**Project Organization and Storage**

Note: though this document

**There are four main pillars of effective data storage**:

1. **Plan for the future**.
   1. Think about when and how you may need to access for your files in the future.
      1. What will be the most efficient folder layout?
      2. What files will you have to access most, if any, after the processing is done?
   2. Think about how others might have to access your files in the future.
      1. Are the naming conventions of your files comprehensible to someone who has not worked on the project?
      2. Does the layout of your files make sense?
      3. And most importantly, documentation.
2. **Document, document, document.**
   1. **It is vital that you write down what you do during your time on a project**. Always assume that someone else – your PI, a new RS, etc. – will need to access your processed data and understand what you did at some point in the future.
   2. You should, at the very least, have a detailed pipeline document
3. **Remove extraneous files**.
   1. If you have duplicate files in your project folder or if you have files in your project folder that are stored elsewhere, consider deleting these files from your project folder before archiving/storing it.
   2. For instance, you may have preprocessed scalar maps both in your project folder as well as in another location. These would be good candidates for removal before long-term storage.
4. **Compress as much as possible**.
   1. Once you have gotten rid of files that you do not need, it is vital to compress the files you will need to store.
   2. This may involve compressing entire folders, individual files, or both.
   3. **Note**: Do not compress already compressed files.
      1. For example: If you have a folder of compressed niftis (with the file extension .nii.gz), do not then compress the entire folder.
   4. In order to do this efficiently and effectively, consider point #1.

**Places to store data**:

1. Box or another cloud storage platform. On these platforms, space is less of an issue, but it is still good practice to delete extraneous files and compress all other files.
2. Your personal repository or archive. In this case, storage is up to you, but the above best practices are still recommended.
3. Your PI’s research repository space or archive. In this case, good storage practices are the most vital as every PI has their own preferences and space is at a premium on places like the /fs/research/ drives and other external drives used for long-term storage.

**Methods of compressing data**:

1. Right click a file or folder and select “Compress”. Only use if you need to compress an entire folder of unzipped files or an individual file.
2. Compress via the Terminal.
   1. Use the command `gzip` as it is the fasted and most space saving bash command for zipping.
   2. Example usage:
      1. gzip -r /path/to/folder
      2. “-r” stands for “recursive”. This will search through all folders and sub folders to zip only files that are unzipped.

**Accessing compressed data**:

1. It is generally recommended to unzip files via command line but using a command similar to the one you used to zip the files: gunzip -r /path/to/folder