

Data Preparation for Neuroimagers: BIDS, mriqc, and fmriprep

Dr. Kyle Kurkela

kkurkela@bu.edu

Research Computing Services
Information Services & Technology
Boston University

Resources

- Tutorial Slides and a Recording of today's tutorial will be made available here:
 - <https://www.bu.edu/tech/support/research/training-consulting/rcs-tutorial-videos-and-third-party-tutorials/>
- Code and examples data will be made available here:
 - https://rcs.bu.edu/examples/imaging/tut_dataprep_scc/
- Questions? Comments? Concerns?
 - help@scc.bu.edu
 - kkurkela@bu.edu

Assumptions

- This tutorial is aimed at a beginner level.
- However, I will assume that you have a basic level of comfort within a Linux environment.
- The Research Computing Services team has many resources for learning Linux:
 - A 20 min Introduction to Linux:
 - <https://www.bu.edu/tech/support/research/training-consulting/rcs-tutorial-videos-and-third-party-tutorials/intro-linux-20min/>
 - A 2 hr Introduction to Linux and the SCC:
 - <https://www.bu.edu/tech/support/research/training-consulting/rcs-tutorial-videos-and-third-party-tutorials/intro-scc/>
 - Linux on the SCC Cheat Sheet:
 - http://scv.bu.edu/documents/Linux_SCC_CheatSheet.pdf

Learning Objectives

- By the end of this tutorial, you should ...
 1. have a basic understanding of the SCC and why it is important.
 2. have a basic understanding of the two major file formats for neuroimaging data: DICOM and NIFTI.
 3. have a basic understanding of how neuroimaging data is organized in BIDS.
 4. be able to convert DICOM files to NIFTI files using *dcm2niix*.
 5. be able to convert DICOM files to BIDS formatted NIFTI files using *dcm2bids*.
 6. be able to run a quality assurance routine using *MRIQC* on the SCC.
 7. be able to run a preprocessing routine using *FMRIprep* on the SCC.

Workshop Overview

1. What is the SCC? Why should you use it as a neuroimager?
2. Connecting to the SCC
3. Organizing Data on the SCC with *dcm2bids*
 - The Brain Imaging Data Structure (BIDS)
4. Assessing the quality of your data with *mriqc*
5. Preprocessing your data with *fmriprep*

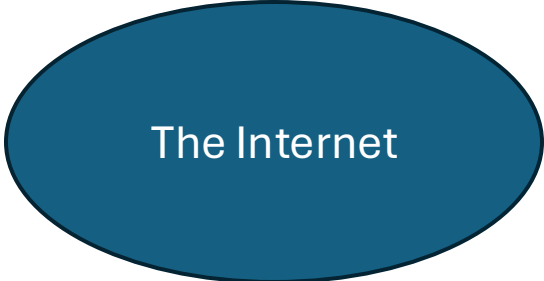
Workshop Overview

1. What is the SCC? Why should you use it as a neuroimager?
2. Connecting to the SCC
3. Organizing Data on the SCC with *dcm2bids*
 - The Brain Imaging Data Structure (BIDS)
4. Assessing the quality of your data with *mriqc*
5. Preprocessing your data with *fmriprep*

What is the SCC?



Your Computer

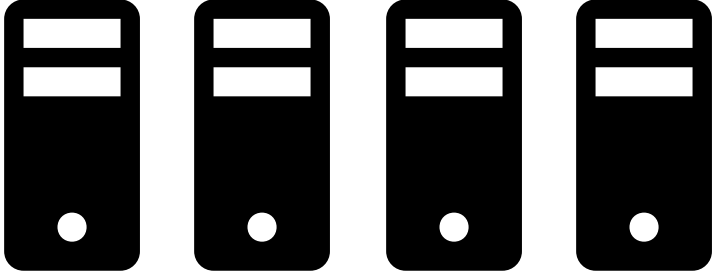


The Internet

Public Network

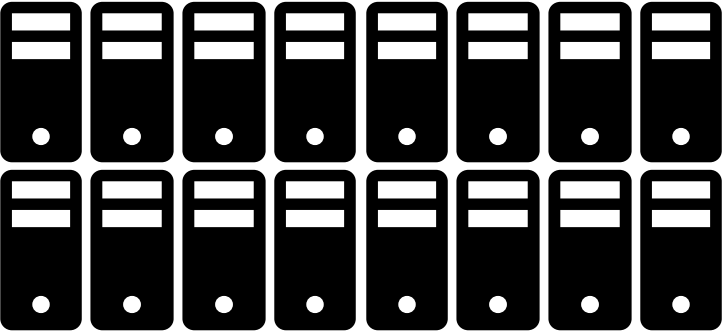


SCC OnDemand



Login Nodes
scc1, scc2, geo, scc4

Public Network

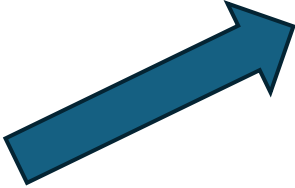
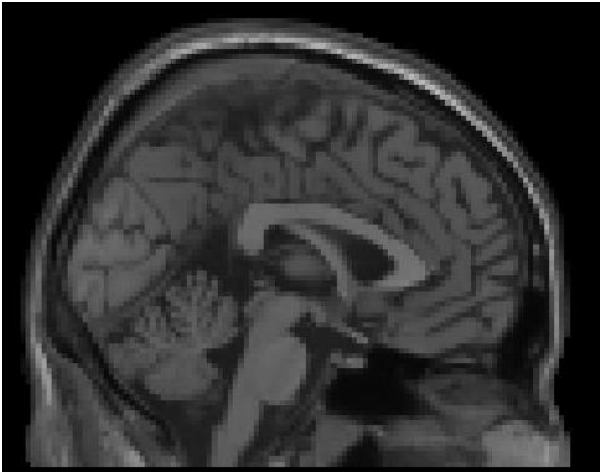


Compute Nodes

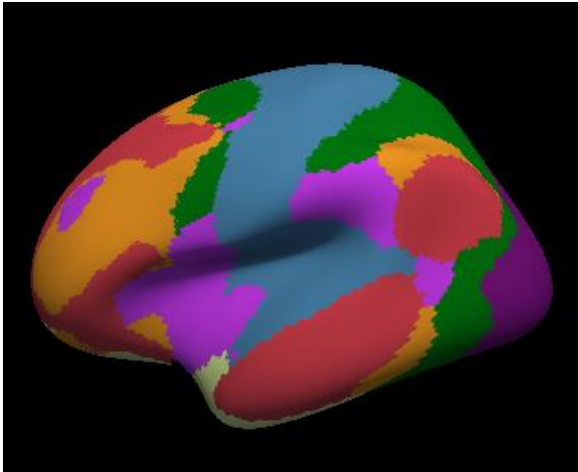
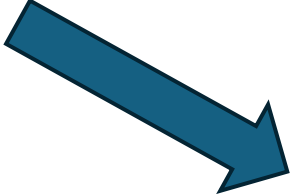
Private Network

Private Network

Why the SCC?



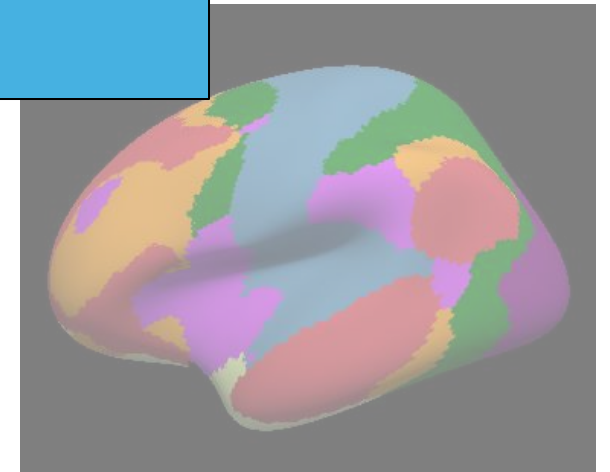
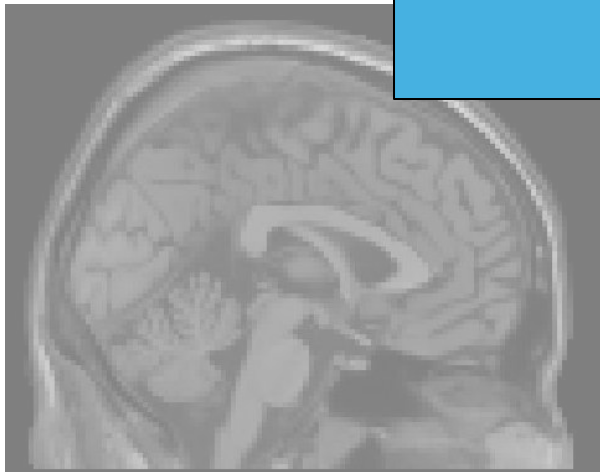
Personal
Computer



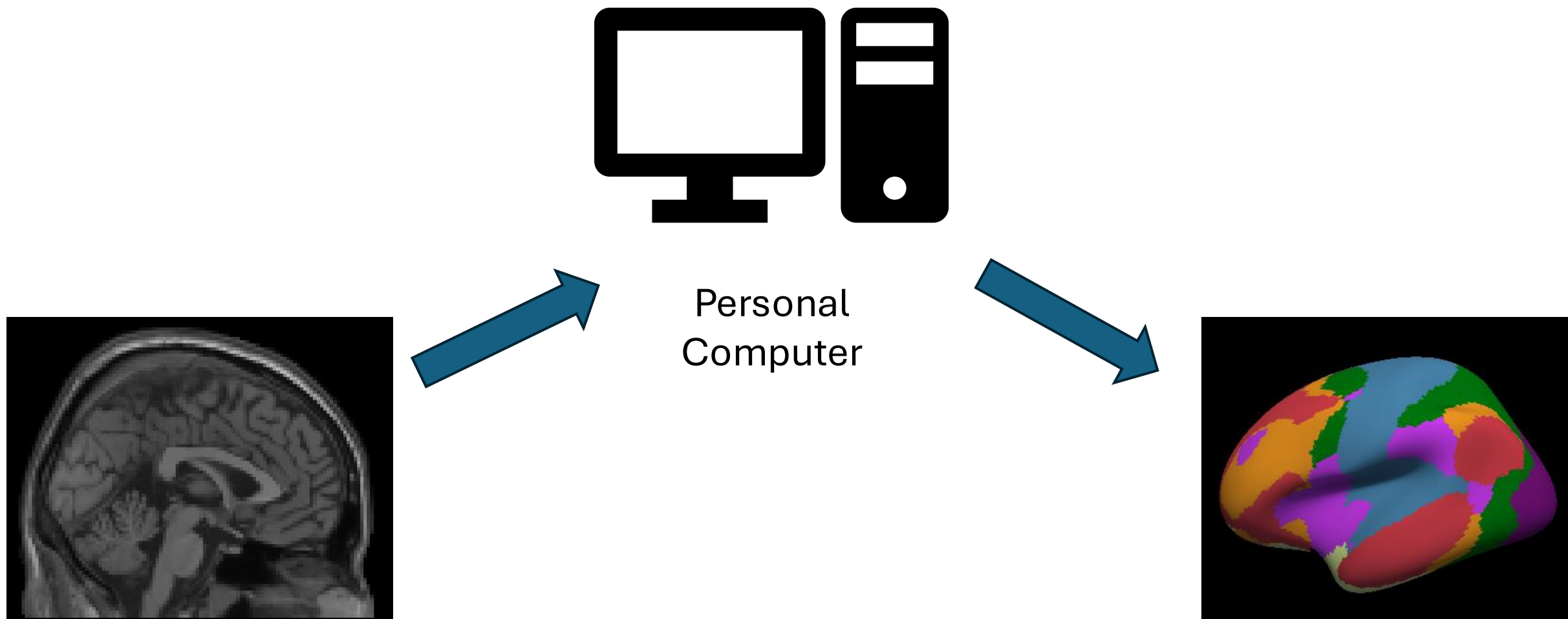
Why the SCC?



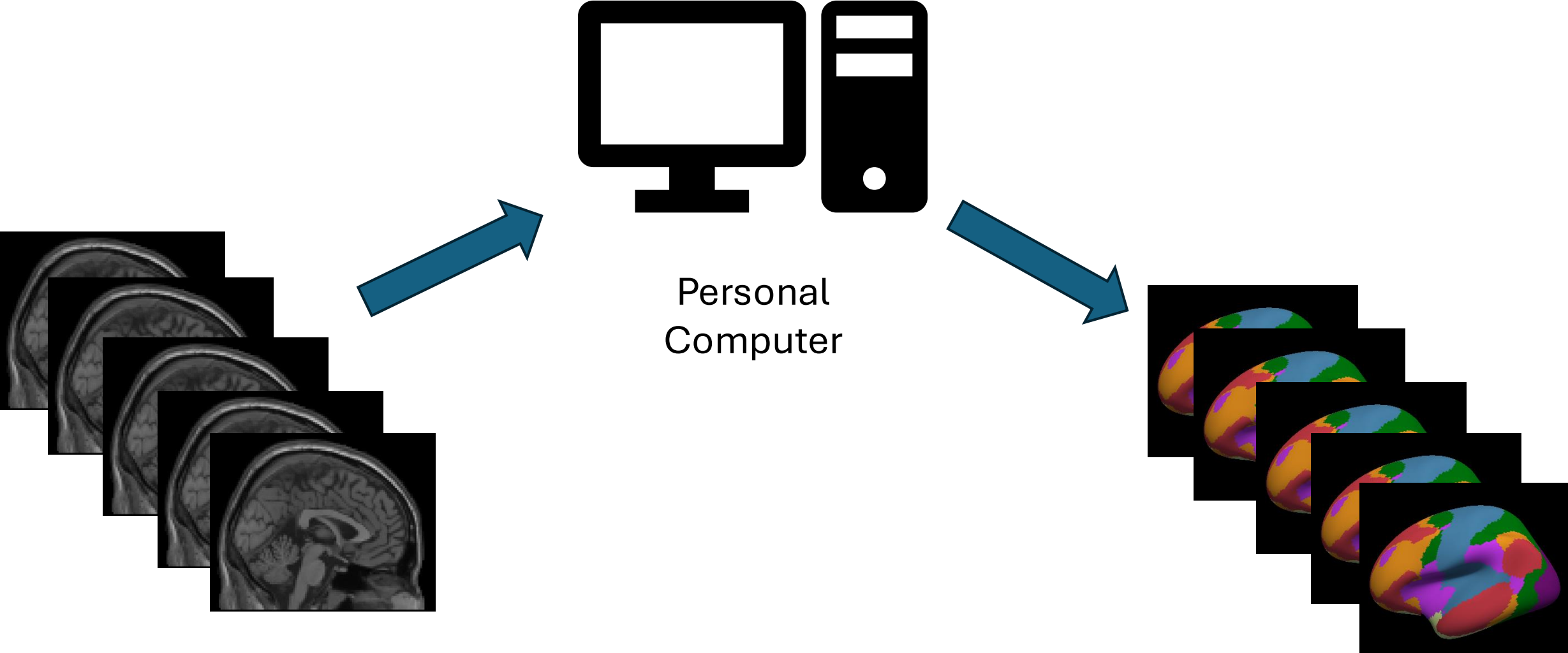
~ 6 Hours



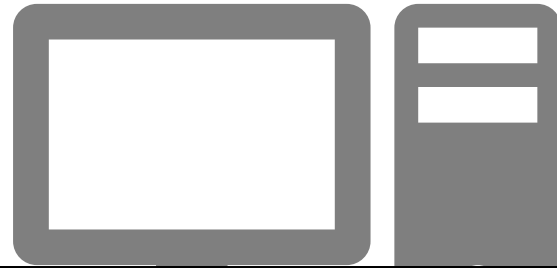
Why the SCC?



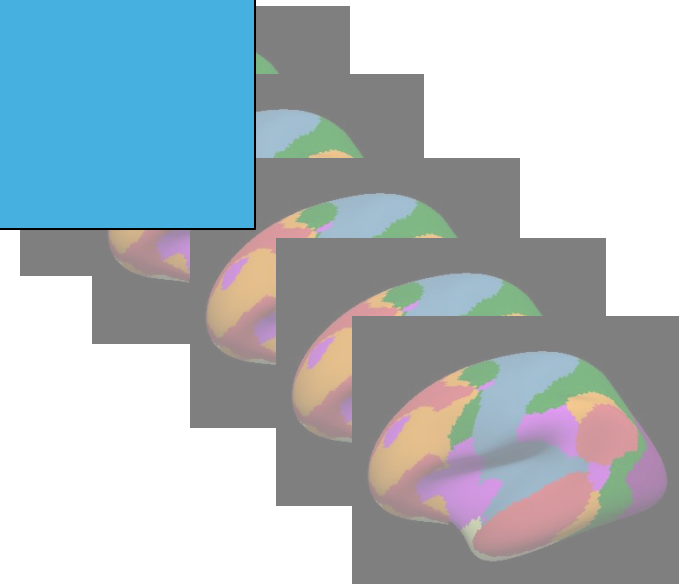
Why the SCC?



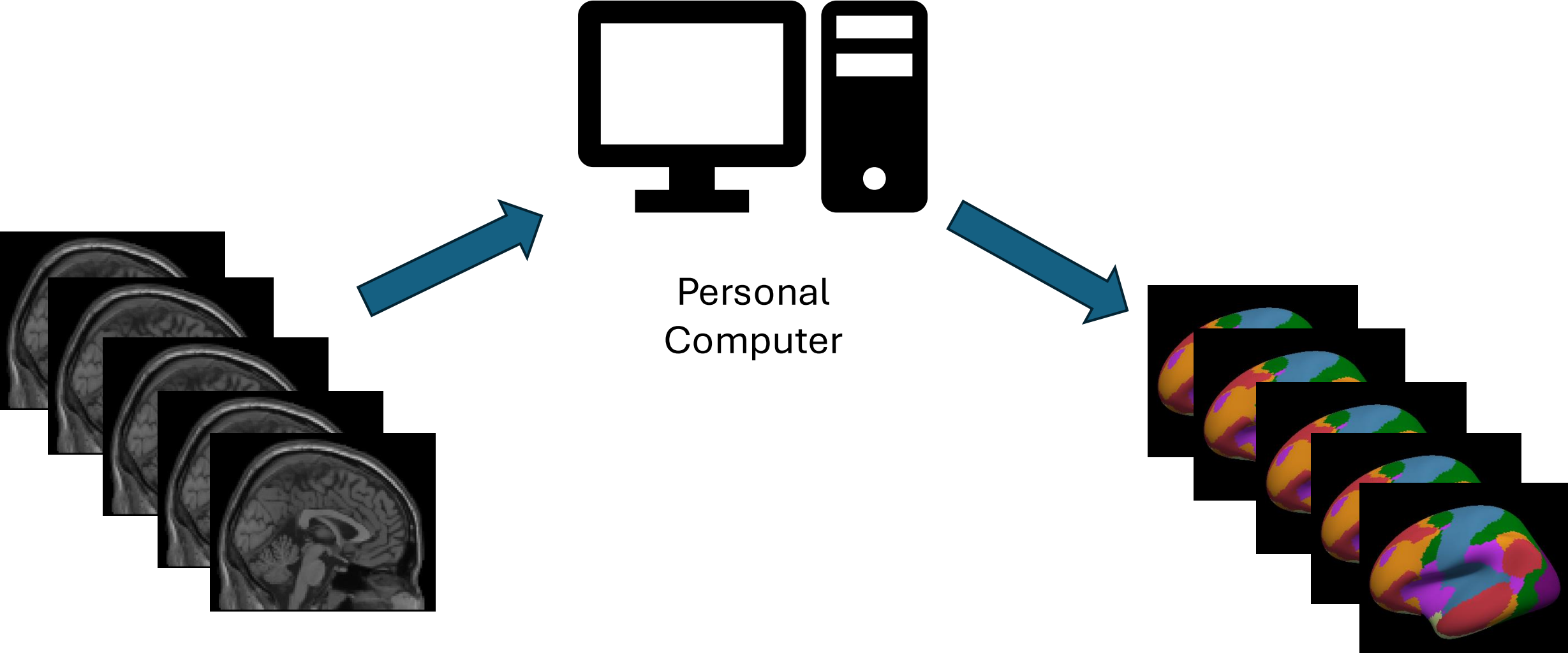
Why the SCC?



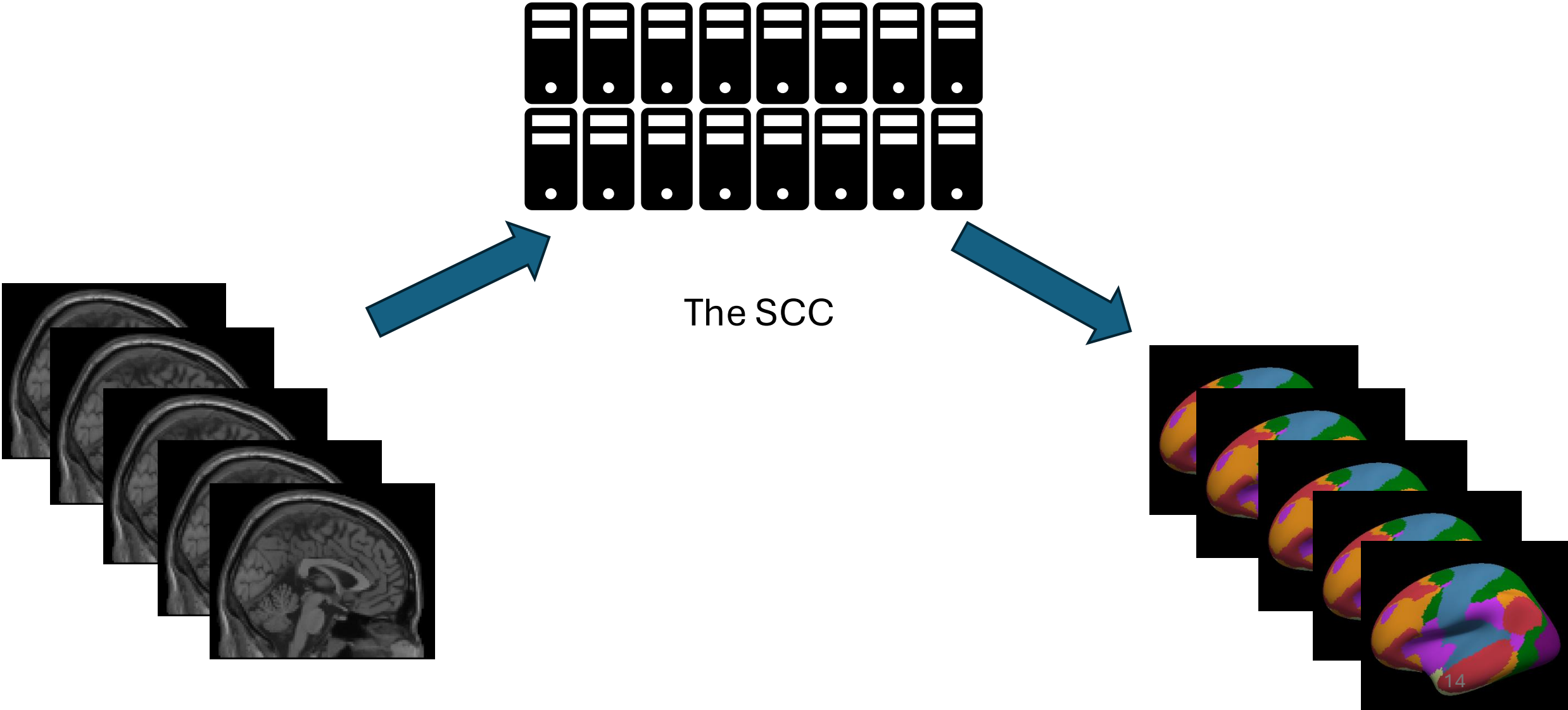
6 Hours x 5 subjects =
30 hours!



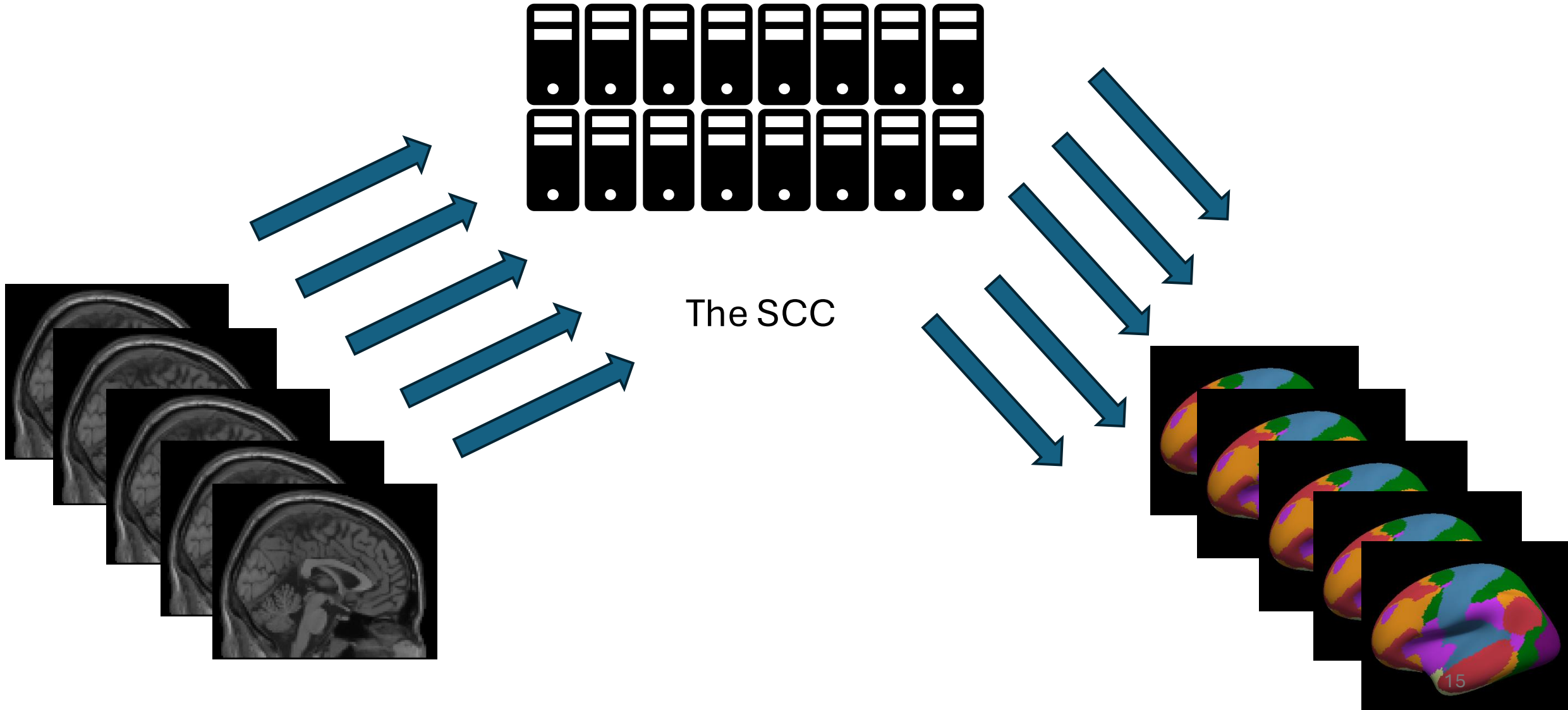
Why the SCC?



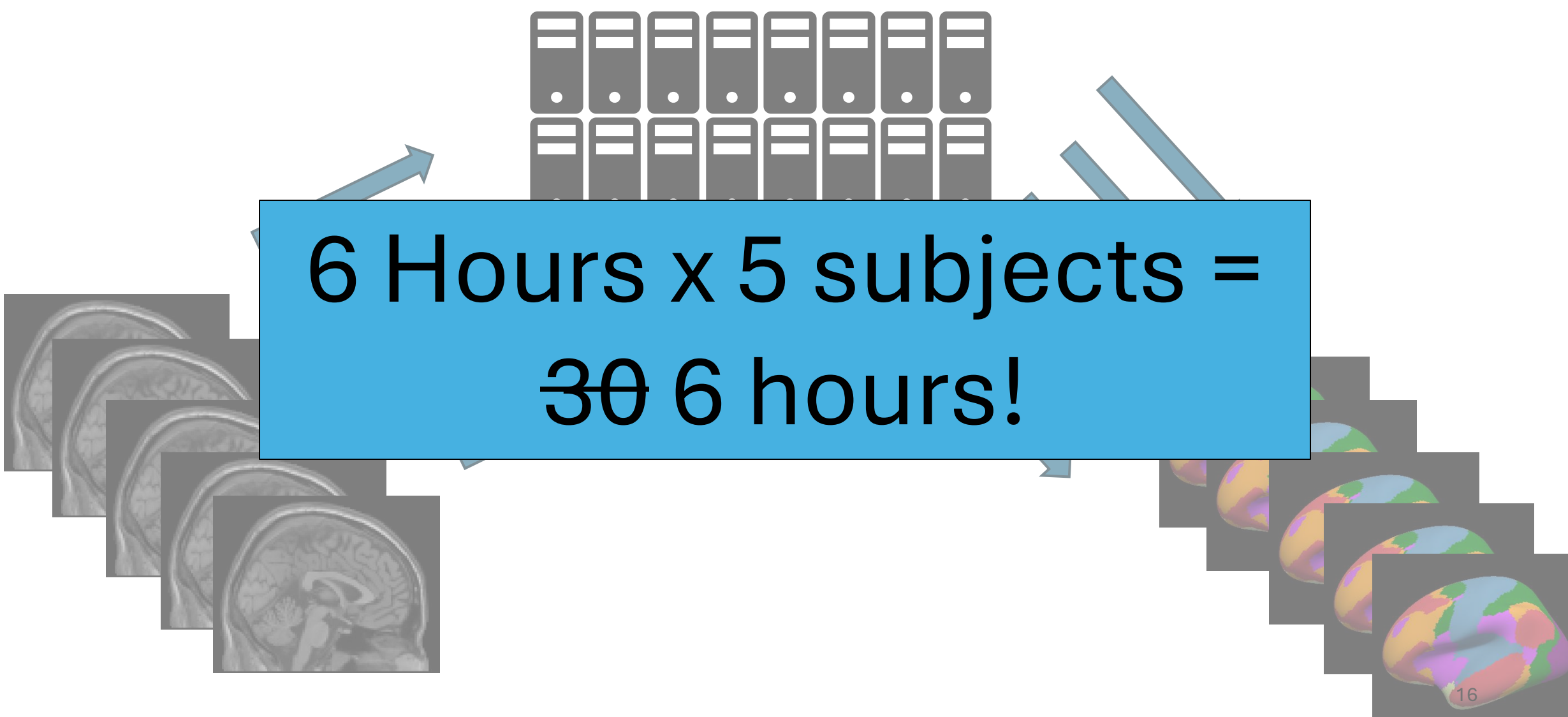
Why the SCC?



Why the SCC?



Why the SCC?

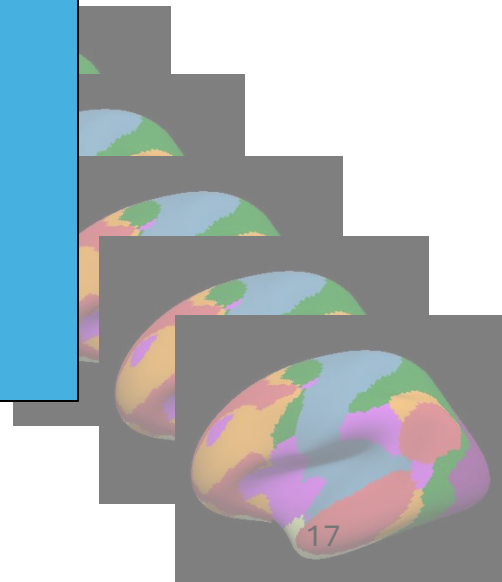


6 Hours x 5 subjects =
~~30~~ 6 hours!

Why the SCC?



The power of the SCC is that it greatly *increases the speed* with which you can process and analyze your large, high dimensional neuroimaging datasets.



Workshop Overview

1. What is the SCC? Why should you use it as a neuroimager?
2. Connecting to the SCC
3. Organizing Data on the SCC with *dcm2bids*
 - The Brain Imaging Data Structure (BIDS)
4. Assessing the quality of your data with *mriqc*
5. Preprocessing your data with *fmriprep*

Workshop Overview

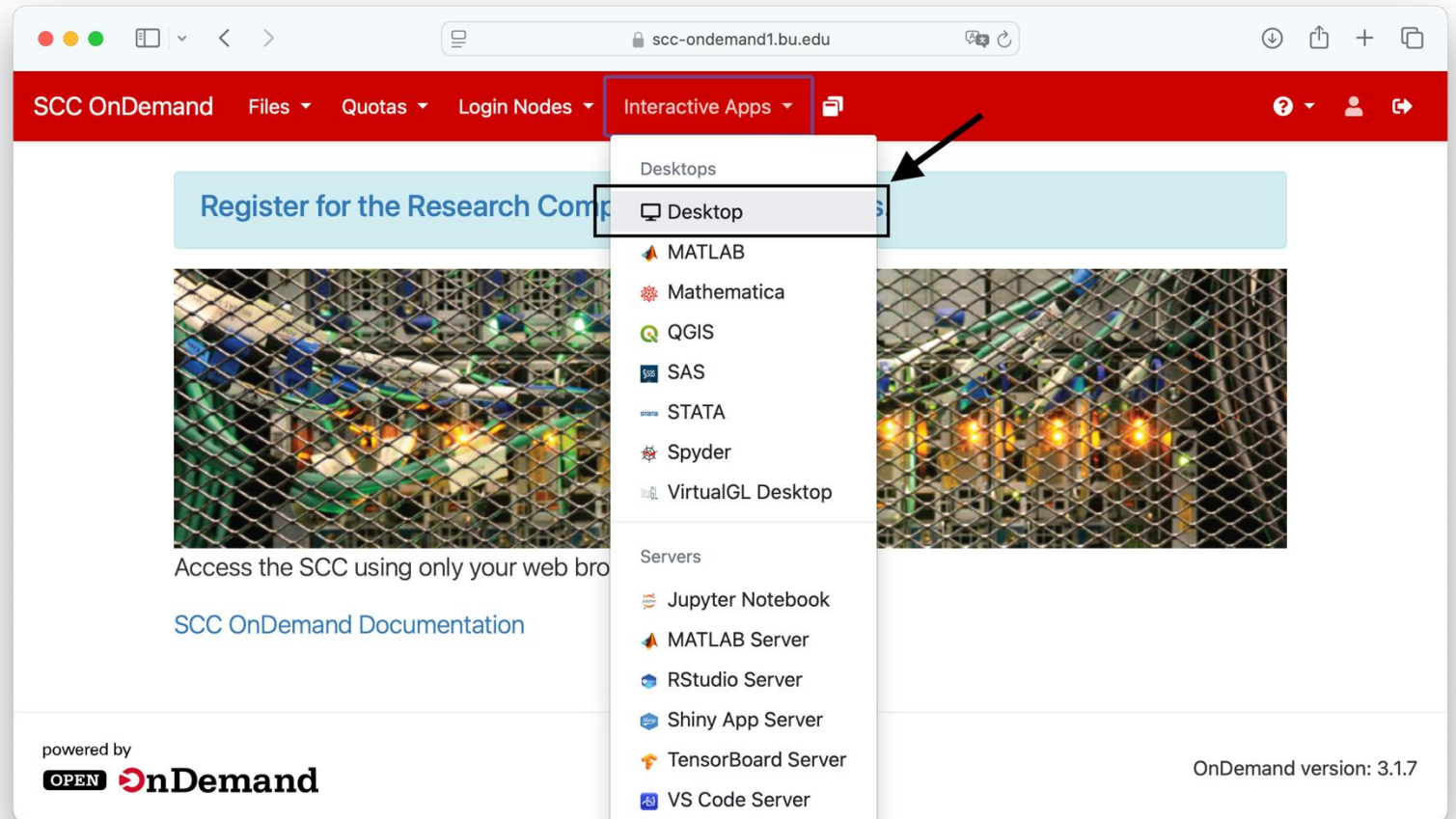
1. What is the SCC? Why should you use it as a neuroimager?
2. **Connecting to the SCC**
3. Organizing Data on the SCC with *dcm2bids*
 - The Brain Imaging Data Structure (BIDS)
4. Assessing the quality of your data with *mriqc*
5. Preprocessing your data with *fmriprep*

Getting an SCC Account

- Today we will be using tutorial accounts. These accounts are temporary and will be inaccessible after today's workshop
- All users of the SCC must be on a Research Project headed by a full-time BU Faculty member with some exceptions:
 - Some academic classes
 - 3 Month Trial Accounts for “trying out” the SCC
- Please see for more detailed information:
 - <https://www.bu.edu/tech/support/research/account-management/>

Accessing Tutorial Accounts

- Open a web browser
- Go to scc-ondemand-tutorial.bu.edu
- Username: tuta#
- Password: *****
- Click “Interactive Apps”
- Click “Desktop”



Copy the tutorial data

- Open up a terminal window and follow along with me.
- NOTE: all commands used in today's tutorial will be in a publicly accessible text file *notes.txt*
 - https://rcs.bu.edu/examples/imaging/tut_dataprep_scc/notes.txt

Workshop Overview

1. What is the SCC? Why should you use it as a neuroimager?
2. Connecting to the SCC
3. Organizing Data on the SCC with *dcm2bids*
 - The Brain Imaging Data Structure (BIDS)
4. Assessing the quality of your data with *mriqc*
5. Preprocessing your data with *fmriprep*

Workshop Overview

1. What is the SCC? Why should you use it as a neuroimager?
2. Connecting to the SCC
3. Organizing Data on the SCC with *dcm2bids*
 - The Brain Imaging Data Structure (BIDS)
4. Assessing the quality of your data with *mriqc*
5. Preprocessing your data with *fmriprep*

MRI Data 101 -- DICOM

- Magnetic Resonance Imaging (MRI) data largely stored in DICOM format
- DICOM
 - Standard medical image format used by most human imaging scanners.
 - Very rich meta data.
 - Supports both 2D and 3D images (3D images are stored as a sequence of 2D images).
 - Not very conducive for image processing.
 - Relatively difficult to compress.
 - See for more information:
 - <https://www.dicomstandard.org>
 - <https://dicom.innolitics.com/ciods/mr-image>

Explore the DICOM files

- Let's return to SCC OnDemand and explore the data a little bit, inspecting the headers of one of the DICOM files. Follow along with me...
- NOTE: all commands used in today's tutorial will be in a publicly accessible text file *notes.txt*
 - https://rcs.bu.edu/examples/imaging/tut_dataprep_scc/notes.txt

MRI Data 101 -- NIFTI

- Raw DICOM image files are usually converted to one of several different file formats for further processing and analysis.
- NIFTI
 - Most popular – the research standard format
 - Relatively limited meta-data
 - Represents 3D volumes (or 4D or 5D) in a single file
 - Easily compressed to reduce storage space
 - Examples: sub_001.nii, subject_001.nii.gz
 - Freesurfer, FSL, AFNI, SPM, CONN

DICOMs -> NIFTIs

- Let's convert the DICOM files we just looked at to NIFTI files and explore them a bit closer. Follow along with me...
- NOTE: all commands used in today's tutorial will be in a publicly accessible text file *notes.txt*
 - https://rcs.bu.edu/examples/imaging/tut_dataprep_scc/notes.txt

MRI Data Organization

- MRI data is numerous and complex – you need an organization system!
- In the good-old-days, every research group, institution, lab would have their own idiosyncratic organization systems. It was a nightmare.
- BIDS came to the rescue!

The Brain Imaging Data Structure (BIDS)

- BIDS – a universal organization system for brain imaging data
- Organizes data into subfolders for data types nested within sessions nested within subjects with specific meta data files required at different stages of a hierarchy depending on data type
- Please see the BIDS data format manual, associated website, and associated paper for further details
 - **Manual:** <https://bids-specification.readthedocs.io/en/stable/>
 - **Website:** <https://bids.neuroimaging.io/specification.html>
 - **Paper:** Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., ... & Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, 3(1), 1-9.

The Brain Imaging Data Structure (BIDS)

- What does a BIDS formatted dataset look like? What are the basics?
- Let's take a look at some publicly available ones on OpenNeuro
- <https://openneuro.org>

The Brain Imaging Data Structure (BIDS)

How do I get my data into BIDS format?

1. Manually renaming and placing files into subdirectories
2. Automatically using your preferred programming language
 - Example: [R Script from Kyle's GithubRepo](#)
3. Using a software package to automate the process:
 - [There are a number packages](#) – Kyle's favorite and recommended tool is [*dcm2bids*](#)

Software Spotlight: *dcm2bids*

- One of a vast array of software packages designed to organize your MRI data into a bids format
- Relies on the very popular *dcm2nix* software package to convert DICOM → NIFIT AND performs data organization all in one go
- You need to carefully design a config.json file to sort your scans into the appropriate BIDS datatypes

DICOMs -> NIFTIs -> BIDS

- Let's convert our example DICOM data into a bids formatted dataset. Follow along with me...
- NOTE: all commands used in today's tutorial will be in a publicly accessible text file *notes.txt*
 - https://rcs.bu.edu/examples/imaging/tut_dataprep_scc/notes.txt

Creating a dcm2bids config file

- The trickiest part about using dcm2bids is [creating the config file](#).
- The config file is a json text file that tells dcm2bids how to map header information from the DICOM file to the appropriate BIDS information
 - For example, which DICOM files represent an ANAT image? A BOLD images? A FMAP image?

Does your dataset meet the BIDS guidelines?

- How do you know if your datasets meets BIDS guidelines?
- You can use the *BIDS validator*.
 - Available as a locally installable utility and as an online web browser utility
 - Web Utility: <https://bids-standard.github.io/bids-validator/>

Workshop Overview

- What is the SCC? Why should you use it as a neuroimager?
- Connecting to the SCC
- Organizing Data on the SCC with *dcm2bids*
 - The Brain Imaging Data Structure (BIDS)
- Assessing the quality of your data with *mriqc*
- Preprocessing your data with *fmriprep*

Workshop Overview

- What is the SCC? Why should you use it as a neuroimager?
- Connecting to the SCC
- Organizing Data on the SCC with *dcm2bids*
 - The Brain Imaging Data Structure (BIDS)
- **Assessing the quality of your data with *mriqc***
- Preprocessing your data with *fmriprep*

Assessing your Data

- Now that you have your data organized, how do you know if it is any good?
- In neuroimaging, there are several common data quality concerns including but not limited to:
 - Singal dropout, typically near the sinuses and ear canals
 - Magnetic Field Inhomogeneities
 - Participant motion

Software Spotlight – *MRIQC*

- MRIQC is a BIDS compatible software package that will create a nicely formatted html report giving a snapshot overview of the quality of an MRI scan
- MRIQC is designed to use the best performing tools from other MRI software packages with minimal data processing time
- Once your data is in compliance with BIDS, the running of the software package is straightforward!
- <https://mriqc.readthedocs.io/en/latest/>

MRIQC

- Let's take a closer look at an example batch script for running MRIQC on the SCC. Follow along with me...
- NOTE: all commands used in today's tutorial will be in a publicly accessible text file *notes.txt*
 - https://rcs.bu.edu/examples/imaging/tut_dataprep_scc/02_mriqc.qsub
 - https://rcs.bu.edu/examples/imaging/tut_dataprep_scc/notes.txt

An Example MRIQC Individual Report

- Let's take a closer look at an example output report:
 - T1 and T2 Images:
 - <https://mriqc.readthedocs.io/en/latest/reports/smri.html>
 - BOLD Images:
 - <https://mriqc.readthedocs.io/en/latest/reports/bold.html>

Workshop Overview

- What is the SCC? Why should you use it as a neuroimager?
- Connecting to the SCC
- Organizing Data on the SCC with *dcm2bids*
 - The Brain Imaging Data Structure (BIDS)
- Assessing the quality of your data with *mriqc*
- Preprocessing your data with *fmriprep*

Workshop Overview

- What is the SCC? Why should you use it as a neuroimager?
- Connecting to the SCC
- Organizing Data on the SCC with *dcm2bids*
 - The Brain Imaging Data Structure (BIDS)
- Assessing the quality of your data with *mriqc*
- Preprocessing your data with *fmriprep*

Preprocessing your Data

- Now that your data are organized and are of high quality, how do you prepare the data for analysis?
- Typically, in MRI studies you want to perform a set of data processing steps prior to analyzing your data. They include:
 - Brain extraction and defacing
 - Motion correction
 - Slice timing correction
 - Smoothing
 - Registration and Normalization
- There are many different preprocessing pipelines each with their own algorithms
 - How do you choose which software package to use? Which ones are the best?

Software Spotlight – *FMRIPREP*

- [FMRIPREP](#) is a BIDS compatible software package that creates a preprocessing pipeline from all the best performing parts of other commonly used MRI processing software
 - For example: a smoothing algorithm from SPM, a normalization algorithm from AFNI, etc, etc.
- FMRIPREP offers a ton of options for tinkering with its internals
- Like MRIQC, creates nicely formatted html reports on the success (or failure) of the preprocessing pipeline
- Once your data is in compliance with BIDS, running FMRIPREP is easy!

FMRIPREP

- Let's return once again to the SCC and examine a batch script designed to run fmriprep. Follow along with me...
- NOTE: all commands used in today's tutorial will be in a publicly accessible text file *notes.txt*
 - https://rccs.bu.edu/examples/imaging/tut_dataprep_scc/notes.txt

An Example FMRIprep Report

- Lets take a look at an example output report for FMRIprep:
 - <https://fmripred.org/en/stable/outputs.html>

Additional Web Resources

- Research Computing Support Pages
 - www.bu.edu/tech/support/research/
- Software Packages Available on the SCC
 - www.bu.edu/tech/support/research/software-and-programming/software-and-applications/
- Brain Imaging Data Standard
 - bids.neuroimaging.io
- *dcm2bids*
 - unfmontreal.github.io/Dcm2Bids/3.2.0/
- *MRIQC*
 - mriqc.readthedocs.io/en/latest/
- *FM RIPREP*
 - fmriprep.org/en/stable/

- Please open the following link in a web browser and fill out our survey!
 - http://scv.bu.edu/survey/tutorial_evaluation.html
- Let us know how we did, how we can improve, and what types of workshops and tutorials you would like to see in the future