

BIDS Extension Proposal xx (BEP0xx): White Matter PASTA Tractometry

version 0.0.4-dev

Available under the CC-BY 4.0 International license.

Extension moderators/leads: Robert Smith <robert.smith@florey.edu.au>, Ariel Rokem <arokem@uw.edu> and Franco Pestilli <pestilli@utexas.edu>

Contributors: Oscar Esteban <Oscar.Esteban@chuv.ch>, Matthew Cieslak <matthew.cieslak@pennmedicine.upenn.edu>, François Rheault <Francois.M.Rheault@usherbrooke.ca>, Sydney Covitz <sydney.covitz@pennmedicine.upenn.edu>, Hamsi Radhakrishnan <radhakrh@pennmedicine.upenn.edu>, Peer Herholz <herholz.peer@gmail.com>

This document contains a draft for BIDS Extension Proposal 36. It is a community effort to define standards in data / metadata storage for diffusion MRI tractography. 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.

Meta-notes

- Pre-existing BIDS Extension Proposal BEP016 (“The diffusion weighted imaging derivatives”) is being [reduced in scope to cover only diffusion models](#), with responsibility for downstream derivatives of diffusion imaging data being moved to separate BEPs. BEP036 is the first of such proposals.
- The novel content of this document relating to the specification of White matter tractometry content (from section 3.1 onwards) is a duplicate of that originally

presented in [version 0.0.1 of BIDS Extension Proposal 016](#).

- This document is not intended to be kept up to date with respect to either content of the proposal or discussion among invested parties. Revision of the specification, and all relevant discussion, will occur on the [relevant git repository](#).

1. Table of contents

1. Table of contents	2
2. Scope	2
2.1. Goals of BIDS - White matter tractometry	3
2.2. Relation of BIDS - White Matter Tractometry to BIDS	3
3. BIDS - PASTA tractometry	4
3.1 Directory structure	4
3.1.1 _pasta.tsv	4
3.1.2 _pasta.json	5
What is needed to finish this BEP	6

2. Scope

This specification extends the Brain Imaging Data Structure (BIDS) Specification to include derivatives that are the representation of the outcomes of Pointwise Assessment of Streamline Tractography Attributes (PASTA; Jones et al., 2005). We limit ourselves here strictly to 1-d PASTA, as in the original PASTA article, acknowledging that there are many other ways to do tractometry (including Bundles Analytics, a.k.a., BUAN, radiomic tractometry, and others), which are out of scope for this BEP.

Ways that PASTA data may come about:

1. From individual streamlines
2. From groups of streamlines
3. From volumetric ROI (e.g., <https://www.science.org/doi/full/10.1126/sciadv.abm1971>)
4. XXX

2.1. Goals

The tabular data that PASTA pipelines generate is an input to further analysis steps that compare different subjects to each other in terms of their brain tissue properties. Expanding BIDS Derivatives to encompass PASTA tractometry data would facilitate the sharing of a greater fraction of the data generated during the course of diffusion MRI experimentation, for validation and re-use by the wider community. It would additionally improve the capability for comparing data generated by different software packages by providing a common intermediate representation.

The input to a white matter tractometry analysis is typically a tractography output or the output of another pipeline that divides the brain white matter into discrete regions of interest and/or trajectories. These inputs are the responsibility of BEP16, [BEP0xx](#) (which focuses on the tractography results), [BEPyy](#) (which will focus specifically on the division of the brain into different structures).

Regarding brain structures: this BEP does not make any assumptions about the ontology that researchers may choose to use in describing brain structures along which PASTA assessments are made. This is because, for many brain structures, and particularly within the white matter, such ontologies are either (1) non-existent, or (2) not widely used. For example, though the white matter query language (WMQL) does define structures in an unambiguous way, these structures may not always align with the definitions inherited from classical neuroanatomy. Furthermore, even those structures that are defined more-or-less clearly by classical neuroanatomy tend to vary widely between researchers in practice (see Rheault et al., 2020). Thus, researchers need to define for themselves how these structures are delineated and communicate clearly about the decisions they made (e.g., in terms of automation, in terms of virtual dissection procedures, etc.).

2.2. Relation to BIDS

Most core principles of the original BIDS-Raw specification are inherited by the BIDS - White matter PASTA tractometry specification, though some special considerations and additional fields are noted below.

Please refer to general BIDS specification document for context and general guidelines (definitions, units, directory structure, missing values, stimulus and event information, etc.): <https://bids-specification.rtfd.io>

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 [\[RFC2119\]](#).

3. BIDS - PASTA tractometry

BIDS contains “required”, “recommended” and “optional” fields. These are indicated throughout the document:

- REQUIRED: essential to be BIDS compliant (i.e. MUST as per RFC2199)
- RECOMMENDED: gives a warning if not present (i.e. SHOULD as per RFC2199)
- OPTIONAL: no warning if missing (i.e. MAY as per RFC2199)

As in BIDS-Raw, the following apply:

- 1) All specifications of paths need to use forward slashes.
- 2) The inheritance principle applies: any metadata file (.json, .tsv, etc.) may be defined at any directory level. The values from the top level are inherited by all lower levels unless they are overridden by a file at the lower level. For details see BIDS-Raw ([The Inheritance Principle](#)).

3.1 Directory structure

For the purpose of this BEP, a directory containing two files is needed to be compliant with BIDS. The directory will contain a data file and a header file. The datafile is a .tsv|.zarr|.h5 marked with the suffix _pasta, i.e., _pasta.json. Below an example of a PASTA directory

```
my_pipeline/  
  sub-<participant_label>/  
    dwi/  
      sub-01_pasta.tsv|.zarr|.h5  
      sub-01_pasta.json
```

3.1.1 _pasta.tsv

- no header to make it interoperable across different file formats, also follow changes introduced through BEP017
 - header information provided in json, ie the “Axes” key outlining what is entailed in each column

Example

anteriorFrontalICC	1	0.001216	0.306	0.911
anteriorFrontalICC	2	0.001222	0.305	0.912
anteriorFrontalICC	3	0.001344	0.304	0.913
anteriorFrontalICC	4	0.001322	0.303	0.914
anteriorFrontalICC	5	0.001301	0.304	0.914
leftSLF1	1	0.001290	0.202	0.913
leftSLF1	2	0.001291	0.201	0.913
leftSLF1	3	0.001293	0.200	0.913
leftSLF1	4	0.001290	0.203	0.915
leftSLF1	5	0.001290	0.203	0.915

3.1.2 _pasta.json

Keys	Values
Software	This is the software used to calculate the PASTA values. Either same for all columns or one per column in which case it has to be in Axes instead of here
Sources	Is a list of files used to generate the current dataset (this is the closest field to tracking provenance)
Axes	List of dictionaries, one dictionary per column including the name of what’s in the column, its type, its levels, units, etc. For what pertains to the first column, which in general will describe the White Matter Tracts a.k.a., structures, the

	Axes will be a dictionary. More specifically, a dictionary where the keys are strings with the names of the structures (one and the values are: orientation of the PASTA values. I.e., where is the first node and where is the last node in the coordinate frame of the brain. For example, for a structure that passes through the corpus callosum, we might indicate that it goes from left to right by entering: "Orientation": "R".
--	---

Example

```

...
"Software": ["DIPY", "AMICO"],
"Axes": [{"Structures":
    {"Left Arcuate Fasciculus": {"Orientation": "S"},
    "Right Arcuate Fasciculus": {"Orientation": "S"},
    "Frontal CC": {"Orientation": "R"},
    "Right Superior Lateral Fasciculus I":
        {"Orientation": "A"}},
    {"Mean Diffusivity": {"Units": "mm^2/s", "Abbreviation": "MD"}},
    {"Fractional anisotropy": {"Units": "Unitless", "Abbreviation": "FA"}},
    {"Orientation Dispersion Index": {"Abbreviation": "ODI"}}]
}, ],
"Sources": ["/[...]/sub-01/ses-01/dwi/sub01-ses01_dwi.nii.gz",
    "/[...]/dwi/sub-01-some_non_bids_thing.tck"],
}
...

```

What is needed to finish this BEP

1. Examples of file representations
2. A BIDS example data repository (empty datafiles)
- 3.

Extension proposal

Extension of the bids-schema to support the BEP

Examples (to test the validator is working properly) added to bids-examples