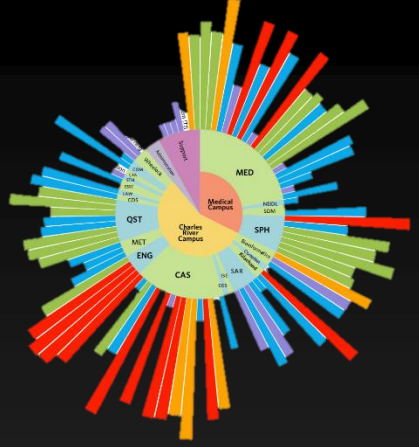


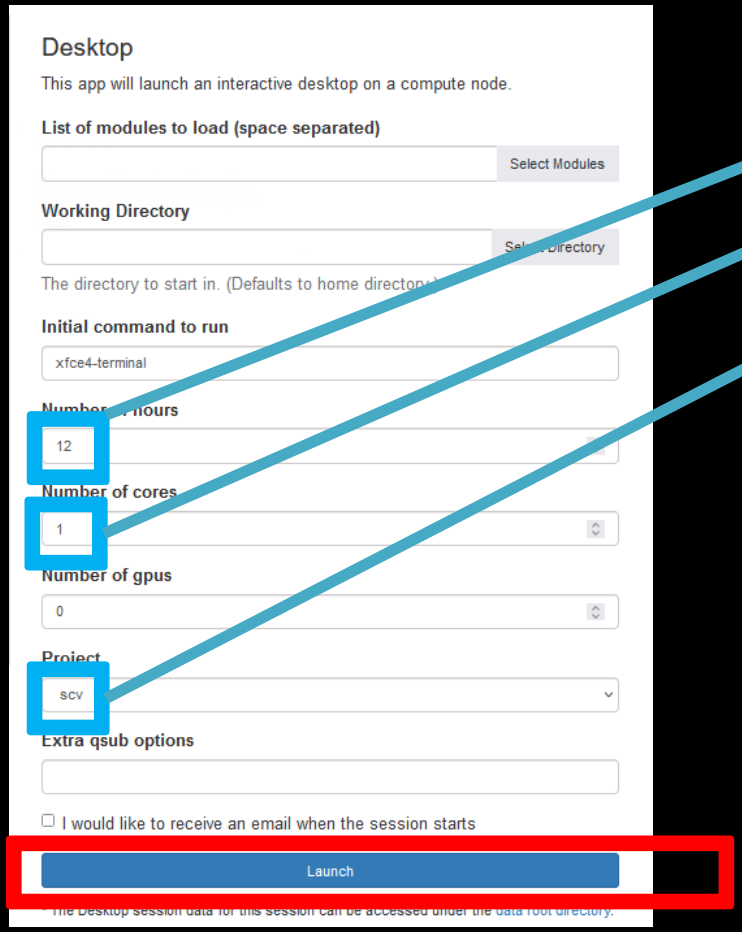
Data Management

Mitchell Horn
mhorn@bu.edu
Research Computing Services
Information Services & Technology
Boston University

BOSTON UNIVERSITY

Open an OnDemand session

1. Go to: scc-ondemand.bu.edu
2. Interactive Apps > Desktop



12 hours

1 core

project-ID

Outline

- Why
- What
- Neuroimaging Example

Why

- Organize complex data
- Provide comprehensive analysis
- Facilitate validating datasets and curation
- Data sharing

Why

```
├── code/
│   ├── code_final/
│   │   ├── final_2/
│   │   │   ├── main_script_fixed.py
│   │   │   └──takethisscriptformostthingsnow.py
│   │   ├── utils_new.py
│   │   ├── main_script.py
│   │   ├── utils_new.py
│   │   ├── utils_2.py
│   │   └── main_analysis_newparameters.py
│   └── main_script_DONTUSE.py
├── data/
│   ├── data_updated/
│   │   └── dataset1/
│   │       └── datafile_a
│   ├── dataset1/
│   │    └── datafile_a
│   ├── outputs/
│   │   ├── figures/
│   │   │   ├── figures_new.py
│   │   │   └── figures_final_forreal.py
│   │   ├── important_results/
│   │   ├── random_results_file.tsv
│   │   ├── results_for_paper/
│   │   ├── results_for_paper_revised/
│   │   └── results_new_data/
│   ├── random_results_file.tsv
│   ├── random_results_file_v2.tsv

[...]
```

BOSTON UNIVERSITY

Why

- Scripts will stop working due to incorrect pointers
- Paths no longer exist
- Misspellings in naming
- Non-comprehensive of actual analysis

What

- Structure study elements (modular)
- Record where you got it from, and where it is now
- Record what you did to it, and with what

What

- Datalad - https://www.datalad.org/

- DVC – https://dvc.org/

- BIDS - https://bids.neuroimaging.io/

What

- Datalad - https://www.datalad.org/

- DVC – https://dvc.org/

- BIDS - https://bids.neuroimaging.io/

BIDS

Brain Imaging Data Structure

- Developed to standardize enormous datasets
- A BIDS “specification” was created, many apps require it
- A requirement for data sharing and papers

BIDS

Brain Imaging Data Structure

- Developed to standardize enormous datasets
- A BIDS “specification” was created, many apps require it
- A requirement for data sharing and papers

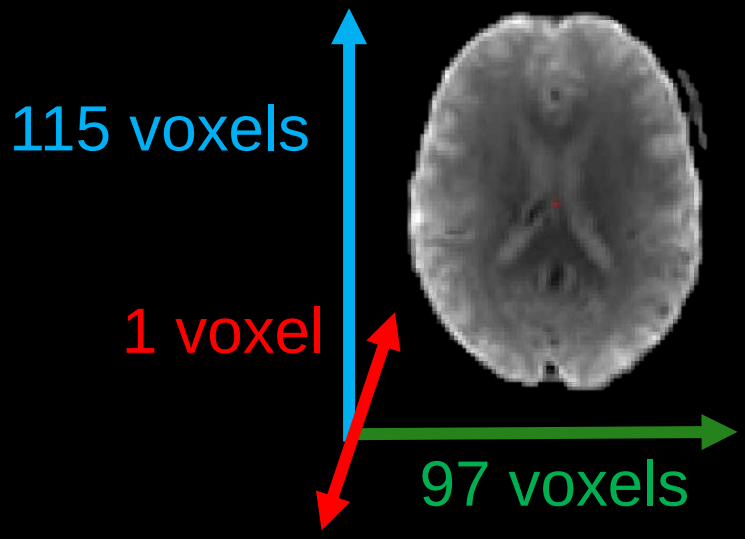
BIDS

MRI outputs data into **DICOMS**

- Digital Imaging and Communications in Medicine
- Standard filetype in Radiology since 1990
- File contains visual and text data

BIDS

Example **single** DICOM



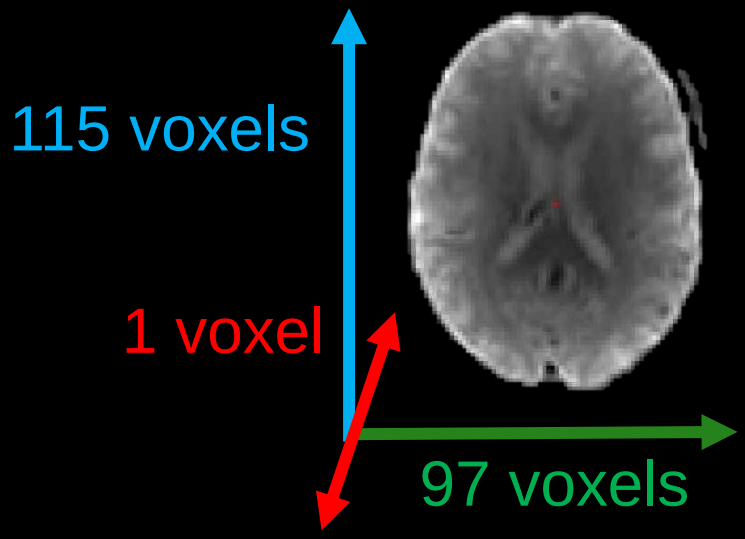
+

0008,0020 Study Date: 20220801
0008,0021 Series Date: 20220801
0008,0022 Acquisition Date: 20220801
0008,0023 Image Date: 20220801
0008,0030 Study Time: 133223.086000
0008,0031 Series Time: 140037.259000
0008,0032 Acquisition Time: 140332.002500
0008,0033 Image Time: 140333.266000
0008,0050 Accession Number:
0008,0060 Modality: MR
0008,0070 Manufacturer: SIEMENS
0008,0080 Institution Name: BU
0008,0081 Institution Address: Commonwealth Ave. 610,E
0008,0090 Referring Physician's Name:
0008,1010 Station Name: AWP166024-4CA74C
0008,1030 Study Description: Workshop
0008,103E Series Description: FacePlace
0008,1040 Institutional Department Name: Research
0008,1050 Attending Physician's Name:
0008,1070 Operator's Name: Shruthi, Stephanie
0008,1090 Manufacturer's Model Name: Prisma
0008,1140 Referenced Image Sequence:

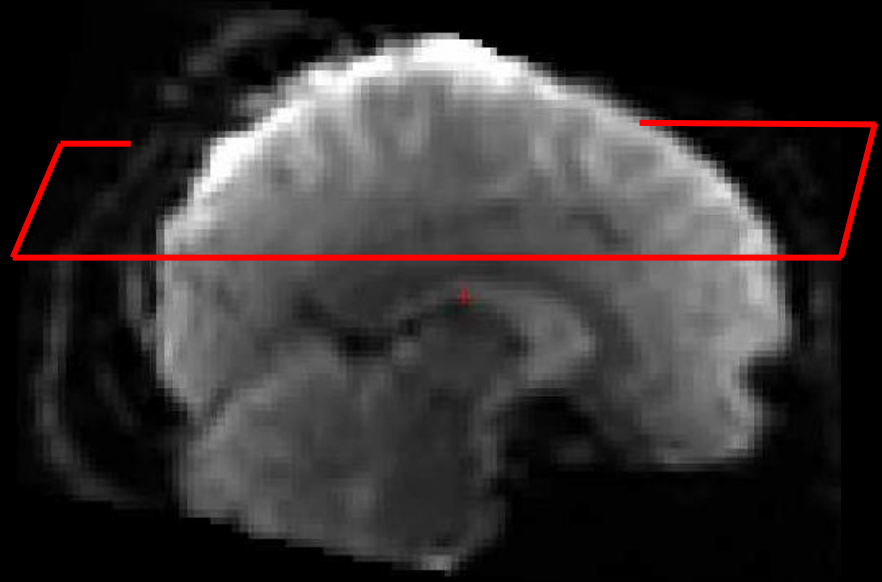
BOSTON UNIVERSITY

BIDS

Example **single** DICOM file



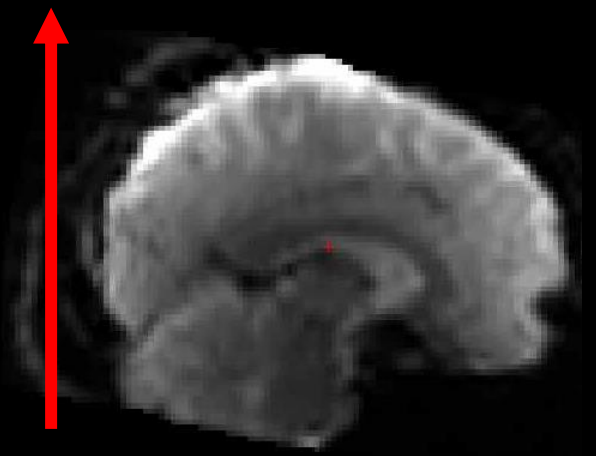
=



BIDS

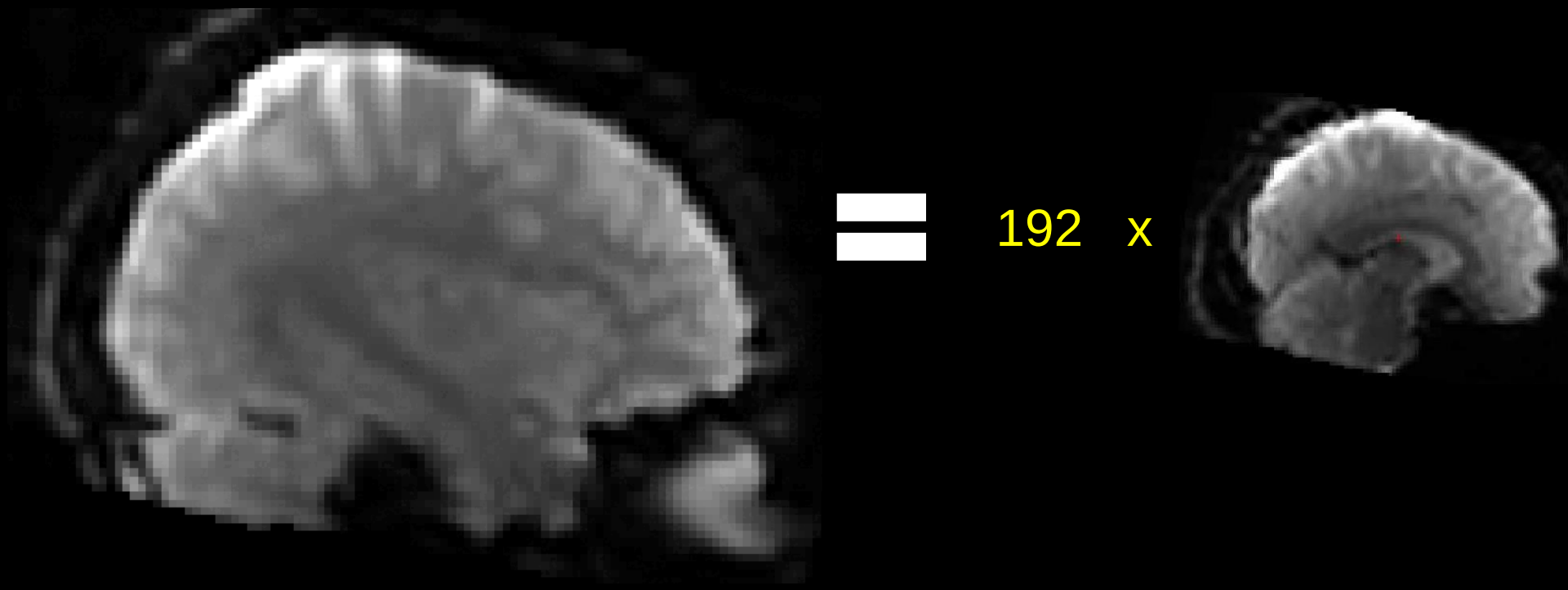
Example **single series of** DICOM files

= 97 slices



BIDS

Example **single scan of** DICOM files

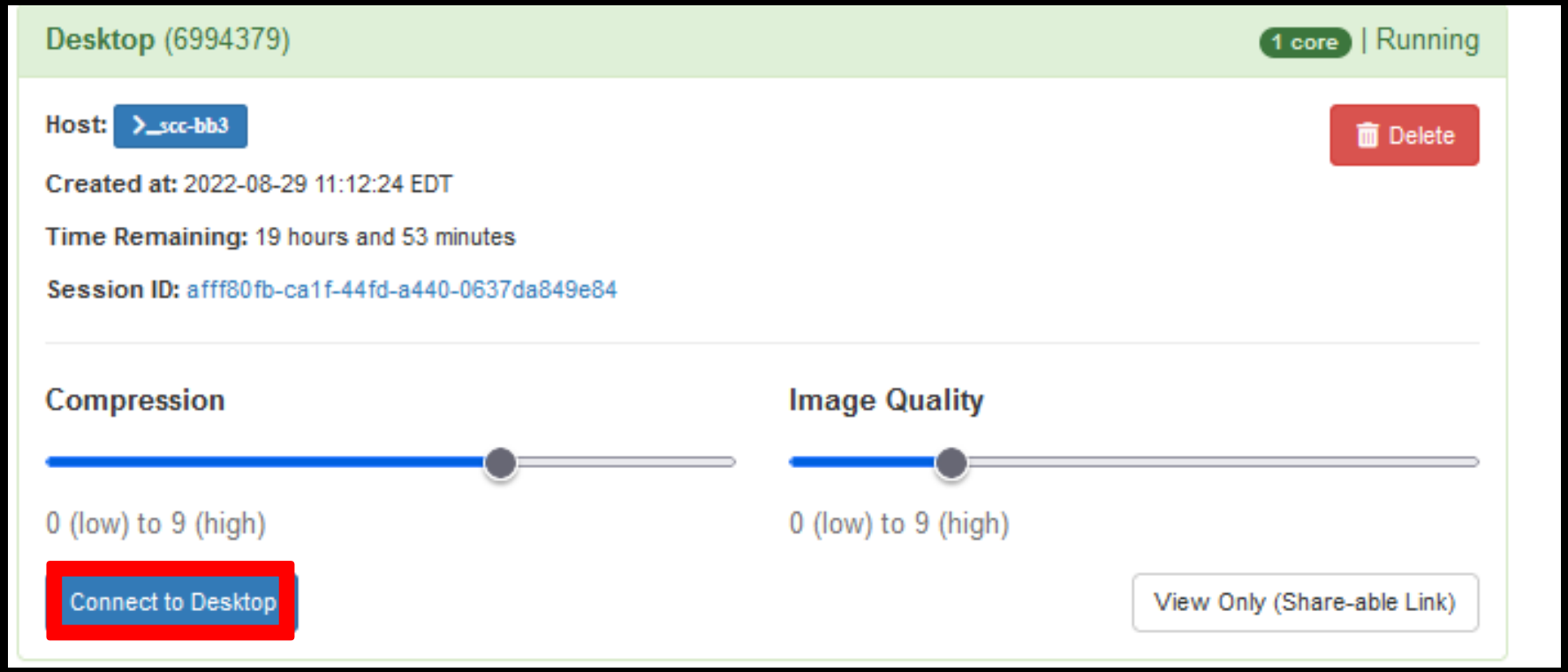


BIDS

Computational Neuroscientist developed new file format:

- NIFTI – Neuroimaging Informatics Technology Initiate (2003)
- maintains orientation information while compressing
- condenses 3D & 4D imaging types to single file

Open an OnDemand session



click Connect to Desktop!

BIDS

- Applications > Firefox > rcs.bu.edu/examples/imaging/
- data.notes

BIDS

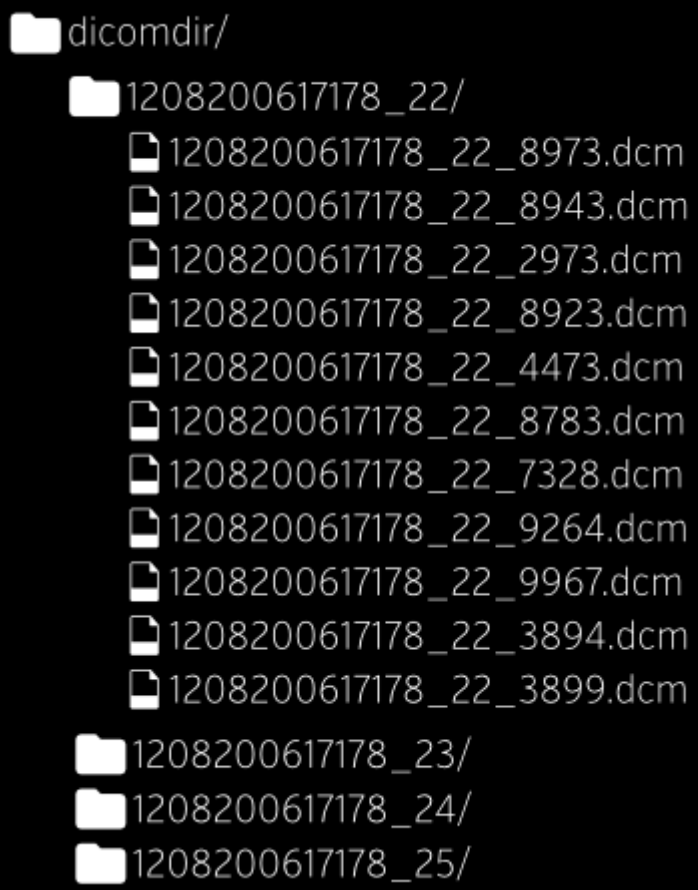
- Applications > Firefox > rcs.bu.edu/examples/imaging/
- data > data.notes
- run step 1: Copy the tutorial data to your local project

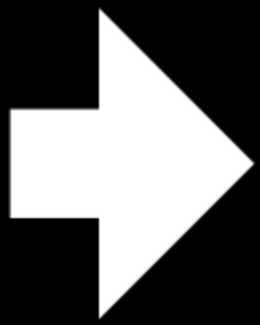
BIDS

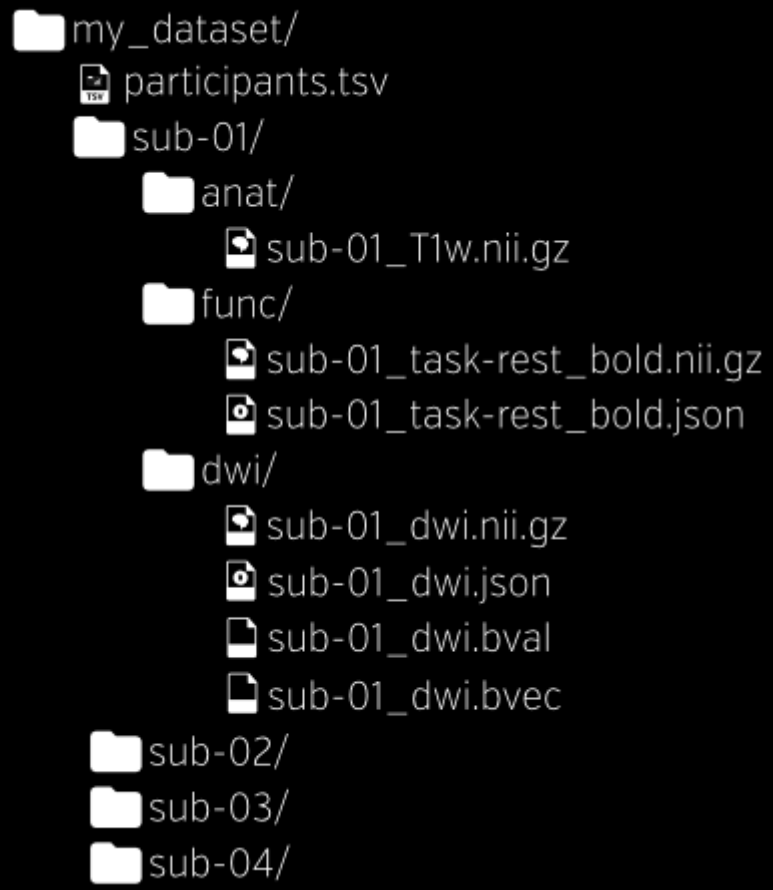
- run step 2: Convert DICOMS to NIFTI

- run step 3: Explore the converted data

BIDS







BOSTON UNIVERSITY

BIDS

Tools for creating the directory tree:

- heudiconv - https://heudiconv.readthedocs.io/en/latest/
- clpipe - https://clpipe.readthedocs.io/en/latest/bids_convert.html
- yaxil - https://yaxil.readthedocs.io/en/latest/arcget.html
- many other open-source packages!

BIDS

Tools for creating the directory tree:

- heudiconv - https://heudiconv.readthedocs.io/en/latest/
- clpipe - https://clpipe.readthedocs.io/en/latest/bids_convert.html
- yaxil - https://yaxil.readthedocs.io/en/latest/arcget.html
- many other open-source packages!

Yaxil

Tool to pull DICOMS from our MRI directly into BIDS

1. know your scans!

2. create a YAML file with descriptors

```yaml
anat:
    T1w:
        - scan: 6
          run: 1
    T2w:
        - scan: 7
          run: 1
func:
    bold:
        - scan: 11
          task: LANG
          run: 1
```

3. run yaxil

```
[~]$ ArcGet.py  -a  xnat  -l  220801_Rise_demo_01  -p  Workshop  -o .
```

BOSTON UNIVERSITY

Yaxil

Requires access to our (BU CNC) MRI data…

…so I did it for you!

- run step 4: Explore BIDS directory structure

BIDS

- Brain Imaging Data Structure

What?

- Three main file types in a BIDS dataset:
  - .JSON – contain metadata as key:value pairs
  - .TSV – contain tables of metadata
  - .NII.GZ – raw data files for fMRI and MRI data

BIDS

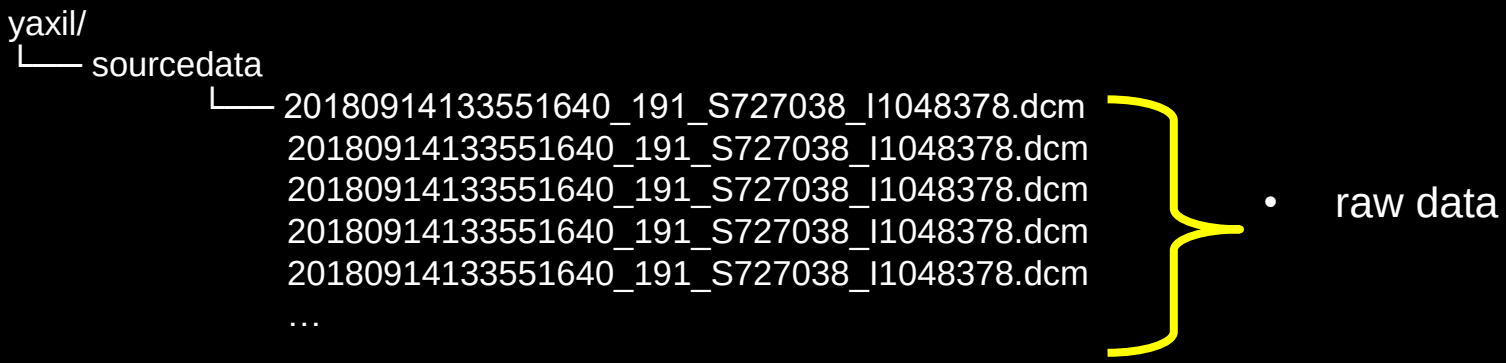
yaxil/ →

- top level directory
- contains the entire imaging study

BIDS

- level-1 subdirectory
- contains source imaging (.dcm, unorganized nifti, etc.)

BIDS

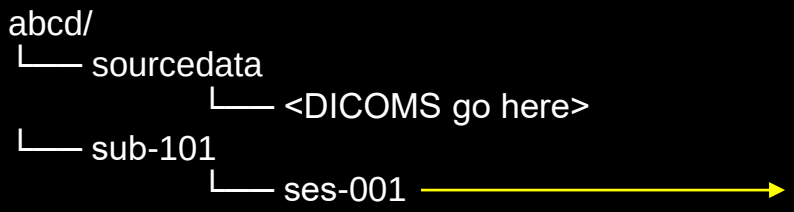


- raw data

BIDS

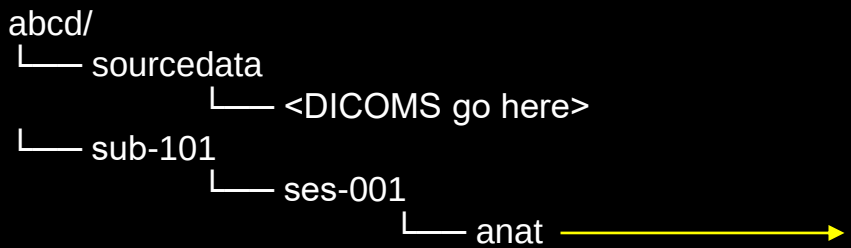
- level-1 subdirectory
- contains all imaging data in BIDS format

BIDS



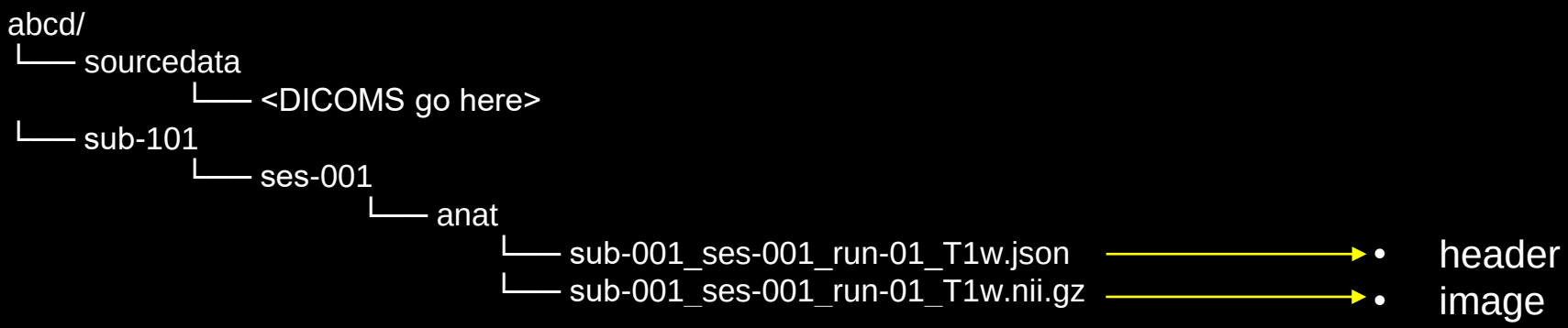
- level-2 subdirectory
- contains all BIDS imaging data from timepoint 001

BIDS

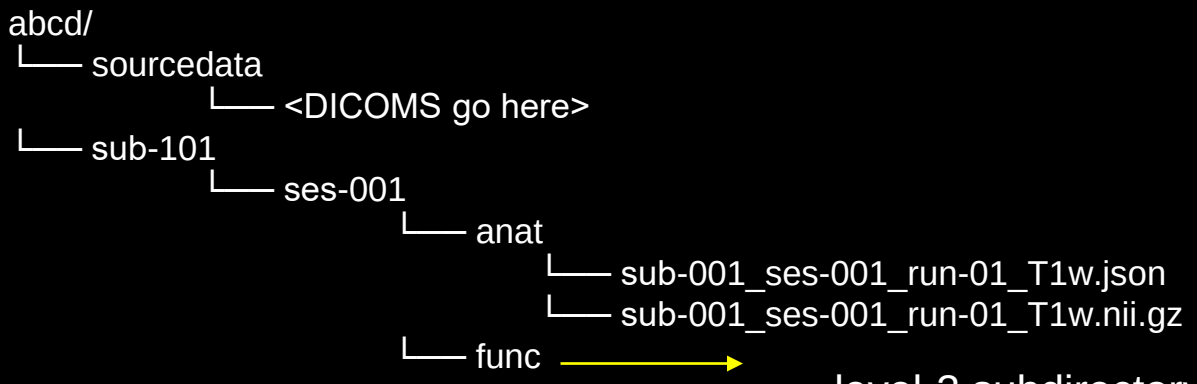


- level-3 subdirectory
- contains only raw anatomical imaging (T1w, T2w)

BIDS

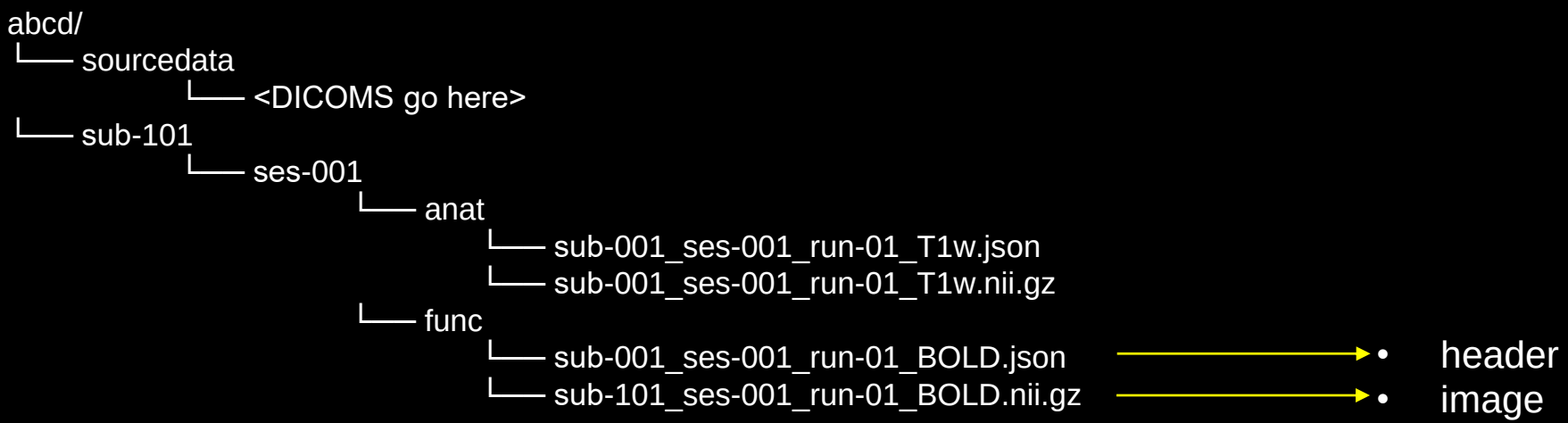

abcd/
sourcedata
<DICOMS go here>
sub-101
ses-001
anat
sub-001_ses-001_run-01_T1w.json
sub-001_ses-001_run-01_T1w.nii.gz
header
image

BIDS



- level-3 subdirectory
- contains only raw functional imaging (BOLD)

BIDS


abcd/
sourcedata
<DICOMS go here>
sub-101
ses-001
anat
sub-001_ses-001_run-01_T1w.json
sub-001_ses-001_run-01_T1w.nii.gz
func
sub-001_ses-001_run-01_BOLD.json
sub-101_ses-001_run-01_BOLD.nii.gz
header
image

BIDS

```
abcd/
└── sourcedata
        └── <DICOMS go here>
└── sub-101
        └── ses-001
                └── anat
                        └── sub-001_ses-001_run-01_T1w.json
                        └── sub-001_ses-001_run-01_T1w.nii.gz
                └── func
                        └── sub-001_ses-001_run-01_BOLD.json
                        └── sub-101_ses-001_run-01_BOLD.nii.gz
```

README	- text file describing the nature of your study
dataset_description.json	- information about the BIDS version, authors, licensing, etc.
participants.tsv	- describes subject characteristics like age, sex, handedness etc.

BIDS Compliancy

- For BIDS.apps to successfully recognize and import data, you can verify organization
- Not always necessary
- http://bids-standard.github.io/bids-validator/

BIDS Apps

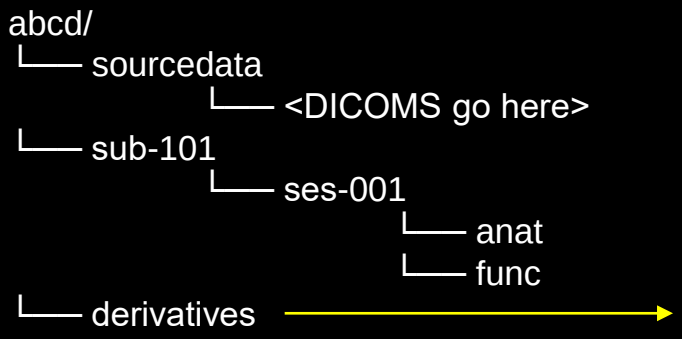
- https://bids-apps.neuroimaging.io/apps/

fMRIPrep

- Workflow taking principal input images and generates a standardized preprocessed output for analysis
- Uses BIDS standard for input and output
- Customize preprocessing via one command line with arguments
- Generate visual outputs for basic QC analysis

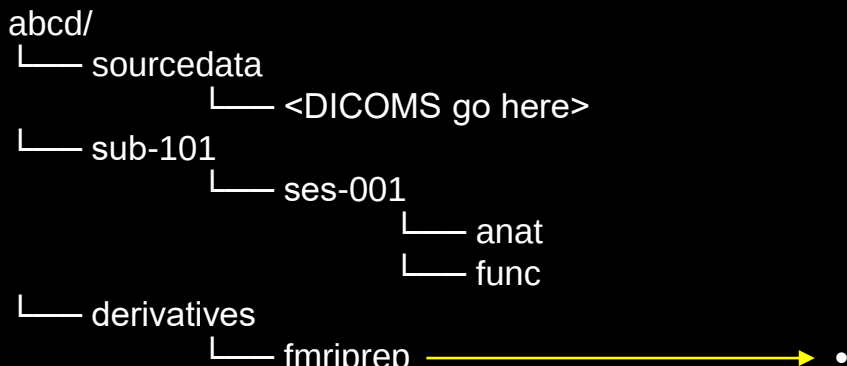
BOSTON UNIVERSITY

BIDS



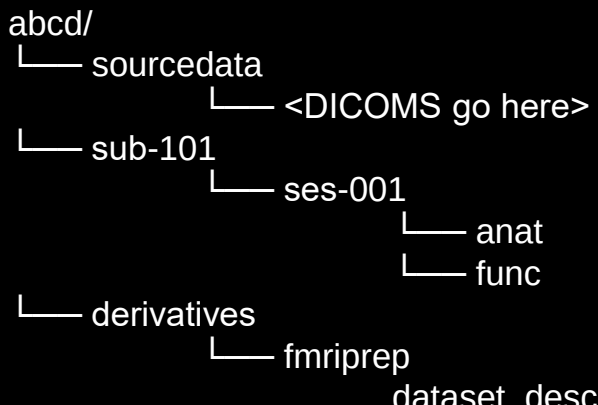
- level-1 subdirectory
- BIDS.apps will store output here
- contains only derived imaging data

BIDS



- level-2 subdirectory
- fMRIPREP will store output here

BIDS



- JSON file
- Information about dataset
  - fmriprep version, authors, funding,

BIDS

```
abcd/
└── sourcedata
        └── <DICOMS go here>
└── sub-101
        └── ses-001
                └── anat
                └── func
└── derivatives
        └── fmriprep
                dataset_description.json
                sub-101.html
```

- .HTML file
- easy-view fMRIPREP report of sub-101

BIDS

```
abcd/
└── sourcedata
        └── <DICOMS go here>
└── sub-101
        └── ses-001
                └── anat
                └── func
└── derivatives
        └── fmriprep
                dataset_description.json
                sub-101.html
                └── sub-101
```

- level-3 subdirectrory
- fMRIPREP output for sub-101

BIDS

```
abcd/
└── sourcedata
        └── <DICOMS go here>
└── sub-101
        └── ses-001
                └── anat
                └── func
└── derivatives
        └── fmriprep
                dataset_description.json
                sub-101.html
                └── sub-101
                        └── anat
```

- level-4 subdirectory
- fMRIPREP all anatomical output for sub-101

BIDS

```
abcd/
└── sourcedata
        └── <DICOMS go here>
└── sub-101
        └── ses-001
                └── anat
                └── func
└── derivatives
        └── fmriprep
                dataset_description.json
                sub-101.html
                └── sub-101
                        └── anat
                        └── figures
```

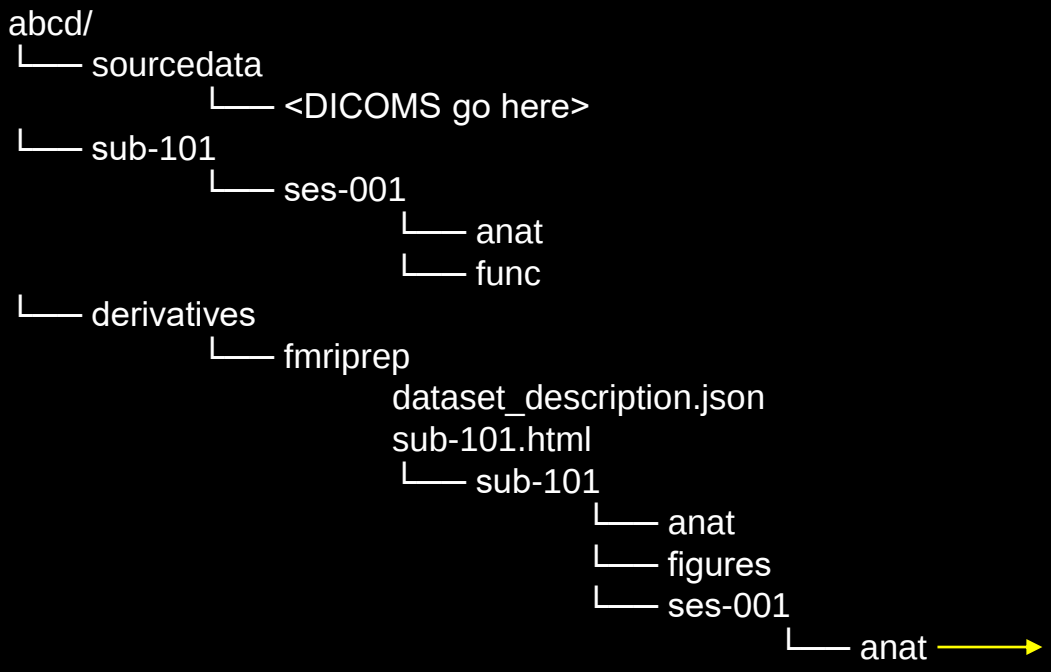
- level-4 subdirectory
- fMRIPREP report images for sub-101

BIDS

```
abcd/
└── sourcedata
        └── <DICOMS go here>
└── sub-101
        └── ses-001
                └── anat
                └── func
└── derivatives
        └── fmriprep
                dataset_description.json
                sub-101.html
                └── sub-101
                        └── anat
                        └── figures
                        └── ses-001
```

- level-4 subdirectory
- fMRIPREP out images for timepoint-1

BIDS



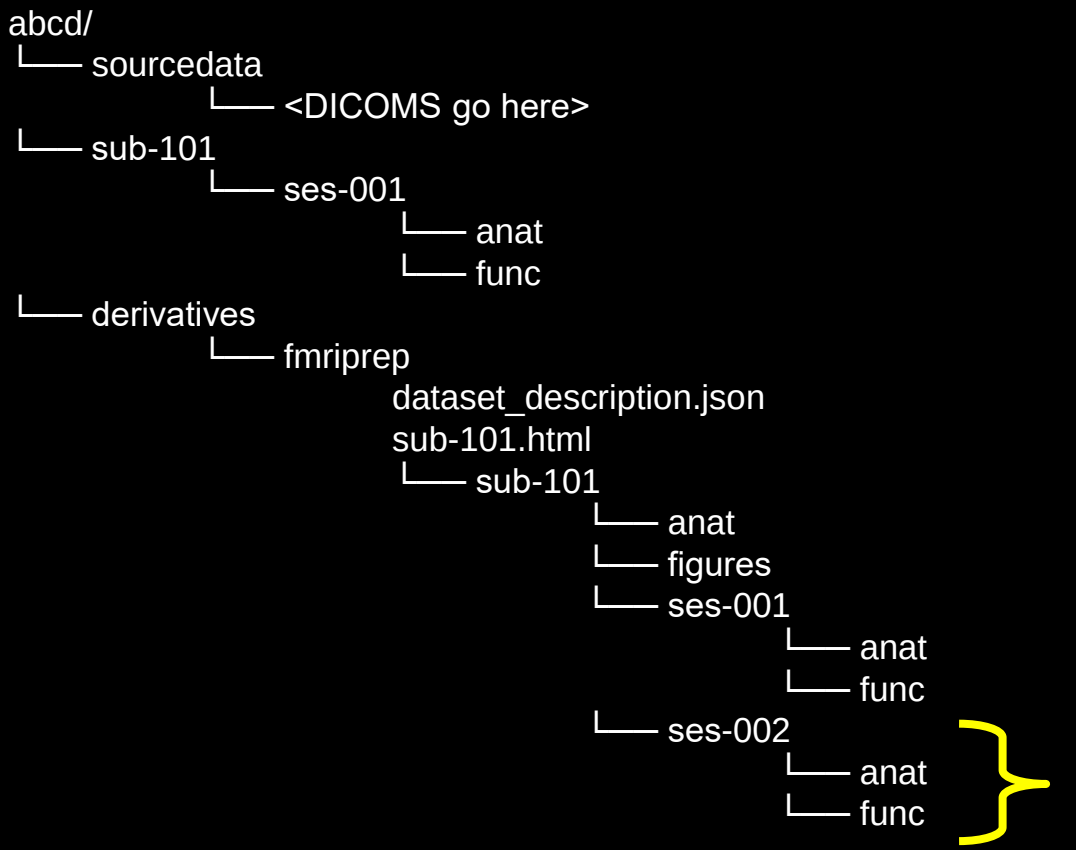
- level-5 subdirectory
- fMRIPREP anatomical images for timepoint-1

BIDS

```
abcd/
└── sourcedata
        └── <DICOMS go here>
└── sub-101
        └── ses-001
                └── anat
                └── func
└── derivatives
        └── fmriprep
                dataset_description.json
                sub-101.html
                └── sub-101
                        └── anat
                        └── figures
                        └── ses-001
                                └── anat
                                └── func
```

- level-5 subdirectory
- fMRIPREP functional images for timepoint-1

BIDS



- repeat for each timepoint