

# fMRIPrep-next: Preprocessing as a fit-transform model

Christopher J Markiewicz<sup>1</sup>, Mathias Goncalves<sup>1</sup>, John Kruper<sup>2</sup>, Ma Feilong<sup>3</sup>, Guarav H Patel<sup>4</sup>, Juan Sanchez-Pena<sup>4</sup>, Robert E Smith<sup>5</sup>, Lea Waller<sup>6</sup>, Joseph Wexler<sup>1</sup>, Russell Poldrack<sup>1</sup>, Oscar Esteban<sup>7</sup>

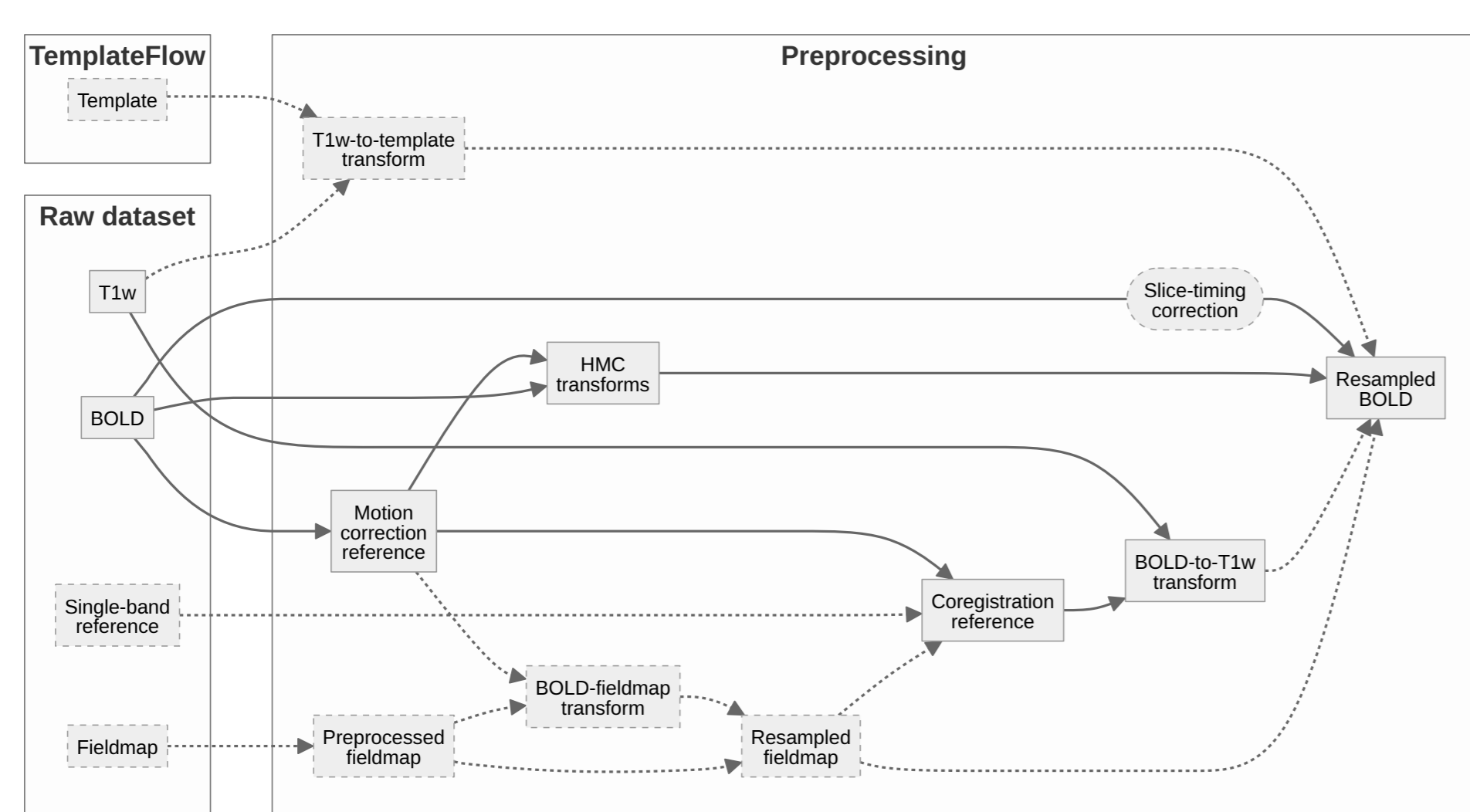
fMRIPrep: <https://fmripred.org>  
NiPreps: <https://nipreps.org>

<sup>1</sup>Stanford University, Stanford, CA, <sup>2</sup>University of Washington, Seattle, WA, <sup>3</sup>Dartmouth College, Hanover, NH, <sup>4</sup>Columbia University, New York, NY, <sup>5</sup>Florey Institute of Neuroscience and Mental Health, Melbourne, VIC, <sup>6</sup>Charité Universitätsmedizin Berlin, Berlin, Berlin, <sup>7</sup>Lausanne University Hospital and University of Lausanne, Lausanne, VD

OHBM Poster #2257

## End-to-end preprocessing pipelines are best suited to small workloads

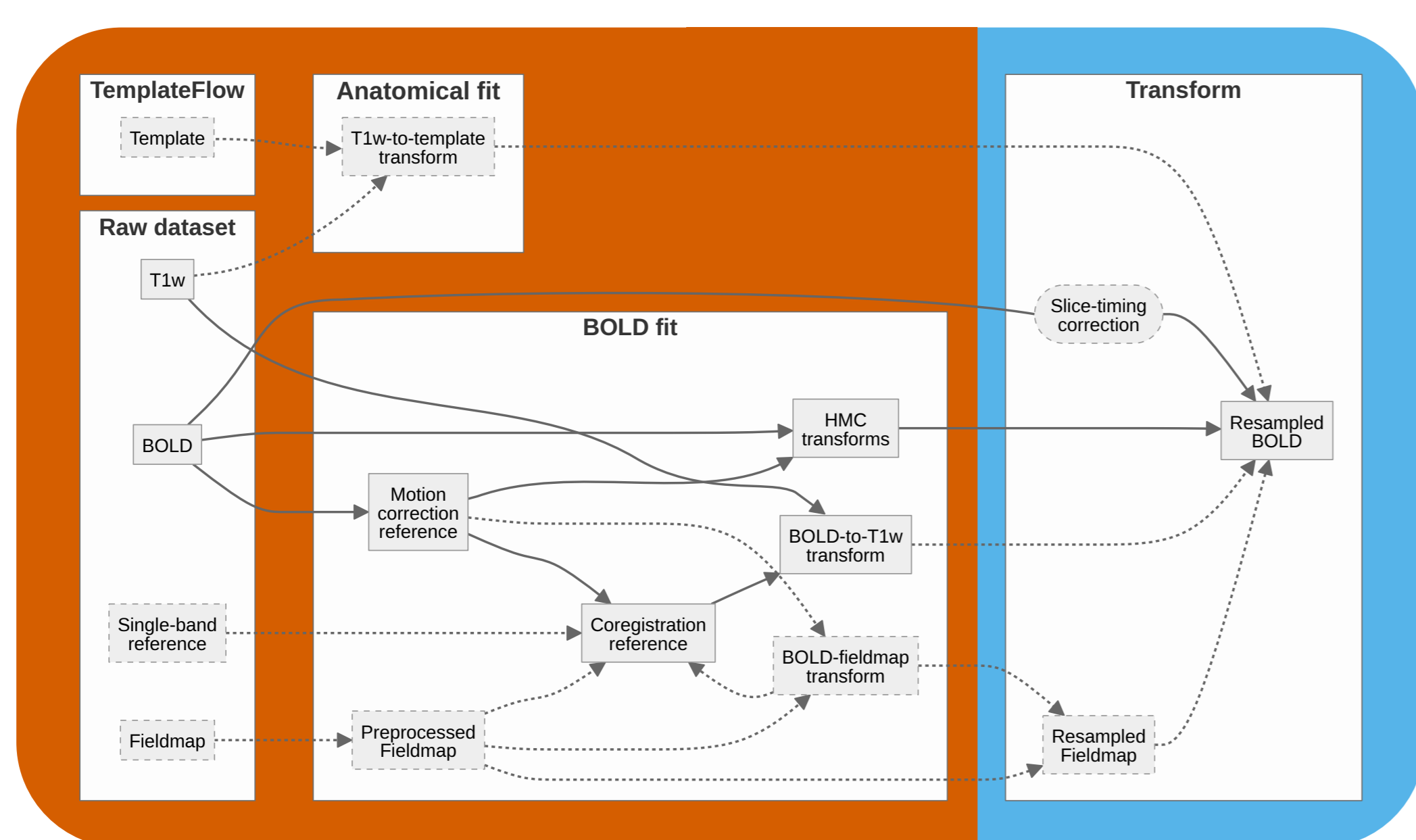
- ▶ Data are preprocessed and analyzed in the same place
- ▶ Subjects can be run all-at-once
- ▶ Little to no customization or modification is needed
- ▶ Derivative BOLD series can easily be stored together



Schematic overview of the preprocessing workflow to generated a BOLD series resampled into a target space. Arrows indicate data dependencies, and dashes indicate optional inputs or steps.

When processing large datasets, or targeting multiple output spaces, generating derivative BOLD series demands significant storage resources.

It is possible to split the workflow to isolate the components that require computation (fit) from those that can be reconstructed deterministically (transform).

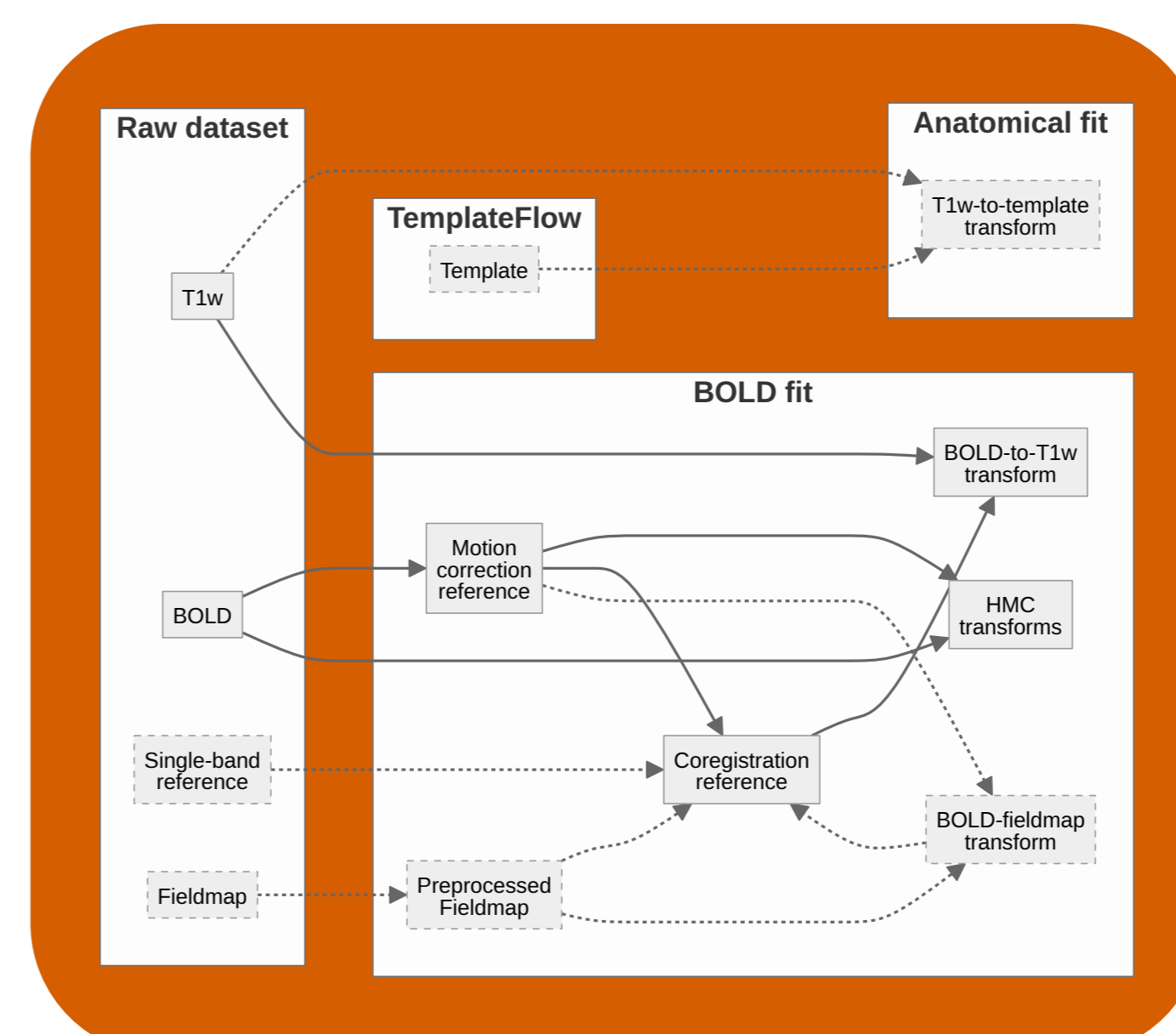


The FIT workflow contains most computationally intensive and all non-deterministic components of the preprocessing workflow. The TRANSFORM workflow generates most of the space-consuming derivatives.

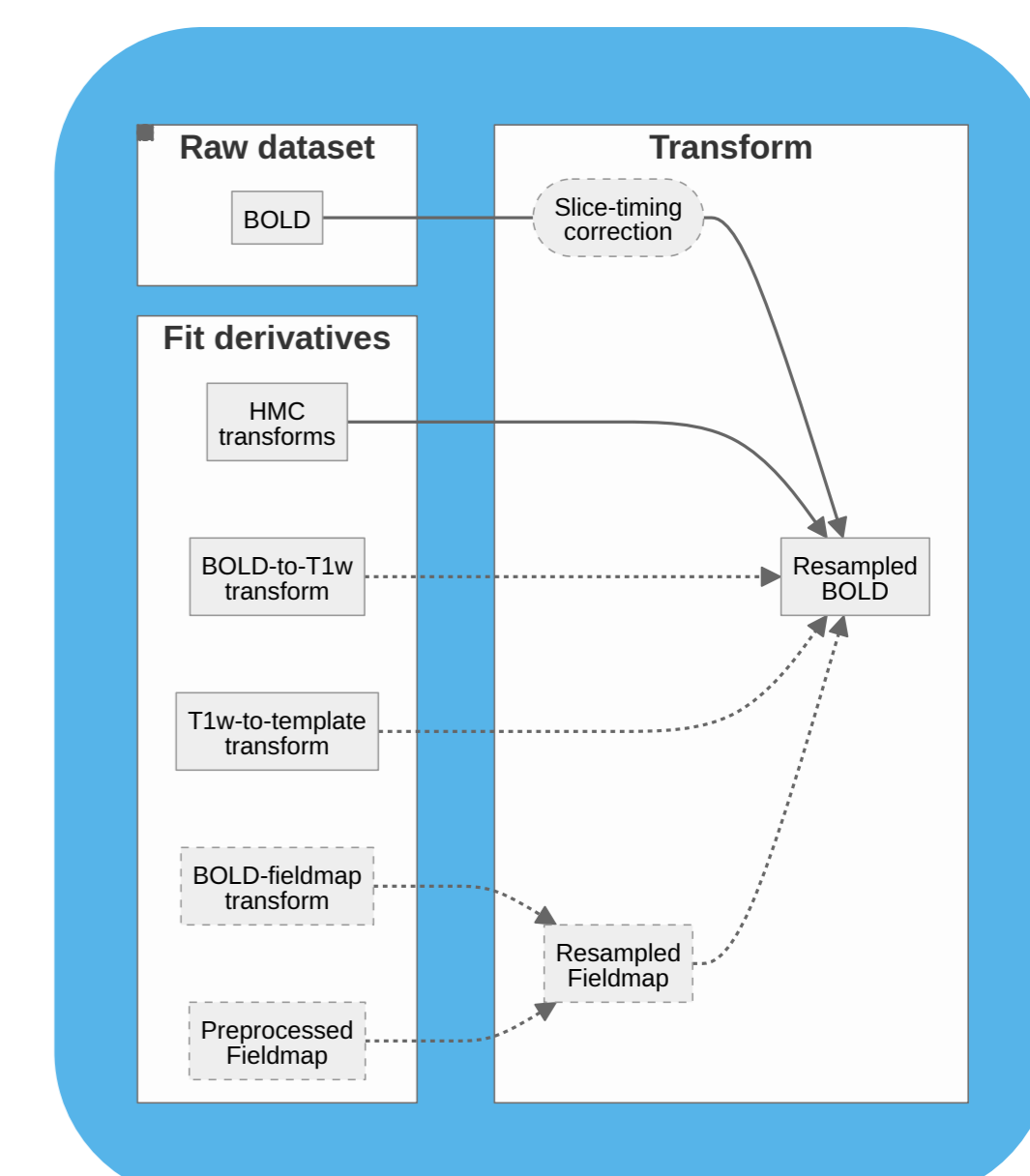
## We rewrote fMRIPrep to create a minimal set of derivatives and then generate the remaining derivatives from these

- ▶ The minimal derivatives for bold series are transforms and the files needed to calculate them.
- ▶ Head motion correction transforms were the main derivatives that needed to be added.
- ▶ Anatomical derivatives are mostly single volumes, and so were purged less ruthlessly.

## Partial derivatives enable new use cases



Running only the FIT workflow produces a compact set of derivatives. More target spaces can be selected with only a small impact on storage requirements.



Running only the TRANSFORM workflow avoids unnecessary recomputation. Resampling targets may be chosen for the specific analysis.

## Use Case: Secondary research

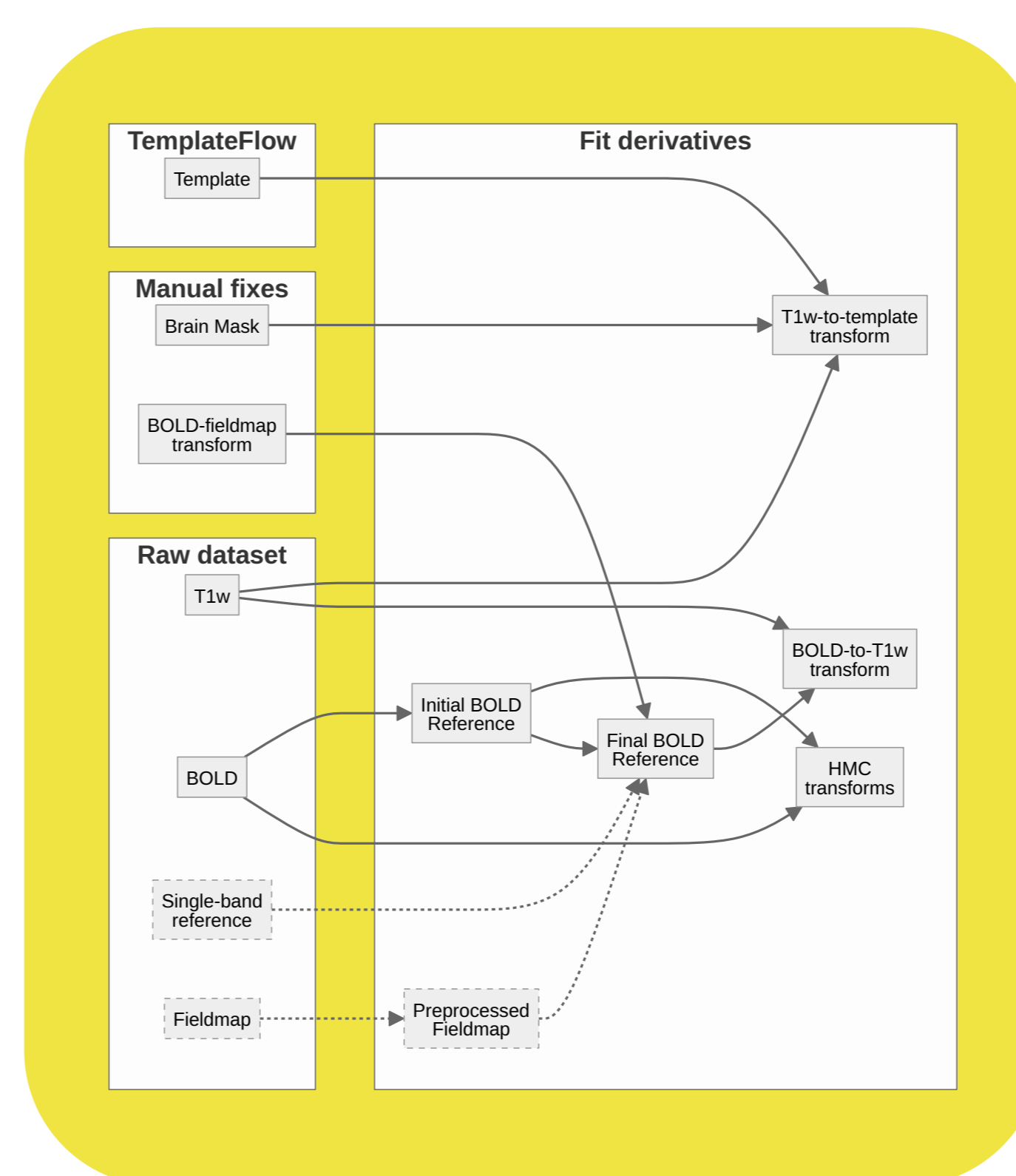
Alice publishes her data and the FIT workflow derivatives to a public archive. She provides transforms to multiple volumetric and surface templates. Bob and Carol each fetch the data and run the TRANSFORM workflow to reconstruct identical BOLD series in MNI space for analysis.

```
fmriprep raw/ out/ participant --level minimal \
--output-spaces ...
```

```
fmriprep raw/ out/ participant --level full \
--derivatives minimal/ --output-spaces ...
```

## Use Case: Manual intervention

Dave uses an alternative brain mask and BOLD-to-fieldmap registration to improve the quality of his preprocessing. He reruns fMRIPrep with these derivatives as inputs.



```
fmriprep raw/ out/ \
participant \
--level minimal \
--derivatives fixes/ \
--output-spaces ...
```

## Future directions

- ▶ Extend the fit-transform model to other nipreps tools. It has been adopted in ASLPrep already.
- ▶ fMRIpost library to generate secondary derivatives on-the-fly in downstream tools.
- ▶ Convert transform workflow into a standalone tool that is less heavy than the fMRIPrep container.

## Acknowledgements

NiPreps development, including fMRIPrep supported by the BRAIN Initiative award 1RF1MH121867-01A1. This research is part of the Frontera computing project at the Texas Advanced Computing Center. Frontera is made possible by National Science Foundation award OAC-1818253.

## Space/time comparisons:

<https://fmripred.org/en/latest/benchmarks.html>

