1 Changelog

1.1.1

- Improved the MEG landmark coordinates description.
- Replaced ManufacturersCapModelName in meg.json with CapManufacturer and CapManufacturersModelName.
- Remove EEGSamplingFrequency and ManufacturersAmplifierModelName from the meg.json.
- Improved the behavioural data description.

1.1.0

- Added support for MEG data (merged BEP008)
- Added SequenceName field.
- Added support for describing events with Hierarchical Event Descriptors [8.5 Task events].
- Added VolumeTiming and AcquisitionDuration fields [8.4 Task (including resting state) imaging data].
- Added DwellTime field.

1.0.2

- Added support for high resolution (anatomical) T2star images [8.3 Anatomy imaging data].
- Added support for multiple defacing masks [8.3 Anatomy imaging data].
- Added optional key and metadata field for contrast enhanced structural scans [8.3 Anatomy imaging data].
- Added DelayTime field [8.4 Task (including resting state) imaging data].
- Added support for multi echo BOLD data [8.4 Task (including resting state) imaging data].

1.0.1

- Added InstitutionName field [8.4 Task (including resting state) imaging data].
- Added InstitutionAddress field [8.4 Task (including resting state) imaging data].
- Added DeviceSerialNumber field [8.4 Task (including resting state) imaging data].
- Added NumberOfVolumesDiscardedByUser and NumberOfVolumesDiscardedByScanner field [8.4 Task (including resting state) imaging data].
- Added TotalReadoutTime to functional images metadata list [8.4 Task (including resting state) imaging data].

1.0.1-rc1

- Added T1 Rho maps [8.3 Anatomy imaging data].
- Added support for phenotypic information split into multiple files [8.11 Participant

key file].
- Added recommendations for multi site datasets
- Added SoftwareVersions
- Added run-<run_index> to the phase encoding maps. Improved the description.
- Added InversionTime metadata key.
- Clarification on the source vs raw language.
- Added trial_type column to the event files.
- Added missing sub-<participant_label> in behavioural data file names
- Added ability to store stimuli files.
- Clarified the language describing allowed subject labels.
- Added quantitative proton density maps.

1.0.0

- Added ability to specify fieldmaps acquired with multiple parameter sets.
- Added ability to have multiple runs of the same fieldmap.
- Added FLASH anatomical images.

1.0.0-rc4

- Replaced links to neurolex with explicit DICOM Tags.
- Added sourcedata.
- Added data dictionaries.
- Be more explicit about contents of JSON files for structural (anatomical) scans.

1.0.0-rc3

- Renamed PhaseEncodingDirection values from “x”, “y”, “z” to “i”, “j”, “k” to avoid confusion with FSL parameters
- Renamed SliceEncodingDirection values from “x”, “y”, “z” to “i”, “j”, “k”

1.0.0-rc2

- Removed the requirement that TSV files cannot include more than two consecutive spaces.
- Refactor of the definitions sections (copied from the manuscript)
- Make support for uncompressed .nii files more explicit.
- Added BIDSVersion to dataset.json
- Remove the statement that SliceEncodingDirection is necessary for slice time correction
- Change dicom converter recommendation from dcmstack to dcm2nii and dicm2nii following interactions with the community (see https://github.com/moloney/dcmstack/issues/39 and https://github.com/neurolabusc/dcm2niix/issues/4).
- Added section on behavioral experiments with no accompanying MRI acquisition
- Add _magnitude.nii[.gz] image for GE type fieldmaps.
- Replaced EchoTimeDifference with EchoTime1 and EchoTime2 (SPM toolbox requires this input).
- Added support for single band reference image for DWI.

- Added DatasetDOI field in the dataset description.
- Added description of more metadata fields relevant to DWI fieldmap correction.
- PhaseEncodingDirection is now expressed in “x”, “y” etc. instead of “PA” “RL” for DWI scans (so it’s the same as BOLD scans)
- Added rec-<label> flag to BOLD files to distinguish between different reconstruction algorithms (analogous to anatomical scans).
- Added recommendation to use _physio suffix for continuous recordings of motion parameters obtained by the scanner side reconstruction algorithms.

1.0.0-rc1

- Initial release

Introduction

Motivation

Neuroimaging experiments result in complicated data that can be arranged in many different ways. So far there is no consensus how to organize and share data obtained in neuroimaging experiments. Even two researchers working in the same lab can opt to arrange their data in a different way. Lack of consensus (or a standard) leads to misunderstandings and time wasted on rearranging data or rewriting scripts expecting certain structure. Here we describe a simple and easy-to-adopt way of organising neuroimaging and behavioural data. By using this standard you will benefit in the following ways:

- It will be easy for another researcher to work on your data. To understand the organisation of the files and their format you will only need to refer them to this document. This is especially important if you are running your own lab and anticipate more than one person working on the same data over time. By using BIDS you will save time trying to understand and reuse data acquired by a graduate student or postdoc that has already left the lab.
- There are a growing number of data analysis software packages that can understand data organised according to BIDS (see http://bids.neuroimaging.io for the most up to date list).
- Databases such as OpenNeuro.org accept datasets organised according to BIDS. If you ever plan to share your data publicly (nowadays some journals require this) you can minimize the additional time and energy spent on publication, and speed up the curation process by using BIDS to structure and describe your data right after acquisition.
- There are validation tools that can check your dataset integrity and let you easily spot missing values.

BIDS is heavily inspired by the format used internally by OpenfMRI.org and has been

supported by the International Neuroinformatics Coordinating Facility and the Neuroimaging Data Sharing Task Force. While working on BIDS we consulted many neuroscientists to make sure it covers most common experiments, but at the same time is intuitive and easy to adopt. The specification is intentionally based on simple file formats and folder structures to reflect current lab practices and make it accessible to a wide range of scientists coming from different backgrounds.

Definitions

The keywords "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

Throughout this protocol we use a list of terms. To avoid misunderstanding we clarify them here.

1. Dataset - a set of neuroimaging and behavioural data acquired for a purpose of a particular study. A dataset consists of data acquired from one or more subjects, possibly from multiple sessions.
2. Subject - a person or animal participating in the study.
3. Session - a logical grouping of neuroimaging and behavioural data consistent across subjects. Session can (but doesn't have to) be synonymous to a visit in a longitudinal study. In general, subjects will stay in the scanner during one session. However, for example, if a subject has to leave the scanner room and then be re-positioned on the scanner bed, the set of MRI acquisitions will still be considered as a session and match sessions acquired in other subjects. Similarly, in situations where different data types are obtained over several visits (for example fMRI on one day followed by DWI the day after) those can be grouped in one session. Defining multiple sessions is appropriate when several identical or similar data acquisitions are planned and performed on all -or most- subjects, often in the case of some intervention between sessions (e.g., training).
4. Data acquisition - a continuous uninterrupted block of time during which a brain scanning instrument was acquiring data according to particular scanning sequence/protocol.
5. Data type - a functional group of different types of data. In BIDS we define five data types: func (task based and resting state functional MRI), dwi (diffusion weighted imaging), fmap (field inhomogeneity mapping data such as field maps), anat (structural imaging such as T1, T2, etc.), meg (magnetoencephalography).
6. Task - a set of structured activities performed by the participant. Tasks are usually accompanied by stimuli and responses, and can greatly vary in complexity. For the purpose of this protocol we consider the so-called “resting state” a task. In the context of brain scanning, a task is always tied to one data acquisition. Therefore, even if during one acquisition the subject performed multiple conceptually different behaviours (with different sets of instructions) they will be considered one (combined) task.

7. Event - a stimulus or subject response recorded during a task. Each event has an onset time and duration. Note that not all tasks will have recorded events (e.g., resting state).
8. Run - an uninterrupted repetition of data acquisition that has the same acquisition parameters and task (however events can change from run to run due to different subject response or randomized nature of the stimuli). Run is a synonym of a data acquisition.

Compulsory, optional, and additional data and metadata

The following standard describes a way of arranging data and writing down metadata for a subset of neuroimaging experiments. Some aspects of the standard are compulsory. For example a particular file name format is required when storing structural scans. Some aspects are regulated but optional. For example a T2 volume does not need to be included, but when it is available it should be saved under a particular file name specified in the standard. This standard aspires to describe a majority of datasets, but acknowledges that there will be cases that do not fit. In such cases one can include additional files and subfolders to the existing folder structure following common sense. For example one may want to include eye tracking data in a vendor specific format that is not covered by this standard. The most sensible place to put it is next to the continuous recording file with the same naming scheme but different extensions. The solutions will change from case to case and publicly available datasets will be reviewed to include common data types in the future releases of the BIDS spec.

Source vs. raw vs. derived data

BIDS in its current form is designed to harmonize and describe raw (unprocessed or minimally processed due to file format conversion) data. During analysis such data will be transformed and partial as well as final results will be saved. Derivatives of the raw data (other than products of DICOM to NIfTI conversion) MUST be kept separate from the raw data. This way one can protect the raw data from accidental changes by file permissions. In addition it is easy to distinguish partial results from the raw data and share the latter. Similar rules apply to source data which is defined as data before harmonization and/or file format conversion (for example E-Prime event logs or DICOM files).

This specification currently does not go into details of recommending a particular naming scheme for including different types of source data (raw event logs, parameter files, etc. before conversion to BIDS) and data derivatives (correlation maps, brain masks, contrasts maps, etc.). However, in the case that these data are to be included:

1. These data MUST be kept in separate `sourcedata` and `derivatives` folders each

with a similar folder structure as presented below for the BIDS-managed data. For example: `derivatives/fmriprep/sub-01/ses-pre/sub-01_ses-pre_mask.nii.gz` or `sourcedata/sub-01/ses-pre/func/sub-01_ses-pre_task-rest_bold.dicom.tgz` or `sourcedata/sub-01/ses-pre/func/MyEvent.sce`.

2. A README file SHOULD be found at the root of the `sourcedata` or the `derivatives` folder (or both). This file should describe the nature of the raw data or the derived data. In the case of the existence of a `derivatives` folder, we RECOMMEND including details about the software stack and settings used to generate the results. Inclusion of non-imaging objects that improve reproducibility are encouraged (scripts, settings files, etc.).
3. We RECOMMEND including the PDF print-out with the actual sequence parameters generated by the scanner in the `sourcedata` folder.

The Inheritance Principle

Any metadata file (`.json`, `.bvec`, `.tsv`, etc.) may be defined at any directory level, but no more than one applicable file may be defined at a given level (Example 1). The values from the top level are inherited by all lower levels unless they are overridden by a file at the lower level. For example, `sub-*_task-rest_bold.json` may be specified at the participant level, setting TR to a specific value. If one of the runs has a different TR than the one specified in that file, another `sub-*_task-rest_bold.json` file can be placed within that specific series directory specifying the TR for that specific run. There is no notion of "unsetting" a key/value pair. For example if there is a JSON file corresponding to particular participant/run defining a key/value and there is a JSON file on the root level of the dataset that does not define this key/value it will not be "unset" for all subjects/runs. Files for a particular participant can exist only at participant level directory, i.e `/dataset/sub-*[/ses-*]/sub-*_T1w.json`. Similarly, any file that is not specific to a participant is to be declared only at top level of dataset for eg: `task-sist_bold.json` must be placed under `/dataset/task-sist_bold.json`

Example 1: Two JSON files at same level that are applicable for NIfTI file.

```
sub-01/
    ses-test/
        sub-test_task-overtverbgeneration_bold.json
        sub-test_task-overtverbgeneration_run-2_bold.json
        anat/
            sub-01_ses-test_T1w.nii.gz
        func/
            sub-01_ses-test_task-overtverbgeneration_run-1_bold.nii.gz
            sub-01_ses-test_task-overtverbgeneration_run-2_bold.nii.gz
```

In the above example, two JSON files are listed under `sub-01/ses-test/`, which are each applicable to `sub-01_ses-test_task-overtverbgeneration_run-2_bold.nii.gz`, violating the constraint that no more than one file may be defined at a given level of the directory

structure. Instead `task-overtverbgeneration_run-2_bold.json` should have been under `sub-01/ses-test/func/`.

Example 2: Multiple run and rec with same acquisition (acq) parameters acq-test1

```
sub-01/
    anat/
    func/
        sub-01_task-xyz_acq-test1_run-1_bold.nii.gz
        sub-01_task-xyz_acq-test1_run-2_bold.nii.gz
        sub-01_task-xyz_acq-test1_rec-recon1_bold.nii.gz
        sub-01_task-xyz_acq-test1_rec-recon2_bold.nii.gz
        sub-01_task-xyz_acq-test1_bold.json
```

For the above example, all NIfTI files are acquired with same scanning parameters (`acq-test1`). Hence a JSON file describing the acq parameters will apply to different runs and rec files. Also if the JSON file (`task-xyz_acq-test1_bold.json`) is defined at dataset top level directory, it will be applicable to all task runs with `test1` acquisition parameter.

Case 2: Multiple json files at different levels for same task and acquisition parameters

```
sub-01/
   sub-01_task-xyz_acq-test1_bold.json
         anat/
         func/
             sub-01_task-xyz_acq-test1_run-1_bold.nii.gz
             sub-01_task-xyz_acq-test1_rec-recon1_bold.nii.gz
             sub-01_task-xyz_acq-test1_rec-recon2_bold.nii.gz
```

In the above example, the fields from `task-xyz_acq-test1_bold.json` file will apply to all bold runs. However, if there is a key with different value in `sub-01/func/sub-01_task-xyz_acq-test1_run-1_bold.json`, the new value will be applicable for that particular run/task NIfTI file/s.

The Inheritance Principle

Any metadata file (`.json`, `.bvec`, `.tsv`, etc.) may be defined at any directory level, but no more than one applicable file may be defined at a given level (Example 1). The values from the top level are inherited by all lower levels unless they are overridden by a file at the lower level. For example, `sub-*_task-rest_bold.json` may be specified at the participant level, setting TR to a specific value. If one of the runs has a different TR than the one specified in that file, another `sub-*_task-rest_bold.json` file can be placed within that specific series directory specifying the TR for that specific run. There is no notion of "unsetting" a key/value pair. For example if there is a JSON file corresponding to particular participant/run defining a key/value and there is a JSON file on the root level of the dataset that does not define this key/value it will not be "unset" for all subjects/runs.

Files for a particular participant can exist only at participant level directory, i.e `/dataset/sub-*[/ses-*]/sub-*_T1w.json`. Similarly, any file that is not specific to a participant is to be declared only at top level of dataset for eg: `task-sist_bold.json` must be placed under `/dataset/task-sist_bold.json`

Example 1: Two JSON files at same level that are applicable for NIfTI file.

```
sub-01/
    ses-test/
        sub-test_task-overtverbgeneration_bold.json
        sub-test_task-overtverbgeneration_run-2_bold.json
        anat/
            sub-01_ses-test_T1w.nii.gz
        func/
            sub-01_ses-test_task-overtverbgeneration_run-1_bold.nii.gz
            sub-01_ses-test_task-overtverbgeneration_run-2_bold.nii.gz
```

In the above example, two JSON files are listed under `sub-01/ses-test/`, which are each applicable to `sub-01_ses-test_task-overtverbgeneration_run-2_bold.nii.gz`, violating the constraint that no more than one file may be defined at a given level of the directory structure. Instead `task-overtverbgeneration_run-2_bold.json` should have been under `sub-01/ses-test/func/`.

Example 2: Multiple run and rec with same acquisition (acq) parameters acq-test1

```
sub-01/
    anat/
    func/
        sub-01_task-xyz_acq-test1_run-1_bold.nii.gz
        sub-01_task-xyz_acq-test1_run-2_bold.nii.gz
        sub-01_task-xyz_acq-test1_rec-recon1_bold.nii.gz
        sub-01_task-xyz_acq-test1_rec-recon2_bold.nii.gz
        sub-01_task-xyz_acq-test1_bold.json
```

For the above example, all NIfTI files are acquired with same scanning parameters (`acq-test1`). Hence a JSON file describing the acq parameters will apply to different runs and rec files. Also if the JSON file (`task-xyz_acq-test1_bold.json`) is defined at dataset top level directory, it will be applicable to all task runs with `test1` acquisition parameter.

Case 2: Multiple json files at different levels for same task and acquisition parameters

```
sub-01/
   sub-01_task-xyz_acq-test1_bold.json
         anat/
         func/
             sub-01_task-xyz_acq-test1_run-1_bold.nii.gz
             sub-01_task-xyz_acq-test1_rec-recon1_bold.nii.gz
             sub-01_task-xyz_acq-test1_rec-recon2_bold.nii.gz
```

In the above example, the fields from `task-xyz_acq-test1_bold.json` file will apply to all

bold runs. However, if there is a key with different value in `sub-01/func/sub-01_task-xyz_acq-test1_run-1_bold.json`, the new value will be applicable for that particular run/task NIfTI file/s.

Extensions

The BIDS specification can be extended in a backwards compatible way and will evolve over time. A number of extensions are currently being worked on:

Extension label	Title	Moderators/leads
BEP001	Structural acquisitions that include multiple contrasts (multi echo, flip angle, inversion time) sequences	Gilles de Hollander
BEP002	The BIDS Models Specification	Tal Yarkoni
BEP003	Common Derivatives	Chris Gorgolewski
BEP004	Susceptibility Weighted Imaging (SWI)	Fidel Alfaro Almagro
BEP005	Arterial Spin Labeling (ASL)	Henk-Jan Mutsaerts and Michael Chappell
BEP006	Electroencephalography (EEG)	Cyril R Pernet, Robert Oostenveld, Stefan Appelhoff
BEP009	Positron Emission Tomography (PET)	Melanie Ganz
BEP010	intracranial Electroencephalography (iEEG)	Dora Hermes and Chris Holdgraf
BEP011	The structural preprocessing derivatives	Andrew Hoopes
BEP012	The functional preprocessing derivatives	Camille Maumet and Chris Markiewicz
BEP013	The resting state fMRI derivatives	Steven Giavasis
BEP014	The affine transformations and nonlinear field warps	Oscar Esteban
BEP015	Mapping file	Eric Earl, Camille Maumet, and Vasudev Raguram
BEP016	The diffusion weighted imaging derivatives	Franco Pestilli and Oscar Esteban
BEP017	Generic BIDS connectivity data schema	Eugene Duff and Paul McCarthy

Extension label	Title	Moderators/leads
BEP018	Genetic information	Cyril R Pernet, Clara Moreau, and Thomas Nichols
BEP019	DICOM Metadata	Satrajit Ghosh
BEP020	Eye Tracking including Gaze Position and Pupil Size(ET)	Benjamin Gagl and Dejan Draschkow
BEP021	Common Electrophysiological Derivatives	Stefan Appelhoff, Cyril Pernet, Robert Oostenveld, Teon Brooks

When an extension reaches maturity it is merged into the main body of the specification. If you would like to contribute to BIDS please consult the BIDS Contributor Guide All of the ideas that are not backwards compatible and thus will have to wait for BIDS 2.0 are listed here

Citing BIDS

When referring to BIDS in context of academic literature please cite:

> Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., Handwerker, D.A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B.N., Nichols, T.E., Pellman, J., Poline, J.-B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J.A., Varoquaux, G., Poldrack, R.A., 2016. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data 3, 160044.

as well as other papers describing specific BIDS extensions (see below).

BIDS has also a Research Resource Identifier (RRID) - `RRID:SCR_016124` - which you can also include in your manuscript in addition to citing the paper.