

# SCIENTIFIC DATA

OPEN  
COMMENT

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

Cyril R. Pernet <sup>1</sup>, Stefan Appelhoff <sup>2</sup>, Krzysztof J. Gorgolewski <sup>3</sup>, Guillaume Flandin<sup>4</sup>, Christophe Phillips <sup>5</sup>, Arnaud Delorme<sup>6,7</sup> & Robert Oostenveld <sup>8,9</sup>

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 organize and share study data within and between laboratories. Here we present an extension to BIDS for electroencephalography (EEG) data, EEG-BIDS, along with tools and references to a series of public EEG datasets organized using this new standard.

EEG was first applied in humans nearly a century ago<sup>1</sup>. It records the electric potential fluctuations at the scalp, primarily from locally synchronous post-synaptic activity in the apical dendrites of pyramidal cells in the cortex. Widely used in both clinical and non-clinical settings, EEG is becoming increasingly important in cognitive neuroscience, with statistics from scientific reports showing that interest in EEG has been growing faster since the early 2000s. This can be attributed to interest in brain–computer interfaces and more sophisticated dynamics measures along with more accurate biophysical models for reconstructing sources. EEG is more versatile than other imaging modalities because (i) it is lightweight and requires relatively low-cost equipment, (ii) it can be used in many different environments (e.g., while sitting in a lab chair, driving, walking, playing a video game, sleeping, interacting with others in social situations, etc.), (iii) it can be used either alone or in conjunction with other imaging modalities, (iv) its task design constraints are less restrictive than metabolic (PET) or hemodynamic (fMRI) imaging methods, and (v) it captures neural activity with millisecond precision, making it possible to record cortical dynamics at the speed of perception, thought and action. Because of this versatility, the field of applications for EEG is broad. In turn, the commercial market for EEG systems is much larger than that of other imaging techniques (e.g., PET, MRI, MEG), resulting in a multitude of equipment manufacturers (more than 10 major manufacturers in neuroscience) building different hardware systems, usually with their own software and proprietary data formats. Manufacturers have little financial incentive to cooperate and provide compatible formats. This resulting diversity of formats is an impediment to reusing data as well as to building large-scale EEG databases.

The Brain Imaging Data Structure, originally proposed for magnetic resonance imaging data (MRI), is a human brain research community standard used for organizing and sharing brain imaging study data within and between laboratories for many (ultimately all) imaging modalities<sup>2</sup>. BIDS primarily addresses the heterogeneity of data organization by following the FAIR principles<sup>3</sup> of findability, accessibility, interoperability, and reusability. BIDS addresses *findability* and *reusability* by providing rich metadata in dedicated sidecar files and *interoperability* by using existing standard data formats. *Accessibility* is not directly addressed by BIDS, but by repositories that build on BIDS, such as OpenNeuro (<https://openneuro.org>). By stipulating how to structure data using naming conventions and dedicated metadata files to store dictionaries (.json) and data (.tsv), BIDS fosters interoperability

<sup>1</sup>Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, Scotland. <sup>2</sup>Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany. <sup>3</sup>Department of Psychology, Stanford University, Stanford, California, USA. <sup>4</sup>Wellcome Centre for Human Neuroimaging, London, United Kingdom. <sup>5</sup>GIGA Institute, University of Liège, Liège, Belgium. <sup>6</sup>Swart Center for Computational Neuroscience, University of California San Diego, San Diego, California, USA. <sup>7</sup>CerCo, CNRS/Université Paul Sabatier, Toulouse, France. <sup>8</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands. <sup>9</sup>NatMEG, Karolinska Institutet, Stockholm, Sweden. These authors contributed equally: Cyril R. Pernet and Stefan Appelhoff. Correspondence and requests for materials should be addressed to C.R.P. (email: [cyril.pernet@ed.ac.uk](mailto:cyril.pernet@ed.ac.uk)) or S.A. (email: [appelhoff@mpib-berlin.mpg.de](mailto:appelhoff@mpib-berlin.mpg.de)) or R.O. (email: [r.oostenveld@donders.ru.nl](mailto:r.oostenveld@donders.ru.nl))

and reuse of acquired datasets. Because BIDS data are structured, BIDS also addresses issues of reproducibility by allowing the creation of fully automated data analysis workflows.

Here we report on the extension of BIDS to EEG data. EEG-BIDS builds upon the MEG-BIDS extension<sup>4</sup> and is presented concurrently with the iEEG-BIDS extension covering human intracranial electrophysiology<sup>5</sup>. In this document we only highlight the main features of EEG-BIDS. The full documentation of the EEG-BIDS extension can be found via the link to the specification on the general BIDS website (<https://bids.neuroimaging.io/>).

### EEG-BIDS Summary

The extension of BIDS to EEG data closely follows the general BIDS specification (see Figure 1). Each subject's data corresponds to a directory of raw data containing subdirectories for each session and data modality. This is accompanied by a “dataset\_description.json” file containing generic information about the dataset and in the case of the EEG modality, a metadata file with the suffix “eeg.json”. The “eeg.json” file exhaustively specifies among other metadata details of the experimental task and the EEG recording system. Optional directories include the “sourcedata” directory, which can be used to supply original non-formatted data. Furthermore, a “stimuli” directory and a “code” directory can be present to allow data conversion and preprocessing to be reproduced, as indicated in the original specification<sup>2</sup>. Within each subject directory, the “eeg” subdirectory contains the EEG and metadata. For instance, for a single session study, “sub-XX” would have subdirectory “eeg” which contains EEG files using the naming pattern “sub-XX\_task-YY\_eeg.<extension>” corresponding to acquisitions of EEG data. In addition, “sub-XX\_task-YY\_channels.tsv” must be specified describing the parameters of the data acquisition and two extra files, “sub-XX\_task-YY\_electrodes.tsv” and “sub-XX\_task-YY\_coordsystem.json”, should be specified if the positions of the electrodes are known (see below).

As in the MRI specification of BIDS, “sub-XX\_task-YY\_events.tsv” files should be present, encoding all of the parameters of the experimental design (onset of events, trial type, duration, responses, etc.). While such information is often present as one or several binary “trigger” channels in the EEG recordings, the representation of events is rarely explicit in the original data (e.g., a numeric code is used to indicate the onset of a given picture presented in a given experimental condition) hence the necessity of these “events.tsv” files. Since the initial specification of BIDS for MRI in 2016, the Hierarchical Event Descriptor (HED) system for precise annotation of events has been integrated, which is particularly useful for electrophysiological data in which dynamics of brain activity are associated with multiple experimental events<sup>6</sup>.

### Specific EEG-BIDS Considerations

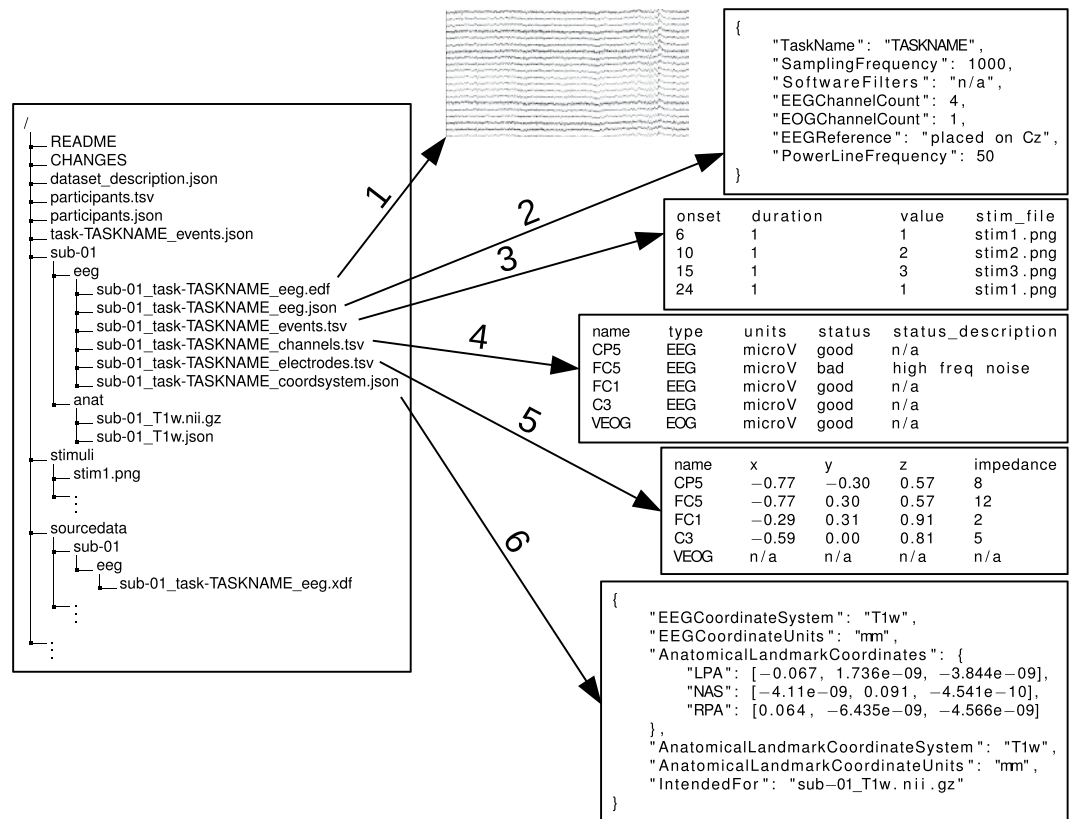
The process of converging on a list of suitable data formats for EEG-BIDS was governed by three major requirements: A suitable data format should (i) address the needs of a large portion of the global EEG community, (ii) be interoperable according to the FAIR principles<sup>3</sup>, and (iii) meet the technical requirements of neuroscientific workflows, such as saving numerical data with high precision.

As a solution to this challenge, the EEG-BIDS specification incorporates only two recommended “official” data formats: The European Data Format (EDF), which is an ongoing international effort to provide a common data format for electrophysiological recordings that began in 1992<sup>7</sup>, and the BrainVision Core Data Format, developed by Brain Products GmbH. While the BrainVision Core Data Format was designed by Brain Products GmbH for its proprietary EEG recording equipment and analysis software, it is based on the Microsoft Windows INI file and has a concise documentation. Both of these formats follow the three requirements for suitable data formats for EEG-BIDS: (i) A recent survey indicates that they are widely used in the community (<https://bids.berkeley.edu/news/bids-megeegieeg-data-format-survey/>), (ii) they have open access documentation and an open source implementation for both reading and writing in at least two programming languages that are widely used in the field (in this case, Python and MATLAB, among others), and (iii) they have high numerical precision (EDF:16 bits, BrainVision Core Data Format:32 bits). To accommodate a larger scientific audience and facilitate adoption, the EEG-BIDS standard also allows two “unofficial” commonly used data formats: The format used by the MATLAB toolbox EEGLAB<sup>8</sup> (“.set” and “.fdt” files), and the Biosemi format (“.bdf”). While not actively encouraged, these two formats are included due to their popularity and their interoperability among the major software packages. Future versions of BIDS may extend the list of “officially” supported data formats, based on the fulfillment of the above mentioned three requirements for suitable data formats. Independently of the raw data format used, critical metadata about the recording are always available in BIDS .tsv and .json files.

In order to provide an unambiguous documentation of EEG data, we are clarifying two sets of terms that are often used interchangeably: Electrodes versus channels, and fiducials versus anatomical landmarks.

We distinguish between electrodes and channels using the following definitions: (i) An EEG electrode is a contact point attached to the skin, (ii) a channel is the combination of the analog differential amplifier and analog-to-digital converter that result in a potential (voltage) difference being stored in the EEG dataset. The “reference” and “ground” electrodes should in general not be referred to as channels and only as electrodes. Some systems (e.g., Biosemi) have an active floating reference, whilst for most of the other systems, the potential at electrodes is neither amplified nor recorded. For EEG-BIDS, researchers must specify a “channels.tsv” file and may in addition specify an “electrodes.tsv” file with an accompanying “coordsystem.json” file.

Furthermore, we distinguish between fiducials and anatomical landmarks using the following definitions: (i) Fiducials are objects with a well defined location used to facilitate the localization of electrodes and co-registration with other geometric data such as the participant's own T1 weighted magnetic resonance head image, a T1 weighted template head image, or a spherical head model. Commonly used fiducials are vitamin E pills, which are clearly visible in an anatomical MR image, or reflective spheres that are localized with an infrared optical tracking system. (ii) Anatomical landmarks on the other hand define locations on a research subject such



**Fig. 1** Exemplary EEG-BIDS dataset with previews of EEG files. The left side of the figure shows a standard BIDS directory tree with the root containing files describing the dataset in general (“README”, “dataset\_description.json”), a file describing the participants (“participants.tsv”), and as several JSON files (“participants.json”, “task-TASKNAME\_events.json”), which contain the description necessary to understand the contents of the TSV files. Note that JSON files at high levels get inherited by lower levels unless overridden (the Inheritance Principle). Next to the files at the root, there is a stimuli and a sourcedata directory that can be used to save the respective study data. Most important are the subject directories named “sub- <sub-label>” for each study participant. Nested in the subject directories are all recorded data split over modalities (eeg and anat, for the EEG and structural MRI data respectively). The right side of the figure shows the contents of the eeg modality directory, including the raw EEG data (1) and associated metadata (2). An events.tsv file (3) specifies all events that were recorded during the session and can reference presented stimuli with the stimuli directory of the dataset (see stim\_file column). A “channels.tsv” file (4) provides further information about the raw EEG data and can contain information not present in the raw EEG data file such as filter settings and channel status (good/bad). Finally, an “electrodes.tsv” file (5) and an accompanying “coordsystem.json” file (6) provide electrode locations and specify which coordinate framework to use to interpret the electrode locations (for example with respect to a T1 weighted MRI scan).

as the nasion, which is the intersection of the frontal bone and two nasal bones of the human skull. Fiducials are typically used in conjunction with anatomical landmarks.

### Public EEG-BIDS Datasets

Several study examples (with zero-byte data files) are available in the BIDS-examples GitHub repository (<https://github.com/bids-standard/bids-examples>). We have also released three full datasets formatted using the EEG-BIDS standard:

- The Matching Pennies dataset<sup>9</sup> is an example of a single recording session per participant. It was collected as part of a student project to replicate a brain-computer interface study of motor intention decoding.
- The Rishikesh dataset<sup>10</sup> is an example of multiple recording sessions per participant. Participants were asked to meditate continuously whilst experience-sampling probe questions were presented at random intervals throughout the duration of the experiment.
- The simultaneous resting-state EEG-fMRI dataset<sup>11</sup> offers resting-state data in EEG and fMRI modalities, and structural T1 weighted data and diffusion data (NODDI sequence).

## Community Tools and Software Support

As part of the BIDS project, datasets formatted to follow the EEG-BIDS standard can be validated using the “bids-validator”, a JavaScript application that runs locally as a command line version (using Node.js) or within an Internet browser (<https://bids-standard.github.io/bids-validator/>). With this validation tool, researchers can check their newly formatted datasets and make full use of the data structure’s strengths for instance, checking for missing data or underspecified metadata.

The BIDS starter kit (<https://github.com/bids-standard/bids-starter-kit>) is a collection of community-driven guides, tutorials, helper scripts, and wiki resources to help researchers get started with BIDS. The resources cover two popular programming languages (Python and MATLAB) and will be extended over time to incorporate additional guides.

We are collaborating with the developers of the most widely used EEG data analysis tools in order to help EEG practitioners convert their existing data to the EDF or BrainVision Core Data Format. Data conversion utilities from many raw EEG data formats to the EDF and BrainVision format are available in MATLAB from the FieldTrip<sup>12</sup> and EEGLAB<sup>8</sup> toolboxes, and in Python from the MNE-Python<sup>13</sup> package. While specific software implementations are beyond the scope of the data specification, we here also want the reader to be aware of further work being undertaken by the open source community to interact with BIDS datasets. For instance, an EEGLAB<sup>8</sup> “study” can now be exported as BIDS (std\_tobids.m), FieldTrip<sup>12</sup> can similarly export (data2bids.m) while SPM12<sup>14</sup> (spm\_bids.m) and MNE-Python<sup>13</sup> (in form of the MNE-BIDS project <https://github.com/mne-tools/mne-bids>) can read any BIDS dataset. Ultimately, reading/writing BIDS dataset will be fully automatic, as it is the case for MRI.

## Data Analysis Pipelines and Beyond Sharing Raw Data

EEG’s long history, versatility and variety of applications makes it a data- and method-rich technique. Recently, the OHBM Committee on Best Practice in Data Analysis and Sharing (COBIDAS) released a guideline for good practice and reproducibility in EEG<sup>15</sup>. According to the guideline, even the simplest processing pipeline already contains eight separate steps, making it clear that while EEG-BIDS helps with data sharing, it will remain non-trivial to develop automated processing pipelines of neurophysiology data such as already available for MRI data<sup>16</sup>. On its own, EEG-BIDS is the first necessary step toward achieving validated and reproducible data analysis through standardizing the complete documentation of a dataset.

This article describes the new EEG extension for BIDS, and has limited itself to sharing raw data using previously developed community standards. Challenges that are specific to EEG, such as support for data formats, are still under active debate, and some additional formats will likely be incorporated once the technical issues are addressed and FAIR standards are achieved. The development of BIDS for EEG derivatives is also already underway (BIDS Extension Proposal 21): This will also make sharing of processed data possible, thereby fostering re-analyses, meta-analyses, and new analyses without the burden of data preparation.

## References

- Berger, H. Über das Elektrenkephalogramm des Menschen. *Arch. Für Psychiatr. Nervenkrankh.* **87**, 527–570 (1929).
- Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* **3**, 160044 (2016).
- Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
- Niso, G. *et al.* MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. *Sci. Data* **5**, 180110 (2018).
- Holdgraf, C. *et al.* BIDS-iEEG: an extension to the brain imaging data structure (BIDS) specification for human intracranial electrophysiology. *Sci. Data* **6**, <https://doi.org/10.1038/s41597-019-0105-7> (2019).
- Bigdely-Shamlo, N. *et al.* Hierarchical Event Descriptors (HED): Semi-Structured Tagging for Real-World Events in Large-Scale EEG. *Front. Neuroinform.* **10**(42) (2016).
- Kemp, B., Värri, A., Rosa, A. C., Nielsen, K. D. & Gade, J. A simple format for exchange of digitized polygraphic recordings. *Electroencephalogr. Clin. Neurophysiol.* **82**, 391–393 (1992).
- Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
- Appelhoff, S., Sauer, D. & Gill, S. S. Matching Pennies: A Brain Computer Interface Implementation Dataset. *Open Science Framework*, <https://doi.org/10.17605/OSF.IO/CJ2DR> (2018).
- Delorme, A. BIDS formatted EEG meditation experiment data. *Zenodo*, <https://doi.org/10.5281/zenodo.2536267> (2018).
- Clayden, J. & Deligianni, F. EEG, fMRI and NODDI dataset. *Open Science Framework*, <https://doi.org/10.17605/OSF.IO/DVMRB> (2016).
- Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
- Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**(267), (2013).
- Litvak, V. *et al.* EEG and MEG Data Analysis in SPM8. *Comput. Intell. Neurosci.* **2011**, 852961 (2011).
- Pernet, C. R. *et al.* Best Practices in Data Analysis and Sharing in Neuroimaging using MEEG. Preprint at <https://doi.org/10.31219/osf.io/a8dhx> (2018).
- Gorgolewski, K. J. *et al.* BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLOS Comput. Biol.* **13**, e1005209 (2017).

## Acknowledgements

We thank the members of the BIDS community for their insights, stimulating discussions and helpful comments on early drafts of the EEG extension. We are grateful to Dora Hermes and Chris Holdgraf for sharing their article drafts of the iEEG specification and to Scott Makeig for discussions on data formats and metadata exchange. We would like to thank Deborah Ain for her very helpful comments and corrections. This work was supported by the EU-H2020 Marie Curie “ChildBrain” Innovative Training Network grant no. 641652 to Robert Oostenveld.

### Author Contributions

C.R.P. Conception and design of the specification, moderating community interactions during the process, preparation of datasets and examples, coding of bids EEGLAB tools, making the figure, writing the manuscript. S.A. Critical review of the specification, moderating community interactions during the process, preparation of datasets and examples, coding of the bids-validator extension, coding of BIDS MNE tools, making the figure, writing the manuscript, preparing the manuscript for resubmission. K.G. Critical review of the specification, coding of the bids-validator, merging of BIDS specifications, critical review and final approval of the version submitted. G.F. Critical review of the specification, coding of BIDS SPM and MATLAB tools, critical review and final approval of the version submitted. C.P. Critical review of the specification, preparation of examples, critical review and final approval of the version submitted. A.D. Critical review of the specification, preparation of datasets and examples, coding of BIDS EEGLAB tools, critical review and final approval of the version submitted. R.O. Conception and design of the specification, moderating community interactions during the process, preparation of datasets and examples, coding of the bids-validator extension, coding of BIDS FieldTrip tools, critical review and final approval of the version submitted.

### Additional Information

**Competing Interests:** K.G. is a part-time Kaggle Research Consultant for Google LLC.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019