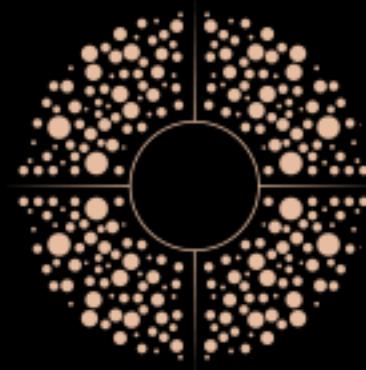




BIDS
BRAIN IMAGING DATA STRUCTURE



open
data



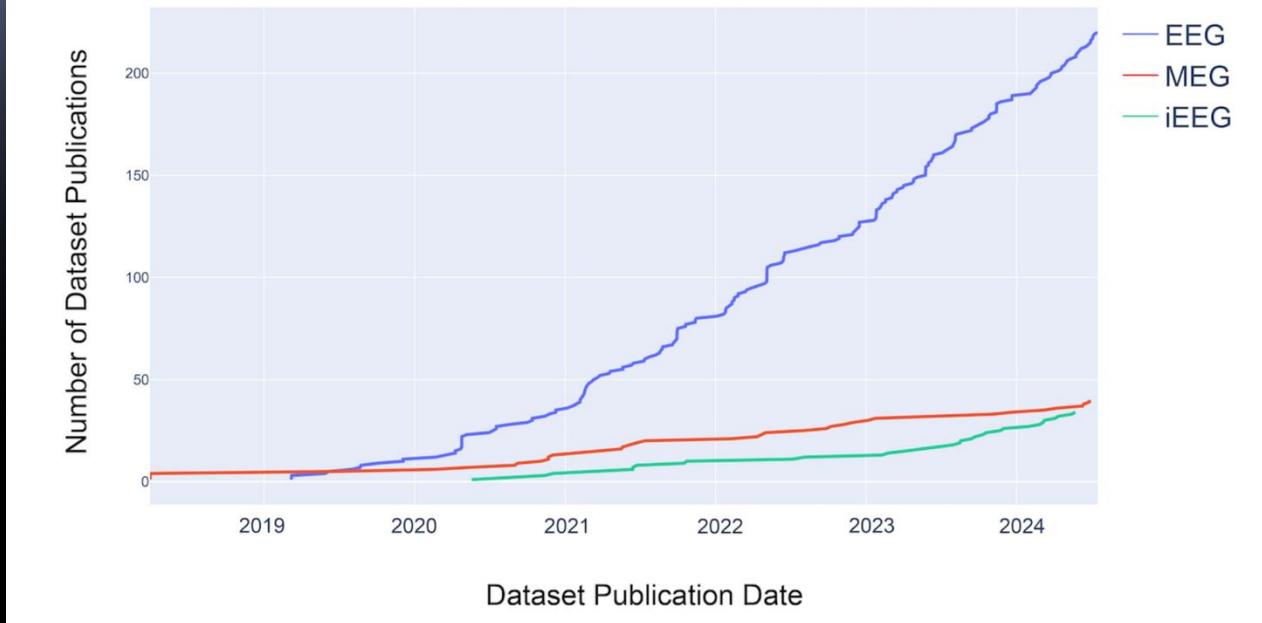
NEMAR

Arnaud Delorme

University of California, San Diego
CerCo CNRS, Toulouse

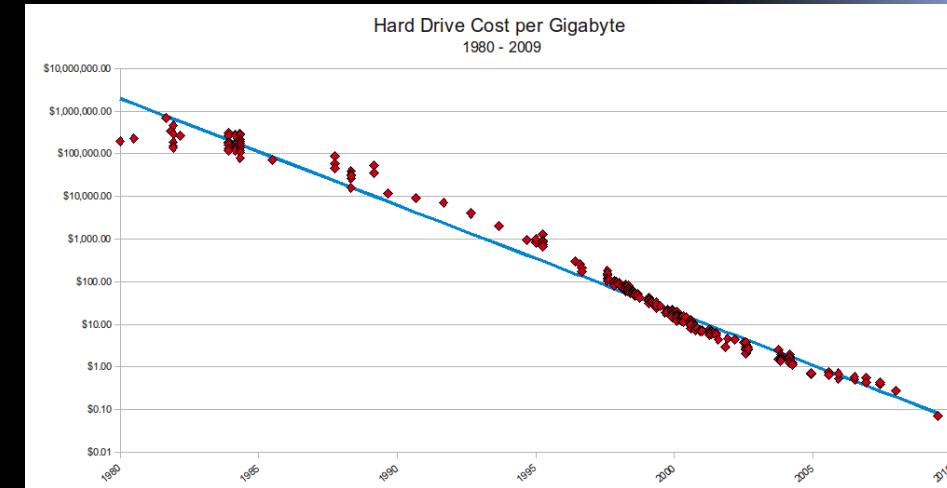


UCSD



Why is sharing increasing?

- ▶ Cost of storage is low and there are plenty of free repository
- ▶ Grant agencies do not want to spend million to acquire data and then throw it away
- ▶ Issue with replicability in Science. Journal require you to post data.
- ▶ Standard pipelines and AI are in the air



<https://mkomo.com/assets/hd-cost-graph.png>

The screenshot shows a Science journal article. The header includes 'Science' and navigation links for 'Contents', 'News', 'Careers', and 'Journals'. The main title of the article is 'Estimating the reproducibility of psychological science'. Below the title, a sub-section titled 'Empirically analyzing empirical evidence' is visible, along with a brief description of the study's purpose and findings.

RESEARCH ARTICLE

Estimating the reproducibility of psychological science

Empirically analyzing empirical evidence

One of the central goals in any scientific endeavor is to understand causality. Experiments that seek to demonstrate a cause/effect relation most often manipulate the postulated causal factor. Aarts *et al.* describe the replication of 100 experiments reported in papers published in 2008 in three high-ranking psychology journals. Assessing whether the replication and the original experiment yielded the same result according to several criteria, they find that about one-third to one-half of the original findings were also observed in the replication study.

Open Data

Findable

Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

Accessible

Data is deposited in a trusted repository.

Interoperable

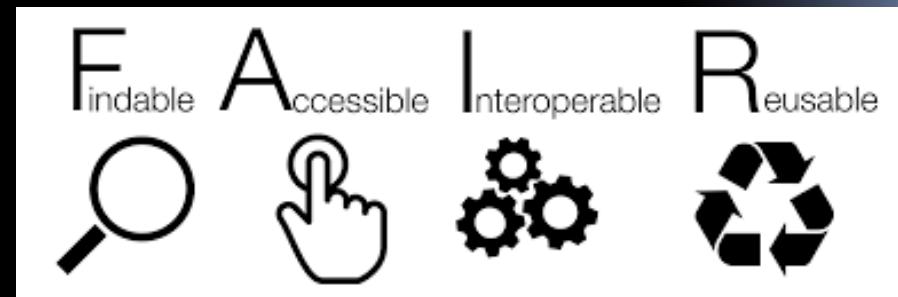
(Meta)data uses a formal, shared, and broadly applicable language or format.

Reusable

Data is described with clear and understandable attributes.

Anonymized.

There should be a clear and acceptable license for re-use.



Open Data - BIDS

Findable

Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

Accessible

Data is deposited in a trusted repository.

Interoperable

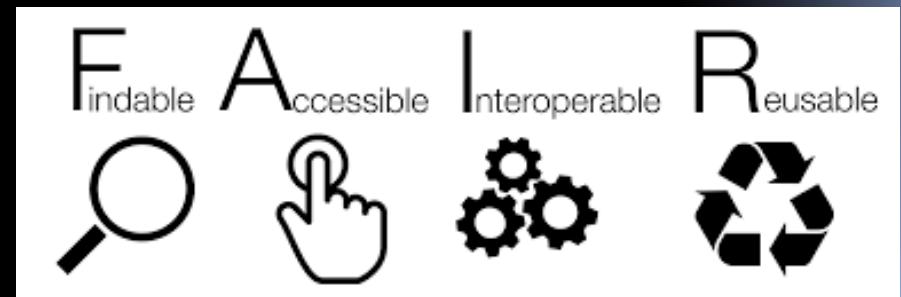
(Meta)data uses a formal, shared, and broadly applicable language or format.

Reusable

Data is described with clear and understandable attributes.

Anonymized.

There should be a clear and acceptable license for re-use.



What is BIDS?

BIDS is a way to organize your existing raw data

- ▶ To improve consistent and complete documentation
- ▶ To facilitate re-use by your future self and others

BIDS is not

- ▶ A new file format
- ▶ A search engine
- ▶ A data sharing tool



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

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

Comment: MEG-BIDS, the brain imaging data structure extended to magnetoencephalography

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

COMMENT

EEG-BIDS, an extension to the brain imaging data structure for electroencephalography

Received: 16 January 2019

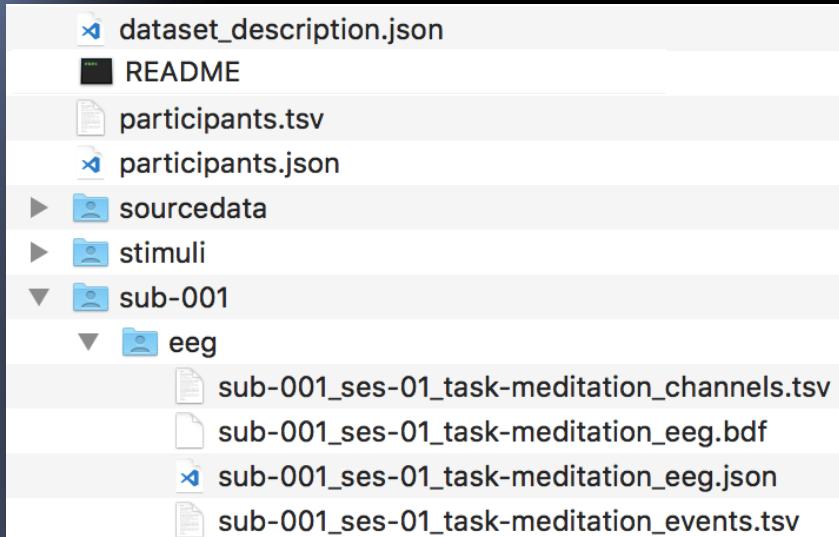
Accepted: 7 May 2019

Published online: 25 June 2019

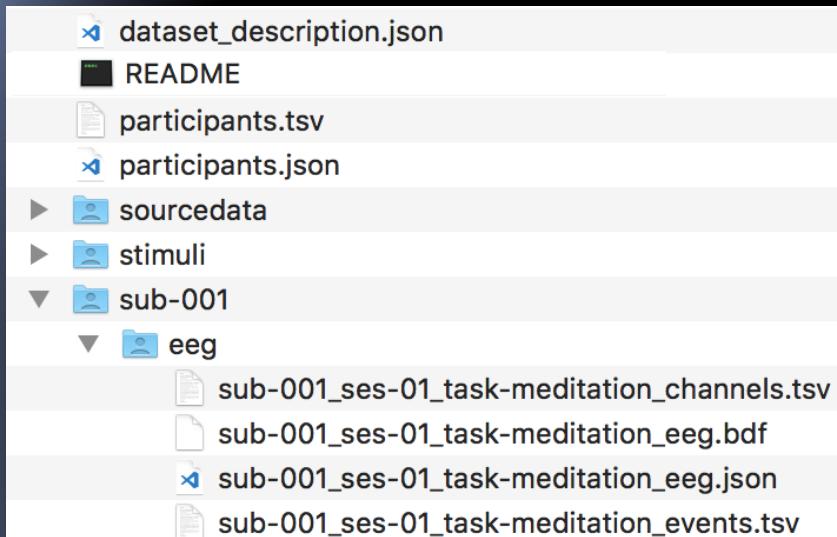
Cyril R. Pernet¹, Stefan Appelhoff², Krzysztof J. Gorgolewski³, Guillaume Flandin⁴, Christophe Phillips⁵, Arnaud Delorme^{6,7} & Robert Oostenveld^{8,9}

The Brain Imaging Data Structure (BIDS) project is a rapidly evolving effort in the human brain imaging research community to create standards allowing researchers to readily

BIDS-EEG



BIDS-EEG

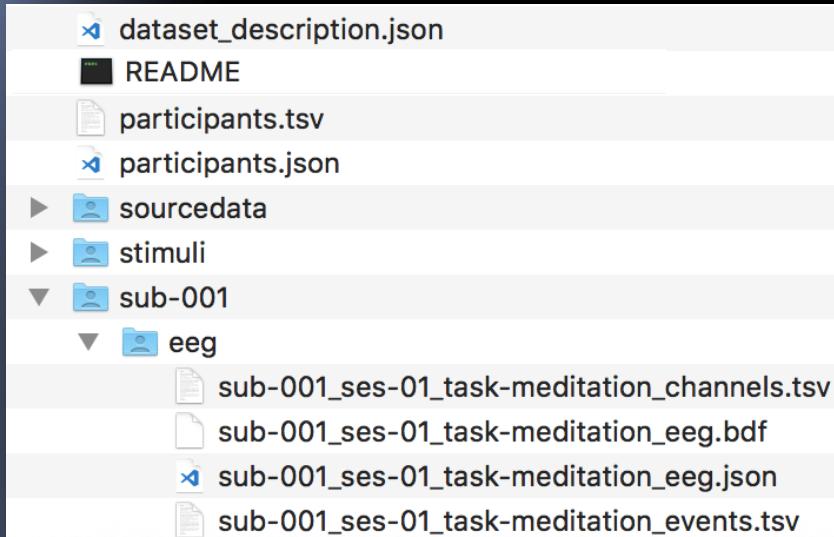


► description of dataset in JSON format

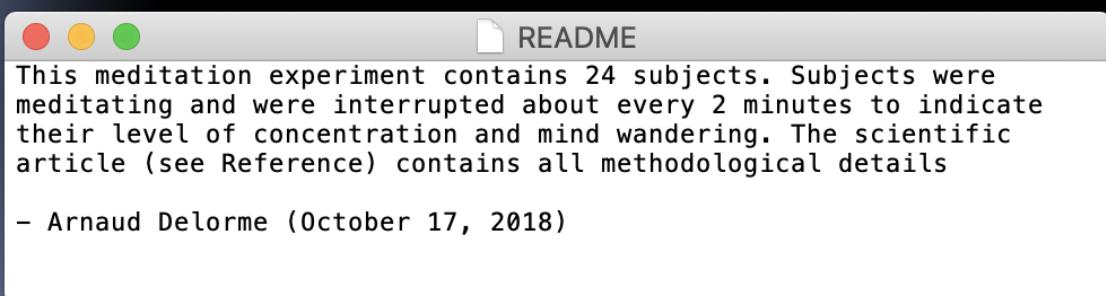
```
dataset_description.json
```

```
{  
  "Name": "Meditation study",  
  "ReferencesAndLinks": ["https://www.ncbi.nlm.nih.gov/pubmed/27815577"],  
  "License": "CC0",  
  "BIDSVersion": "1.1.1"  
}
```

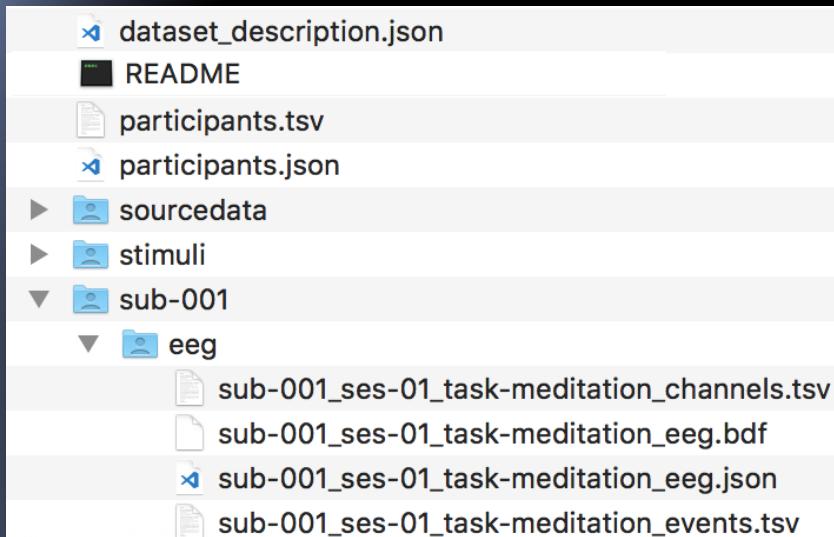
BIDS-EEG



- ▶ description of dataset in JSON format
- ▶ readme file for users



BIDS-EEG

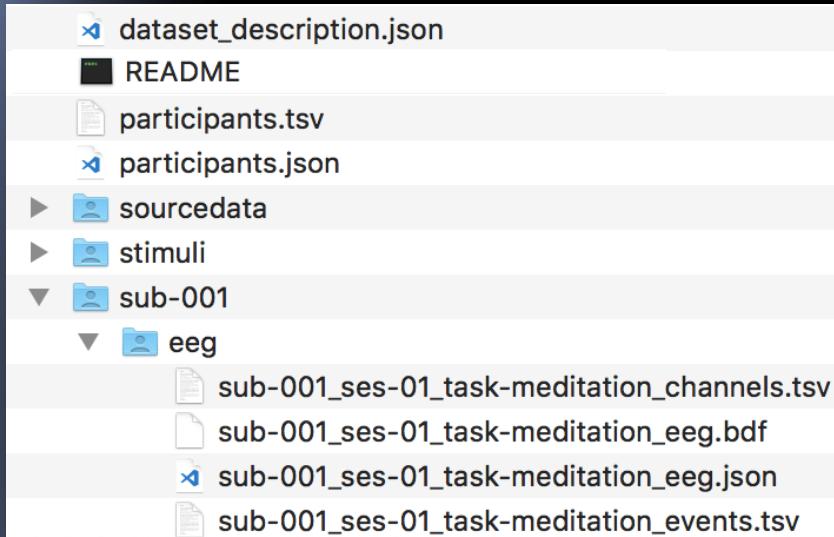


- ▶ description of dataset in JSON format
- ▶ readme file for users
- ▶ participants table files in tab delimited format

A screenshot of a terminal window showing the contents of the participants.tsv file. The file is displayed in a table format with four columns: participant_id, gender, age, and group.

participant_id	gender	age	group
sub-001	M	32	expert
sub-002	M	35	expert
sub-013	F	47	novice
sub-016	M	77	novice

BIDS-EEG



- ▶ description of dataset in JSON format
- ▶ readme file for users
- ▶ participants table files in tab delimited format
- ▶ description of participants table file columns (above)

A screenshot of a JSON editor window titled "participants.json". The content of the file is:

```
{  
    "participant_id": {  
        "Description": "unique participant identifier"  
    },  
    "gender": {  
        "Description": "sex of the participant",  
        "Levels": {  
            "M": "male",  
            "F": "female"  
        }  
    },  
    "age": {  
        "Description": "age of the participant"  
    }  
}
```

BIDS-EEG

dataset_description.json	▶ description of dataset in JSON format
README	▶ readme file for users
participants.tsv	▶ participants table files in tab delimited format
participants.json	▶ description of participants table file columns (above)
▶ sourcedata	▶ original raw data if converted to a supported format
▶ stimuli	▶ original stimuli (sound files and images)
▶ sub-001	▶ anonymized subject 1 folder
▶ eeg	
sub-001_ses-01_task-meditation_channels.tsv	
sub-001_ses-01_task-meditation_eeg.bdf	
sub-001_ses-01_task-meditation_eeg.json	
sub-001_ses-01_task-meditation_events.tsv	

BIDS-EEG

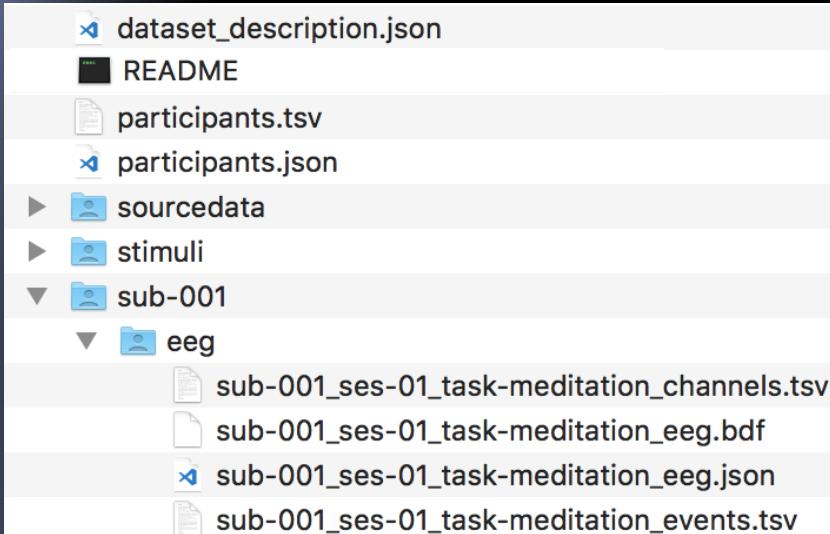
dataset_description.json	▶ description of dataset in JSON format
README	▶ readme file for users
participants.tsv	▶ participants table files in tab delimited format
participants.json	▶ description of participants table file columns (above)
▶ sourcedata	▶ original raw data if converted to a supported format
▶ stimuli	▶ original stimuli (sound files and images)
▶ sub-001	▶ anonymized subject 1 folder
▶ eeg	
sub-001_ses-01_task-meditation_channels.tsv	▶ file describing channels
sub-001_ses-01_task-meditation_eeg.bdf	
sub-001_ses-01_task-meditation_eeg.json	
sub-001_ses-01_task-meditation_events.tsv	

- ▶ description of dataset in JSON format
- ▶ readme file for users
- ▶ participants table files in tab delimited format
- ▶ description of participants table file columns (above)
- ▶ original raw data if converted to a supported format
- ▶ original stimuli (sound files and images)
- ▶ anonymized subject 1 folder
- ▶ file describing channels

sub-001_ses-01_task-meditation_channels.tsv

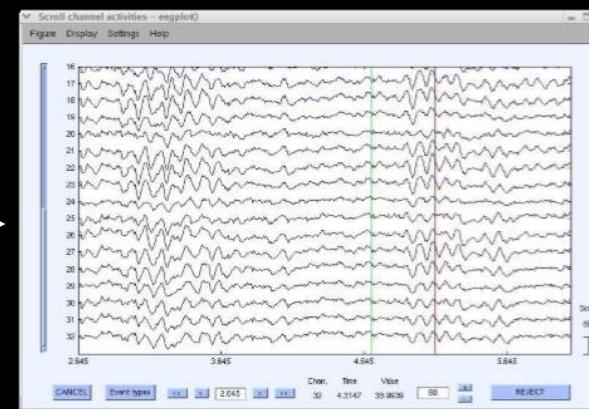
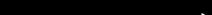
name	type	units
Fp1	EEG	microV
AF7	EEG	microV
AF3	EEG	microV
F1	EEG	microV
F3	EEG	microV
F5	EEG	microV
F7	EEG	microV
FT7	EEG	microV
FC5	EEG	microV
FC3	EEG	microV
FC1	EEG	microV

BIDS-EEG

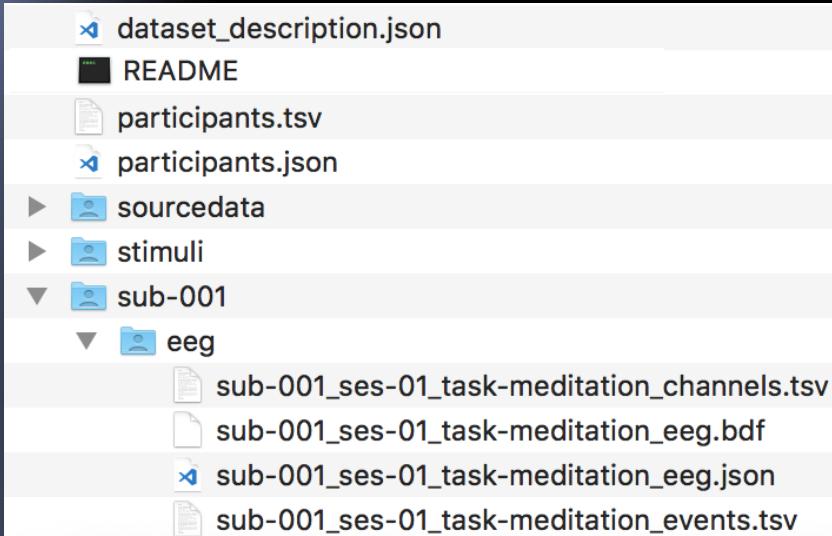


- ▶ description of dataset in JSON format
 - ▶ readme file for users
 - ▶ participants table files in tab delimited format
 - ▶ description of participants table file columns (above)
 - ▶ original raw data if converted to a supported format
 - ▶ original stimuli (sound files and images)
 - ▶ anonymized subject 1 folder
- ▶ file describing channels
 - ▶ raw EEG data file (not all raw formats are possible)

```
00000000 00000000 00000000 11101000 10110111 01000010 00000000 00000000  
10001011 01010100 00100100 00010000 01101000 01101000 00001101 01000011  
00000000 01010010 100001101 01000100 001000100 01110000 01010000 11101000  
01110011 01001010 00000000 00000000 11000110 10000100 00100100 11011000  
00000000 00000000 00000000 00001001 01010111 10001101 01001100 00100100  
01101100 11101000 10010001 01000010 00000000 10001011 01001100  
01001000 00010000 01010001 10001101 01010100 00100100 01001000 01010010  
11101000 00110010 01000001 00000000 11000110 10000100 00100100  
11011000 00000000 00000000 00001010 10111001 00000000 10111011  
01000011 00000000 11101000 00110000 01100010 00000000 00000000 01010000  
10001101 01001100 00100100 11101000 10000110 00111010 00000000  
00000000 10000111 00100100 00110100 10110100 10000110 00001101  
01000011 00000000 01010000 01000110 01001100 00101100 01010001  
11101000 00100010 01001010 00000000 00000000 11000110 10000100 00100100  
11011000 00000000 00000000 00000000 00001011 01010011 10001101 01001100  
00100100 01010100 11101000 01000010 00000000 10000100 10000101
```



BIDS-EEG



- ▶ description of dataset in JSON format
- ▶ readme file for users
- ▶ participants table files in tab delimited format
- ▶ description of participants table file columns (above)
- ▶ original raw data if converted to a supported format
- ▶ original stimuli (sound files and images)
- ▶ anonymized subject 1 folder
- ▶ file describing channels
- ▶ raw EEG data file (not all raw formats are possible)
- ▶ amplifier and recording information

A screenshot of a JSON editor window titled "sub-001_ses-01_task-meditation_eeg.json". The JSON object contains the following properties:

```
{  
  "InstitutionAddress": "Place du Docteur Baylac, Pavillon Baudot, 31059 Toulouse, France",  
  "InstitutionName": "Paul Sabatier University",  
  "InstitutionalDepartmentName": "Centre de Recherche Cerveau et Cognition",  
  "PowerLineFrequency": 50,  
  "ManufacturersModelName": "ActiveTwo",  
  "TaskName": "meditation",  
  "EEGReference": "CMS/DRL",  
  "Manufacturer": "BIOSEMI",  
  "EEGChannelCount": 64,  
  "MiscChannelCount": 15,  
  "RecordingType": "continuous",  
  "RecordingDuration": 2721,  
  "SamplingFrequency": 256,  
  "EOGChannelCount": 0,  
  "ECGChannelCount": 0,  
  "EMGChannelCount": 0,  
  "SoftwareFilters": "n/a"  
}
```

BIDS-EEG

dataset_description.json	► description of dataset in JSON format
README	► readme file for users
participants.tsv	► participants table files in tab delimited format
participants.json	► description of participants table file columns (above)
▷ sourcedata	► original raw data if converted to a supported format
▷ stimuli	► original stimuli (sound files and images)
▷ sub-001	► anonymized subject 1 folder
▷ eeg	
sub-001_ses-01_task-meditation_channels.tsv	► file describing channels
sub-001_ses-01_task-meditation_eeg.bdf	► raw EEG data file (not all raw formats are possible)
sub-001_ses-01_task-meditation_eeg.json	► amplifier and recording information
sub-001_ses-01_task-meditation_events.tsv	► events in tabular format

- description of dataset in JSON format
- readme file for users
- participants table files in tab delimited format
- description of participants table file columns (above)
- original raw data if converted to a supported format
- original stimuli (sound files and images)
- anonymized subject 1 folder
- file describing channels
- raw EEG data file (not all raw formats are possible)
- amplifier and recording information
- events in tabular format

sub-001_ses-01_task-meditation_events.tsv — Edited						
onset	duration	trial_type	response_time	sample	value	
71.3867187500	n/a	stimulus	n/a	18275	128	
75.7304687500	n/a	response	n/a	19387	2	
79.7734375000	n/a	response	n/a	20422	2	
125.6093750000	n/a	stimulus	n/a	32156	128	
179.8007812500	n/a	stimulus	n/a	46029	128	
183.0976562500	n/a	response	n/a	46873	2	
185.6328125000	n/a	response	n/a	47522	4	
284.8593750000	n/a	stimulus	n/a	72924	128	
287.7578125000	n/a	response	n/a	73666	2	
289.9609375000	n/a	response	n/a	74230	2	

BIDS specification:

<https://bids-specification.readthedocs.io/en/latest/>

This example:

<https://openneuro.org/datasets/ds001787>

EEG-BIDS (BIDS export/import)

Figure 2: Edit BIDS participant info - pop_participantinfo

The screenshot shows three windows related to BIDS participant information:

- BIDS task information -- pop_taskinfo()**: This window contains fields for Dataset name (Auditory Oddball Task) and Task name (AuditoryOddball). It also includes a README section with experimental details and a Participant task description.
- Participant information**: A table showing participant details: participant_id (s1, s2), Gender (M, F), Age (1, 2), and Group (1, 2).
- BIDS metadata for participant fields**: A table defining participant fields: participant_id (Unique participant label, n/a), Gender (Participant gender, F), Age (Participant age (years), n/a, years), and Group (Participant group label, Click to specify).

Status
EEG import
BEH import
iEEG import
MEG import
Motion-cap import
Eye-tracking import
EEG export
Eye-tracking export
BEH export
MRI export
Motion-cap export
iEEG export
MEG export

Task and event details

Capturing the nature of events and event context using Hierarchical Event Descriptors (HED)

 Kay Robbins,  Dung Truong,  Stefan Appelhoff,  Arnaud Delorme,  Scott Makeig
doi: <https://doi.org/10.1101/2021.05.06.442841>

CTagger

File

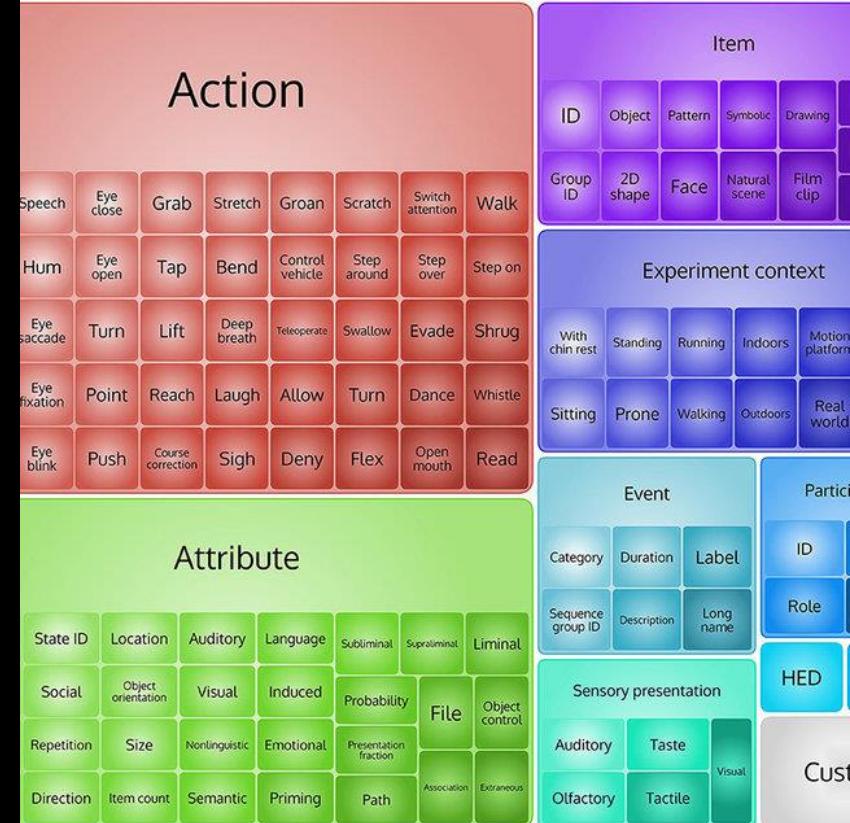
Tagging field: value

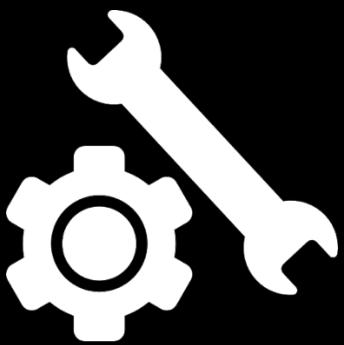
HED tags Save annotations Show HED schema

Field levels

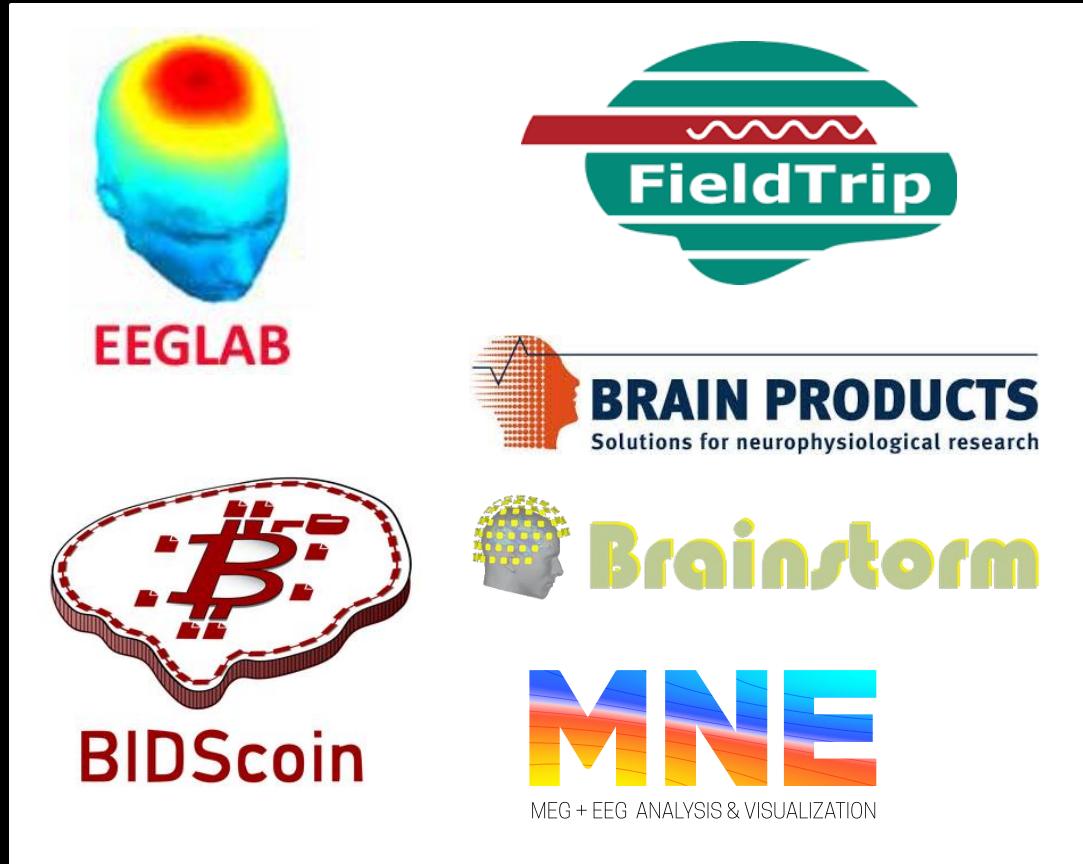
- animal_distractor
- animal_target
- correct
- incorrect

Cancel Done





TOOLS



<https://bids.neuroimaging.io/benefits.html>

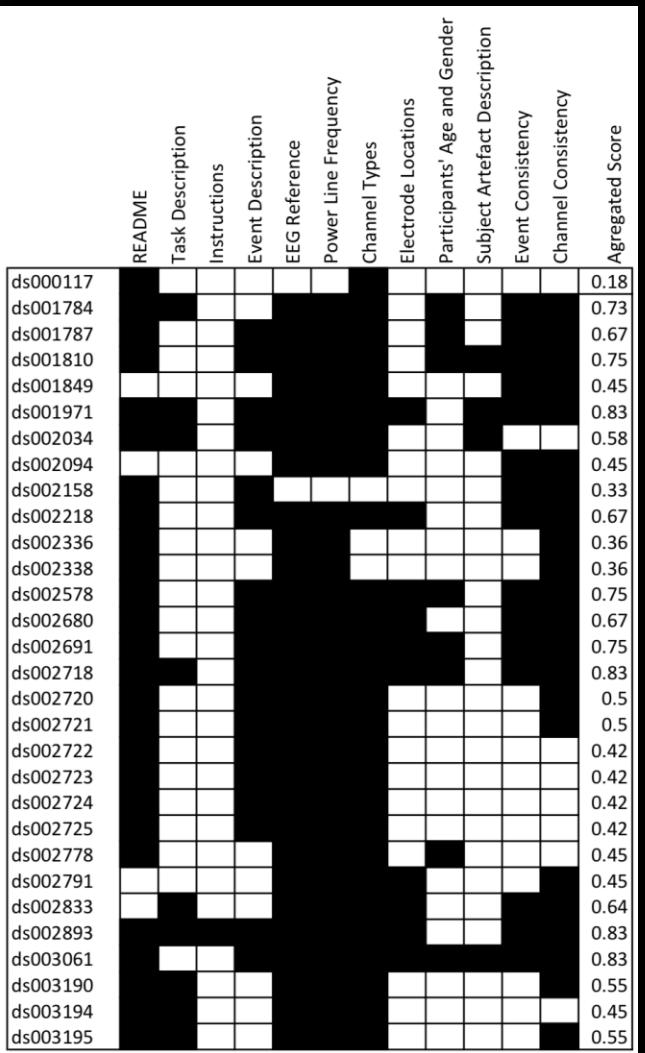
Neuroelectromagnetic Data Archive and Tools Resource



NEMAR

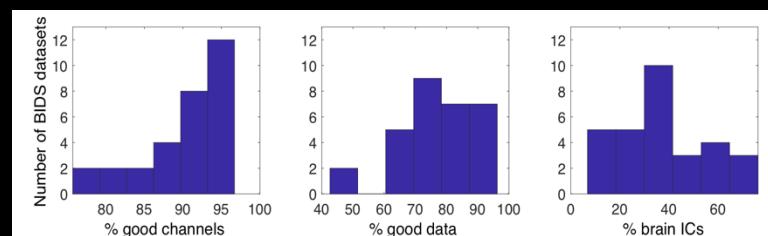
The screenshot shows the OpenNEURO platform interface. At the top, there's a navigation bar with links for 'MY DASHBOARD', 'PUBLIC DASHBOARD', 'SUPPORT', 'FAQ', and 'UPLOAD DATASET'. Below this is a section titled 'Versions' showing two versions: 'Draft' (2019-03-08) and '1.0.0' (2019-03-08). The main content area displays a dataset titled 'EEG meditation study' uploaded by Arnaud Delorme. It includes a 'BIDS Validation' section indicating 'Valid' status with 2 warnings, a 'Dataset File Tree' showing directory structure and files like 'participants.json', 'participants.tsv', and 'README', and a large list of subjects from 'sub-001' to 'sub-012'. There are also sections for 'README', 'AUTHORS', 'DATASET DOI', 'LICENSE', 'ACKNOWLEDGEMENTS', 'HOW TO ACKNOWLEDGE', 'FUNDING', and 'REFERENCES AND LINKS'.

The screenshot shows the NEMAR homepage. The header features the 'NEMAR' logo with a brain icon. Below the header, there's a large call-to-action text: 'Search, Visualize, Analyze, Download'. To the right, there's a large image of a brain with a grid overlay. Below the brain, there's a search bar with the placeholder 'Search for Datasets and Tools' and a 'Search' button. The footer contains links for 'About', 'Discover', 'Community', 'Support', 'Login', and a 'Reading List' link.

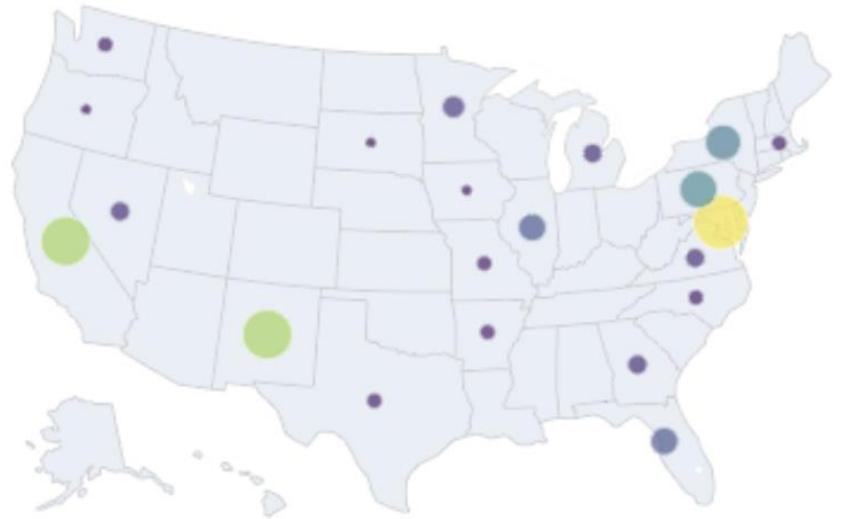


Delorme et al. (2021) Tools for Importing and Evaluating BIDS-EEG Formatted Data. NER IEEE conference on biomedical engineering. May, online.

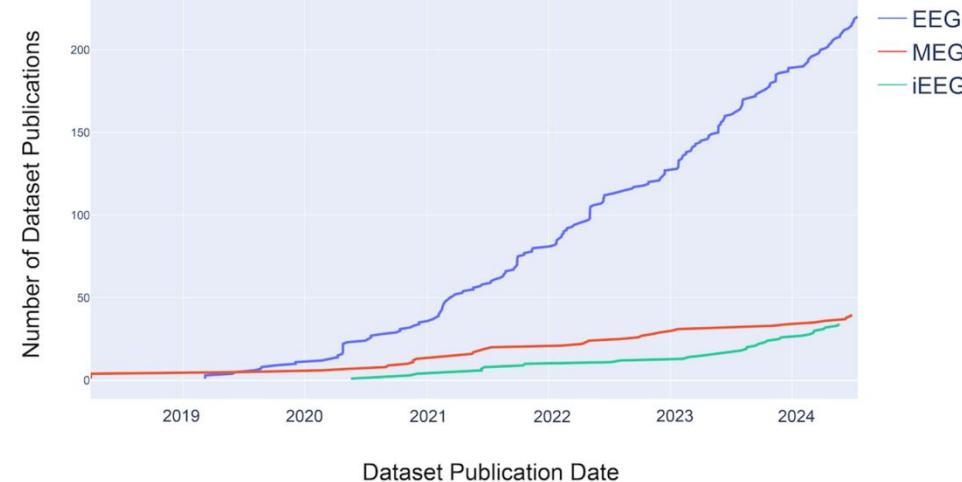
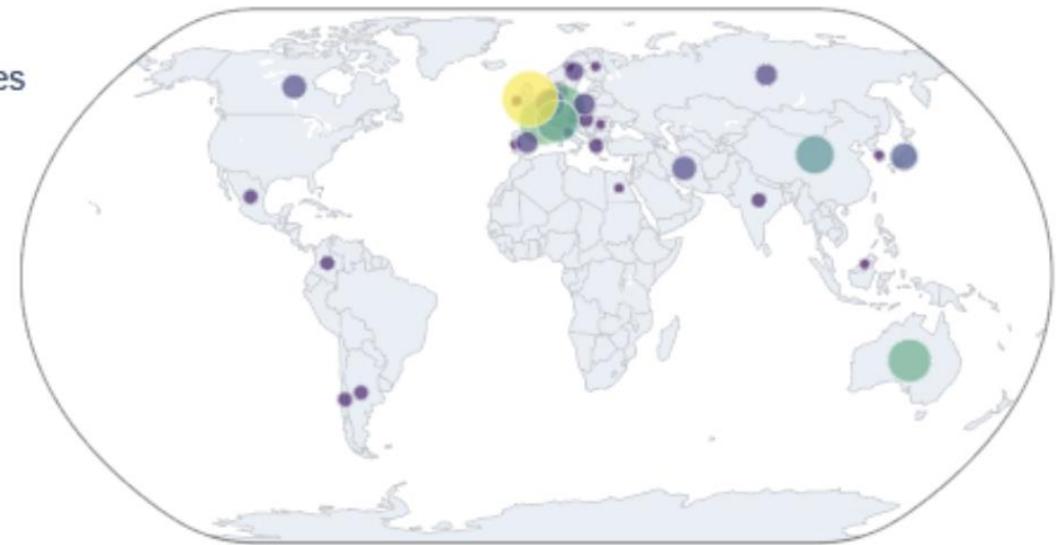
Dataset	n	Flat	Chan.	Good chan.	Good data	Brain ICs
ds000117	84	-	74	89 - 90	68 - 78	28 - 32
ds001784	30	-	60	85 - 92	63 - 77	5 - 11
ds001787	40	-	64	93 - 95	85 - 90	24 - 30
ds001810	263	-	64	78 - 80	71 - 74	35 - 37
ds001849	120	-	30	86 - 88	66 - 70	41 - 45
ds001971	273	-	112*	91 - 95	81 - 84	8 - 10
ds002034	167	21	62	95 - 98	47 - 52	29 - 33
ds002094	43	-	30	88 - 93	66 - 77	37 - 45
ds002158	8	-	63	80 - 86	19 - 66	11 - 23
ds002218	18	-	32	94 - 97	66 - 78	35 - 46
ds002336	54	1	63	81 - 83	74 - 89	33 - 39
ds002338	85	-	63	74 - 77	87 - 95	24 - 28
ds002578	2	-	18	78 - 89	96 - 97	46 - 60
ds002680	350	-	19	92 - 93	81 - 84	53 - 56
ds002691	20	-	32	93 - 97	82 - 87	30 - 40
ds002718	18	-	74	95 - 96	58 - 76	20 - 28
ds002720	165	-	19	93 - 95	68 - 72	52 - 56
ds002721	185	-	19	95 - 96	67 - 72	45 - 49
ds002722	94	-	32	92 - 95	63 - 70	36 - 39
ds002723	44	-	32	95 - 97	66 - 76	36 - 40
ds002724	96	-	32	94 - 96	72 - 77	35 - 39
ds002725	105	-	31	90 - 94	80 - 85	53 - 58
ds002778	46	-	32	90 - 95	68 - 78	41 - 48
ds002791	92	-	256	90 - 92	62 - 70	6 - 7
ds002833	80	-	256	91 - 94	82 - 88	6 - 8
ds002893	55	12	36	89 - 93	76 - 83	29 - 37
ds003061	39	-	64	84 - 89	86 - 92	22 - 28
ds003190	384	2	8	81 - 83	89 - 91	74 - 78
ds003194	29	-	19	93 - 97	90 - 95	72 - 79
ds003195	20	-	19	90 - 97	89 - 94	65 - 76



United States

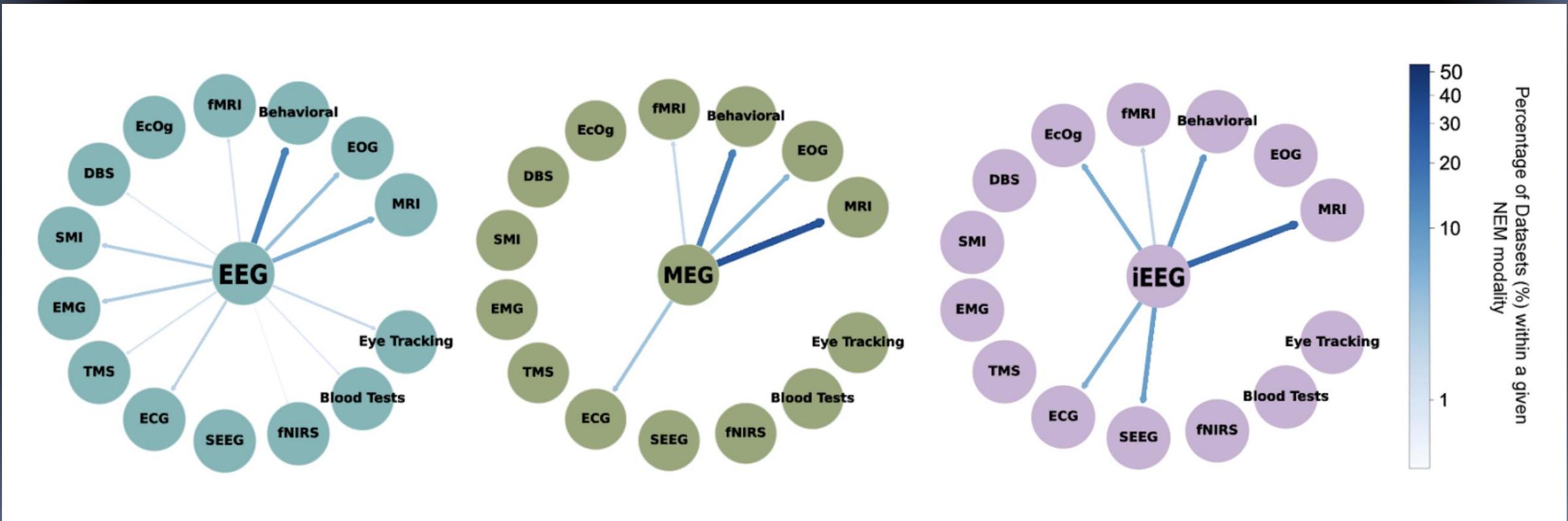
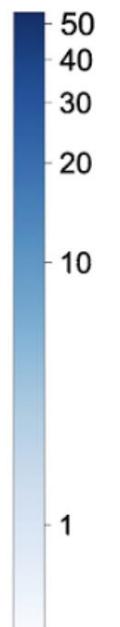


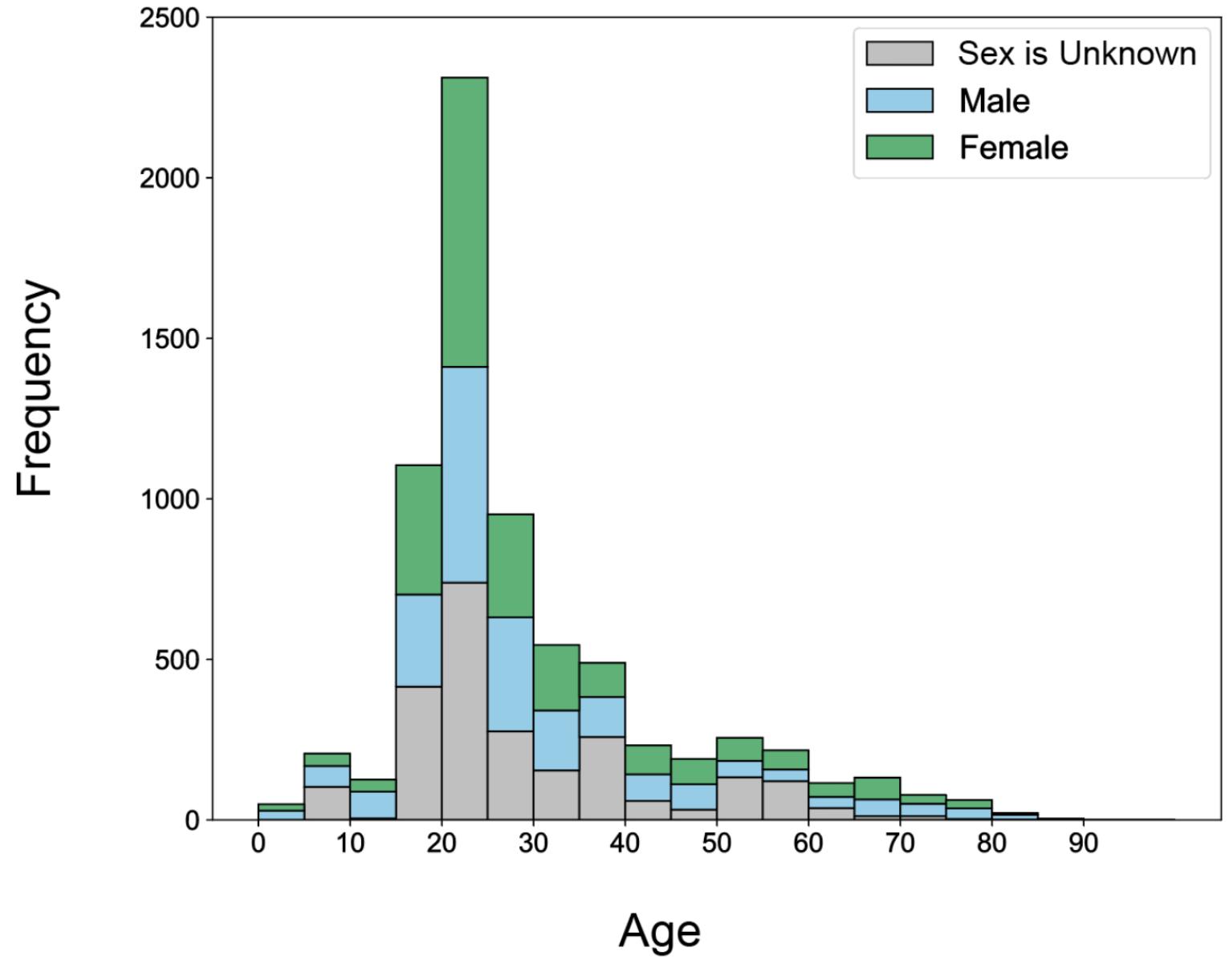
World (Excluding US)



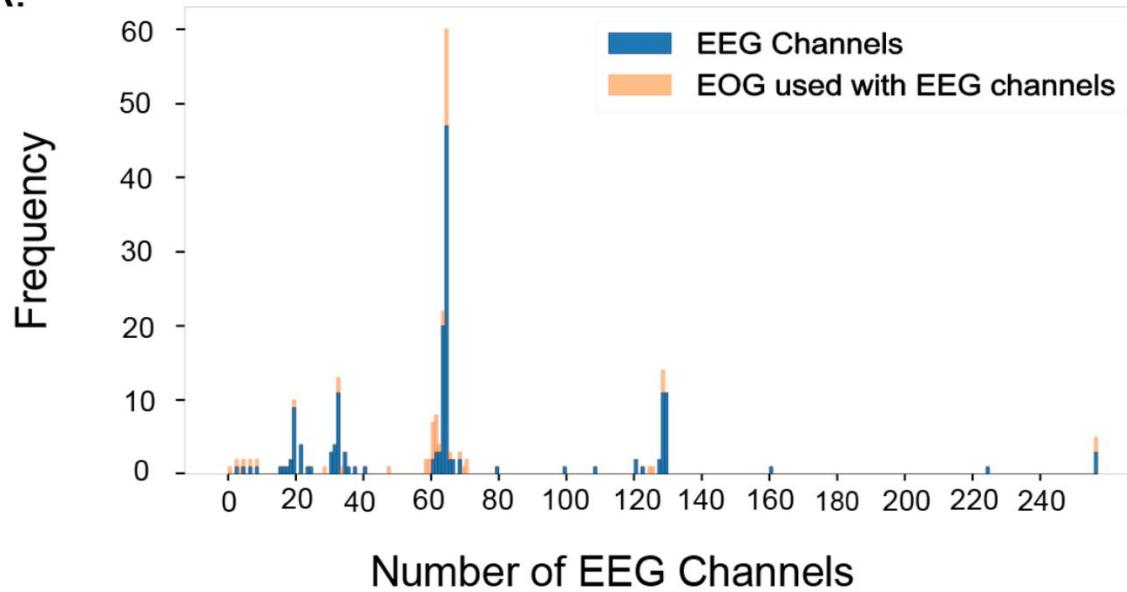
Brandmeyer, T. ... Delorme, A. (2024). Neuroelectromagnetic Data Archive and Repository - Open Source Platform Analysis. Manuscript submitted for publication.

Percentage of Datasets (%) within a given NEM modality

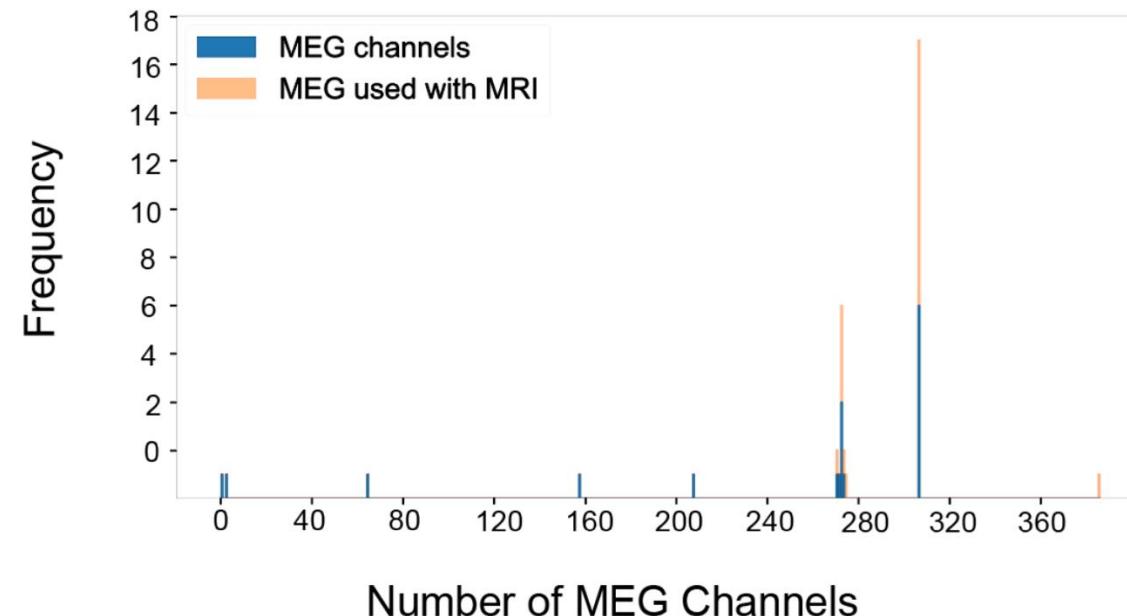


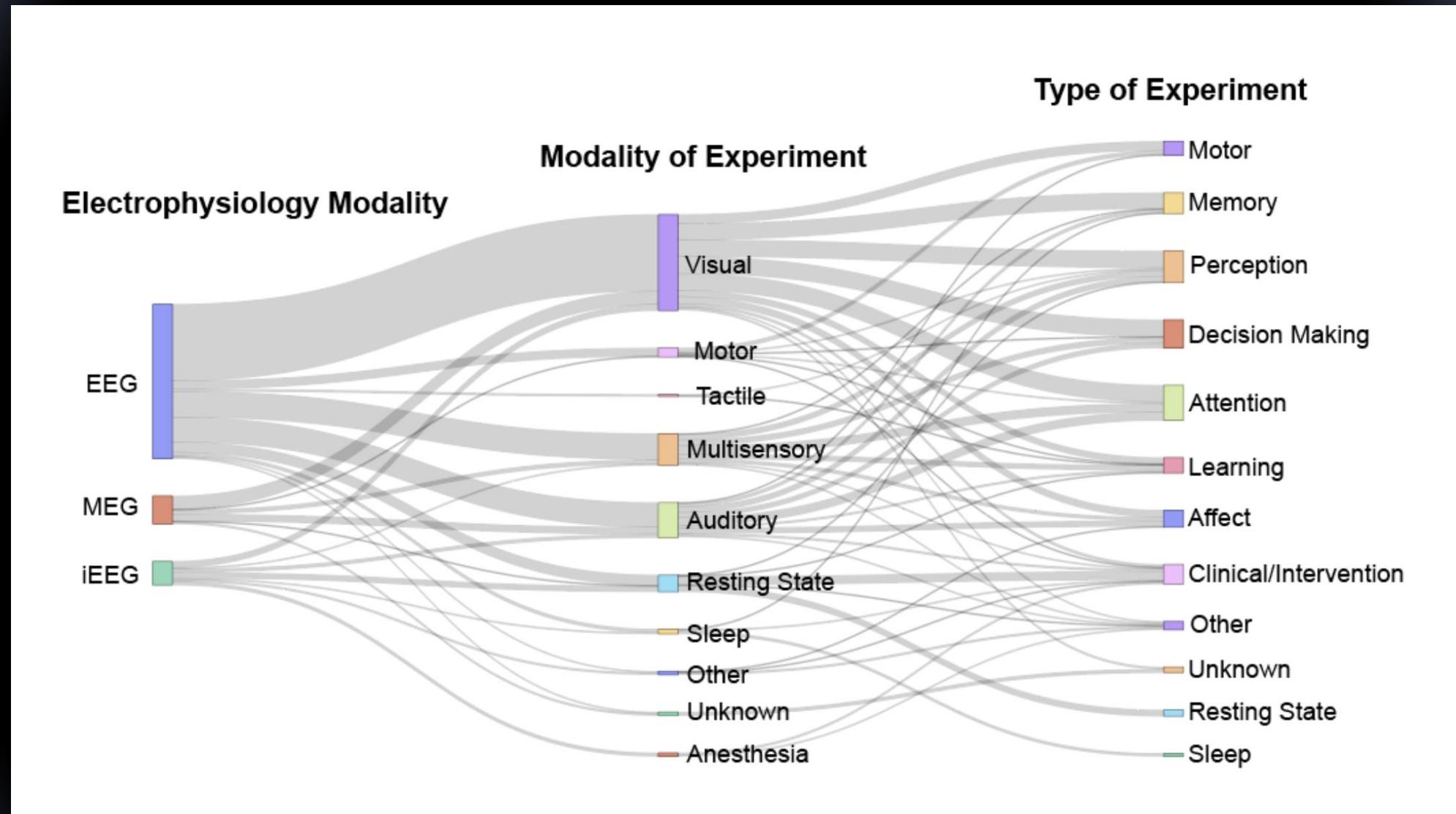


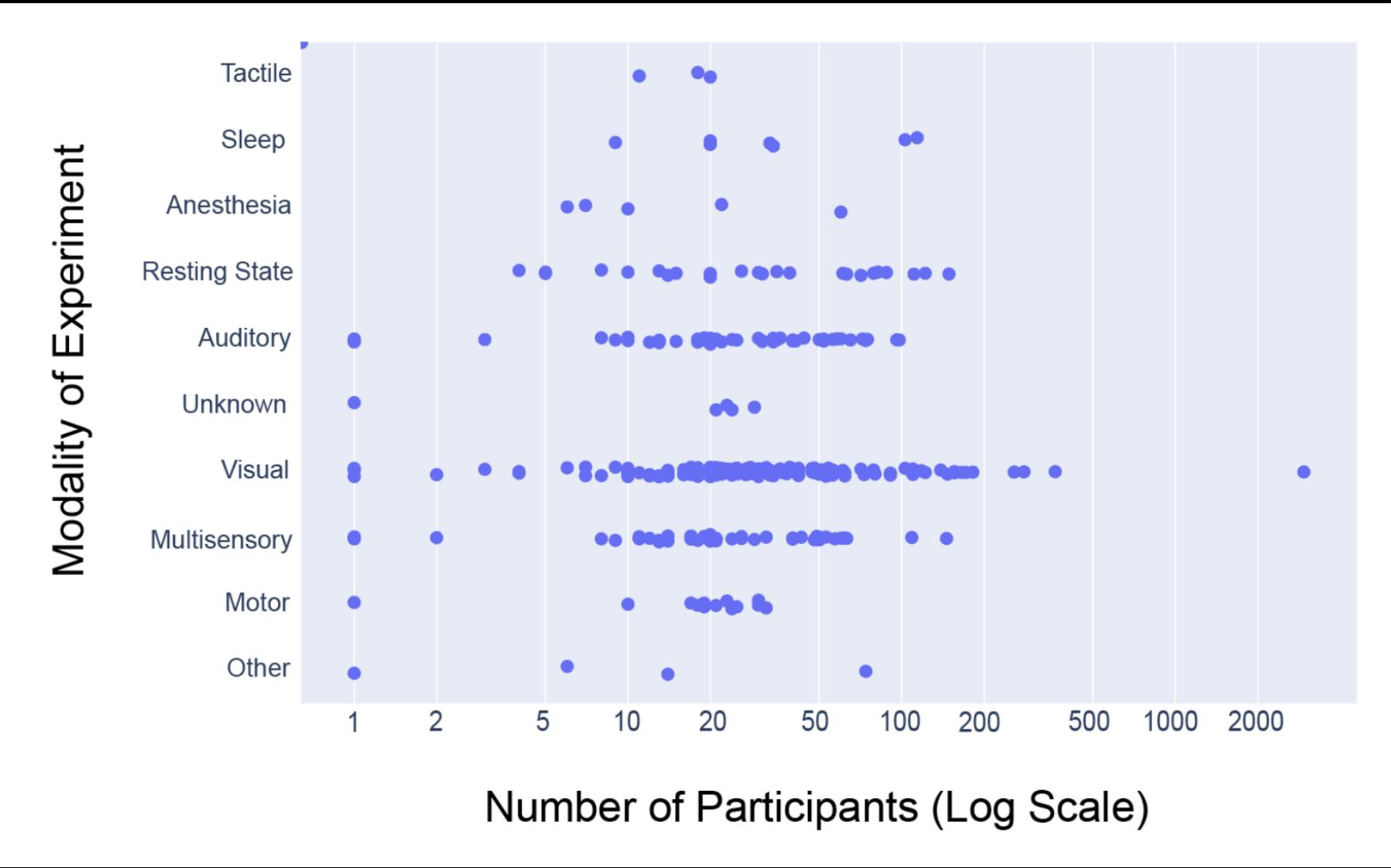
A.

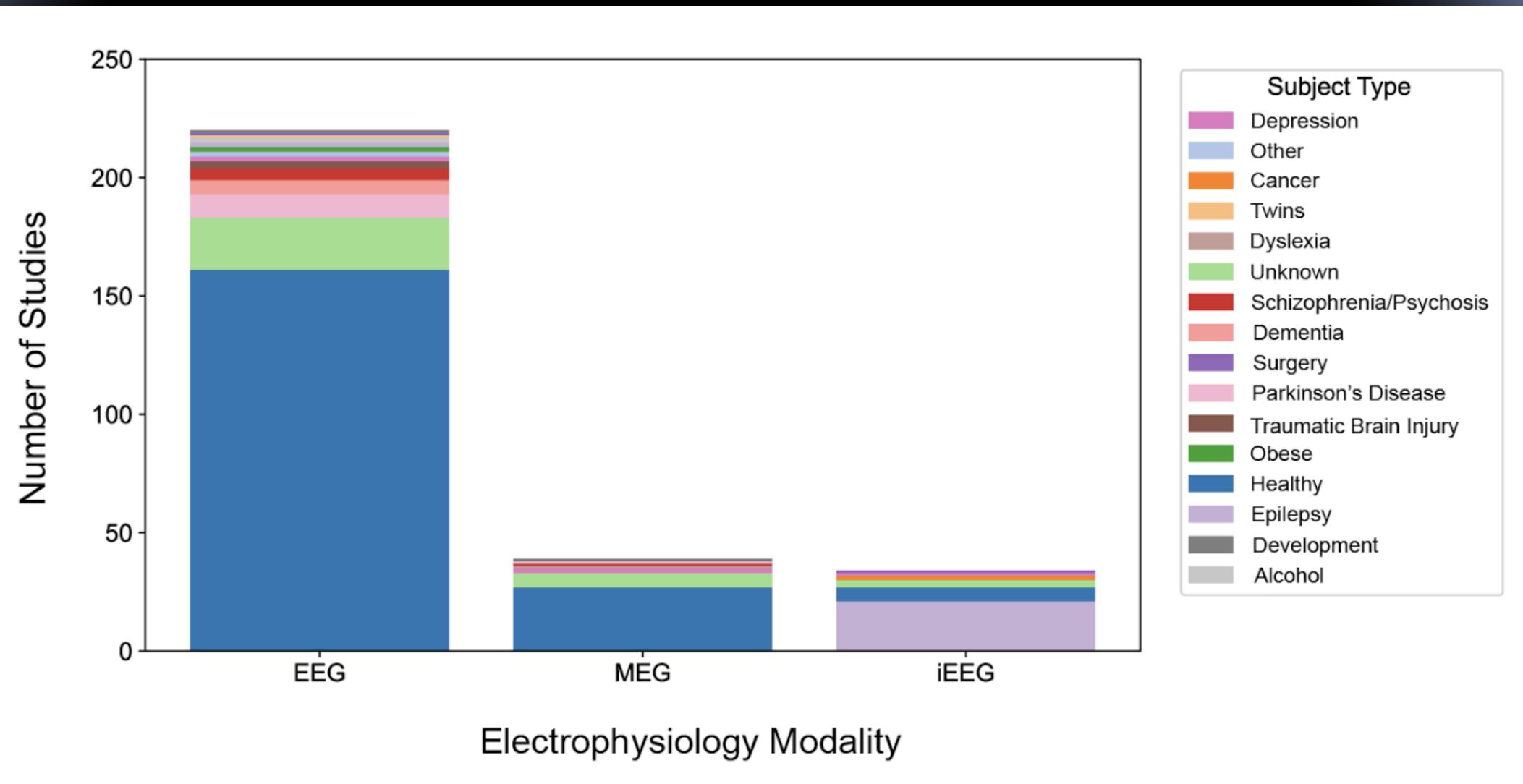


B.











ACCESS

EXPANS-

No wait
48-hour max



GPU Nodes: 208 NVIDIA V100 GPUs (32 GB)

CPU Nodes: Intel Xeon (x40)

Memory per Node: 728 compute nodes
(128 cores and 256 Gb of RAM)

Computational resources

