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

version 0.0.1 (working copy)

Extension moderator/lead: Andrew Hoopes <ahoopes@mgh.harvard.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 BIDS Common Derivatives proposal, 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, for Common Derivatives, 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.

## Segmentation Metadata

The `desc-manual` key-value pair should be used to specify whether a segmentation (or parcellation) has been created manually. For automatic segmentations, the `atlas` key should indicate which atlas, if any, was used to create the labels.

## Discrete Segmentations

Discrete segmentations of brain tissue represent each tissue class with a unique integer label in a 3D volume.

Template:

```
<pipeline_name>/
        sub-<participant_label>/
                func|anat|dwi/
                        <source_file>[_space-<space>]_dseg.nii.gz
```

Example:

```
pipeline/sub-001/
```

```
        anat/
                sub-001_space-ACPC_dseg.nii.gz
                sub-001_space-ACPC_dseg.json
```

A segmentation could be a binary mask that functions as a discrete label for a single structure. In this case, the `label` key must be used to specify the corresponding structure. For example:

```
pipeline/sub-001/
        anat/
                sub-001_space-ACPC_label-GM_dseg.nii.gz
```

## Probabilistic Segmentations

Probabilistic segmentations of brain tissue represent a single tissue class with values ranging from 0 to 1 in individual 3D volumes or across multiple frames. Similarly to a discrete, binary segmentation, the `label` key can be used to specify the corresponding structure.

Template:

```
<pipeline_name>/
        sub-<participant_label>/
                func|anat|dwi/
                        <source_file>[_space-<space>][_label-<label>]_probseg.nii.gz
```

Example:

```
pipeline/sub-001/
        anat/
                sub-001_space-ACPC_label-BG_probseg.nii.gz
                sub-001_space-ACPC_label-WM_probseg.nii.gz
```

A 4D probabilistic segmentation, in which each frame corresponds to a different tissue class, must provide a label mapping in its JSON sidecar. For example:

```
pipeline/sub-001/
        anat/
                sub-001_space-ACPC_probseg.nii.gz
                sub-001_space-ACPC_probseg.json
```

The JSON sidecar must include the `label-map` key that specifies a tissue label for each frame:

```
{
        "label-map": [
                "BG",
                "WM",
                "GM"
        ]
}
```

## Anatomical Labels

BIDS supplies a standard, generic label-index dictionary, defined in the table below, that contains common tissue classes and can be used to map segmentations (and parcellations) between lookup tables.

| Index | Name | Abbreviation |
|---|---|---|
| 0 | Background | BG |
| 1 | Grey Matter | GM |
| 2 | White Matter | WM |
| 3 | Cerebrospinal Fluid | CSF |
| 4 | Grey and White Matter | GWM |
| 5 | Bone | B |
| 6 | Soft Tissue | ST |
| 7 | Non-brain | NB |
| 8 | Lesion | L |
| 9 | Cortical Grey Matter | CGM |
| 10 | Subcortical Grey Matter | SCGM |
| 11 | Brainstem | BS |
| 12 | Cerebellum | CBM |

These definitions can be overridden (or added to) by providing custom labels in a sidecar `<matches>_labels.tsv` file, in which `<matches>` corresponds to segmentation filename.

Example:

```
pipeline/sub-001/
        anat/
                sub-001_space-ACPC_dseg.nii.gz
                sub-001_space-ACPC_dseg_labels.tsv
```

Definitions can also be specified with a top-level `labels.tsv`, which propagates to segmentations in relative subdirectories.

Example:

```
pipeline/
```

```
       labels.tsv
       sub-001/anat/
              sub-001_space-ACPC_dseg.json
```

These tsv lookup tables should contain the following columns:

| **Column name** | **Description** |
|---|---|
| `index` | The label integer index |
| `name` | The unique label name |
| `abbr` | OPTIONAL The unique label abbreviation |
| `mapping` | OPTIONAL Corresponding integer label in the standard BIDS label lookup |
| `color` | OPTIONAL Label color for visualization |

An example, custom `labels.tsv` that defines three labels:

```
index  name                 abbr   color         mapping
100    "Grey Matter"        GM     #ff53bb       1
101    "White Matter"       WM     #2f8bbe       2
102    "Brainstem"          BS     #36de72       11
```

When specifying a particular label within a filename or JSON, it's acceptable to supply either the label index, abbreviation, or name (with whitespace removed).

## Reconstructed cortical surfaces

Reconstructed cortical surfaces should be stored as GiFTI files, and each hemisphere should be stored separately.

Template:

```
<pipeline_name>/
       sub-<participant_label>/
              func|anat|dwi/
                     <source_file>_hemi-{L|R}[_space-<space>]_[surftype].surf.gii
```

Example:

```
pipeline/sub-001/
       anat/
              sub-001_hemi-L_pial.surf.gii
              sub-001_hemi-R_pial.surf.gii
```

The supported surface types are:

| **Surface type** | **Description** |
|---|---|
| `wm` | The gray matter / white matter border for the cortex |
| `smoothwm` | The smoothed gray matter / white matter border |
| `pial` | The gray matter / pial matter border |
| `midthickness` | The midpoints between wm and pial surfaces |
| `inflated` | An inflation of the midthickness surface (useful for visualization) |
| `sphere` | The sphere (used for registration - see transforms for nomenclature) |
| `flat` | The flattened surface (used for visualization) |

Note: reconstructed cortical surfaces are unique in that they contain both 2-dimensional and 3-dimensional elements of "space". More details on this concept are given in BEP14 Spaces and Transforms.

## Surface-Mapped Anatomical Scalar Derivatives

Surface-mapped scalar overlays should be stored as either GiFTI or CIFTI files (which allow for the combination of left and right hemispheres).

Template:

```
<pipeline_name>/
        sub-<participant_label>/
                func|anat|dwi/
                        <source_file>[_hemi-{L|R}][_space-<space>]_<suffix>.<ext>
```

The preferred extension for scalar GiFTI files is `.shape.gii`. The `hemi` tag is required for GiFTI files. For example:

```
pipeline/sub-001/
        anat/
                sub-001_hemi-L_curv.shape.gii
                sub-001_hemi-R_curv.shape.gii
```

The preferred extension for scalar CiFTI files is `.dscalar.nii`. For example:

```
pipeline/sub-001/
        anat/
```

```
            sub-001_curv.dscalar.nii
```

The file suffix should concisely describe the parameter that is represented in the overlay, and while the suffix can be individually customized, the following values should be reserved for their common use-cases:

| **Suffix** | **Description** |
|---|---|
| `curv` | Cortical surface curvature indices |
| `thickness` | Cortical thickness |
| `area` | Discretized surface area across regions |
| `dist` | Distance from a point |
| `defects` | Marked regions with surface defects |
| `sulc` | Sulcal depth |
| `myelinmap` | Myelin map calculated from T1w to t2 ratio |
| `distortion` | Distortion map calculated from a surface registration |

## Surface Parcellations

Discrete parcellations (surface segmentations) of cortical structures should also be stored as GiFTI or CiFTI files, and like for volumetric segmentations, parcellations should inherit the appropriate label-names from top-level or sidecar `labels.json` files.

Template:

```
<pipeline_name>/
      sub-<participant_label>/
            func|anat|dwi/
                  <source_file>[_hemi-{L|R}][_space-<space>]_dparc.<ext>
```

The preferred extension for GiFTI parcellations is `.label.gii`. The `hemi` tag is required for GiFTI files. For example:

```
pipeline/sub-001/
      anat/
            sub-001_hemi-L_dparc.label.gii
            sub-001_hemi-R_dparc.label.gii
```

The preferred extension for CiFTI parcellations is `.dlabel.nii`. For example:

```
pipeline/sub-001/
```

```
	anat/
		sub-001_dparc.dlabel.nii
		sub-001_dparc.dlabel.nii
```

## Statistics

Structural statistics produced by segmentation routines should be stored within tsv files, which could contain common parameters specified in the table below.

Template:

```
<pipeline_name>/
	sub-<participant_label>/
		func|anat|dwi/
			<source_file>[_desc-<label>]_stats.tsv
```

Example:

```
pipeline/sub-001/
	anat/
		sub-001_desc-volumetric_stats.tsv
```

| **Column name** | **Description** |
|---|---|
| `index` | RECOMMENDED Label integer index |
| `name` | RECOMMENDED Structure name |
| `centroid` | Center coordinate of structure |
| `volume` | Volume of structure |
| `intensity` | Intensity of voxels within structure |
| `thickness` | Thickness of cortical structure |
| `area` | Surface area of cortical structure |
| `curv` | Curvature index of cortical structure |

Some parameters might require unit specification or have multiple associated statistics (such as `avg`, `std`, `min`, `max`, `range`). The suggested syntax for this is `<parameter>[-<stat>][-<units>]`. An example volumetric stats file might look something like this:

```
index	name			volume-mm3	intensity-avg		intensity-std
11	Brainstem		23415.9		80.11			3.40
32	Left-Hippocampus	5349.7		75.23			2.27
32	Right-Hippocampus	4112.1		76.98			4.01
```