Open science efforts for PET neuroimaging data

Melanie Ganz, Assoc. Prof.

Department of Computer Science, University of Copenhagen, Denmark

Neurobiology Research Unit, Copenhagen University Hospital, Rigshospitalet, Denmark

What is the Brain Imaging Data Structure (BIDS)?

- It's a data structure ; nothing to do with format per se

- It's about: - how you organize data in a folder
  - how you name files
  - how you document metadata
  - using community standards and dictionaries to do all the above

- It cares about imaging data but also behaviour/cognition

bids.neuroimaging.io

BIDS slides courtesy of Cyril Pernet

BIDS is a widely supported "best practice" (becoming *a posteriori*, the "standard", rather than being created *de novo*)

Oostenveld & Zwiers – 2020 BIDS talk

Evolution of BIDS

1. Kickoff meeting at Stanford in Spring 2015
2. Meeting at OHBM 2015 (June)
3. Introduced to neuroinformatics community at INCF Congress 2015 (August)
4. First release candidate and public call for comments (September)
5. Version 1.0.0 published along the introductory paper

→ Initially covered structural MRI and fMRI, now all sort of MRI, PET, EEG, MEG, iEEG, extensions for animal, connectivity, imaging genomics ..

CREDITS

Data Sharing Task Force

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

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

Received: 18 December 2015
Accepted: 19 May 2016
Published: 21 June 2016

The Poldrack Lab @ Stanford

Brain Imaging Data Structure v1.6.0

The BIDS Specification
- Introduction
- Common principles
- Modality agnostic files
- Modality specific files
- Derivatives
- Longitudinal and multi-site studies
- BIDS Extension Proposals
- Appendix
- Changelog

The BIDS Starter Kit

Datatype specific publications

EEG

- Pernet, C. R., Appelhoff, S., Gorgolewski, K.J., Flandin, G., Phillips, C., Delorme, A., Oostenveld, R. (2019). **EEG-BIDS, an extension to the brain imaging data structure for electroencephalography**. Scientific data, 6 (103). doi:10.1038/s41597-019-0104-8

iEEG

- Holdgraf, C., Appelhoff, S., Bickel, S., Bouchard, K., D'Ambrosio, S., David, O., Devinsky, O., Dichter, B., Flinker, A., Foster, B. L., Gorgolewski, K. J., Groen, I., Groppe, D., Gunduz, A., Hamilton, L., Honey, C. J., Jas, M., Knight, R., Lauchaux, J.-P., Lau, J. C., Lee-Messer, C., Lundstrom, B. N., Miller, K. J., Ojemann, J. G., Oostenveld, R., Petridou, N., Piantoni, G., Pigorini, A., Pouratian, N., Ramsey, N. F., Stolk, A., Swann, N. C., Tadel, F., Voytek, B., Wandell, B. A., Winawer, J., Whitaker, K., Zehl, L., Hermes, D. (2019). **iEEG-BIDS, extending the Brain Imaging Data Structure specification to human intracranial electrophysiology**. Scientific data, 6 (102). doi:10.1038/s41597-019-0105-7

MEG

- Niso Galan, J.G., Gorgolewski, K.J., Bock, E., Brooks, T.L., Flandin, G., Gramfort, A., Henson, R.N., Jas, M., Litvak, V., Moreau, J., Oostenveld, R., Schoffelen, J.-M., Tadel, F., Wexler, J., Baillet, S. (2018). **MEG-BIDS, the brain imaging data structure extended to magnetoencephalography**. Scientific Data, 5 (180110). doi:10.1038/sdata.2018.110

Table of contents

Motivation

Extensions

Citing BIDS
- Original publication
- Datatype specific publications
  - EEG
  - iEEG
  - MEG
  - PET
  - Genetics
- Research Resource Identifier (RRID)

Brain Imaging Data Structure v1.6.0

The BIDS Specification
- Introduction
- Common principles
- Modality agnostic files
- Modality specific files
- Derivatives
- Longitudinal and multi-site studies
- BIDS Extension Proposals
- Appendix
- Changelog

The BIDS Starter Kit

PET

- Knudsen GM, Ganz M, Appelhoff S, Boellaard R, Bormans G, Carson RE, Catana C, Doudet D, Gee AD, Greve DN, Gunn RN, Halldin C, Herscovitch P, Huang H, Keller SH, Lammertsma AA, Lanzenberger R, Liow JS, Lohith TG, Lubberink M, Lyoo CH, Mann JJ, Matheson GJ, Nichols TE, Nørgaard M, Ogden T, Parsey R, Pike VW, Price J, Rizzo G, Rosa-Neto P, Schain M, Scott PJH, Searle G, Slifstein M, Suhara T, Talbot PS, Thomas A, Veronese M, Wong DF, Yaqub M, Zanderigo F, Zoghbi S, Innis RB. (2020). **Guidelines for Content and Format of PET Brain Data in Publications and in Archives: A Consensus Paper**. Journal of Cerebral Blood Flow and Metabolism, 2020 Aug; 40(8): 1576-1585. doi:10.1177/0271678X20905433

Genetics

- Clara Moreau, Martineau Jean-Louis, Ross Blair, Christopher Markiewicz, Jessica Turner, Vince Calhoun, Thomas Nichols, Cyril Pernet (2020). **The genetics-BIDS extension: Easing the search for genetic data associated with human brain imaging**. GigaScience, 9 (10). doi:10.1093/gigascience/giaa104

Research Resource Identifier (RRID)

BIDS has also a Research Resource Identifier (RRID), which you can also include in your citations in addition to relevant publications (see above):

Table of contents

Motivation
Extensions
Citing BIDS
- Original publication
- Datatype specific publications
  - EEG
  - iEEG
  - MEG
  - PET
  - Genetics
- Research Resource Identifier (RRID)

v: stabl

The BIDS goal is to make more data accessible to more researchers

Getting lost in your data

COBIDAS MEEG – figure 1

- **Multitude of techniques** (MEG, EEG, PET, MRI, NIRS, TMS, etc …) and applications.
- Despite similarities in experimental designs and data types **each researcher tends to organize and describe their data in their own way**.

Making more data accessible

- for yourself in 6 months time

- to people in the lab
(new students, collaborators, governance)

- to other researchers
(data sharing)

BIDS data structure basics

Essence of BIDS

- Directory hierarchy and file naming system
  - Conveys essential meta-data in filenames alone
- Remaining meta-data in “sidecar” files

A BIDS folder (any/all modality)

Name	Date modified	Type
stimuli	21/03/2018 21:58	File folder
sub-002	24/03/2018 08:56	File folder
sub-003	24/03/2018 08:56	File folder
sub-004	24/03/2018 08:56	File folder
sub-005	24/03/2018 08:56	File folder
sub-006	24/03/2018 08:56	File folder
sub-007	24/03/2018 08:56	File folder
sub-008	24/03/2018 08:56	File folder
sub-009	24/03/2018 08:56	File folder
sub-010	24/03/2018 08:56	File folder
sub-011	24/03/2018 08:56	File folder
sub-012	24/03/2018 08:56	File folder
sub-013	24/03/2018 08:56	File folder
sub-014	24/03/2018 08:56	File folder
sub-015	24/03/2018 08:56	File folder
sub-016	24/03/2018 08:56	File folder
sub-017	24/03/2018 08:56	File folder
sub-018	24/03/2018 08:56	File folder
sub-019	24/03/2018 08:56	File folder
dataset_description.json	15/03/2018 11:30	JSON File
participants.tsv	19/03/2018 20:21	TSV File
README.txt	15/03/2018 11:33	TXT File

- **source** (optional)
- **stimuli** (optional)
- **derivatives** (optional)
- **sub-XXX**

Anyone can now find his/her way around data !

rigid folder structure and naming convention

Metadata as text files (tsv, json) with standard dictionary

Name	Date modified	Type
stimuli	21/03/2018 21:58	File folder
sub-002	24/03/2018 08:56	File folder
sub-003	24/03/2018 08:56	File folder
sub-004	24/03/2018 08:56	File folder
sub-005	24/03/2018 08:56	File folder
sub-006	24/03/2018 08:56	File folder
sub-007	24/03/2018 08:56	File folder
sub-008	24/03/2018 08:56	File folder
sub-009	24/03/2018 08:56	File folder
sub-010	24/03/2018 08:56	File folder
sub-011	24/03/2018 08:56	File folder
sub-012	24/03/2018 08:56	File folder
sub-013	24/03/2018 08:56	File folder
sub-014	24/03/2018 08:56	File folder
sub-015	24/03/2018 08:56	File folder
sub-016	24/03/2018 08:56	File folder
sub-017	24/03/2018 08:56	File folder
sub-018	24/03/2018 08:56	File folder
sub-019	24/03/2018 08:56	File folder
dataset_description.json	15/03/2018 11:30	JSON File
participants.tsv	19/03/2018 20:21	TSV File
README.txt	15/03/2018 11:33	TXT File

```
{
    "Name": "",
    "BIDSVersion": "",
    "License": "",
    "Authors": "",
    "Acknowledgements": "",
    "HowToAcknowledge": " ",
    "Funding": "",
    "ReferencesAndLinks": "",
    "SourceDatasetsURLs": ""
}
```

Your data are identifiable and citable

Metadata as text files (tsv, json) with standard dictionary

Name	Date modified	Type
stimuli	21/03/2018 21:58	File folder
sub-002	24/03/2018 08:56	File folder
sub-003	24/03/2018 08:56	File folder
sub-004	24/03/2018 08:56	File folder
sub-005	24/03/2018 08:56	File folder
sub-006	24/03/2018 08:56	File folder
sub-007	24/03/2018 08:56	File folder
sub-008	24/03/2018 08:56	File folder
sub-009	24/03/2018 08:56	File folder
sub-010	24/03/2018 08:56	File folder
sub-011	24/03/2018 08:56	File folder
sub-012	24/03/2018 08:56	File folder
sub-013	24/03/2018 08:56	File folder
sub-014	24/03/2018 08:56	File folder
sub-015	24/03/2018 08:56	File folder
sub-016	24/03/2018 08:56	File folder
sub-017	24/03/2018 08:56	File folder
sub-018	24/03/2018 08:56	File folder
sub-019	24/03/2018 08:56	File folder
dataset_description.json	15/03/2018 11:30	JSON File
participants.tsv	19/03/2018 20:21	TSV File
README.txt	15/03/2018 11:33	TXT File

participant_id	age	sex
sub-002	34	M
sub-003	12	F
sub-004	33	F

Subjects info are shared at the root – easy to figure populations, age, and other basic demographics

Human and machine readable!

A BIDS folder (any/all modality)

Name redundancy – harder to make mistakes ; files comes with metadata (json)

Inside folder sub-002

→ anat folder
sub-002_T1w.nii.gz

→ eeg folder
????

→ pet folder
????

A BIDS folder (any/all modality)

Name redundancy – harder to make mistakes ; files comes with metadata (json)

Inside folder sub-002

→ anat folder
sub-002_T1w.nii.gz

→ eeg folder
sub-002_something_eeg.set

→ pet folder
sub-002_something_pet.nii

What BIDS is NOT

- The Brain Imaging Data *Structure*, is not a new file format
- BIDS is not a database tool, but it helps archiving and major imaging databasing tools (XNat, Loris, etc) are BIDS compliant
- BIDS is not a search engine, but it facilitates data discovery for search engines
- BIDS is not a sharing tool, but it facilitates sharing
  - with yourself next time you look at old data,
  - with members of the lab (new students, new collaborators)
  - with the community

Standard nomenclature

Journal of Cerebral Blood Flow & Metabolism (2007) 27, 1533–1539
© 2007 ISCBFM All rights reserved 0271-678X/07 $30.00
www.jcbfm.com

Review Article

Consensus nomenclature for *in vivo* imaging of reversibly binding radioligands

Robert B Innis[1], Vincent J Cunningham[2], Jacques Delforge[3], Masahiro Fujita[1], Albert Gjedde[4], Roger N Gunn[5], James Holden[6], Sylvain Houle[7], Sung-Cheng Huang[8], Masanori Ichise[9], Hidehiro Iida[10], Hiroshi Ito[11], Yuichi Kimura[12], Robert A Koeppe[13], Gitte M Knudsen[14], Juhani Knuuti[15], Adriaan A Lammertsma[16], Marc Laruelle[2], Jean Logan[17], Ralph Paul Maguire[18], Mark A Mintun[19], Evan D Morris[20], Ramin Parsey[9], Julie C Price[21], Mark Slifstein[9], Vesna Sossi[22], Tetsuya Suhara[11], John R Votaw[23], Dean F Wong[24] and Richard E Carson[25]

[1]National Institute of Mental Health, Bethesda, Maryland, USA; [2]GlaxoSmithKline and Imperial College, London, UK; [3]CEA/DSV/SHF, Orsay, France; [4]University of Aarhus, Aarhus, Denmark; [5]GlaxoSmithKline and University of Oxford, London, UK; [6]University of Wisconsin, Madison, Wisconsin, USA; [7]Centre for Addiction and Mental Health & University of Toronto, Toronto, Ontario, Canada; [8]UCLA School of Medicine, Los Angeles, California, USA; [9]Columbia University, New York, New York, USA; [10]National Cardiovascular Center Research Institute, Suita City, Osaka, Japan; [11]National Institute of Radiological Sciences, Chiba, Japan; [12]Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan; [13]University of Michigan, Ann Arbor, Michigan, USA; [14]Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark; [15]Turku PET Centre, Turku, Finland; [16]VU University Medical Centre, Amsterdam, The Netherlands; [17]Brookhaven National Laboratory, Upton, New York, USA; [18]Pfizer Global R&D, Groton, Connecticut, USA; [19]Washington University School of Medicine, St Louis, Missouri, USA; [20]Indiana University-Purdue University, Indianapolis, Indiana, USA; [21]University of Pittsburgh, Pittsburgh, Pennsylvania, USA; [22]University of British Columbia, Vancouver, British Columbia, Canada; [23]Emory University, Atlanta, Georgia, USA; [24]Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; [25]Yale University, New Haven, Connecticut, USA

Consensus on publishing PET studies

- Replication in science can be improved with standards for reporting and sharing of primary research data

Opinion

JCBFM

Journal of Cerebral Blood Flow & Metabolism
2020, Vol. 40(8) 1576–1585
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0271678X20905433
journals.sagepub.com/home/jcbfm
SAGE

Guidelines for the content and format of PET brain data in publications and archives: A consensus paper

Gitte M Knudsen[1], Melanie Ganz[1], Stefan Appelhoff[2], Ronald Boellaard[3], Guy Bormans[4], Richard E Carson[5], Ciprian Catana[6], Doris Doudet[7], Antony D Gee[8], Douglas N Greve[6], Roger N Gunn[9], Christer Halldin[10], Peter Herscovitch[11], Henry Huang[5], Sune H Keller[12], Adriaan A Lammertsma[3], Rupert Lanzenberger[13], Jeih-San Liow[14], Talakad G Lohith[15], Mark Lubberink[16], Chul H Lyoo[17], J John Mann[18], Granville J Matheson[10], Thomas E Nichols[19], Martin Nørgaard[1], Todd Ogden[20], Ramin Parsey[21], Victor W Pike[14], Julie Price[6], Gaia Rizzo[9], Pedro Rosa-Neto[22,23], Martin Schain[20], Peter JH Scott[24], Graham Searle[9], Mark Slifstein[21], Tetsuya Suhara[25], Peter S Talbot[26], Adam Thomas[27], Mattia Veronese[28], Dean F Wong[29], Maqsood Yaqub[3], Francesca Zanderigo[30], Sami Zoghbi[14] and Robert B Innis[14]

From OpenNeuro ...

- Official repository for BRAIN Initiative
- Part of the Amazon Public Datasets project
- 627 public datasets
- 21.756 subjects / ~16 TB
- 10-20 new dataset uploads per
- month
- Serving 1000 + downloads/month (almost 20TB of data)
- Over 8K users/month

...to OpenNeuroPET

- Establish PET archive as an extension of OpenNeuro
  - Standard format and content
  - "Best Practices" for pipelines and QC checks
- Educate and seek feedback from the PET user community
- Establish average images of receptor density, connecting to the larger fMRI community

OpenNeuroPET

- Principles
  - Standard format to meaningfully share and combine data
  - Follow Guidelines, include “best practices” for data pipelines, and sample data sets for QC
- Benchmarks of success
  - Number of scans deposited
  - Number of resulting publications
- Community engagement
  - Provide help to sites
  - Integrate with commonly used software
  - Stimulate projects

OpenNeuroPET setup

OpenNeuroPET setup

PI: Robert Innis

Funding: NIMH via BRAIN Initiative (Brain Research through Advancing Innovative Neurotechnologies)

Duration: Oct 2021 – Sept 2026

Collaborator: NRU, Rigshospital; MGH

Consultant: Russell Poldrack (Stanford)

PI: Gitte Moos Knudsen

Funding: Novo Nordisk Foudation (Research Infrastructure program)

Duration: Jan 2021 – Dec 2025

Collaborator: NIMH; MGH

Consultant: Russell Poldrack (Stanford)

NIH National Institute of Mental Health

Sharing data using 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.

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

https://www.force11.org/group/fairgroup/fairprinciples

Delorme – 2020 BIDS talk

Data Sharing in EU

- **Declaration of Helsinki:** generate, collect, use, process data
- **Declaration of Taipei:** share data

- Consider using 2 separate consent forms, or a 2-parts consent form: one for your study and one for data sharing.
- GDPR means you need a lawful basis – as researchers this is our job to generate, collect, use and process, and arguably to share ('public task')
- Within GDPR, sharing/transferring data is considered to be a type of processing – as data controller you can ask OpenNEuroPET, a data processor, to share your data securely (after signing a data transfer agreement contract)

Proposed legal solution

open-brain-consent.readthedocs.io/en/stable/gdpr/index.html

Open Brain Consent

stable

Search docs

Sample consent forms
Recommendations
Ultimate consent form
GDPR edition
Context
GDPR Documents
Ultimate consent form - GDPR edition
Data Privacy Impact assessment (DPIA)
Data User Agreement
Anonymization tools
Contribute
Contact information
Discussions

GDPR edition

Context

Responsible and ethical sharing of data and code that underlie the results of scientific work is an important step towards improving research transparency, fostering inclusivity and building public trust in science. At the same time, privacy of sensitive personal data, including neuroimaging data, is highly important. Ethical review boards at research institutions are responsible for reviewing a study protocol and deciding whether it can continue based on its adherence to the relevant ethical and research integrity principles, which typically include regulations on personal data privacy. In the European Union, such data privacy requirements are subject to the General Data Protection Regulation (GDPR) as implemented by its member countries. Despite the increased importance that funders and institutions are currently placing on open science practices, no clear, thorough and openly available guides exist for publicly sharing neuroimaging data under the GDPR. Our goal is to share community-contributed templates for consent forms and other documentation required for ethical approval of brain research data processing and sharing under the GDPR.

History and links

Several people and groups have contributed to this project over the past years. Below are some links pointing to events/initiatives where work was done on the templates.

Consent from – inform people on:

- *'While the collection, use and storage of your data are done for the purpose of conducting the study to which you are currently participating, these data might also be used for other future research projects in the field of medical and cognitive neuroscience. '* = strike the balance in purpose

- Confidentiality of your data, Access to your data for verification, Retention period of your data, Withdrawing consent,
- Passing on to countries outside the European Union (EU)
- Information about your rights regarding processing of the data

FAIR and Secure: nEUrothenticate

FAIR and Secure: nEUrothenticate

- **Governance** (how it is handled, by whom)
- **Findability and Accessibility** (make dataset level metadata available thx to DataLad – read by OpenNeuro, CONP, EBRAINS)
- **Interoperable and Reusable**: BIDS
- **Protection**

➢ Identity and Access Management (IAM)

➢ Data encryption (at rest and in transit)

- **Auditability** (keep trace of users, downloads, any authorized and (tentative of) non-authorized access, etc)

Thank you!

novo nordisk fonden

NIH National Institute of Mental Health

Athinoula A. Martinos Center For Biomedical Imaging

Harvard–MIT Health Sciences & Technology

MASSACHUSETTS GENERAL HOSPITAL

Stanford University

REGION H Rigshospitalet