Robust web-based parallel-optimized minimal pre-processing and analysis pipeline for MRI big data

Xiaowei Song [a,c,d], Xue Wang [a], Kate Alpert [b], Lei Wang [b], Todd Parrish [a]

[a] Dept. of Radiology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[b] Dept. of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[c] Neuroimaging Research Branch, NIDA, NIH, Baltimore, MD, USA
[d] Dept. of CSEE, University of Maryland, Baltimore County, Baltimore, MD, USA

NORTHWESTERN UNIVERSITY

Introduction

The current trend in neuroscience studies is to collect large samples of structural and functional data, which results in having to deal with Big Data. Manual pre-processing and analysis is very tedious and error-prone. Here we report a web-based processing pipeline with quality assurance. It requires minimal user input once the processing parameters have been set. It aims to provide an easy automated environment to analyze resting-state and task functional data for all types of subjects (normal or those with lesions present), while also having flexibility and extend-ability for users with specific software needs.

Methods

1. ANATOMIC: Robust structural volume pre-processing (Figure 1)
Skull-stripping is the first step of structural data pre-processing. While there has been much effort in developing skull-stripping methods, no clear solution has emerged. Compared to BEaST [1], FSL BET and FreeSurfer watershed method, ROBEX [2] was chosen as the default option since it rarely fails and can finish in 1 minute for a typical human head with minimal errors. For failure-proof, user-defined skull-stripping command can be specified. Optional non-local means (NLM) filtering to smooth the image without blurring the edges, plus optional non-uniform intensity bias field correction by ANTS N4 are also provided.
Then robust MRI registration [3] to normalize the subject's brain to the MNI template space, followed by optional FSL FNIRT medium degree-of-freedom (DoF) non-linear registration, or SPM low-DoF (1K) non-linear alignment. And by FSL FAST to generate white matter partial volume is included to prepare boundary-based co-registration (BBR) [4], plus FSL FIRST to obtain 7 sub-cortical structures.
For flexibility, the SPM VBM8-toolbox is implemented as a compatible alternative for anatomical normalization, which can normalize by high-DoF (6.4M) DARTEL.
If a subject has a lesion present (chronic stroke), the lesion will be identified automatically. After editing the mask or approved directly, to minimize the lesion affect for normalization, especially for non-linear registration, we can choose cost function masking or enantiomorphic mapping.

2. BOLD: Functional task/resting-state pre-processing (Figure 2.A)
Once the time series data has passed the QA metrics, the BOLD data can be processed in the following order: optional despike, optional slice-timing correction based on DICOM header or user input, motion correction with middle time frame as reference, quadric trend removal, optional nuisance factors regression (Friston 24, FD, AFNI Euclidean and DVARS, WM/CSF principal components [5] etc.) , optional band pass filtering, optional percent signal conversion, optional motion scrubbing based on FD, AFNI Euclidean and DVARS threshold, 1-step normalization to template space (affine or non-linear) after BBR way co-registered to anatomical space, optional smoothing.

3. 1ST LEVEL STATISTICS
1st level analysis includes regional homogeneity and functional connectivity based on ROIs, MNI coordinates, or masks. For task fMRI, FSL FEAT or SPM batch job by file name mapping can be specified to estimate BOLD activation. The file name mapping mechanism enables hundreds of subjects' batch job to run in parallel on headless HPC nodes.

4. 2ND LEVEL STATISTICS
Group statistical processing is available by using file name mapping for SPM batch job, after 1st level analysis has been completed. ROI values from multiple contrasts can be exported to csv format ready for REDCap group analysis with subjects' behavior index.

5. Web interface and PDF result report (Figure 2.B)
Parameters specification and pipeline launching are available through web interface built into NUNDA (Northwestern Neural imaging Data Archive). The parameters are utilized by initial file (.ini) which can ease editing, repeated use for consistency, and for archival purpose. Final PDF reports are generated for preprocessing and 1st level analysis. The PDF report plus optional compression of intermediate files, allows users to reanalyze the data from a particular step and makes quality assurance checking easy.
To ensure the pipeline operation, we only change the code in the development version and merge to a stable version after passing all system tests. To ensure reproducibility of the data processing, the entire command line entries for each step of the pipeline are also included in PDF.
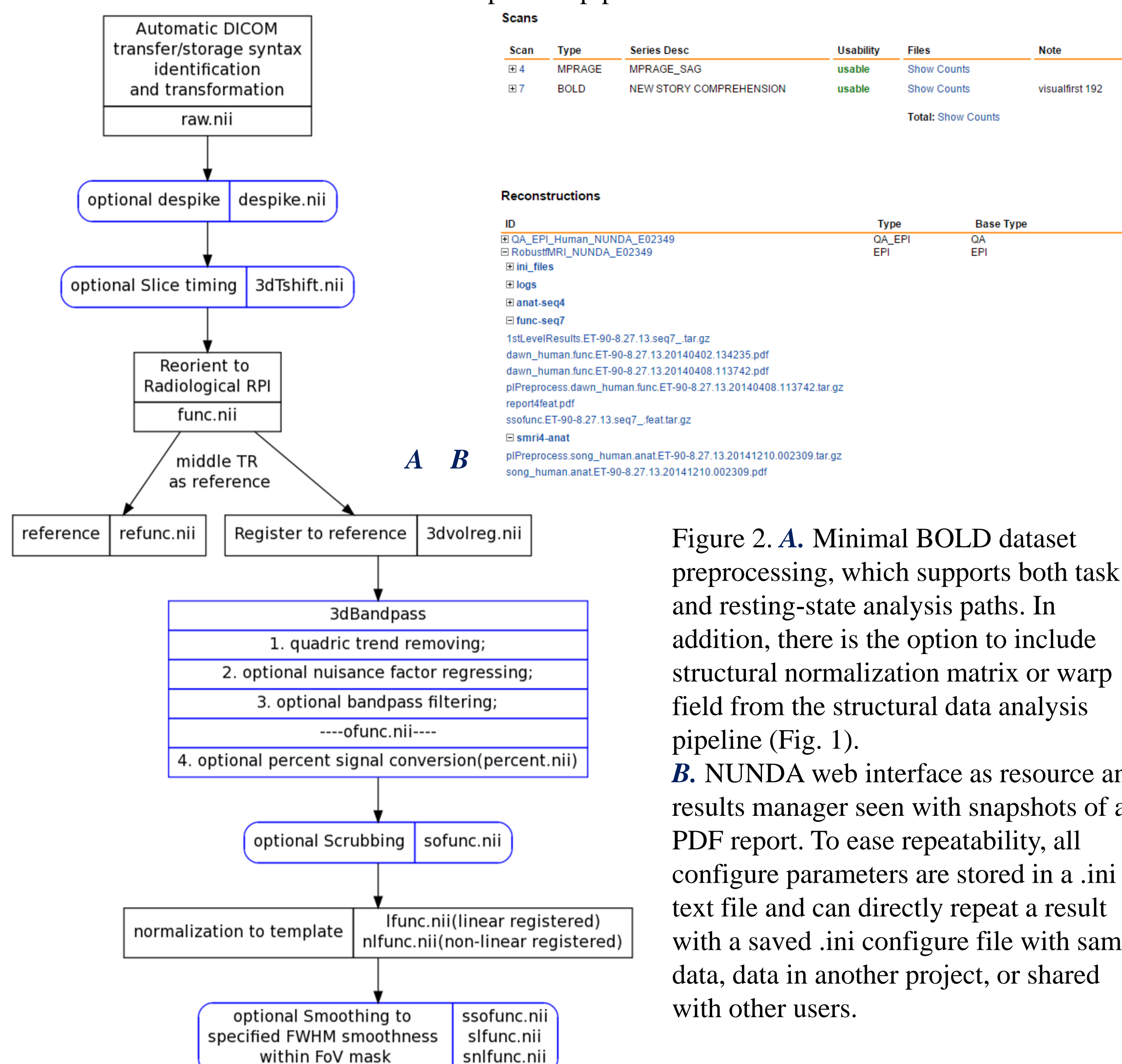

Figure 2. **A.** Minimal BOLD dataset preprocessing, which supports both task and resting-state analysis paths. In addition, there is the option to include structural normalization matrix or warp field from the structural data analysis pipeline (Fig. 1).
B. NUNDA web interface as resource and results manager seen with snapshots of a PDF report. To ease repeatability, all configure parameters are stored in a .ini text file and can directly repeat a result with a saved .ini configure file with same data, data in another project, or shared with other users.
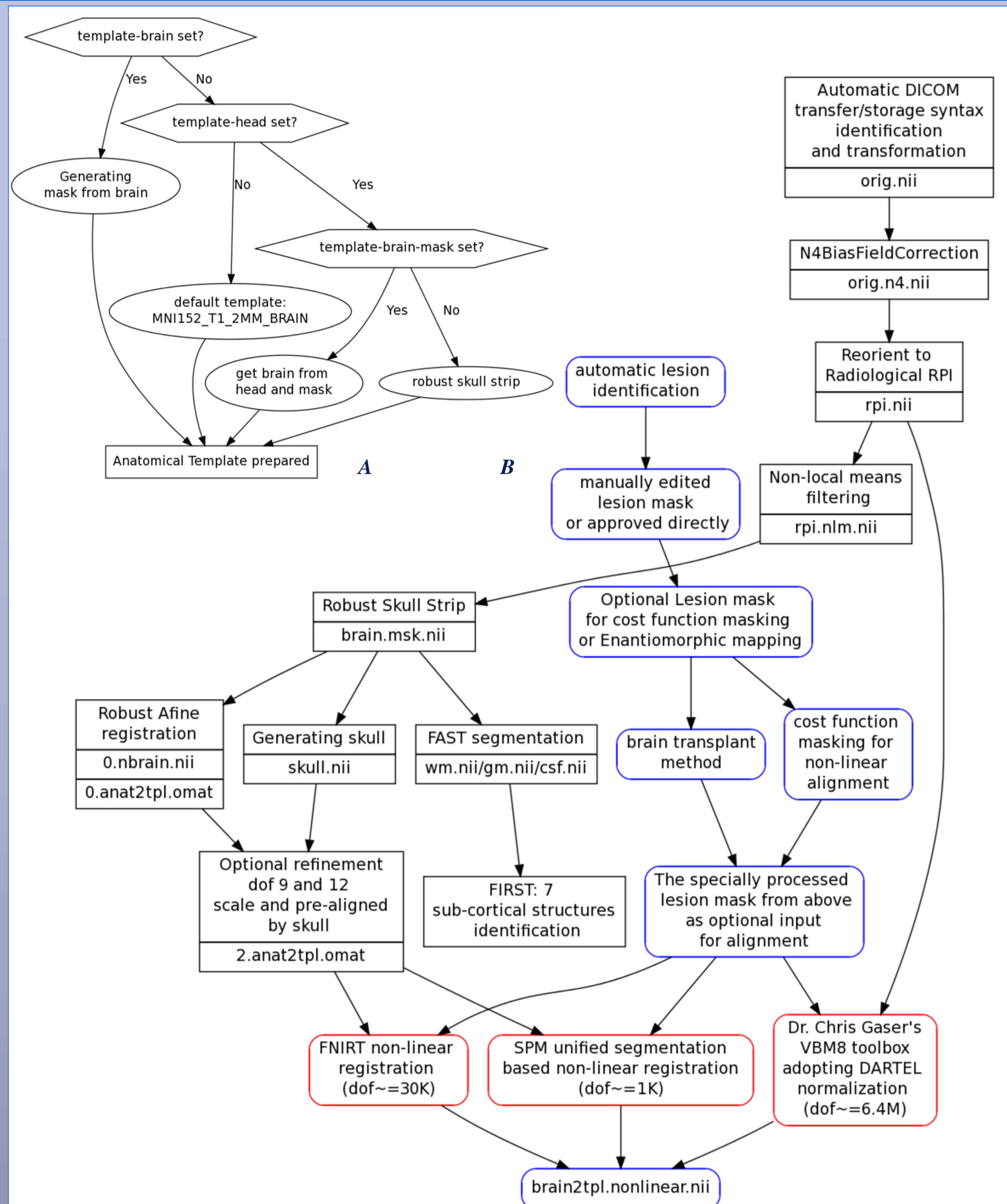

Figure 1. **A.** Automatic adaptive anatomical template preparation. **B.** Robust anatomical MPRAGE pre-processing supporting lesion based cost functional masking or enantiomorphic mapping, plus supporting multi-level dof non-linear registration after robust affine registration.

Projects	Subjects	Sessions
54	785	1892

Table 1. There have been hundreds of subjects fully automatically processed by the reported pipeline to June 2015.

Results & Conclusions

NUNDA is an XNAT based web interface for DICOM and processed data. We have built several processing pipelines which are robust, flexible and extendable, and have successfully processed thousands of sessions' data (Table 1). In the future, we will also incorporate FreeSurfer to make surface-based analysis available.
Acknowledgment: This study was supported by NIDCD grant 1P50DC012283-01A1.

References:

1. Eskildsen et.al. (2012), "BEaST: Brain Extraction Based on Nonlocal Segmentation Technique." NeuroImage 59, 3: 2362–73.

2. Iglesias et.al. (2011), "Robust Brain Extraction across Datasets and Comparison with Publicly Available Methods." IEEE Transactions on Medical Imaging 30, 9: 1617–34.

3. Reuter et.al. (2009). "Highly Accurate Inverse Consistent Registration: A Robust Approach." NeuroImage 53, 4: 1181–96.

4. Greve et. al. (2009). "Accurate and Robust Brain Image Alignment Using Boundary-Based Registration." NeuroImage 48, 1: 63–72.

5. Behzadi et. al. (2007). "A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI." NeuroImage 37, 1: 90–101.