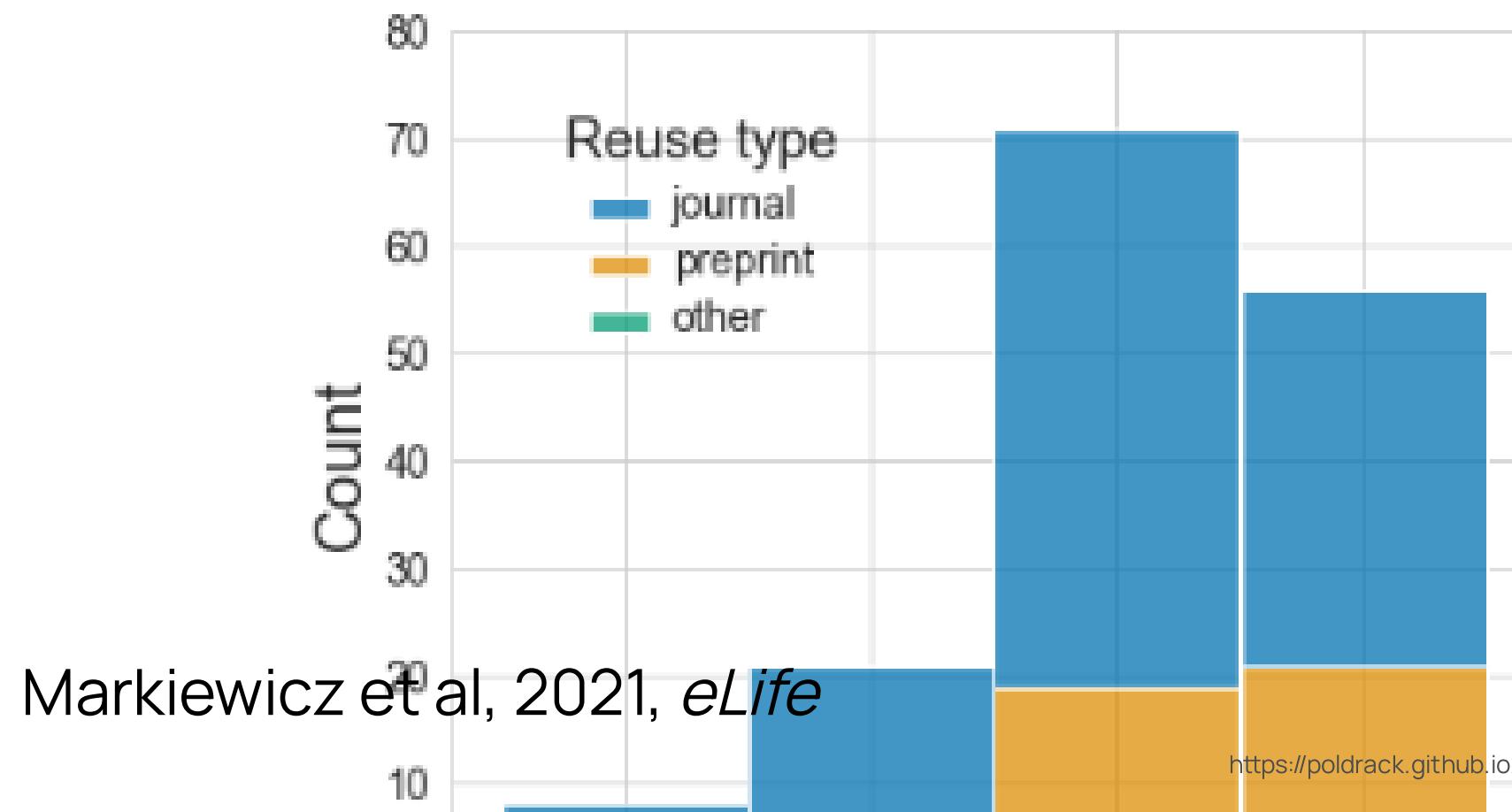


Figure 5. Published reuses of OpenNeuro datasets, split by the type of reuse. Note that the final bar includes only reuses identified through June 2021.

Processing of OpenNeuro data

brainlife.io: processing of MRI data

The screenshot shows the brainlife.io web application. The left sidebar has a blue header with the brainlife logo and navigation links: DATASETS (selected), PROJECTS, APPS, PUBLICATIONS, and DATATYPES. The main area is titled "PUBLIC/PROTECTED PROJECTS". It lists two main sections: "HCP 3T / Diffusion" and "HCP 7T / Diffusion".

- HCP 3T / Diffusion:** Contains a summary card with "1112 sub | 6880 objs (4.51 TB)". Below it is another card for "HCP 3T Retest / Diffusion" with "45 HCP 3T subjects retested".
- HCP 7T / Diffusion:** Contains a summary card with "150 sub | 300 objs (22.56 MB)". Below it is a card for "O3D" (Open Diffusion Data and Derivative) with the text "O3D (Open Diffusion Data and Derivative) is a reference repository for precision".

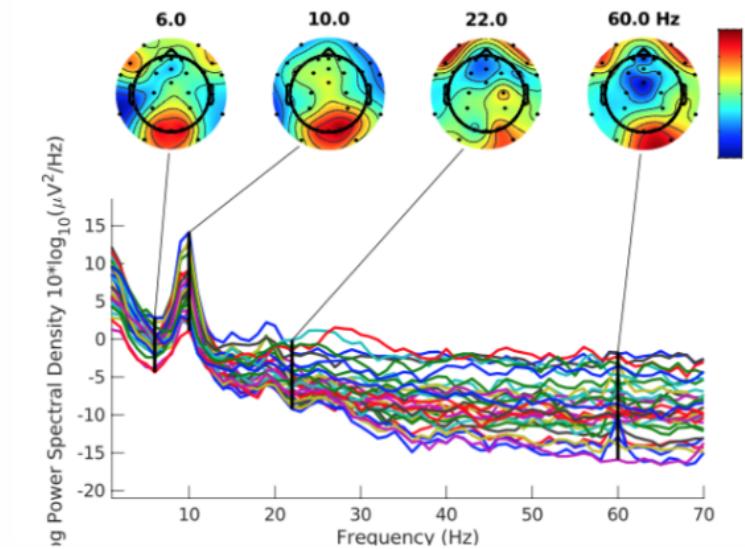
At the bottom, there are two footer cards: "45 sub | 315 objs (101.23 GB)" and "12 sub | 2708 objs (2.28 TB)". A URL "https://poldrack.github.io/talks-OpenScienceEcosystem/" is also visible at the bottom right.

NEM

Data Summary

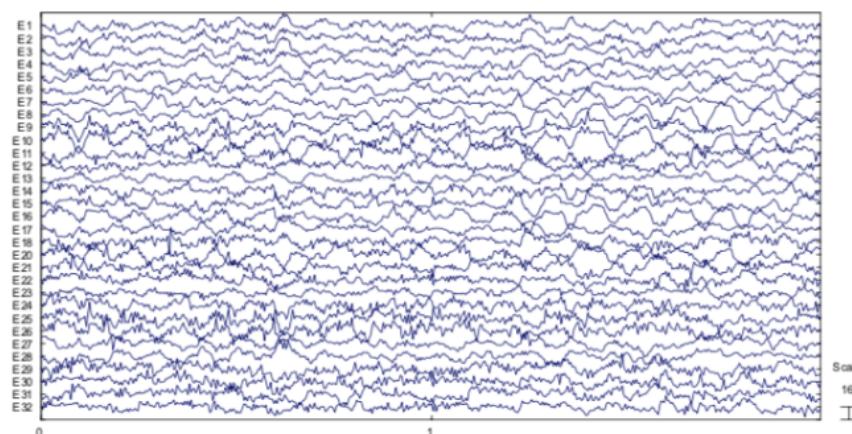
Subject: sub-001
Session(s): 1, **Run(s):** 1
Data size: 32 channels, 298k frames
Acceptable scalp channels: 100.0% (32 of 32)
Acceptable data points channels: 90.9% (271k of 298k)
Source quality metric based on independent component: 48.4%

Scalp channel log spectra



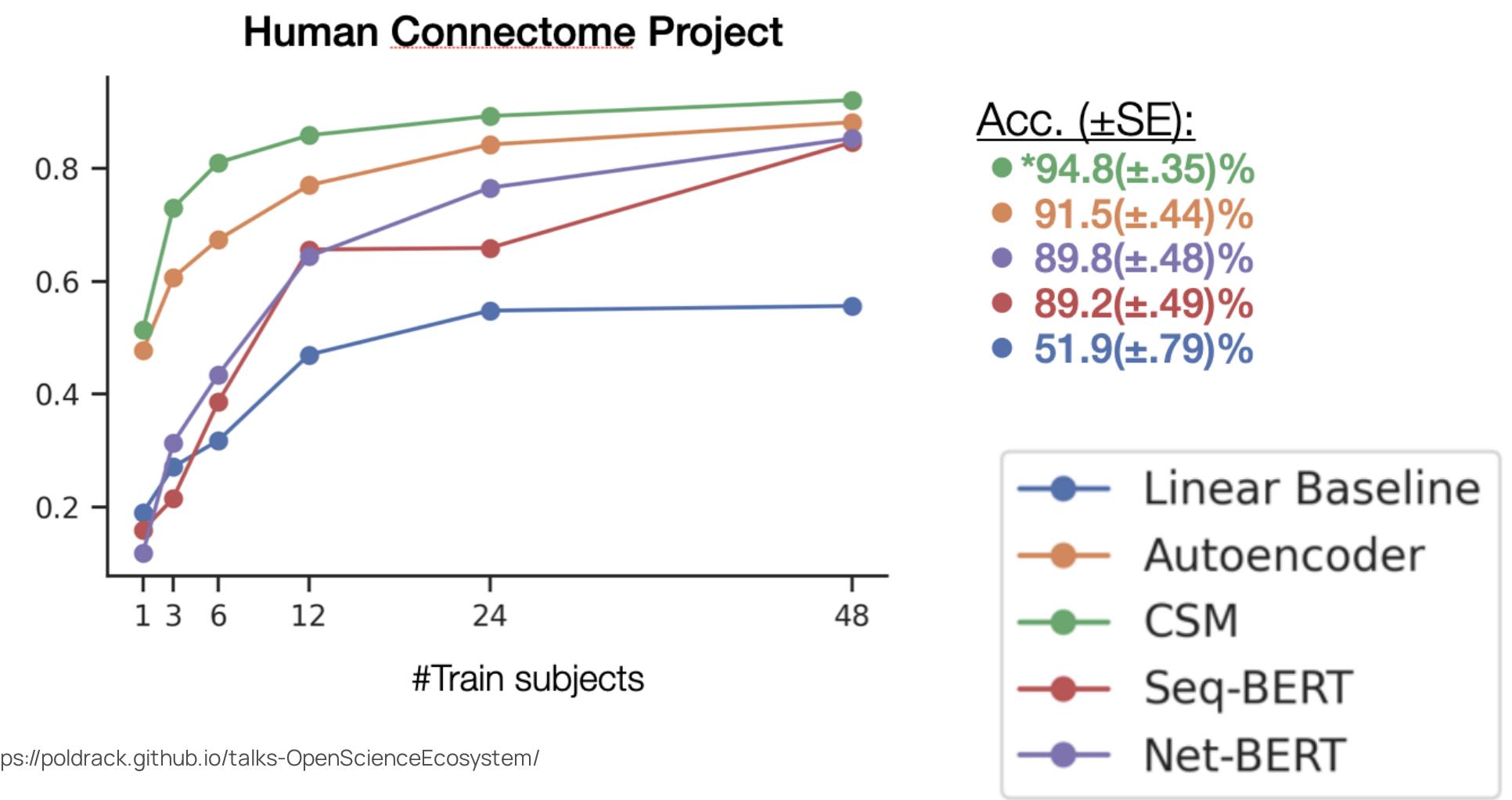
3 data

Sample scalp channel data (mid 2 seconds)



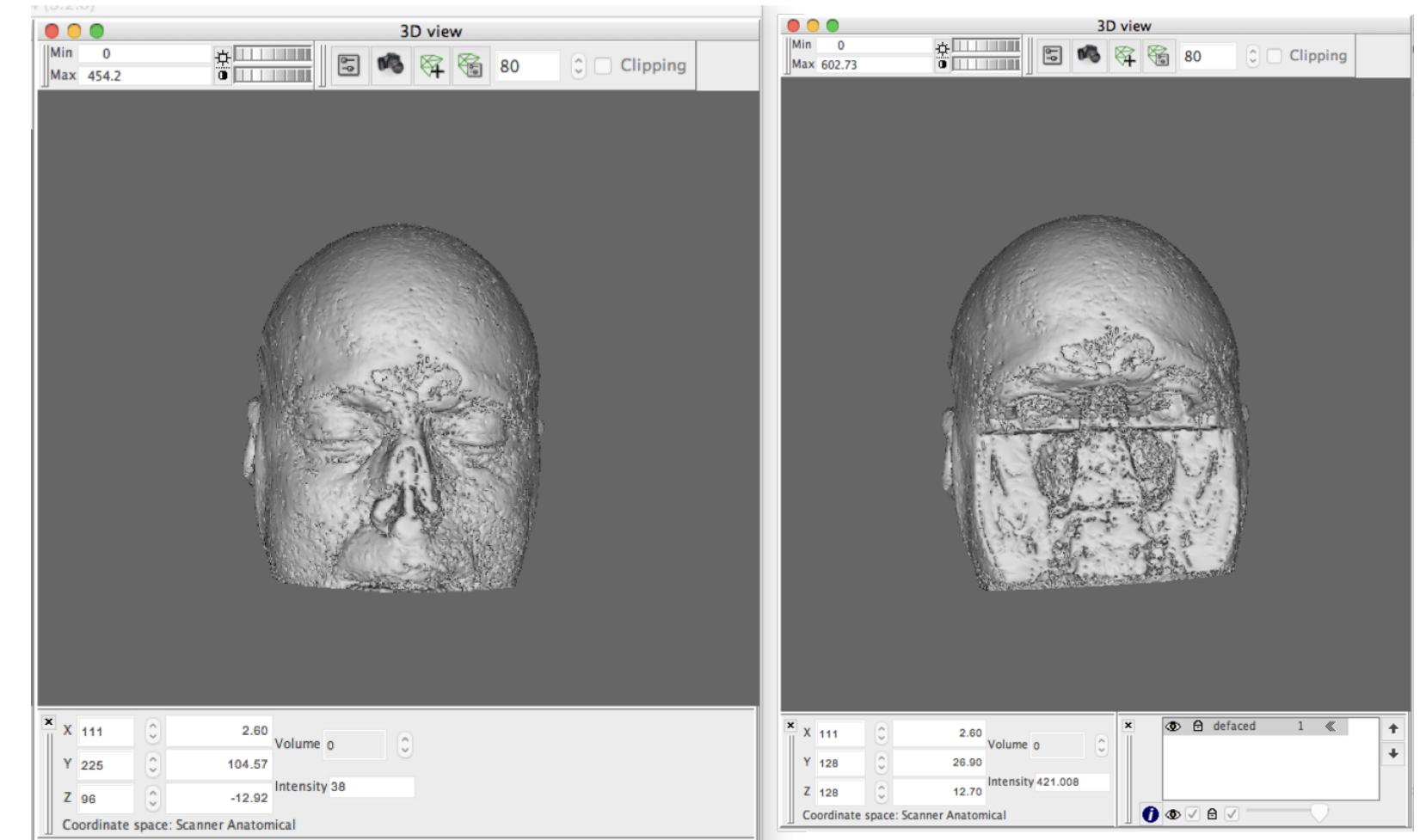
Example of OpenNeuro reuse

- 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 Thomas, Re, & Poldrack, 2022, *NeurIPS* 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

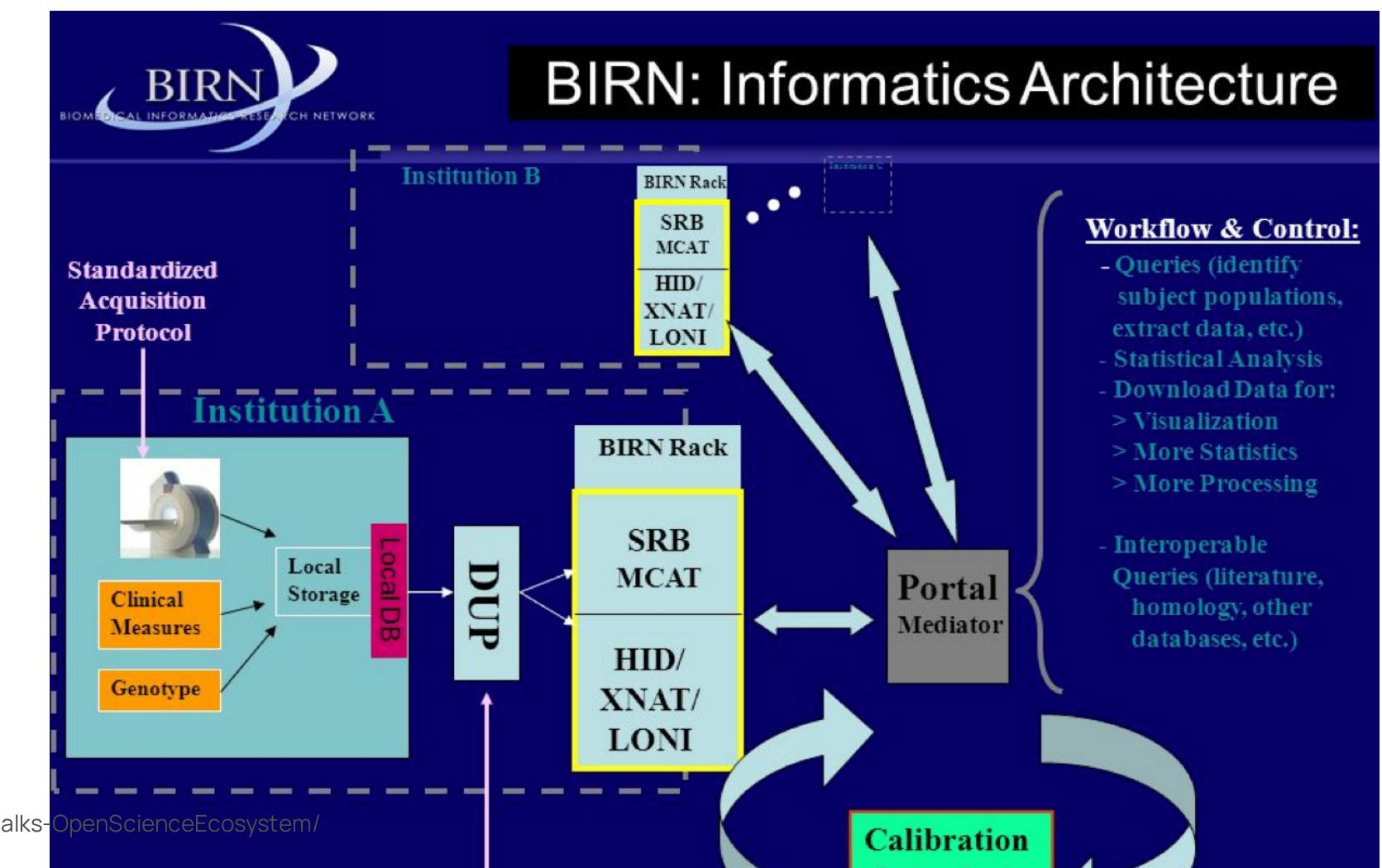
- *Data are digital end-to-end*
 - Minimizes manual steps in the process
- *Standardized file formats and data standards*
 - Makes data immediately usable by anyone
 - Reduces burden of curation and preparation
- *Demonstrated scientific utility*
- *Numerous success stories*

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

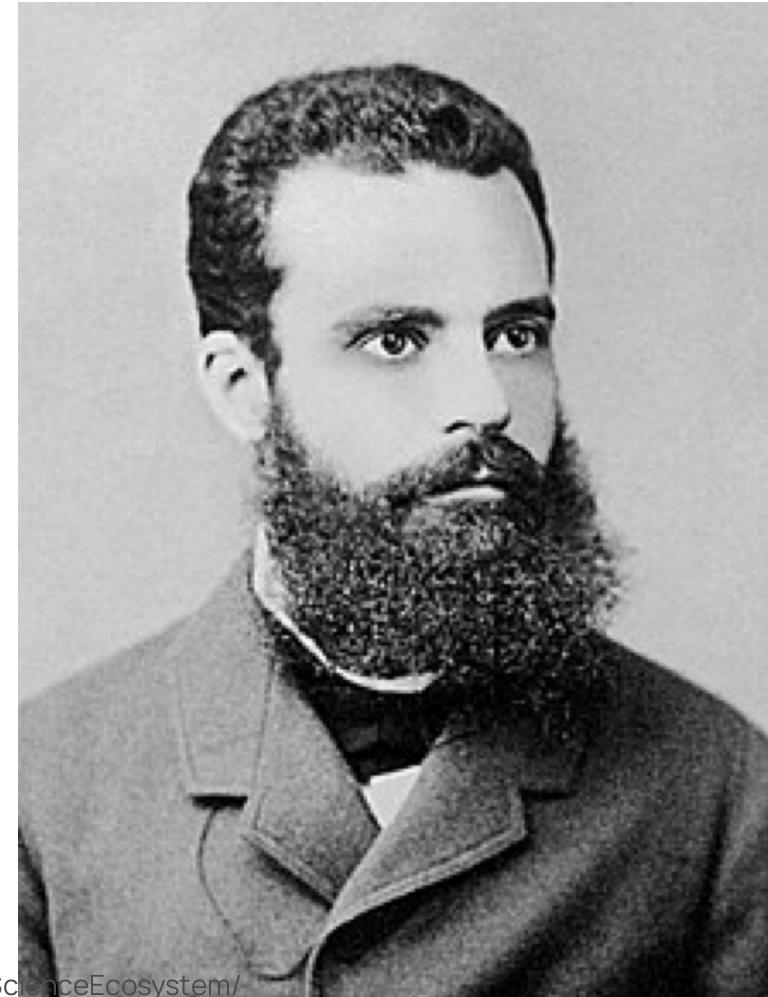
Lessons learned

- 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

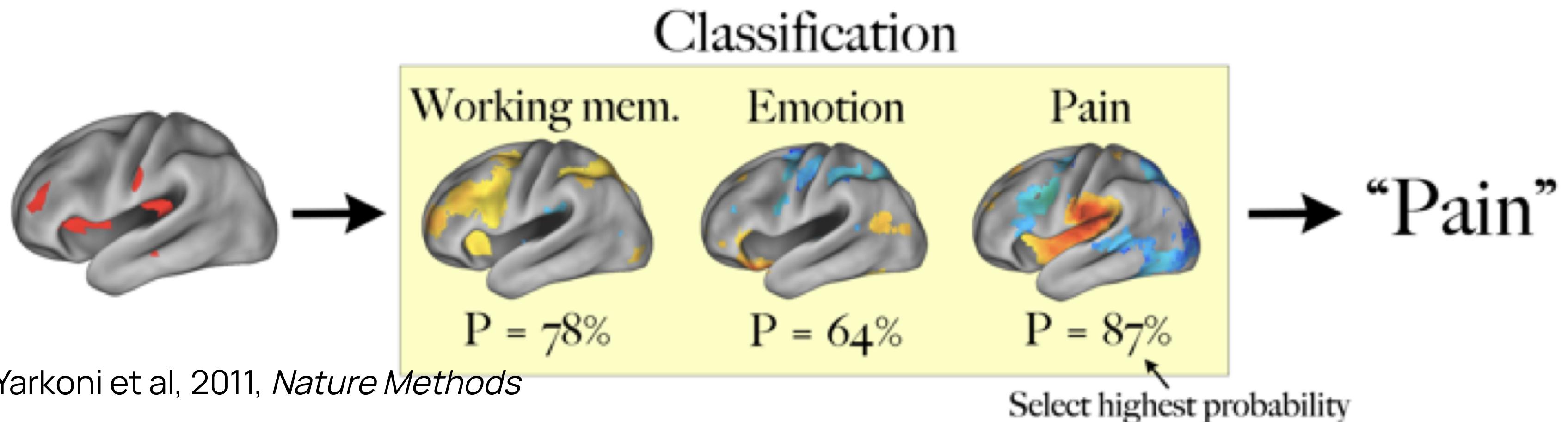


OpenNeuro Team



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*

