STANDARDIZED IMAGING PIPELINES: BIDS, FMRIPREP, MRIQC, AND OTHER TOOLS

Björn Schiffler
Post-doc in Karin Jensen's group

OUTLINE

1. The Brain Imaging Data Structure (BIDS) format
2. Getting your raw Dicom data to BIDS structured Nifti files
3. The "container" philosophy of reproducibility (Docker/Singularity)
4. Quality control of raw data with mriqc
5. Preprocessing neuroimaging data with fmriprep

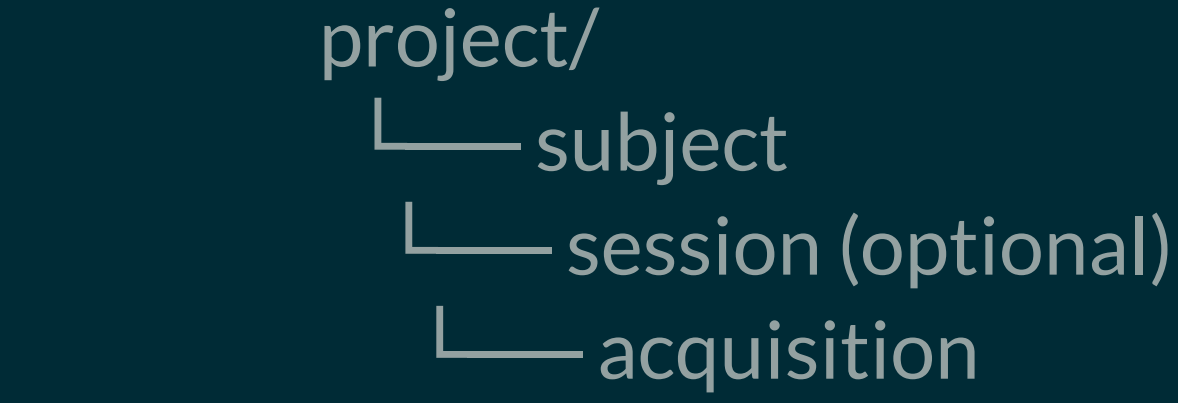
WHY STANDARDIZE YOUR DATA STORING?

App name	Description	Applicable modalities	References
example	Example App that also serves as a template for new apps. Calculates intracranial volume.	T1w	n/a
Freesurfer	Surface extraction, longitudinal pipeline and study specific template calculation using FreeSurfer.	T1w	[19,20]
ndmg	One-click reliable and reproducible pipeline for T1w + DWI weighted MRI connectome estimation.	T1w, DWI	[21,22]
BROCCOLI	Fast fMRI analysis on many-core CPUs and GPUs.	T1w, fMRI	[23]
FibreDensityAndCrosssection	Fixel-Based Analysis (FBA) of Fibre Density and Fibre Cross-section.	DWI	[24,25]
SPM	Statistical Parametric Mapping.	T1w, fMRI	[26]
MRIQC	Quality Assessment of structural and functional MRI.	T1w, fMRI	[52]
FMRIPREP	A generic fMRI preprocessing pipeline providing results robust to the input data quality as well as informative reports.	T1w, fMRI	In preparation
Quality Assessment Protocol	Quality Assessment of structural and functional MRI.	T1w, fMRI	[27]
Configurable Pipeline for the Analysis of Connectomes	Pipeline for high throughput processing and analysis of structural and functional MRI data.	T1w, fMRI	[28]
Hyperalignment	Computes hyperalignment transformations for functional alignment.	fMRI	[29]
mindboggle	Pipeline to improve the accuracy, precision, and consistency of automated labeling and shape analysis of human brain image data.	T1w	[30]
MRtrix3 connectome	Robust generation and statistical analysis of structural connectomes estimated from diffusion tractography.	T1w, DWI	[31]
nilearn	Extraction of time-series and connectomes for population analysis.	fMRI	[32]
nipypelines	Preprocessing of functional time series for resting or task analysis	T1w, fMRI	[16]
automatic analysis (aa)	Neuroimaging pipeline system written in Matlab.	T1w, T2w, fMRI, DWI	[33]
Niak Preprocessing	Noise reduction, segmentation, coregistration, motion estimation, resampling.	fMRI	[34]
HPCPipelines	Anatomical and functional preprocessing pipelines used in the Human Connectome Project.	T1w, T2w, fMRI	[35,36]
BrainIAK-SRM	Functional alignment using Shared Response Model implementation from the Brain Imaging Analysis Kit.	fMRI	[37]
OPPNI	Optimization of Preprocessing Pipelines for NeuroImaging, for analysis of fMRI data.	fMRI	[38–41]
MAGeTbrain	Multiple Automatically Generated Templates brain segmentation algorithm	T1w	[42–45]
tracula	Automatic reconstruction of a set of major white-matter pathways from diffusion-weighted MR images	T1w, DWI	[46]

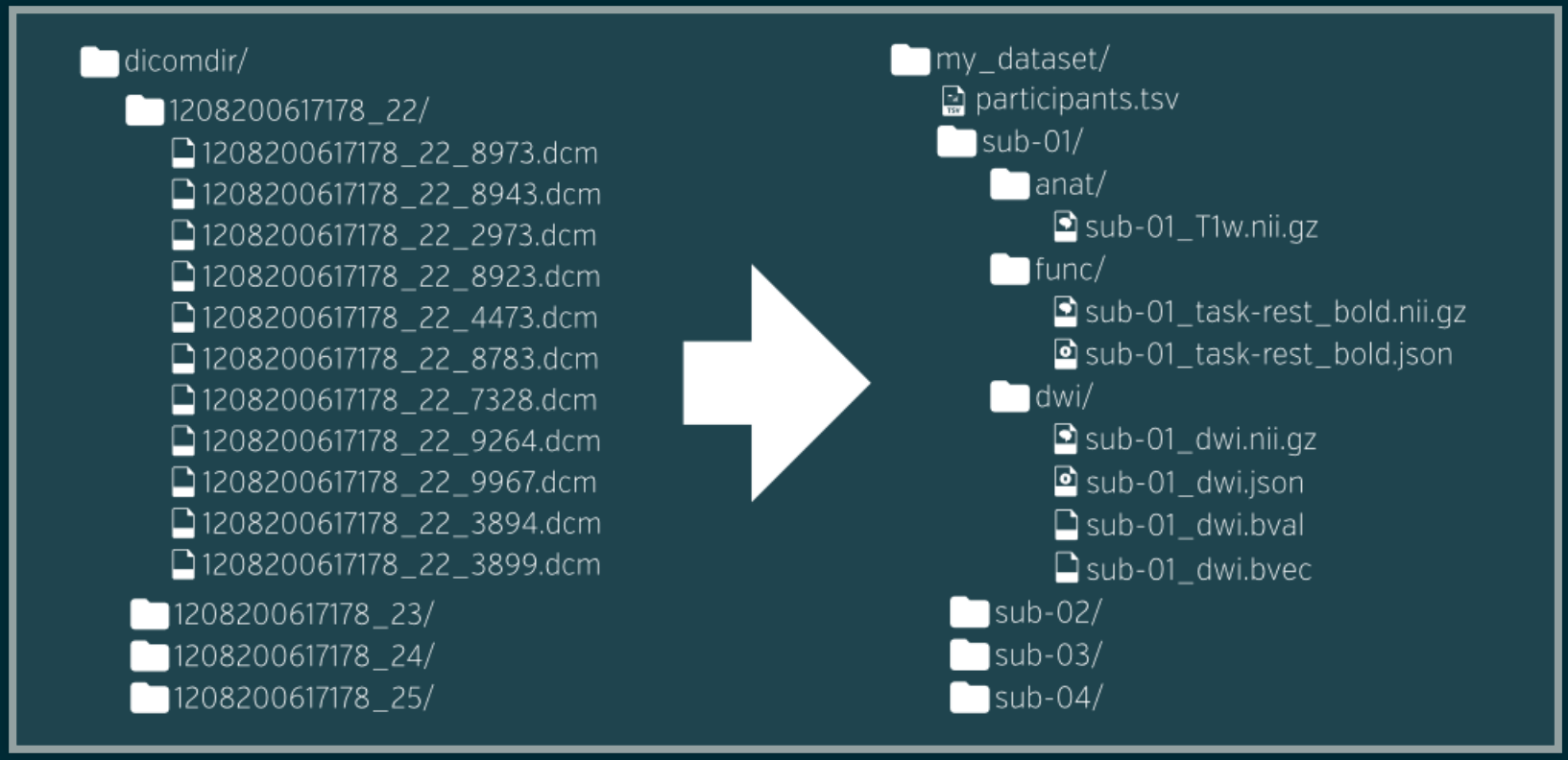
https://doi.org/10.1371/journal.pcbi.1005209.t001

Source: Gorgolewski et al., 2016, Scientific Data

THE BRAIN IMAGING DATA STRUCTURE (BIDS) FORMAT



THE BRAIN IMAGING DATA STRUCTURE (BIDS) FORMAT



Dicom header information gets stored in .json files

BIDS specification PDF

HOW DO I GET MY DATA INTO BIDS FORMAT?

- Various tools available which help with that, e.g. *heudiconv* (**heu**ristic **di**com **con**verter) or *dcm2bids*
- Conversion from .dcm to .nii on the fly usually included
- Worst case, getting data into BIDS format can also be done manually
- Check if correctly BIDS formatted via BIDS validation tools

HEUDICONV

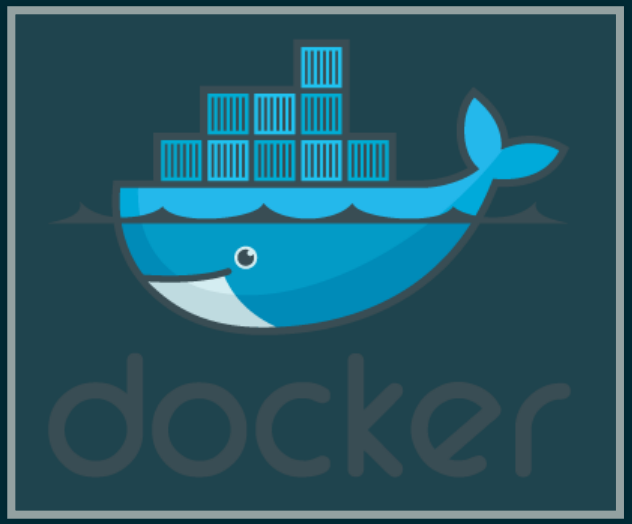
- Heudiconv uses heuristics to find the dicoms corresponding to anatomical, functional, diffusion-weighted etc. sequences
- These heuristics need to be adapted to each study
- E.g. "fmri_1" could be an fMRI session which has "fMRI" in the description and consists of 13650 dicoms
- "fmri_2" could have "fMRI" in the description and consist of 16044 dicoms
- Use what you know about your sequences to build a heuristic!

EXAMPLE HEURISTIC FOR HEUDICONV

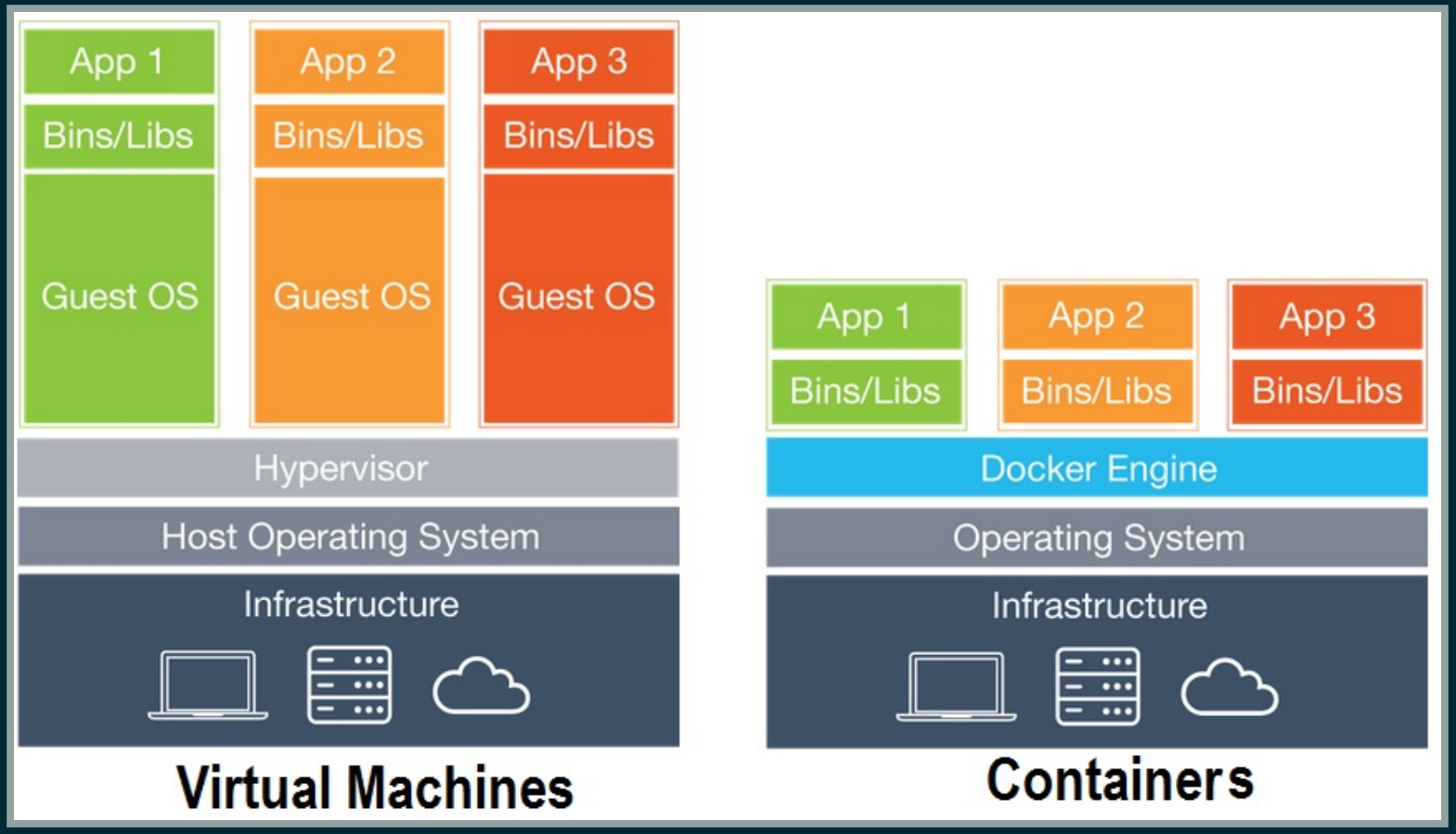
```
t1w = create_key('sub-{subject}/anat/sub-{subject}_T1w')
dwi = create_key('sub-{subject}/dwi/sub-{subject}_run-{item:01d}_
pressure1 = create_key('sub-{subject}/func/sub-{subject}_task-pre
pressure2 = create_key('sub-{subject}/func/sub-{subject}_task-pre

info = {t1w: [], dwi: [], rest: []}

for s in seqinfo:
    if (s.dim3 == 176) and ('t1' in s.series_description or 'T1'
        info[t1w] = [s.series_id]  # assign if a single series me
    if (s.dim3 == 3900) and (s.dim4 == 1) and ('dti' in s.series_
        info[dwi].append(s.series_id)  # append if multiple serie
    if (s.dim3 == 13650) and ('fmri_1' in s.series_description):
        if s.is_motion_corrected:  # exclude non motion corrected
            info[pressure1].append({'item': s.series_id, 'rec': '
        else:
```

DOCKER CONTAINERS

- Container softwares such as Docker bundle all relevant software for processing
- Similar to Virtual Machines, but still rely on some OS subprocesses
- Having exact version numbers of bundled software allows for reproducibility



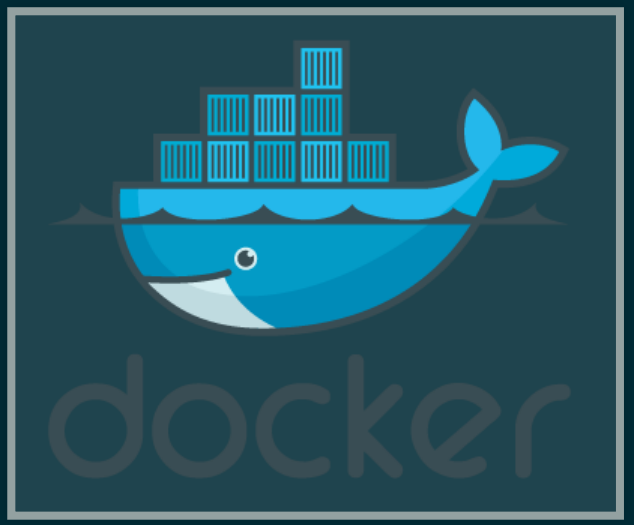
CONTAINERS VS. VIRTUAL MACHINES



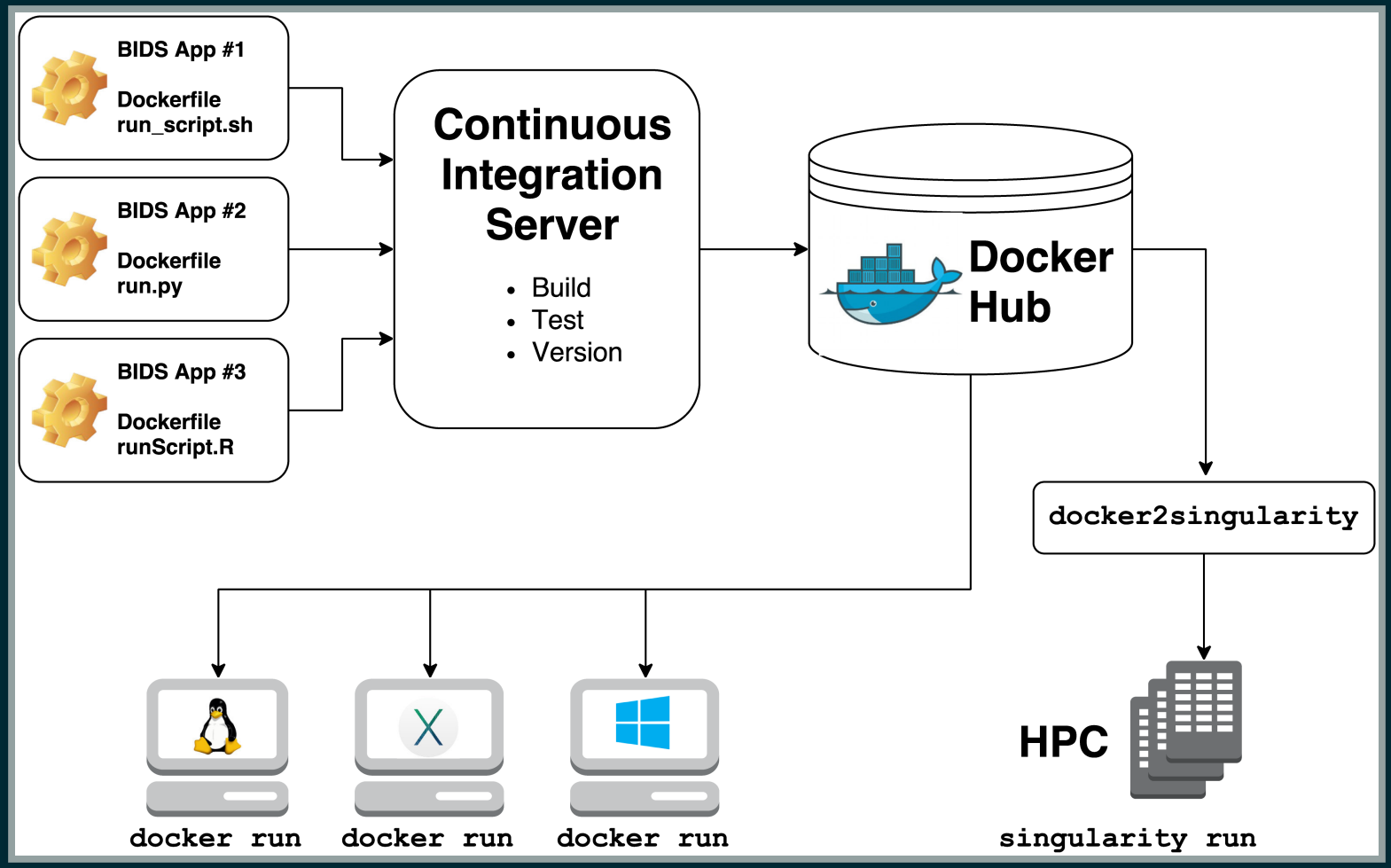
Source: Infoworld

DOCKER CONTAINERS

- Example: *fmriprep* bundles e.g. Freesurfer, FSL, AFNI, ANTs, among other tools
- You can make your own container when you have finished an analysis to archive everything needed to reproduce, e.g. package versions



CONTAINERS



Source: Gorgolewski et al., 2017, PLoS Computational Biology

SINGULARITY CONTAINERS

- Docker usually needs certain admin rights
- Singularity has been developed as a high performance cluster implementation for container based processing
- Singularity is installed on Akalla

SINGULARITY WORKFLOW

1. Obtain the relevant container in the version that you want
2. Either directly from Singularity Hub (no conversion necessary) or from Docker Hub (docker2singularity conversion necessary)
3. Conversion with docker2singularity requires a local installation of Docker
4. Run the container

CONTAINER MAINTENANCE

- Containers can be archived in a central location on the server so that everyone has access to all of them
- Containers can be quite large (e.g. >10 GB per container for fmriprep), so deleting old/unused containers is important

QUALITY CONTROL OF IMAGING DATA WITH MRIQC

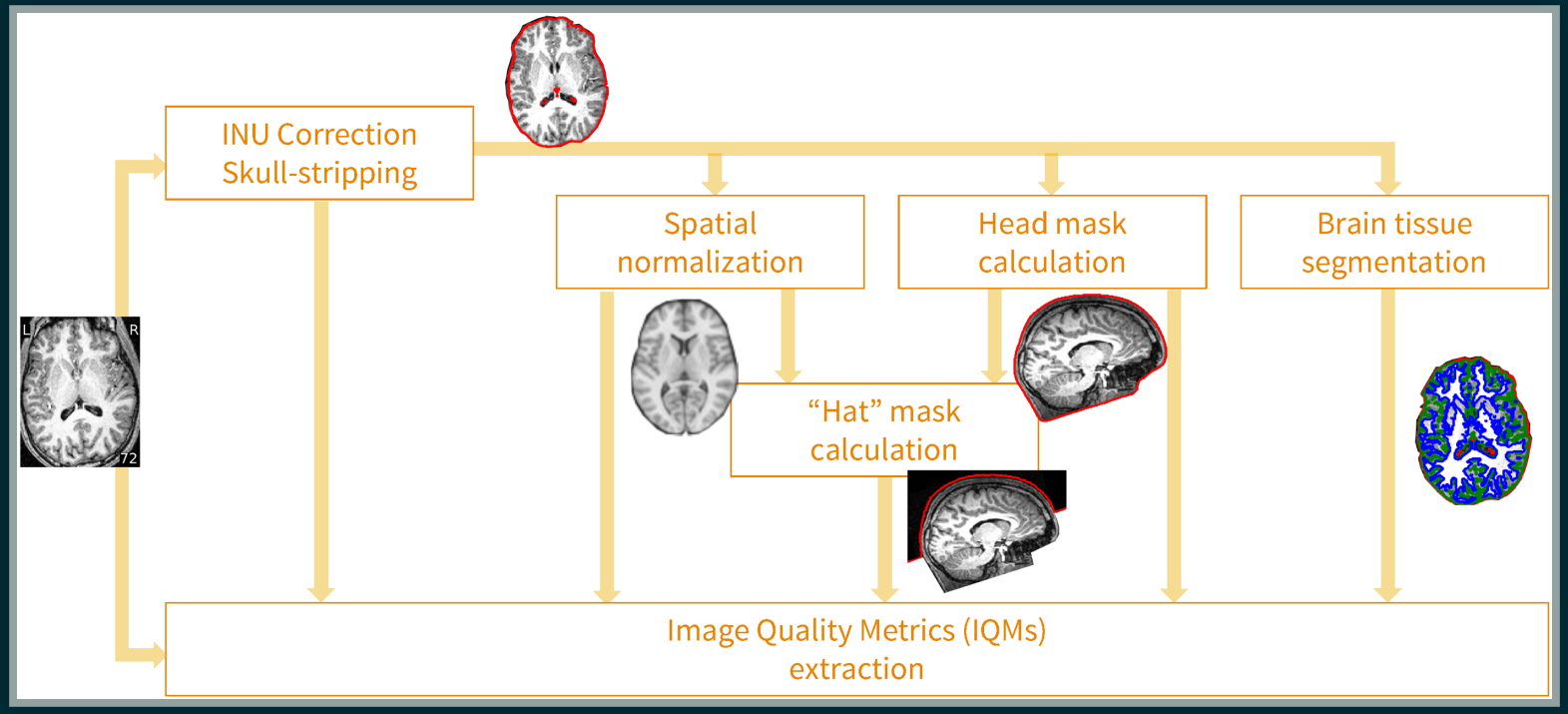
- How do you currently check the quality of your raw data? Visual inspection?
- mriqc is a BIDS tool to compare data quality across your dataset and helps to spot systematic errors during data acquisition

QUALITY CONTROL OF IMAGING DATA WITH MRIQC

With data in BIDS format, running mriqc is simple:

```
singularity run mriqc_container input_folder output_folder \
participant --no-sub -w work_folder --verbose-reports
```

QUALITY CONTROL OF IMAGING DATA WITH MRIQC



Source: Esteban et al., 2017, PLoS ONE

QUALITY CONTROL OF IMAGING DATA WITH MRIQC

- Overview of image quality metrics (IQMs) such as Framewise Displacement (FD), DVARS, SNR etc. in your raw data
- Both for anatomical as well as functional MRI data

QUALITY CONTROL OF IMAGING DATA WITH MRIQC

Measures based on noise measurements	
CJV	The coefficient of joint variation of GM and WM was proposed as objective function by Ganzetti et al. [30] for the optimization of INU correction algorithms. Higher values are related to the presence of heavy head motion and large INU artifacts.
CNR	The contrast-to-noise ratio [31] is an extension of the SNR calculation to evaluate how separated the tissue distributions of GM and WM are. Higher values indicate better quality.
SNR	MRIQC includes the the signal-to-nose ratio calculation proposed by Dietrich et al. [32], using the air background as noise reference. Additionally, for images that have undergone some noise reduction processing, or the more complex noise realizations of current parallel acquisitions, a simplified calculation using the within tissue variance is also provided.
QI_2	The second quality index of [12] is a calculation of the goodness-of-fit of a χ^2 distribution on the air mask, once the artifactual intensities detected for computing the QI_1 index have been removed. The description of the QI_1 is found below.
Measures based on information theory	
EFC	The entropy-focus criterion [33] uses the Shannon entropy of voxel intensities as an indication of ghosting and blurring induced by head motion. Lower values are better.
FBER	The foreground-background energy ratio [14] is calculated as the mean energy of image values within the head relative the mean energy of image values in the air mask. Consequently, higher values are better.
Measures targeting specific artifacts	
INU	MRIQC measures the location and spread of the bias field extracted estimated by the inu correction. The smaller spreads located around 1.0 are better.
QI_1	The first quality index of [12] measures the amount of artifactual intensities in the air surrounding the head above the nasio-cerebellar axis. The smaller QI_1, the better.
WM2MAX	The white-matter to maximum intensity ratio is the median intensity within the WM mask over the 95% percentile of the full intensity distribution, that captures the existence of long tails due to hyper-intensity of the carotid vessels and fat. Values should be around the interval [0.6, 0.8].
Other measures	
FWHM	The full-width half-maximum [34] is an estimation of the blurriness of the image using AFNI's `3dFWHMx`. Smaller is better.
ICVs	Estimation of the icv of each tissue calculated on the FSL `FAST`'s segmentation. Normative values fall around 20%, 45% and 35% for cerebrospinal fluid (CSF), WM and GM, respectively.
rPVE	The residual partial volume effect feature is a tissue-wise sum of partial volumes that fall in the range [5%-95%] of the total volume of a pixel, computed on the partial volume maps generated by FSL `FAST`. Smaller residual partial volume effects (rPVEs) are better.
SSTATs	Several summary statistics (mean, standard deviation, percentiles 5% and 95%, and kurtosis) are computed within the following regions of interest: background, CSF, WM, and GM.
TPMs	Overlap of tissue probability maps estimated from the image and the corresponding maps from the ICBM nonlinear-asymmetric 2009c template [35].

https://doi.org/10.1371/journal.pone.0184661.t002

PREPROCESSING YOUR DATA WITH FMRIPREP

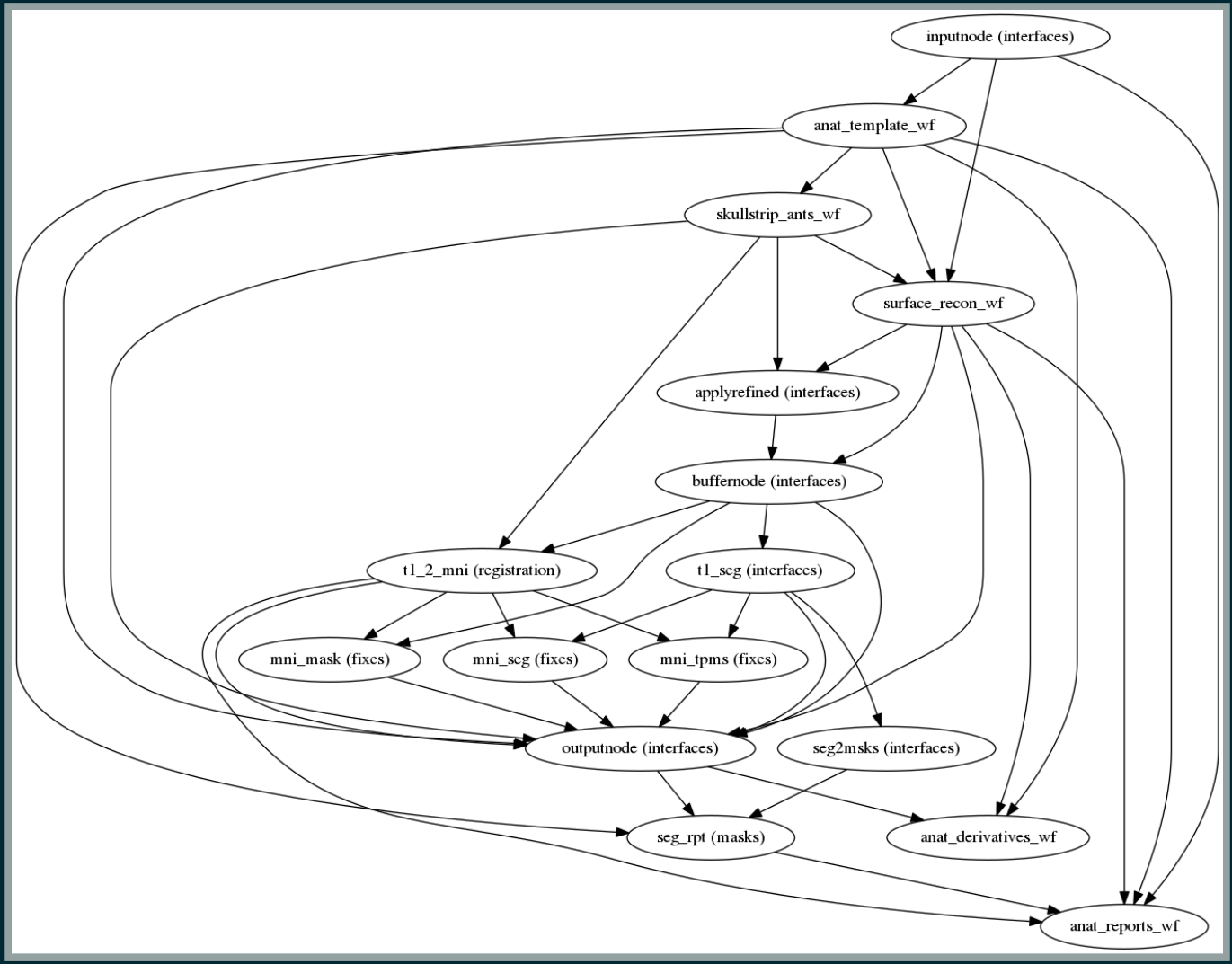
- Preprocessing pipeline for (f)MRI data
- Performs minimal preprocessing (skull stripping, motion correction, segmentation, coregistration, normalization etc.)
- No "controversial" steps such as smoothing
- Integration of Freesurfer for surface based processing (optional)
- Generates preprocessing quality reports in .html files

PREPROCESSING YOUR DATA WITH FMRIPREP

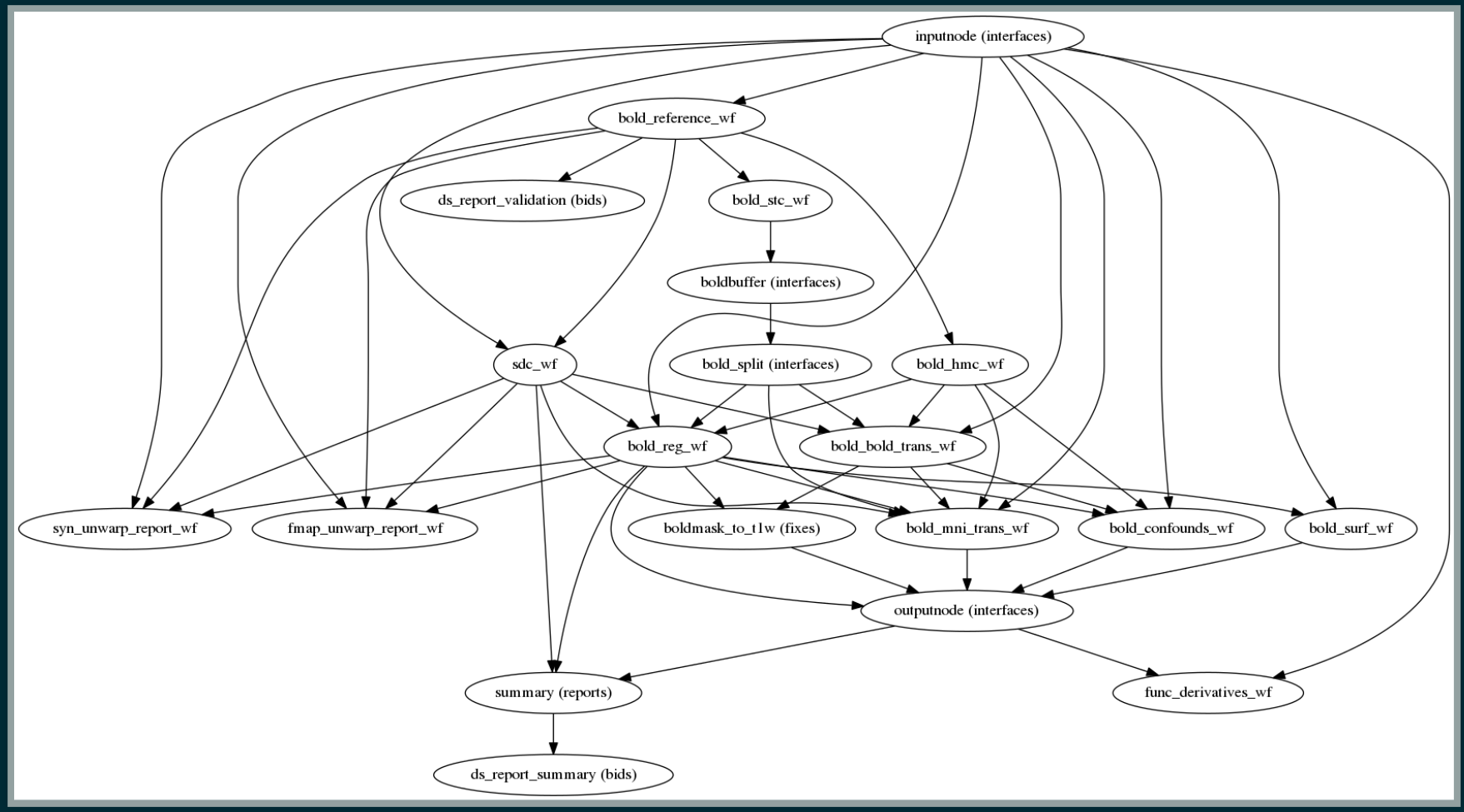
With data in BIDS format, running fmriprep is simple:

```
singularity run fmriprep_container input_folder \
output_folder participant -w work_folder
```

PREPROCESSING YOUR DATA WITH FMRIPREP - T1



PREPROCESSING YOUR DATA WITH FMRIPREP - FMRI



PREPROCESSING YOUR DATA WITH FMRIPREP

Example output

HIGH LEVEL SUMMARY

1. Adapt heudiconv heuristic.py to your study
2. .dcm to .nii in BIDS with heudiconv
3. Validate your BIDS folder structure
4. Quality control with mriqc
5. Preprocessing with fmriprep
6. Quality check the .html output

QUESTIONS?

- Next: Going through a step-by-step tutorial