



HAL
open science

Transfer, Collection and Organisation of Electrophysiological and Imaging Data for Multicentre Studies

Nicolas Roehri, Samuel Medina Villalon, Aude Jegou, Bruno Colombet,
Bernard Giusiano, Aurélie Ponz, Fabrice Bartolomei, Christian-George Bénar

► **To cite this version:**

Nicolas Roehri, Samuel Medina Villalon, Aude Jegou, Bruno Colombet, Bernard Giusiano, et al..
Transfer, Collection and Organisation of Electrophysiological and Imaging Data for Multicentre Stud-
ies. Neuroinformatics, 2021, 10.1007/s12021-020-09503-6 . hal-03139022

HAL Id: hal-03139022

<https://hal.science/hal-03139022>

Submitted on 11 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Transfer, collection and organisation of electrophysiological and imaging data for multicentre studies

Nicolas Roehri¹ · Samuel Medina Villalon¹ · Aude Jegou¹ · Bruno Colombet¹ · Bernard Giusiano^{1,2} · Aurélie Ponz¹ · Fabrice Bartolomei^{1,3} · Christian-George Bénar^{1*}

Accepted: 1 December 2020

Abstract Multicentre studies are of utmost importance to confirm hypotheses. The lack of established standards and the ensuing complexity of their data management often hamper their implementation. The Brain Imaging Data Structure (BIDS) is an initiative for organizing and describing neuroimaging and electrophysiological data. Building on BIDS, we have developed two software programs: *BIDS Manager* and *BIDS Uploader*. The former has been designed to collect, organise and manage the data and the latter has been conceived to handle their transfer and anonymisation from the partner centres. These two programs aim at facilitating the implementation of multicentre study by providing a standardised framework.

Keywords Multicenter · Data Management · BIDS · Electrophysiology · Neuroimaging

Introduction

Multicentre research is carried out by two or more research institutions which share resources such as techniques, methods or data to answer a common question. Such studies are crucial to validate hypotheses, models or theories and provide many benefits over single-centre studies such as increased sample size and better generalisability of the results (Cheng et al., 2017; Dashevsky et al., 2018). They are however accompanied by several challenges. The major challenge is to coherently transfer and collect data in a robust and error minimizing way, which seamlessly blend in the workflow of the researchers conducting the

1. Aix Marseille Univ, Inserm, INS, Institut de Neurosciences des Systèmes, Marseille, France

2. APHM, Timone Hospital, Public Health Department, Marseille, France

3. APHM, Timone Hospital, Clinical Neurophysiology, Marseille, France

*corresponding author(s):

Christian-George Bénar E-mail: christian.benar@univ-amu.fr

study (Irving and Curley, 2008). The lack of standards is another barrier as it leads to a waste of time in developing specific scripts, programs and interfaces.

The Brain Imaging Data Structure (BIDS) (Gorgolewski et al., 2016) is an initiative aiming at establishing a common standard to describe data and its organisation on disk for both neuroimaging and electrophysiological data (Niso et al., 2018; Holdgraf et al., 2019; Pernet et al., 2019). Datasets following BIDS are composed of three folders: *raw*, *sourcedata* and *derivatives* which contain the data in an open file format, original format and processes form, respectively. Standardised metadata are stored in either dictionary (key/value) JSON files or tabulation-separated value (TSV) files. Each file is named according to a strict but explicit convention in a series of key-values (sub-01_ses-01_task-rest_run-01_ieeg.vhdr). Notably, the *dataset_description.json* and *participants.tsv* files at the root of the dataset contain information about the dataset (e.g. name, project, author) and details about the participants (e.g. identifier, age, sex, group), respectively. This standard does however not provide features for a data management practice, namely, a consistent identification and tracking system, and fixed naming variables with associated values (Schmitt and Burchinal, 2011).

An identifier (ID) which uniquely and consistently identifies a subject is critical to ensure data quality and follow the progression of data collections, especially when handling anonymised data. This is of utmost importance for clinical data as a great care is required to protect the patient identity. Such environment may be more prone to missing elements and variability (unit specific vocabulary and differences in electrode placing and naming), hence the importance of storing naming variables and allowed values in a *master file* (Schmitt and Burchinal, 2011). For instance, there is no standard naming of the intracranial EEG electrodes and their name is entered manually (sometimes several times for the same patient) which may lead to inconsistent naming across recording sessions and unnoticed error during analysis of the data.

In this work, we proposed a standardised framework for multicentre studies handling multimodal dataset. Building on the concept of the Brain Imaging Data Structure, we have developed, within the PHRC SPREAD and RHU EPINOV (two ongoing multicentre studies gathering eleven epilepsy centres in France), two software programs to tackle these issues. These programs aims to provide: an interface to anonymise and securely send the data, a unique and consistent ID for the subjects, a BIDS converter and an interface for BIDS-related data management and issue solving. *BIDS Manager* has been designed to collect, organise and manage the data in the investigating centre and *BIDS Uploader* has been conceived to handle the transfer of data from the partner centres.

Methodology

Context

Our team was in charge of developing a solution to transfer clinical electrophysiological and imaging data from a dozen French epilepsy units for two multicentre clinical studies (one being a clinical trial). Both studies had similar constraints, namely the transfer and storage of pseudo-anonymised data which must be readily accessible for analysis pipeline to generate a report returned back to the epilepsy unit for surgical decision-making. The emphasis was put on the ability to track the patient's data while preserving his or her identity in a reversible manner so that the clinicians of the corresponding epilepsy unit could link the report to the patient.

The type of data handled within these studies are anatomical magnetic resonance imaging (T1-, T2-weighted, FLAIR), computerized tomography scans, diffusion weighted imaging and intracranial recording – stereoelectroencephalography and electrocorticography – (which are all covered by BIDS standards (Gorgolewski et al., 2016; Niso et al., 2018; Holdgraf et al., 2019; Pernet et al., 2019)).

Rather than developing tools specific to these projects, we conceived flexible programs relying on a configurable JavaScript Object Notation File (JSON) file (see Systematic nomenclature). Also, we focused on graphical interfaces in order for the software to blend in the research routine.

Implementation

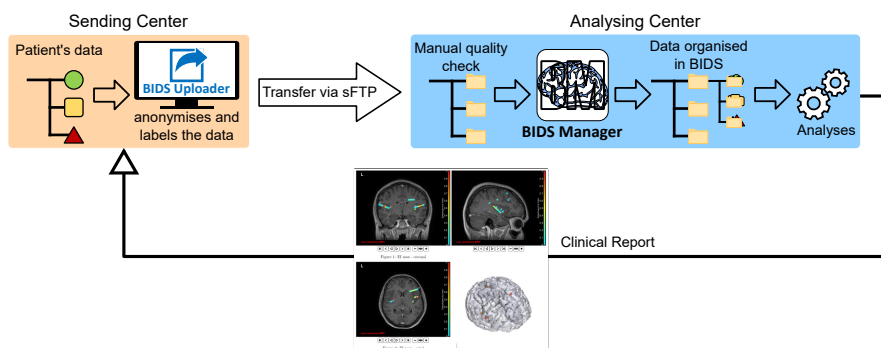


Fig. 1 Illustration of the implementation of *BIDS Uploader* and *BIDS Manager* in one partner center and the analysing center.

BIDS Uploader and *BIDS Manager* are two python-based programs available for Windows, Linux, and Mac OS X platforms.

BIDS Uploader is installed in each partner centre, which allows the user to

select, tag and send files of interest via a secured File Transfer Protocol (FTP) to the server of the analysing centre. Its interface has been designed in a sequential fashion so that each of its sections can only be accessed once the previous section has been completed, in order to reduce the risks of missing values.

BIDS Manager permits to explore and dynamically organise the data according to the BIDS standard whilst tracking the subject data readiness and integrity. *BIDS Manager* depends on two pieces of software to convert the electrophysiological and imaging data: AnyWave¹ (Colombet et al., 2015) and dicm2nii² (by Xiangrui Li) respectively.

Figure 1 illustrates a possible implementation of the programs within different centers.

Results

Illustrative Example

Let us use a toy example to describe our tools. The main objective of this example project is to study seizure recordings and the ancillary one is to compare a biomarker of epilepsie between resting and sleep. The researchers thus want to collect at least one seizure recorded with intracranial EEG (iEEG) as well as an hour of resting state and slow wave sleep data along with structural T1 weighted imaging before and after electrode implantation.

Some characteristics about the subjects also need to be stored, such as age, sex and handedness (optional) as well as the centre in which their data were acquired.

Data acquired before the implantation will be stored in the session labeled 'preop', and for those acquired after in the session labeled 'postimp'. The BIDS required field 'task' is used to describe the type of iEEG recording with the following values: 'seizure', 'sleep' and 'rest'. To store the subject information, we add columns in the *participants.tsv* file. The BIDS folder corresponding to this dataset is shown in Fig. 2.

Systematic nomenclature

To ensure a systematic nomenclature across centres, we designed a key/values file named *requirements.json* (our *master file*), based on the strict naming conventions proposed by BIDS, which characterises the dataset of the given study. This file is stored in the 'code' folder at the root of the dataset directory and describes the subject and modality types. Under "Subject", "keys" lists the subject attributes (e.g. the age, sex and handedness) followed by their allowed values ("F" or "M" for the sex) and "required_keys" states whether an

¹ available here: <http://meg.univ-amu.fr/wiki/AnyWave>

² <https://mathworks.com/matlabcentral/fileexchange/42997-xiangruili-dicm2nii>

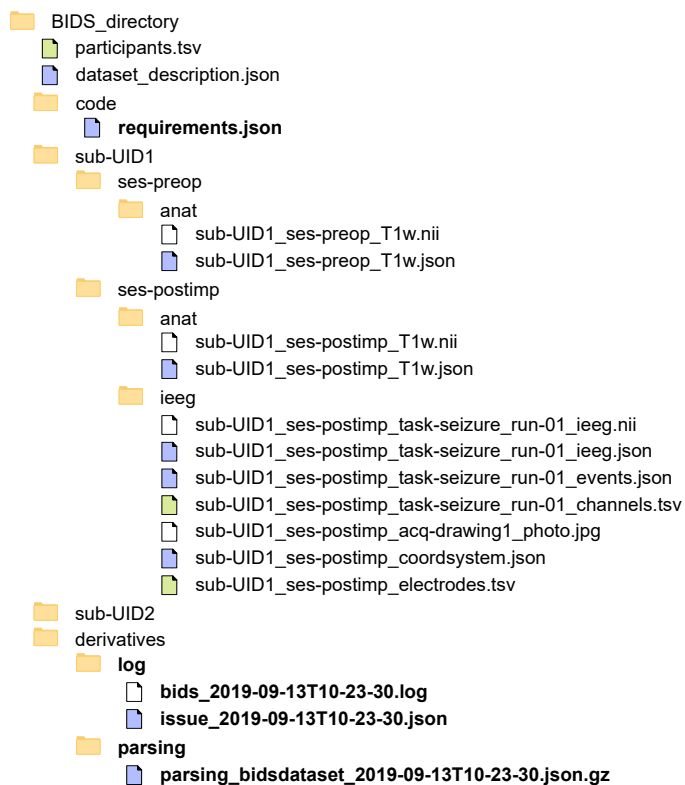


Fig. 2 Illustration of the BIDS structured dataset corresponding to the example. This dataset contains the data and sidecar files (.json and .tsv files containing metadata) required by the BIDS standard as well as some files – written in bold – specific to our proposed framework.

attribute is required. The dictionaries following the modality keys (e.g. "Anat" or "Ieeg") list the required files (e.g. the iEEG seizure recording). The following dictionaries describe the different modality types. Particularly, the attributes must be keys which are allowed by BIDS. There are no "required_keys" for the modality types as these are defined in the BIDS specifications³.

Below is the *requirements.json* file corresponding to our example. Notably, only the seizure files are required for the iEEG as the others are in the ancillary study.

³ <https://bids-specification.readthedocs.io/en/stable/>

```

"Subject": {
  "keys": {"age": 0, "sex": ["F","M"], "handedness": ["R","L"],
    "centre": ["Mars", "Lyon", "Stras", "Paris"]} ,
  "required_keys": ["age", "sex", "centre"],
  "Anat": [
    {"type":
      {"ses": "preop", "modality": "T1w"}, "amount": 1},
    {"type":
      {"ses": "postimp", "modality": "T1w"}, "amount": 1}
  ],
  "Ieeg": [
    {"type":
      {"ses": "postimp", "task": "seizure", "run": "_",
        "modality": "ieeg"}, "amount": 1}]
  },
  "Anat": {
    "keys": {"ses": ["preop","postimp"]}
  },
  "Ieeg": {
    "keys": {
      "ses": ["postimp"], "task": ["seizure","rest","sleep"]
    }
  }
}

```

BIDS Uploader

Prior to selecting the files to be transmitted using *BIDS Uploader* (Fig. 3), the user must first fill in the information concerning the subject: surname, first name, sex and date of birth plus all the other information written in the *requirements.json* specific to the project (Fig. 3a). *BIDS Uploader* only proposes to the user to select values from within the corresponding list in the *requirements.json* (e.g. for handedness, it proposes "R" or "L" and not "right" or "left", see Systematic nomenclature).

Once these elements have been entered, the program assigns a unique and specific twelve-alphanumeric character code to the subject (the *UID* in Fig. 2) by encrypting (using the SHA-256 algorithm) the surname, first name, date of birth and a protocol specific keyword (kept secret). Every time the user types in the information of the same subject, let us say John Doe born on 01/09/1983 with the protocol specific keyword 'test', the unique and specific identifier will always be 3a8573a45f91.

The program then invites the user to select the files to be imported (Fig. 3b) according to the modalities described in the *requirements.json* (i.e. in this example anatomical data ("Anat") and intracranial recordings ("Ieeg")) and tag them. Similarly, the software only proposes allowed tags from the previously chosen key list.

Prior to pseudo-anonymisation, *BIDS Uploader* verifies whether the identity of the subject input by the user matches the information in the selected files. If mismatches are found, the program asks the user to either modify the subject

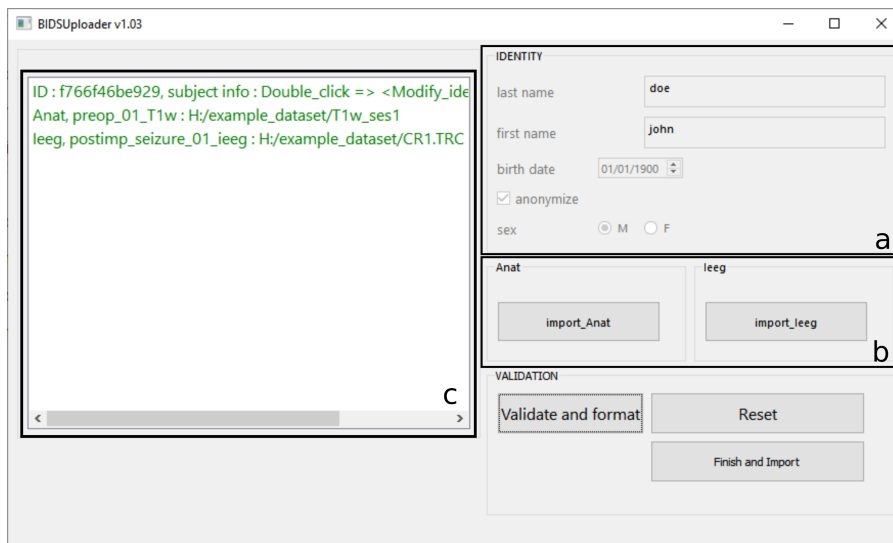


Fig. 3 Interface of *BIDS Uploader*. Information concerning the subject’s identity is first entered in frame **a**. Once filled, frame **b** is accessible and the user can select the type of data to be imported and tag them. The type of data and the values of the tags are proposed according to the *requirements.json* file. Frame **c** lists the selected files and corresponding tags. After clicking on the ‘*Validate and format*’ button, *BIDS Uploader* changes the color of each line in red in presence of identity mismatch in the file or in green in absence of mismatch.

information, remove the file or force the validation in case the data were already anonymised (Fig. 3c). The definition of pseudo-anonymisation used here is the one of the General Data Protection Regulation (GDPR, a regulation in EU law on data protection).⁴

Once all files have passed the validation stage, the software copies in a temporary folder the data, anonymises them, writes a metadata file (*data2import.json*), containing the different tags, anonymised subject information and file names, and sends them to the server of the analysing centre via a secure FTP (Secure Shell, SSH). During the anonymisation procedure, the software removes all sensitive fields from the header of the imaging and electrophysiological data (e.g. first name, family name, date of birth, address of the patient). When the transfer is finished, the interface is reset and the temporary folder is deleted. An example of the *data2import.json* file is given below. It illustrates a case of a transfer of data, namely a preoperative *T1w* image and an iEEG recording containing a *seizure* acquired in *Paris*, of a *right-handed 43-year-old male* patient identified as *f766f46be929* and intended for the protocol *YourProtocol*. Notably, the first name and surname of the patient is not present in this file

⁴ GDPR defines in Article 3 pseudo-anonymisation as “the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information.” Such “additional information” must be “kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person.”

only the SHA-identifier ensuring anonymity.
Importantly, *BIDS Uploader* complies with GDPR criterion

```

"Subject": [
  {
    "sub": "f766f46be929",
    "Anat": [
      {
        "sub": "f766f46be929",
        "ses": "preop",
        "acq": "",
        "ce": "",
        "rec": "",
        "run": "01",
        "mod": "",
        "modality": "T1w",
        "fileLoc": "preop_01_T1w",
        "AnatJSON": {}
      }
    ],
    "Ieeg": [
      {
        "sub": "f766f46be929",
        "ses": "postimp",
        "task": "seizure",
        "acq": "",
        "run": "01",
        "proc": "",
        "modality": "ieeg",
        "fileLoc": "CR1.TRC",
        "IeegJSON": {},
        "IeegChannelsTSV": [],
        "IeegEventsTSV": []
      }
    ],
    "age": "43",
    "sex": "M",
    "handedness": "R",
    "center": "Paris"
  }
],
"Derivatives": [],
"DatasetDescJSON": {
  "Name": "YourProtocol",
  "BIDSVersion": "1.0.1",
}

```

BIDS Manager

Basic functionalities

BIDS Manager allows the user to prepare a folder for a new BIDS dataset using a user-friendly interface. Via this interface, the user can create the *dataset_description.json* file which describes the dataset (required by BIDS

standard) and the *requirements.json* (described in Systematic nomenclature). The user can explore the dataset and open files without going through all the nested folders. Moreover *BIDS Manager* stores the state and metadata of the dataset in a key/value file with the corresponding date (parsing_dataset_yyyy-mm-ddTHH-MM-DD.json.gz in Fig. 2), in order to monitor any modification. Any action is stored in a log file in a specific folder in the BIDS dataset.

Importing data

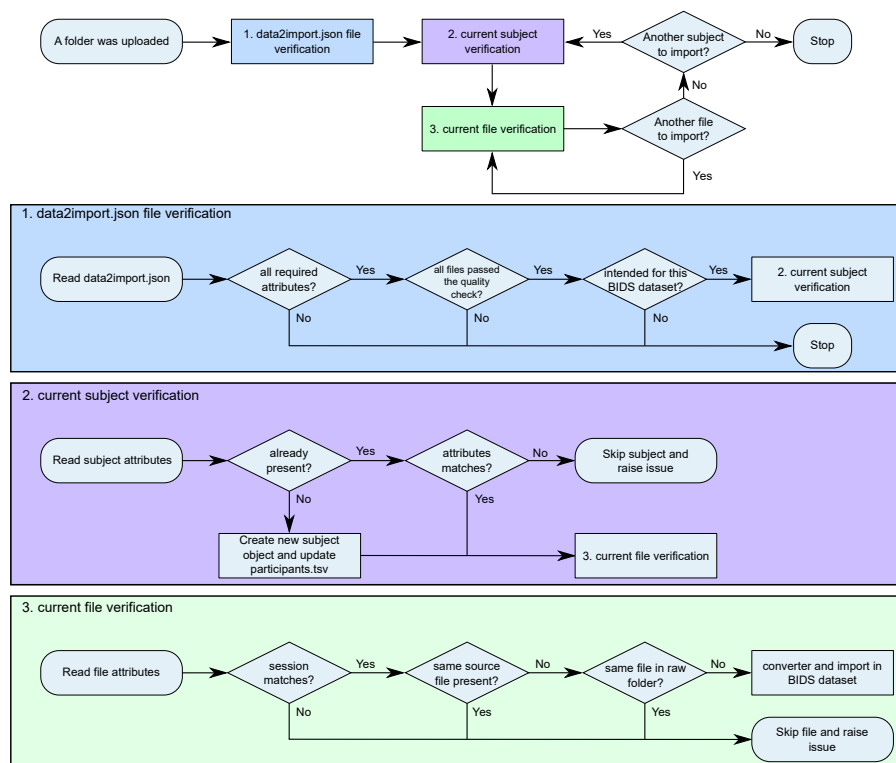


Fig. 4 Flow chart of the importation procedure. Once a new folder has been sent, it goes through three verification stages before being added to the dataset.

To import data which were transferred by *BIDS Uploader*, *BIDS Manager* first verifies three key points (Fig. 4):

1. *data2import.json file verification*: the metadata sent by *BIDS Uploader* must have all the BIDS required keys, all files to be imported must have been verified manually (e.g. for file corruption or missing slices in DICOM folder) and the folder must be intended for the protocol of the chosen BIDS directory otherwise the program raises an error.

2. *current subject verification*: If the subject is not already in the BIDS folder, the *participants.tsv* is updated with the subject information. If the subject is already present, the software checks that the characteristics of the subject to be imported match the characteristics of the subject already stored in the BIDS folder (e.g. sex, age and centre). If mismatches are found, an error is raised.
3. *current file verification*: The session should correspond to previous stored sessions, the source file name should be different from the source file names already present and there should not be a file with the same name in the raw folder. Otherwise an error is raised.

If no error was generated, *BIDS Manager* launches AnyWave and dicm2nii. AnyWave converts the electrophysiological files into BrainVision Core Data Format (one format accepted by both BIDS-iEEG (Holdgraf et al., 2019) and BIDS-EEG (Pernet et al., 2019)) and dicm2nii converts the imaging data into Nifti (Cox et al., 2004) format. Both programs create all required sidecar file containing the metadata which are also imported and renamed according to BIDS standard.

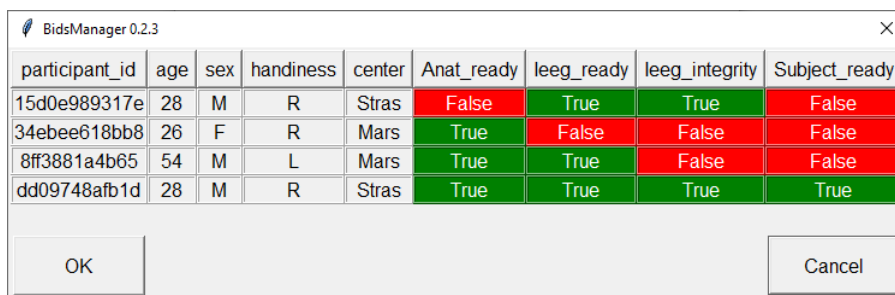
If errors were raised, these are stored in a JSON file named *issue.yyyy-mm-ddTHH-MM-DD.json* with the corresponding date in the log folder of the BIDS directory (Fig. 2). The user can solve these issues via the interface by either correcting the mismatches with the correct values proposed by *BIDS Manager* or removing the file from the list of files to be imported. If the user needs to contact the sending centre, the issues can be solved in another session as they are stored in the specific issue file.

Tracking system and electrode naming integrity

BIDS Manager adds new columns to the *participants.tsv* file in order to allow researchers to track the subject's data readiness and, in the specific case of iEEG datasets, guarantee the concordant naming of electrodes. Following our example, the software adds the 'Anat_ready', 'Ieeg_ready', 'Ieeg_integrity' and 'Subject_ready' columns (Fig. 5).

Whenever data are imported, *BIDS Manager* compares the list of files within the subject folder with the required files from the *requirements.json* (see Systematic nomenclature). When all data of a modality type have been imported for a given subject, *BIDS Manager* updates the column corresponding to the modality type in the *participants.tsv*. For instance, when the seizure recording of a given participant is imported, the value in column 'Ieeg_ready' turns from 'False' to 'True'.

To ensure the same electrode naming in the iEEG recordings of a given subject, we use the BIDS *electrodes.tsv* file as reference which contains the electrode coordinates, additional information (e.g. material and hemisphere), and the electrode names. To create this file, we have modified our in-house program



participant_id	age	sex	handedness	center	Anat_ready	leeg_ready	leeg_integrity	Subject_ready
15d0e989317e	28	M	R	Stras	False	True	True	False
34ebee618bb8	26	F	R	Mars	True	False	False	False
8ff3881a4b65	54	M	L	Mars	True	True	False	False
dd09748afb1d	28	M	R	Stras	True	True	True	True

OK Cancel

Fig. 5 Widget of *BIDS Manager* which displays the participants.tsv and the completeness of the data. For instance, the first three subject's data are not ready as anatomical, electro-physiological data are missing or some channels are mislabelled. The last participant's data are ready to be analysed.

GARDEL⁵ (Medina Villalon et al., 2018) which extracts the electrode coordinates from anatomical imaging data. If discrepancies between the electrode names in the iEEG files and the reference electrodes.tsv are found, *BIDS Manager* stores these mismatches in the issue file (described in Importing data) and proposes the user to rename the electrodes with an allowed names or change its type if the electrodes is not an iEEG electrodes. *BIDS Manager* turns the value in 'leeg_integrity' to 'True' when no mismatch remains. Finally, when all flags have turned 'True', the value in 'Subject_ready' is updated to 'True'.

Discussion

We propose a framework to facilitate the implementation of multicentre research handling human electrophysiological and neuroimaging data. Specifically, we provide a standardised way of transferring and organising data. This framework is based on the Brain Imaging Data Structure and two developed software packages: *BIDS Uploader* and *BIDS Manager* which work conjointly.

The decision of building it on BIDS (a standardised structured dataset) rather than on a database format is twofold. First, scientific data management platforms such as XNAT (Marcus et al., 2007) and COINS (Scott et al., 2011) have specifically been developed for neuroimaging data (and have now been extended to other modalities) but do not permit to store iEEG data with standardised metadata. On the other hand, BIDS has been widely adopted and extended to electrophysiological data (iEEG-, EEG- and MEG-BIDS (Holdgraf et al., 2019; Pernet et al., 2019; Niso et al., 2018)) offering standardised metadata and file naming for both electrophysiological and imaging data (the aforementioned platforms are planning to support BIDS as an export and/or import option). Secondly, a dedicated application programming interface should have been developed to get data from these databases. Most

⁵ available at: <http://meg.univ-amu.fr/wiki/GARDEL:presentation>

available tools (Oostenveld et al., 2011; Gramfort et al., 2013; Litvak et al., 2011; Delorme and Makeig, 2004; Tadel et al., 2011; Colombet et al., 2015; Gorgolewski et al., 2017) are however already BIDS compatible, making previous analysis pipeline readily applicable.

While BIDS already has a strict naming convention, we have appended a second layer which ensures a systematic nomenclature for a given project thanks to the *requirements.json* file. Once the collaborators have chosen the naming variables as well as their allowed values, the *requirements.json* file storing these information is used by the two programs to only propose values from these lists. This drastically decreases the risks of entry errors (Atkinson, 2012; Paulsen et al., 2012; Kupzyk and Cohen, 2015), use of vague terms, homonyms or synonyms (Schmitt and Burchinal, 2011) for both the files and the participant characteristics. Moreover, this file can be specifically adjusted for the needs of any study which makes these controls generic. In case of pseudo-anonymised data, the risk of subject re-identification increases with the number of subject characteristics stored in *participants.tsv* file. The researchers should therefore only store participant characteristics which are essential to the study and use generalised characteristics such as age rather than date of birth.

The creation by *BIDS Uploader* of a unique and specific twelve-alphanumeric character subject code permits to track the subject’s data while preserving its anonymity as his or her surname and first name are not send with the data. This also allows removing data from participants requesting that their data be no longer archived. The protocol specific keyword used to produce this code avoids identifying the same subject in independent studies. Importantly, each time the same subject information are input the same code is generated. This makes it easy for the user to send the data at different time points and is especially suitable for longitudinal studies. Encrypting these information based on manual entries is at risk as one simple typing error yields an entirely different code due to the avalanche effect (Feistel, 1973). To control that, *BIDS Uploader* matches the subject information in the selected files (as described in *BIDS Uploader*). This also ensure that the data to be sent correspond to the chosen subject and not another one as this is impossible to recover after anonymisation.

The tracking system implemented in *BIDS Manager* allows the researchers to follow the progress (or stagnation) of data collection and is a crux of data management practices (Schmitt and Burchinal, 2011). The consecutive stages of verification at import protect the data from being undesirably overwritten or wrongly labeled. As iEEG electrode names are manually typed at recording, they may differ between files. This type of silent error may have dramatic consequences in the analysis process. Indeed, computing the average rate of a given activity for each electrode across files with inconsistent electrode naming yields absurd and unusable results which could be carried on throughout the pipeline. *BIDS Manager*, by ensuring the cohesion of the electrode names across subject’s files, assuredly reduces these silent errors.

The proposed framework does not intend to replace the case report form (CRF) of clinical research (Latha et al., 2014) as their goal is different but may share common information (e.g. participant characteristics contained in both the CRF and the *participants.tsv*). Moreover, we did not address the open-access sharing of the finalised dataset since multicentre studies correspond more to a controlled-access model of sharing (Keerie et al., 2018). As the dataset is already structured in BIDS it requires a minimal amount of development to modify the data to reach the appropriate level of anonymisation. Furthermore, we did not detail how the data were stored. Each BIDS dataset with personal data should be stored on encrypted disks with limited access to the data (handled by the competent personnel of the institution).

The current programs handle all imaging and iEEG data. Further development will include EEG- and MEG-BIDS extensions, and the possibility to handle analysis pipelines.

Information Sharing Statement

BIDS Manager (RRID:SCR_019228) is available to download from https://github.com/Dynamap/BIDS_Manager or as a Windows 10 executable from <https://doi.org/10.6084/m9.figshare.11728872>. The software is free and distributed under the terms of the GNU General Public Licence as published by the Free Software Foundation. *BIDS Manager* is written for Python 3.7 or above, and needs two pieces of software (AnyWave available to download from <http://meg.univ-amu.fr/wiki/AnyWave> and a compiled version of dicm2nii from https://github.com/Dynamap/BIDS_Manager). To install our program, the user has to copy *BIDS Manager* folder on its computer disk and run the setup.py file for the python package. To run the software, the user can either double-click on the executable icon or launch it with python. We have also made available a example dataset for the user to test the software: <https://doi.org/10.6084/m9.figshare.11687064>. For further information on how to use the software, please see *BIDS Manager* online manual: https://github.com/Dynamap/BIDS_Manager or the tutorial video: <https://youtu.be/HvJjr6WZNQA>

Acknowledgements NR was supported by the French national clinical project SPREAD [PHRCN-16-0685 sponsored by University Hospitals of Strasbourg, IDRCB No. 2017-A00497-46]. This work was partially supported by the RHU EPINOV [A*MIDEX project (ANR-17-RHUS-0004) funded by the 'Investissements d'Avenir' French Government]. AJ has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under Specific Grant Agreement No. 785907 (Human Brain Project SGA2). We would like to thank Dr Olivier David who shared the code of the uploader used in the F-TRACT project (f-tract.eu) which served as the starting point of *BIDS Uploader* described above.

Author contributions

NR drafted the manuscript. NR, SMV, BC, BG, AP, FB and CGB designed the framework. NR developed *BIDS Manager* and SMV developed *BIDS Uploader*. AJ contributed in programming *BIDS Manager*. All the authors contributed to reviewing and editing the manuscript.

Competing interests

The authors declare no competing interests.

References

- Atkinson I (2012) Accuracy of data transfer: Double data entry and estimating levels of error. *Journal of Clinical Nursing* 21(19-20):2730–2735, DOI 10.1111/j.1365-2702.2012.04353.x
- Cheng A, Kessler D, Mackinnon R, Chang TP, Nadkarni VM, Hunt EA, Duval-Arnould J, Lin Y, Pusic M, Auerbach M (2017) Conducting multicenter research in healthcare simulation: Lessons learned from the INSPIRE network. *Advances in Simulation* 2(1):1–14, DOI 10.1186/s41077-017-0039-0
- Colombet B, Woodman M, Badier J, Bénar CG (2015) AnyWave: A cross-platform and modular software for visualizing and processing electrophysiological signals. *Journal of Neuroscience Methods* 242:118–126, DOI 10.1016/j.jneumeth.2015.01.017, URL <http://linkinghub.elsevier.com/retrieve/pii/S0165027015000187>
- Cox RW, Ashburner J, Breman H, Fissell K, Haselgrove C, Holmes CJ, Lancaster JL, Rex DE, Smith SM, Woodward JB, Strother SC (2004) A (Sort of) New Image Data Format Standard: {NIfTI-1}. *NeuroImage* 22(May 2014):1, URL <http://nifti.nimh.nih.gov/dfwg/NIFTI-1-Abstract.html>
- Dashevsky BZ, Bercu ZL, Bhosale PR, Burton KR, Chatterjee AR, Frigini LAR, Heacock L, Herskovits EH, Lee JT, Subhas N, Wasnik AP, Gyftopoulos S (2018) Multicenter Research Studies in Radiology. *Academic Radiology* 25(1):18–25, DOI 10.1016/j.acra.2017.05.019, URL <https://doi.org/10.1016/j.acra.2017.05.019>
- Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods* 13:9–21, DOI <http://dx.doi.org/10.1016/j.jneumeth.2003.10.009>, URL <http://www.sccn.ucsd.edu/eeglab/> <https://cloudfront.escholarship.org/dist/prd/content/qt52k1t4sz/qt52k1t4sz.pdf>, j.jneumeth.2003.10.009.
- Feistel H (1973) Cryptography and Computer Privacy. *Scientific American* 228(5):15–23, DOI 10.1038/scientificamerican0573-15
- Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, Flandin G, Ghosh SS, Glatard T, Halchenko YO, Handwerker DA, Hanke M, Keator

- D, Li X, Michael Z, Maumet C, Nichols BN, Nichols TE, Pellman J, Poline JB, Rokem A, Schaefer G, Sochat V, Triplett W, Turner JA, Varoquaux G, Poldrack RA (2016) The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data* 3:1–9, DOI 10.1038/sdata.2016.44
- Gorgolewski KJ, Alfaro-Almagro F, Auer T, Bellec P, Capotă M, Chakravarty MM, Churchill NW, Cohen AL, Craddock RC, Devenyi GA, Eklund A, Esteban O, Flandin G, Ghosh SS, Guntupalli JS, Jenkinson M, Keshavan A, Kiar G, Liem F, Raamana PR, Raffelt D, Steele CJ, Quirion PO, Smith RE, Strother SC, Varoquaux G, Wang Y, Yarkoni T, Poldrack RA (2017) BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Computational Biology* 13(3):1–16, DOI 10.1371/journal.pcbi.1005209
- Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, Hämäläinen M (2013) MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* 7(7 DEC):1–13, DOI 10.3389/fnins.2013.00267
- Holdgraf C, Appelhoff S, Bickel S, Bouchard K, D’Ambrosio S, David O, Devinsky O, Dichter B, Flinker A, Foster BL, Gorgolewski KJ, Groen I, Groppe D, Gunduz A, Hamilton L, Honey CJ, Jas M, Knight R, Lachaux JP, Lau JC, Lee-Messer C, Lundstrom BN, Miller KJ, Ojemann JG, Oostenveld R, Petridou N, Piantoni G, Pigorini A, Pouratian N, Ramsey NF, Stolk A, Swann NC, Tadel F, Voytek B, Wandell BA, 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(1):102, DOI 10.1038/s41597-019-0105-7, URL <http://www.nature.com/articles/s41597-019-0105-7>
- Irving SY, Curley MA (2008) Challenges to Conducting Multicenter Clinical Research. *AACN Advanced Critical Care* 19(2):164–169, DOI 10.1097/01.aacn.0000318119.67061.0f
- Keerie C, Tuck C, Milne G, Eldridge S, Wright N, Lewis SC (2018) Data sharing in clinical trials - practical guidance on anonymising trial datasets. *Trials* 19(1):1–8, DOI 10.1186/s13063-017-2382-9
- Kupzyk KA, Cohen MZ (2015) Data Validation and Other Strategies for Data Entry. *Western Journal of Nursing Research* 37(4):546–556, DOI 10.1177/0193945914532550
- Latha M, Bellary S, Krishnankutty B (2014) Basics of case report form designing in clinical research. *Perspectives in Clinical Research* 5(4):159, DOI 10.4103/2229-3485.140555
- Litvak V, Mattout J, Kiebel S, Phillips C, Henson R, Kilner J, Barnes G, Oostenveld R, Daunizeau J, Flandin G, Penny W, Friston K (2011) EEG and MEG data analysis in SPM8. *Computational Intelligence and Neuroscience* 2011, DOI 10.1155/2011/852961
- Marcus DS, Olsen TR, Ramaratnam M, Buckner RL (2007) The Extensible Neuroimaging Archive Toolkit An Informatics Platform for Managing, Exploring, and Sharing Neuroimaging Data. *Neuroinformatics* 5(4):11–

- 33, DOI 10.1385/NI:5:1:11, URL <http://www.springerlink.com/index/X74873723718574K.pdf>
- Medina Villalon S, Paz R, Roehri N, Lagarde S, Pizzo F, Colombet B, Bartolomei F, Carron R, Bénar CG (2018) EpiTools, A software suite for presurgical brain mapping in epilepsy: Intracerebral EEG. *Journal of Neuroscience Methods* 303:7–15, DOI 10.1016/j.jneumeth.2018.03.018, URL <https://doi.org/10.1016/j.jneumeth.2018.03.018>
- Niso G, Gorgolewski KJ, Bock E, Brooks TL, Flandin G, Gramfort A, Henson RN, Jas M, Litvak V, T Moreau J, Oostenveld R, Schoffelen JM, 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, URL <http://www.nature.com/articles/sdata2018110>
- Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience* 2011, DOI 10.1155/2011/156869
- Paulsen A, Overgaard S, Lauritsen JM (2012) Quality of data entry using single entry, double entry and automated forms processing-an example based on a study of patient-reported outcomes. *PLoS ONE* 7(4):1–6, DOI 10.1371/journal.pone.0035087
- Pernet CR, Appelhoff S, Gorgolewski KJ, Flandin G, Phillips C, Delorme A, Oostenveld R (2019) EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data* 6(1):103, DOI 10.1038/s41597-019-0104-8, URL <http://www.nature.com/articles/s41597-019-0104-8>
- Schmitt CP, Burchinal M (2011) Data management practices for collaborative research. *Frontiers in Psychiatry* 2(JUL):1–8, DOI 10.3389/fpsy.2011.00047
- Scott A, Courtney W, Wood D, de la Garza R, Lane S, King M, Wang R, Roberts J, Turner JA, Calhoun VD (2011) Coins: An innovative informatics and neuroimaging tool suite built for large heterogeneous datasets. *Frontiers in Neuroinformatics* 5(December):1–15, DOI 10.3389/fninf.2011.00033
- Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM (2011) Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience* 2011, DOI 10.1155/2011/879716