

BIDS Extension Proposal 3 (BEP003):

Common Derivatives

version 0.0.1 (working copy)
Available under the CC-BY 4.0 International license.

Extension moderator/lead: Chris Gorgolewski <chrisgor@stanford.edu>

This document contains a draft of the Brain Imaging Data Structure standard extension. It is a community effort to define standards in data/metadata. This is a working document in draft stage and any comments are welcome.

This specification is an extension of BIDS, and general principles are shared. The specification should work for many different settings and facilitate the integration with other imaging methods.

To see the original BIDS specification, see [this link](#). This document inherits all components of the original specification (e.g. how to store imaging data, events, stimuli, and behavioral data), and should be seen as an extension of it, not a replacement.

Table of contents

[Table of contents](#)

[Goals and scope of this extension](#)

[Relationship between BIDS-Derivatives datasets and BIDS-Raw datasets](#)

[General Principles](#)

[Filenaming conventions](#)

[Metadata conventions](#)

[Derivatives related subgroups](#)

[Guidelines for adding new data types](#)

[Storage of derived datasets](#)

[Metadata organization \(file level\)](#)

[Analysis levels](#)

[Compliance with BIDS-Raw](#)

[Derived dataset and pipeline description](#)

[Common Data types](#)

[Preprocessed, coregistered and/or resampled volumes](#)

[Brain mask](#)

[Regions of Interest](#)

[Future releases / work-product](#)

[Atlases](#)

[Visual reports](#)

Goals and scope of BIDS Derivatives

BIDS Derivatives is a representation of the outputs of common processing pipelines, capturing data and meta-data sufficient for a researcher to understand and (critically) reuse those outputs in subsequent processing. It is motivated by [use cases](#) where formalized machine-readable access to processed data enables higher level processing.

The relationship between BIDS-Derivatives datasets and BIDS-Raw datasets

The BIDS-Derivatives specification extends the BIDS-Raw specification. That is, **every BIDS-Derivative dataset must fully comply with the BIDS-Raw spec unless explicitly extended, negated or contradicted in this document or one of the domain-specific extensions**. In case of any apparent conflict between BIDS-Raw and BIDS-Derivatives, the latter always takes precedence. This inheritance principle (i.e., BIDS-Derivative inherits all BIDS-Raw rules) applies to *all* aspects of the spec. Some examples to illustrate:

- BIDS-Derivatives filenames must follow all BIDS-Raw file naming conventions. E.g., BOLD run 1 with task “rest” for subject “01” must begin with “sub-01_task-rest_run-1”.
- If a pipeline produces preprocessed versions of 2 different runs in each of 2 sessions for subject 01, the BIDS-Raw directory structure must be respected, producing paths like “/deriv_directory/sub-01/ses-1/sub-01_ses-1_run-1...”
- Every derivatives directory must include a `dataset_description.json` file at the root level.

General Principles

Filenaming conventions

- Filenames that are permissible under BIDS-Raw have a privileged status. Any *modification* of BIDS-Raw files must use a modified filename that does not conflict with the BIDS-Raw filename. Further, any files created as part of a BIDS-Derivative dataset must not match a permissible filename of a valid BIDS-Raw dataset. Stated equivalently, if any filename in a BIDS-Derivative dataset has a name permissible under BIDS-Raw, then that file must be an identical copy of that file in the associated BIDS-Raw dataset.
- Each Derivatives filename MUST be of the form:
`<BIDSRawBase>[_keyword-<value>]_<suffix>.<ext>`
 - When the derivatives chain involves outputs derived from a single BIDS-Raw input, `BIDSRawBase` MUST be the entire filename from the BIDS-Raw input, including any optional components that were included in the BIDS-Raw filename, *except* that the `suffix` must always come at the end of the filename (and just before the extension). One exception to this rule is filename keywords that are no longer relevant (such as `sub-` in the case of group level derivatives). Depending on the nature of the derivative file, the `suffix` can either be the same as the original BIDS-Raw input if that suffix is still appropriate, or a new appropriate value selected from the controlled list.
 - When the derivatives output involves multiple BIDS-Raw inputs, `BIDSRawBase` should preserve the elements in common between the input filenames in the process of creating a sensible `BIDSRawBase`.
- The filename ends with a single `_suffix` tag and no other `_suffix` tags are allowed elsewhere in the filename. Only suffixes defined in this document (Table XX), the BIDS-Raw specification, or one of the other derivatives specifications are allowed. An arbitrary number of keyword-value pairs may be included, but all included keywords must be (a) either required or allowed by the BIDS-Raw specification (e.g., `sub`, `run`, `task`, etc.) or (b) explicitly allowed by this document or one of the derivatives extensions. In other words, users MUST NOT include arbitrary keyword-value pairs that are not explicitly allowed for the type of file in question.
- Unless specified otherwise, the `value` component of a `_keyword-value` pair is freeform and should be set appropriately by the user.
- Other than the privileged status of BIDS-Raw filenames, there is no prohibition against identical filenames in different BIDS-Derivatives directories, although users should be aware of the potential ambiguity this can create and use the sidecar JSON files to detail the specifics of individual files.

- When necessary to distinguish two files, the `_desc-<value>` keyword-value should be used. This includes the cases of needing to distinguish both differing inputs and differing outputs [e.g., `_desc-T1w` and `desc-T2w` to distinguish brain mask files derived from T1w and T2w images; or `_desc-sm4` and `_desc-sm8` to distinguish between outputs generated with two different levels of smoothing].

Metadata conventions

- Unless specified otherwise, individual sidecar JSON files and all metadata fields within are optional. However, the appropriate use of these files and pertinent fields is very valuable and thus encouraged. Moreover, for some types of files, there may be one or more required metadata fields, in which case at least one metadata file containing that field must be located somewhere within the file's hierarchy (per the Inheritance Principle defined in BIDS-Raw).
- When chaining derivative pipelines, any JSON fields that were specified as mandatory in the input files should be propagated forward in the output file's JSON provided they remain valid. Non-required JSON fields can be propagated, and are highly useful, but it is the pipeline's responsibility to ensure that the values are still relevant and appropriate to the type of output data.
- of a given metadata field MUST NOT be interpreted as an indication that a given field is set to false.

Derivatives related subgroups

This document concerns only those elements of the BIDS-Derivatives specification that are common to multiple extensions/domains within the overall BIDS scope. Domain-specific specifications may be found in individual BIDS Extension Proposals (BEPs) listed below:

[Copy of this document before the split](#)

[BIDS Extension Proposal 11 \(BEP011\): The structural preprocessing derivatives](#)

[BIDS Extension Proposal 12 \(BEP012\): The functional preprocessing derivatives](#)

[BIDS Extension Proposal 14 \(BEP014\): Spaces and Mappings](#)

[BIDS Extension Proposal 16 \(BEP016\): The diffusion weighted imaging derivatives](#)

Guidelines for adding new data types

When developing domain-specific extensions that require the introduction of new data types or file formats, developers are strongly encouraged to reuse file formats adopted in the main BIDS spec (NIFTI, TSV, JSON) or already proposed for a different domain/modality. When selecting a

new file format for data, open and widely used should be preferred (GIFTI, CIFTI), with a bias to most widely used tools.

Storage of derived datasets

Derivatives can be stored/distributed in two ways:

1. Under a `derivatives/` subfolder in the root of the BIDS-Raw dataset folder to make a clear distinction between raw data and results of data processing.

Each pipeline has a dedicated directory under which it stores all of its outputs. There is no restriction on the directory name; however, it is RECOMMENDED to use the format `<pipeline>-<variant>` in cases where it is anticipated that the same pipeline will output more than one variant (e.g., “AFNI-blurring”, “AFNI-noblurring”, etc.).

For example:

```
<dataset>/derivatives/fmripipeline_01/sub-0001
```

```
<dataset>/derivatives/spm/sub-0001
```

```
<dataset>/derivatives/vbm/sub-0001
```

2. As a standalone dataset independent of the source BIDS-Raw or BIDS-Derived dataset. This way of specifying derivatives is particularly useful when the original BIDS-Raw is provided with read-only access, and for publishing derivatives as independent bodies of work, or for describing derivatives that were created from more than one dataset.

The rest of the document assumes Case 1, but Case 2 applies after removing `/derivatives/<pipeline_name>` from template names.

Analysis levels

Compliance with BIDS-Raw

As noted above, the BIDS-Derivative standard extends the BIDS-Raw standard. Consequently, files should be organized to comply with BIDS-Raw to the full extent possible (i.e., unless explicitly contradicted by BIDS-Derivatives). Any subject-specific derivatives should be housed within each subject’s directory; if session-specific derivatives are generated, they should be deposited under a session subdirectory within the corresponding subject directory; and so on.

In addition to the structure defined in the BIDS-Raw specification, the BIDS-Derivatives specification provides additional flexibility intended to account for derivatives that do not fit within the standard BIDS-Raw project layout. In particular, pipelines may generate derivatives that do not strictly belong to any single participant. To write group-level results, the `group/` folder under `derivatives/<pipeline_name>/` can be used:

<dataset>/derivatives/<pipeline_name>/group/

Derived dataset and pipeline description

In keeping with the BIDS-Raw standard, a `dataset_description.json` file MUST be found at the top level of the particular pipeline:

<dataset>/derivatives/<pipeline_name>/dataset_description.json

Keys. The `dataset_description.json` file in a BIDS-Derivatives dataset can contain arbitrary key-value pairs that encode metadata relevant to understanding the structure and contents of the directory. Although there is no strictly controlled vocabulary, users are strongly encouraged to use keys defined in the Raw-BIDS spec when available. In addition, the BIDS-Derivatives spec introduces a number of other required or recommended keys:

<code>PipelineDescription.Name</code>	REQUIRED
<code>PipelineDescription.Version</code>	
<code>PipelineDescription.CodeURL</code>	
<code>PipelineDescription.DockerHubContainerTag</code>	
<code>PipelineDescription.SingularityContainerURL</code>	
<code>PipelineDescription.SingularityContainerVersion</code>	
<code>SourceDatasets</code>	A list of objects specifying the locations and relevant attributes of all source datasets. Valid fields in each object include <code>url</code> , <code>doi</code> , and <code>version</code> .

Metadata organization (file level)

Each derivative file SHOULD be described by a JSON file provided as a sidecar or higher up in the hierarchy of the derived dataset (according to Inheritance Principle - [section 3.5 of the main specification](#)) unless a particular derivative includes REQUIRED metadata fields in which case a JSON file is also REQUIRED. Each derivative type defines their own set of fields, but all of them share the following (non-required) ones:

Description	RECOMMENDED. Freeform natural language description of the nature of the file.
Sources	OPTIONAL. A list of paths relative to dataset root pointing to the file(s) that were directly used in the creation of this derivative. For example in a chain of A->B->C, "C" should only list "B" as Sources, and "B" should only list "A" as Sources. However in case X and Y jointly contribute to Z, then "Z" should list "X" and "Y" as Sources.
RawSources	OPTIONAL. A list of paths relative to dataset root pointing to the BIDS-Raw file (s) that were used in the creation of this derivative.
CustomMetadata	OPTIONAL. Custom user-defined fields and/or nested dictionaries that are not part of the BIDS specification and are not intended to be incorporated into future releases of the BIDS specification.

Common Datatypes

Processed, coregistered and/or resampled volumes

Template:

```
<pipeline_name>/
  sub-<participant_label>/
    func|anat|dwi/
      <BIDSRawBase>[_space-<space>][_desc-<label>][_suffix>].<ext>
```

Processing in this context means transformations of data that does not change the number of dimensions of the input and are not explicitly covered by other Datatypes in the specification.

Examples:

- Motion-corrected, temporally denoised, and transformed to MNI space bold files.
- Inhomogeneity corrected and skull stripped T1w files.
- Motion-corrected dwi files.

The `space` keyword denotes the space of the file with a controlled vocabulary of values - see [BEP014](#) for details. The `desc` keyword is a general purpose field with freeform values. To distinguish between multiple different versions of processing for the same input data the `desc` keyword should be used. **Note that even though `space` and `desc` are optional at least one of them needs to be defined to avoid name conflict with the BIDS-Raw file.**

Examples:

```
pipeline1/sub-001/
  func/
    sub-001_task-rest_run-1_space-MNI305_bold.nii.gz
    sub-001_task-rest_run-1_space-MNI305_bold.json
```

```

pipeline1/sub-001/
  func/
    sub-001_task-rest_run-1_desc-MC_bold.nii.gz
    sub-001_task-rest_run-1_desc-MC_bold.json
  OR
    sub-001_task-rest_run-1_desc-fmriprip_bold.nii.gz
    sub-001_task-rest_run-1_desc-fmriprip_bold.json

```

All REQUIRED metadata fields coming from a derivative file's source file(s) MUST be propagated to the JSON description of the derivative **unless the processing makes them invalid** (e.g., if a source 4D image is averaged to create a single static volume, a `SamplingFrequency` property would no longer be relevant). In addition, all processed files include the following metadata JSON fields:

Key	Description	Value for unprocessed data
Space	REQUIRED. Keyword denoting which space the processed file is in see BEP014 for details.	'orig'
SkullStripped	REQUIRED. Whether the volume was skull stripped (non-brain voxels set to zero) or not.	false

Masks

Template:

```

<pipeline_name>/
  sub-<participant_label>/
    func|anat|dwi/
      <BIDSRawBase>[_space-<space>][_desc-<label>]_mask.nii.gz

```

A binary (1 - inside, 0 outside) mask in the space defined by `<space>` (see Table targets/spaces). By default (i.e., if no transformation has taken place) the value of `space` should be set to "orig".

Key	Description	
RawSources	Elevated from OPTIONAL to REQUIRED	

Space	REQUIRED. Keyword denoting which space the processed file is in see BEP014 for details.	'orig'
Type	RECOMMENDED. Short identifier of the mask. Restricted values: <ul style="list-style-type: none"> - "Brain" - brain mask - "Lesion" - lesion mask - "Face" - face mask - "ROI" - ROI mask 	

Examples:

```
func_loc/
  sub-001/
    func/
      sub-001_task-rest_run-1_space-MNI305_desc-PFC_mask.nii.gz
```

```
manual_masks/
  sub-001/
    anat/
      sub-001_T1w_label-desc_mask.nii.gz
```

Future releases / work-product / not to be included in the first release

[BIDS Extension Proposal 21 \(BEP021\): Common Electrophysiological Derivatives](#)

[BIDS Extension Proposal 15 \(BEP015\): Mapping file](#)

[BIDS Extension Proposal 22 \(BEP022\): PET preprocessing derivatives](#)

[BIDS Extension Proposal 23 \(BEP023\): MRSI derivatives](#)

[BIDS Extension Proposal 17 \(BEP017\): Generic BIDS connectivity data schema](#)

Atlases

```
reference_atlases/  
  sub-001/  
    anat/  
      sub-001_t1w_label-AAL_atlas.nii.gz
```

Label name	Description
AAL2	Updated version of Automated Anatomical Labeling [Tzourio-Mazoyer, 2002]) where 120 regions are extracted using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.
Neuromorphometrics	Maximum probability tissue labels derived from the “MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling” containing 140 cortical and subcortical regions in MNI space. http://www.neuromorphometrics.com/
AICHA	Atlas of Intrinsic Connectivity of Homotopic Areas, a functional brain ROIs atlas based on resting-state fMRI data acquired in 281 individuals containing 384 regions of the whole brain in MNI space. http://www.gin.cnrs.fr/AICHA-344?lang=en
LPBA40	Atlas produced from a set of whole-head MRI of 40 human volunteers [Shattuck, 2008]. Each MRI was manually delineated to identify a set of 56 cortical and subcortical regions in MNI space.

JHU DTI81	Also known as ICBM-DTI-81 white-matter labels atlas [Wakana et al., 2007; Hua et al., 2008]. 48 white matter tract labels were created by hand segmentation of a standard-space average of diffusion MRI tensor maps from 81 subjects; mean age 39 (18:59), M:42, F: 39. http://www.loni.usc.edu/ICBM/Downloads/Downloads_DTI-81.shtml
JHUTracts[0 25 50]	Also known as JHU white-matter tractography atlas [Mori et al., 2005]. 20 structures were identified probabilistically by averaging the results of running deterministic tractography on 28 normal subjects (mean age 29, M:17, F:11). https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases
Desikan	
Destrieux	

Visual reports

Visual reports for the assessment of processing/analysis of individual participants are stored at the top level of the corresponding participant:

```
<pipeline_name>/
  sub-<participant_label>/
    sub-<participant_label>[_T1w].html
```

When the visual report corresponds to a higher level of analysis, then the group/ folder should be used:

```
<pipeline_name>/
  group/
    group_T1w.html
```