

BIDS Derivatives

Christopher J. Markiewicz
Stanford University

Outline

The Evolution of BIDS Derivatives

Overview of BIDS Common Derivatives

Other BEPs and Open Questions

The Evolution of BIDS Derivatives

A brief history of derivatives

Feb 2016: Working document

- Initially covers coregistered volumes, affine/warp transforms, confound regressors, brain masks

Mar 2016 - Mar 2017

- Expansion: estimated field maps, artifact corrections, ROI masks, discrete and probabilistic segmentations, tracts/tractograms, diffusion maps
- Consolidation: artifact corrections and coregistrations become “minimally processed”
- Pipeline descriptions add coarse provenance

Apr 2017

- Dubbed “BEP 003: The BIDS Derivatives Specification”

Aug 2017: (BIDS Apps Workshop @ Stanford)

- BEP 003 becomes “Common Derivatives”, BEPs 011-016 are spawned

Interlude - Sub-BEPs

- BEP 011 - Structural MRI preprocessing derivatives
- BEP 012 - Task fMRI preprocessing derivatives
- BEP 013 - Resting fMRI preprocessing derivatives
- BEP 014 - Affine transformation and nonlinear warp fields
- BEP 015 - Derivative mapping file
- BEP 016 - Diffusion weighted imaging derivatives

Later, related BEPs

- BEP 017 - Generic BIDS connectivity data schema
- BEP 021 - Common electrophysiological derivatives
- BEP 023 - PET preprocessing derivatives
- BEP 028 - BIDS Provenance
- BEP 038 - Atlases
- BEP 039 - Dimensionality reduction-based networks

More history

Aug 2018: (BIDS Derivatives Workshop @ Stanford)

- Sub-BEPs worked on mostly independently
- “Common Derivatives” refined by polling sub-BEPs, finding the points of agreement

Dec 2018

- Release Candidate 1 ([bids-spec#109](#)) with large components of structural, functional, diffusion

Apr 2019

- Move to community-managed branch/PR ([bids-spec#207](#))
- This turned out to be unreviewable, so...

Jul 2019

- Common Derivatives and BEPs 011, 012 and 016 return

Jun 2020

- Common derivatives are included in BIDS 1.4.0 release

Common Derivatives

Changes to Common Principles

Dataset organization

Sub-dataset

```
my_dataset/  
  derivatives/  
    preprocessed/  
      analysis/  
        sub-01/  
        ...
```

Sibling dataset

```
my_study/  
  raw_data/  
    sub-01/  
    ...  
  derivatives/  
    preprocessed/  
      analysis/
```

Super-dataset (new)

```
my_analysis/  
  sourcedata/  
    raw/  
    preprocessed/  
      sub-01/  
      ...
```

Deeply nested (**YODA**)

```
my_analysis/  
  sourcedata/  
    preprocessed/  
      sourcedata/  
        raw/  
        sub-01/  
        ...  
        sub-01/  
        ...  
        sub-01/
```

Non-compliant derivatives

Nothing in this specification should be interpreted to disallow the storage/distribution of non-compliant derivatives of BIDS datasets. In particular, if a BIDS dataset contains a **derivatives/** subdirectory, the contents of that directory may be a heterogeneous mix of BIDS Derivatives datasets and non-compliant derivatives.

Many public datasets already include non-conformant derivatives. This balances attempts to define validatable derivatives while not penalizing people for sharing before the spec existed.

Dataset description

- New fields were introduced for *all* datasets:
 - **DatasetType** indicates "raw" or "derivative" data.
 - **GeneratedBy** contains references to the code (including versions) that produced the derivative dataset.
 - **SourceDatasets** is a list of references to the specific version of the dataset analyzed.

```
1  {
2    "Name": "FMRIprep Outputs",
3    "BIDSVersion": "1.4.0",
4    "DatasetType": "derivative",
5    "GeneratedBy": [
6      {
7        "Name": "fmriprep",
8        "Version": "1.4.1",
9        "Container": {"Type": "docker", "Tag": "poldracklab/fmriprep:1.4.1"}
10      },
11      {
12        "Name": "Manual",
13        "Description": "Re-added RepetitionTime metadata to bold.json files"
14      }
15    ],
16    "SourceDatasets": [
17      {
18        "DOI": "doi:10.18112/openneuro.ds000114.v1.0.1",
19        "URL": "https://openneuro.org/datasets/ds000114/versions/1.0.1"
```

Derivatives file-naming

- Form: <source_entities>[_<key>-<value>]<suffix>.<ext>
- Rules:
 - Derivative files must have some way, such as a new entity or suffix, to distinguish them from raw files.
 - The free-form desc-<label> entity may be used in any derivative file.
 - “Source entities” should be preserved unless they no longer apply.
 - For example, combining across runs, the run-<index> entity should be dropped.
 - Raw files may be included in a derivatives dataset without additional entities.

Additionally, the space-<label> entity is almost always permissible to indicate resampled data.

Derivatives common metadata

- Most metadata is optional but encouraged.
- Raw metadata that still applies to derivatives should be propagated.
- Free-text "**Description**" is always RECOMMENDED.
- "**Sources**" is always at least OPTIONAL (specific derivatives can make RECOMMENDED or REQUIRED).

Spatial references

- "**SpatialReference**" provides an explicit link to local or remote files.
- BIDS defines standard template identifiers (see [Appendix: Coordinate systems](#)). `space-MNI305`, for example, does not require "**SpatialReference**".
- In the absence of explicit entities/metadata, an image is in space "`orig`" and is its own reference.
- If the `<label>` in `space-<label>` is not part of the BIDS-defined standard template identifiers, **SpatialReference** is required.

Preprocessed or cleaned data

Data is considered to be *preprocessed* or if the data type of the input, as expressed by the BIDS **suffix**, is unchanged.

Template:

```
1 <pipeline_name>/  
2     sub-<label>/  
3         <datatype>/  
4             <source_entities>[_space-<space>][_desc-<label>]_<suffix>.<ext>
```

Examples:

```
pipeline1/  
  sub-01/  
    anat/  
      sub-01_space-MNI305_T1w.nii.gz  
      sub-01_space-MNI305_T1w.json  
  eeg/  
    sub-01_task-rest_desc-filtered_eeg.edf  
    sub-01_task-rest_desc-filtered_eeg.json  
  func/  
    sub-01_task-rest_space-MNI305_desc-preproc_bold.nii.gz  
    sub-01_task-rest_space-MNI305_desc-preproc_bold.json
```

Imaging derivatives

Some derivatives common to MRI modalities were encoded. With the addition of electrophysiology and PET, these were reclassified from “common” to “imaging”.

Resampled volumes

Adds resolution (`res-<label>`) and density (`den-<label>`) entities:

```
fmriprep/  
  sub-01/  
    anat/  
      sub-01_space-MNI152NLin2009cAsym_res-2_desc-preproc_T1w.nii.gz  
    func/  
      sub-01_task-rest_space-MNI152NLin2009cAsym_res-2_desc-preproc_bold.nii.gz
```

Binary masks

```
manual_masks/  
  sub-01/  
    anat/  
      sub-01_desc-lesion_mask.nii.gz  
      sub-01_desc-lesion_mask.json  
localizer/  
  sub-01/  
    func/  
      sub-01_task-loc_space-individual_label-FFA_mask.nii.gz  
      sub-01_task-loc_space-individual_label-FFA_mask.json
```

Imaging derivatives (continued)

Discrete and probabilistic segmentations

Example:

```
tissue_segmentation/
  desc-AAL_dseg.tsv
  desc-AAL_probseg.json
  sub-01/
    func/
      sub-01_task-rest_desc-AAL_dseg.nii.gz
      sub-01_task-rest_desc-AAL_probseg.nii.gz
```

Discrete segmentations are 3D images with integer values, following either a default label convention or as specified in **dseg.tsv**:

```
desc-AAL_dseg.tsv
index  name
1      Precentral_L
2      Precentral_R
3      Frontal_Sup_L
4      Frontal_Sup_R
...
```

Probabilistic/partial-volume segmentations are 4D images with values between 0 and 1, with labels identified in a corresponding JSON file:

```
desc-AAL_probseg.json
{"LabelMap": ["Precentral_L", "Precentral_R", "Frontal_Sup_L", ...]}
```

Other BEPs and Open Questions

Possibly relevant precedents

BEP011: Anatomical derivatives

- Surface-sampled measures (`_<suffix>.shape.gii`)
- Tabular morphometry statistics (`_morph.tsv`)

BEP012: Functional derivatives

- Summary statistics (`_<mean|std|...>.nii.gz`)
- Time series
 - Generic, including ROI summary, time series (`_timeseries.<tsv|nii[.gz]>`)
 - Special cases (`_motion.tsv`, `_outliers.tsv`)
 - Spatiotemporal decompositions (`_<components|micing>.<tsv|nii[.gz]>`)

BEP014: Transformations and mappings

- Affine matrices (`_xfm.<ext>`)

Possibly relevant precedents

BEP016: Diffusion derivatives

- Model fit parameters (`_model.nii.gz`)
- Model-derived parameters (`_mdp.nii.gz`)
- `model-<label>` and `param-<label>` entities

BEP038: Atlases

- Promotes `atlas-<label>` from [Imaging data types - Segmentations](#) to something subject-like
- Introduces some summary statistics in passing (c.f. BEP012)

BEP039: Dimensionality reduction-based networks

- `group/` and `group-<label>/` replacing `sub-<label>/`
- `model-<label>/` directory under datatype (`anat/`, `pet/`, etc.)

BEP028: BIDS Provenance

- JSON-LD files at file (`*_prov.jsonld`) and/or dataset (`/prov.jsonld`) level.

Open questions

- Group “subjects”
 - Status quo: Drop subject label when combining across subjects
 - Bare `group/` with no subject labels?
 - `group-<label>/` that acts like a subject?
 - `group-<label>/` that is an additional level (e.g., `group-A/[sub-01/]`)?
- Additional levels in the hierarchy
 - BEP-039 introduces `model-<label>/` at bottom of hierarchy.
 - FitLins (BEP-002 Stats Models implementation) uses `node-<label>/` at top of hierarchy.
- N-dimensional array files ([bids-spec#197](#))
 - Not obviously relevant to PET, but it may come up

Opinionated advice

- There is a tendency to try to specify everything, and pull back to the least common denominator.
 - There will be things you think are important that have to get dropped.
 - For example, we had individual derivatives for motion correction, bias field correction, resampling. They are now covered by `space-<label>` and `desc-<label>`.
- Every rule needs implementing at least twice.
 - Prioritize validator, then producers, then consumers.
 - Minimizing new concepts eases implementation.
- Detailed provenance should probably target BEP028. Resist the urge to specify everything in JSON.
 - You need enough for humans to follow along, and downstream tools to make decisions.
 - Tools can still add what their authors and users think is important.

Thank you