Human Time Data

Working with fMRI Data

An Introduction

Table of Contents

Introduction 02
MRI Terminology 05
Conducting a Study 07
Pretests/Pilot Tests 08
Data Capture 09
- Equipment 09
- Raw Data 10
- EEG Raw Data 12
- Transfer of Data 12

Behavioral Data 13
- E-prime 13
- Psych Toolbox 14
- Presentation 14
- PsychoPy 15

Analysis 17
Preprocessing 18
Analysis Software 21
- SPM12 22
- AFNI 24
- FREESURFER 25
- FSL 30

Metadata 32
Anonymization 34
Data Archives 35
Resources 36

Working with fMRI Data

Introduction

Magnetic resonance imaging (MRI) is a non-invasive neuroimaging technique that can capture both structural and functional aspects of neuroactivity. Different sequences of structural MRI can emphasize aspects of physiological features within in the brain important to the aims of a study, for example, gray matter and white matter tract density, cortical thickness, and blood flow to various areas of the brain. T1-weighted and T2-weighted images are the most often collected structural images used in neuroimaging studies. Both offer different forms of contrast between the types of tissues. Another example of structural imaging is diffusion weighted imaging (DWI), which studies directional water molecule diffusion and can provide imaging of the white matter tracts and their directionality.

Functional magnetic resonance imaging (fMRI) specifically refers to the use of MRI for measuring functional activity in the brain while completing tasks related to experiments, or, in the case of resting state fMRI, where focus is undirected to

studying the resting/default mode network activity in the brain. In most fMRI research, what we call activations are actually a measure of the brain's metabolic activity, use of oxygen and blood flow, which we call the blood oxygen level dependent (BOLD) signal.

When performing structural scans of the brain, precise, high-resolution images are taken, however fMRI uses this same technology to take rapid fire volumes (full images of the brain at a point in time). Because of the speed required and sheer volume of the data that is collected, fMRI images appear blurry and require specific preprocessing steps (discussed later in this booklet) in order to be properly analyzed. For example, even when structural traits are not a focus of an fMRI study, T1 images are obtained and used to register with the functional images in order to aid localization of activated areas.

fMRI has limitations that make pairing it with other complimentary modalities lucrative. While fMRI's strengths lie in localization of neural activation, due to the nature of the modality, it is not as clear exactly when an activation occurred, and thus its weaknesses are temporal in nature. Because of this limitation, some studies have paired fMRI with EEG when timing aspects of an experiment are of importance. Those considering utilizing EEG along with fMRI should consult our guide *Working with EEG Data*.

When performing structural scans of the brain, precise, high-resolution images are taken. fMRI uses this same technology to take rapid fire volumes (full images of the brain at a point in time). Because of the speed required and sheer volume of the data that is collected, fMRI images appear blurry and require specific pre-processing steps (discussed later in this booklet) in order to be properly analyzed. For example, even when structural traits are not a focus of an fMRI study, T1 images are obtained and used to register with the functional images in order to aid localization of activated areas.

Both an exhaustive exploration of the physics involved in MRI and an in-depth description of the mechanisms measured by MRI/fMRI are beyond the scope of this booklet. Many excellent resources on these particular topics can be found at the end of the guide. What we intend to offer in this booklet is guidance for understanding and managing of data related to these studies. It is aimed at the new beginner who perhaps has just been given access to an existent dataset and is struggling to understand what the different file types represent. It is also aimed at helping new researchers at UiO to understand what software, storage and analysis options are available to them and the data management issues common to this particular modality.

MRI Terminology

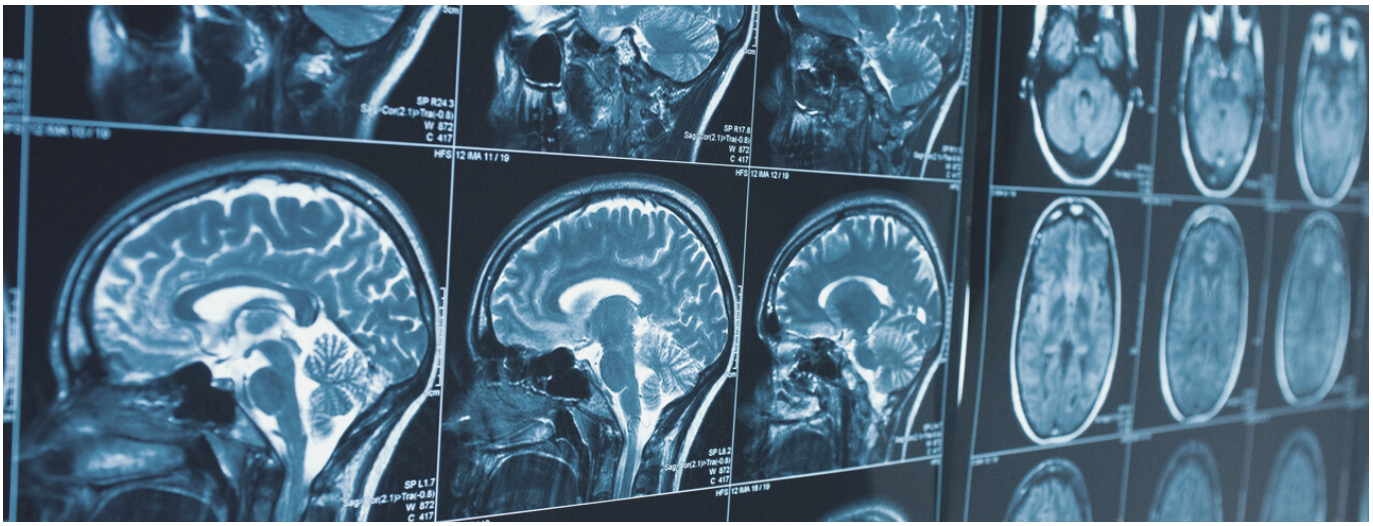
The following is a list of fMRI terminology commonly used to describe stimulation and task parameters and protocols. The list follows the terminology used by the Brain Imaging Data Structure (BIDS - https://bids.neuroimaging.io/) for structural and functional MRI.

- **Dataset.** A collection of neuroimaging and behavioral data acquired for use in a specific study. The dataset may consist of data acquired from multiple subjects, possibly over the course of many sessions.
- **Data acquisition.** An uninterrupted period of time in which the scanner was acquiring data according to a particular scanning sequence/protocol.
- **Data type.** A grouping of different types of related data. These may include: func (for functional, task-based and resting state functional MRI), dwi (diffusion weighted imaging), fmap (field inhomogeneity mapping data like field maps), anat (for anatomical, which includes structural imaging such as T1, T2, weighted images etc.), meg (magnetoencephalography), beh (for behavioral data, for example collected from tasks using hand grips, eye tracking data, pain stimuli response, etc.).
- **Session.** A grouping of neuroimaging and behavioral data that is consistent across participants. A session includes the time it takes to complete all experimental tasks. This usually begins when a participant enters the research environment for the day or segment of the day until they leave it. However, if a subject must leave the scanner room and then be re-positioned on the scanner bed, the MRI acquisitions will still

be considered to be one session and match sessions acquired in other subjects (although this should be documented in the lab book, as it may impact pre-processing and analysis). In situations where different data types are obtained over several visits (for example fMRI on one day and EEG on the next) they may be grouped in one session.

A session typically starts with obtaining informed consent and completing the MRI safety checklist. It generally ends when the participant is removed from the scanner, but can also include a number of pre- or post-observations and measurements (e.g., additional behavioral or clinical testing, blood tests, questionnaires, EEG, etc.).

- **Task.** A set of activities performed by the participant while in the scanner. Tasks usually involve stimuli and responses. Resting state scans should also be considered a task. A task is always performed in connection to one data acquisition. Even if during one acquisition the subject performed multiple conceptually different behaviors (with different sets of instructions) they will be considered one (combined) task.
- **Run.** A continuous period of data acquisition that has the same acquisition parameters and task (however events may change from one run to another in relation to different subject responses or because of randomized presentation of the stimuli). A run is essentially the same as data acquisition.
- **Event.** An isolated occurrence of a stimulus being presented, or a response being made. It is essential to have exact onset timing and duration information in addition to identify the events and when they occurred. Some tasks will not have events however (for example, resting state).



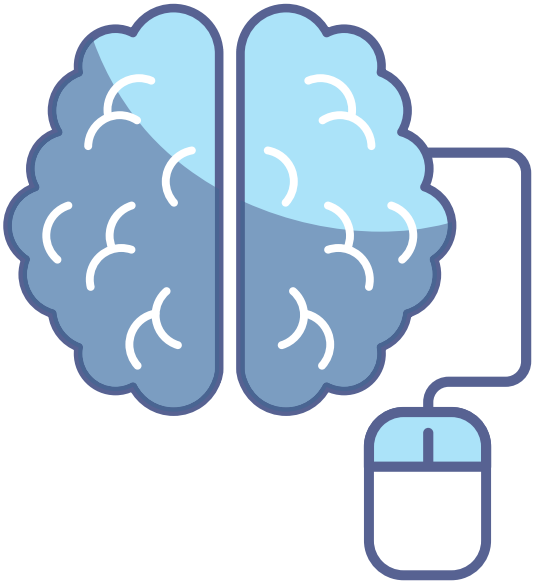
Conducting an fMRI Study

A neuroimaging study typically follows specific steps. First one sets up the experiment itself, using various software detailed elsewhere in this handbook. Then pre- and pilot testing is followed by the actual data acquisition and finishing with the analysis.

Experiment Setup

While setting up an experiment one must decide which software to use for stimulus presentation. At UiO researchers are presented with the choice between three different options: E-prime, MATLAB/Psychtoolbox and Python/PsychoPy (although some may prefer Presentation, it is less commonly used). The following table gives a short overview of the three most frequently used programs.

	E-prime	MATLAB/Psychtoolbox	Python/ PsychoPy
GUI/ Script	Yes/ limited	No/ Yes	Yes/ Yes
Price	Free at UiO; otherwise ~$1000	MATLAB free at UiO; otherwise $50-500;	Free/ open source
Ease of use	Easy	Moderate	Moderate
Flexibility	Moderate	High	High
Timing control	Moderate	High	High



Pre-tests

Pre-tests are preliminary tests of sequences, equipment and stimuli scripts performed at the scanner prior to actual pilot testing.

Pilot tests

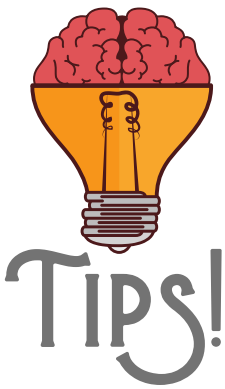
Before collecting data for the experiment, it is essential to discuss the experiment design with colleagues and have them perform the task to provide feedback on its design. Pilot test your experiment behaviorally to make sure you can obtain the predicted behavioral effect before proceeding to collect brain data.

After testing the experiment behaviorally, data from 2-5 subjects are usually collected and analyzed, including the analysis corresponding to the main hypotheses.

Pilot tests may be performed on a phantom or healthy volunteer at the scanner, depending on what the study entails. Volunteers for pilots must go through the same screening and consent process as regular participants. In addition to fulfilling the need for informed consent, we also must ensure that pilot participants do not have any health conditions or prosthetics that contraindicate MRI scanning. Pilot tests should not only include the testing of scanning sequences, tasks and stimuli, but also tests of the pre-processing and analysis pipelines should be performed at this stage. After the analysis of these datasets, flaws, suboptimal design features, extraneous conditions or the need for additional conditions may be found.

Pilot datasets can be used for the final analyses if the following criteria are met:

- The procedure/design is satisfactory.
- The experiment will not be changed after the pilot.
- The pilot participants fulfill the study's inclusion criteria.



Data Capture

Equipment

Neuroimaging studies at PSI are primarily performed using the Philips Ingenia 3T scanner at The Intervention Centre at OUS Rikshospitalet. The lab is also equipped with a 32-channel EEG amplifier, eye tracking camera and a MRI-compatible active sound system. Several devices from Nordic Neuro Labs are available for use in tasks while participants lie in the scanner, including an in-room screen for participant viewing of stimuli, response grips, a response pad and joysticks. Goggles may also be used to present stimuli during fMRI tasks. PSI is responsible for maintenance and operation of all the equipment that is not directly part of the MRI scanner.

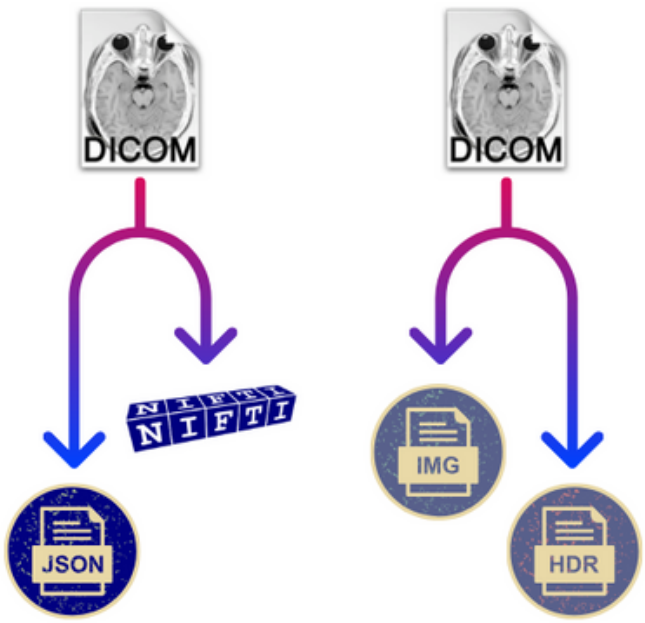
Image © Philips.

Raw data types and structure

DICOM, NIfTI & ANALYZE file formats. The most common raw form of data collected from the scanner is in DICOM file format. DICOMs from Philips scanners come in two formats, classic and enhanced. The format one wishes to use must be designated when files are transferred from the scanner to hard drive.

DICOMS are most often converted to Neuroimaging Informatics Technology Initiative (NIfTI) format prior to analysis. Most pre-processing and analysis programs use this format and it is also the format supported by the Brain Imaging Data Structure (BIDS). NIfTI files come in two forms, as two separate .img and .hdr files (.img.gz and .hdr.gz in their compressed form) or as a single file, .nii (or .nii.gz in its compressed form).

There may be notable differences in the conversion of enhanced DICOM to NIfTI format. Issues may arise when converting the raw data to files that conform to BIDS (see https://github.com/rordenlab/dcm2niix/issues/170).



It is advisable to maintain the original raw data even after NIfTI conversion. Some data repositories may prefer DICOMs over NIfTI format. Full re-conversion from NIfTI to DICOM format is not possible due to the complex nature of DICOMs and the metadata lost when converting to NIfTI format. BIDS makes up for the loss of this metadata with the creation of JSON sidecar files at the point of conversion.

It is also possible to download raw data in the NIfTI format directly from the scanner, however, most projects opt to transfer the complete DICOM files and convert them after. This is in part because DICOM files contain rich metadata embedded in the headers which are not present in the headers associated with other file types. DICOMs are stored in folders which contain images for each individual slice, while

NIfTI compiles these images into one cohesive file (except when using a BIDS-compatible convertor, in which case a JSON file is also created). The headers for DICOMs are embedded within in the file and can be read in neuroimaging software and MATLAB by calling the dicom_info function. The NIfTI compiled file format also has an embedded header, albeit lacking much of the details found in DICOM metadata. As previously stated, JSON sidecar files may be created to preserve this metadata. The two-file NIfTI format comes in .img files with the actual image data and .hdr files with the header information, but has similar metadata limitations to the .nii file. For more information on how to understand NIfTI headers, see: https://brainder.org/2012/09/23/the-nifti-file-format/.

Prior to conversion to NIfTI, DICOMS may need to be renamed and sorted. It is important to keep in mind that they may have the same name as other DICOM files when taken directly from the scanner. This can lead to some files being mistakenly overwritten if transferred to the same folders. Occasionally, researchers have also experienced that the individual DICOM files may be extracted from the scanner in an unexpected order, which will result in the need for resorting to prevent error messages when converting to other file formats. For the sake of the safety of the raw data, new, renamed files should always be directed at a different folder, while maintaining the original DICOMS in the previous folder. We recommend automating this process. MATLAB and Pydicom (python-based) are good options for this.

EEG Raw Data

Those using other modalities along with MRI will have other raw data forms to maintain. For EEG, these files are generated on the EEG laptop located at the scanner. These files are the result: the header, marker, and a binary data file. The files from the EEG at the scanner are proprietary Brain Vision files. Brain Vision is on of the file formats supported by BIDS for EEG. The other most common file type used in EEG is the proprietary Biosemi file format. For more information on these data types, please see our guide *Working with EEG Data.*

File	Extension	Type	Content
Brain Products header file	*.vhdr	Text	Recording parameter; meta-data
Brain Products marker file	*.vmrk	Text	Information about events in the data
Brain Products data file	*.eeg	Binary	Voltage values of EEG recording.
Biosemi BDF	*.bdf	Multiple	Multiple; see https://www.biosemi.com/faq/file_format.htm for full explanation

Transfer of Raw Data

Data from the MR scanner, EEG and behavioral data are transported from the hospital to the department via portable, password-protected encrypted disks. It is important that transfer of the DICOM images occurs at the time of data capture, as data is not kept long term at the scanner.

Data is transferred to Lagringshotell or TSD using the following steps. First, the responsible party contacts the data manager. The password-protected encrypted disk is then delivered to the data manager (you must also provide the password). The images are then uploaded and made available to you in Lagringshotell. General info on the usage of Lagringshotell can be found here (in Norwegian): https://www.uio.no/tjenester/it/hosting/storage/lagringshotell/.

Behavioral Data

The type of files generated when collecting behavioral data differ depending on which program was used. This guide will briefly discuss the different data types produced by the most commonly used programs: E-Prime, PsychoPy, Presentation and PsychToolbox.

E-Prime

E-Prime is a suite of programs which simplify the creation, execution and analysis of psychology experiments. It uses a number of proprietary and standard file types in creation of experiments, result reporting and analysis. File extensions in E-prime

File	Extension	Content
E-Data Aid file	.edat .edat2 .edat3	Data entry and tracking. Saves important tabular experiment data such as participant responses and event onset times. Can be opened in StatView, SPSS, Excel and E-Data Aid for analysis.
E-Studio experiment file	.es .es2 .es3	E-Studio experiment design data for E-Prime experiment files. When an .es file is changed, a new E-run file must be generated reflecting these changes.
E-Run script file	.ebs .ebs2 .ebs3	Script file used to run an E-prime experiment. While it is possible to run an experiment from the .es file, E-Run tends to be more stable.
E-Merge file	.emrg .emrg2 .emrg3	Contains merged data compiled from .edat files. Can be analyzed in E-Data Aid similarly to .edat files.
E-Prime package file	.epk .epk2 .epk3	Used by E-Studio to integrate package file scripts and routines into your experiment with package call objects. Can be edited and viewed in Package File Editor.
E-Prime start up info file	.startupinfo .startupinfo2 .startupinfo3	Used in loading name/value pairs into a context object used in an experiment.
CD file	.cd .cd2 .cd3	E-Data Aid analysis file containing column definitions.
ANL file	.anl	Used in E-Data Aid and in analysis to show pre-determined columns, rows and data.
Text file	.txt	Text files are created to hold data during an experiment. The .edat file is created from the .txt file at experiment completion. If missing, E-Recovery can create a new .edat from the .txt file.
WNDPOS	.wndpos	Stores states during E-Prime session including open windows/positions.
XML	.xml	Format E-Prime uses to generate reports.

will vary depending on the version of E-Prime used to write the file. Newer E-Prime files with 2.x and 3.x suffixes cannot be used in legacy E-Prime versions (1.x). E-Prime will ask if you wish to convert the files when used with a newer version.

Psych Toolbox

Psych Toolbox, another popular experiment and stimuli presentation program, is built as an open access add-on to MATLAB and GNU Octave. As such, the files that you will encounter, as well as the syntax, are all the same as those encountered in MATLAB. These files' content in the context of Psych Toolbox are detailed below:

File	Extension	Content
M file	.m	M files are text files which contain MATLAB executable programs in this context.
Mex file	.mex	A binary executable extension file specific to the Mac version of Psychtoolbox and MATLAB.
DLL file	.dll	A binary executable extension file specific to the Windows version of Psychtoolbox and MATLAB.
Mat file	.mat	The file format MATLAB uses for recording values. The mat file will contain response times and other important inputs from the experiment.

Presentation

Presentation is popular proprietary software option for the creation of neuroscience experiments. It has three primary file types that you will encounter. The program uses two proprietary, easy to learn languages, Presentation Control Language (PCL) and Scenario Description Language (SDL). SDL is a language used to create experiment design elements like variables or duration times. PCL is the actual programming language used for writing the actual scripts.

SDL elements are read and held in the memory prior to the execution of the code written in PCL.

File	Extension	Content
Scenario file	.sce	Scenario files are text files that present sequences of tasks related to experimental conditions. The file must have a text header which defines parameters. The files are written in PCL and/or SDL.
Template file	.tem	Template files are text files that contain trial object definitions to be read in relation to a scenario file written in the form of SDL variables.
PCL program	.pcl	PCL program files contain the actual scripts for programs. Programs can run from their own file or be embedded within a scenario file.

Data outputs from presentation come in the form of log files and output text files:

File	Extension	Content
Log file	.log	Output from the behavioral experiment about events that occurred during a scenario. Contains the event table, the stimulus table, and the video table data.
Output text file	.txt	Additional data outputs that can be designated by the researcher which are not automatically included in the log file.

PsychoPy

PsychoPy is a free and open source presentation software which is functional across operating systems. Researchers can create their experiments using its Builder interface or by writing code in Python. PsychoPy additionally makes it possible to conduct experiments online. One benefit of using PsychoPy is that it is free and doesn't require a license for costly programs like MATLAB or E-Prime. Additionally. PsychoPy is also a great alternative for those who wish to adhere to the open science initiative

because it is accessible for all when sharing experiments along with datasets. It uses basic xml files which can be read in other programs in addition to PsychoPy. The experiment files generated by the program or researcher are:

File	Extension	Content
XML file	.pyexp	Text file created by the PsychoPy Builder. Contains the experiment.
XML file	.xml	Text file created by the PsychoPy Builder. Contains the experiment.
Python script	.py	Executable file containing a Python script. Researchers who prefer to write their own programs may save them in this format. PsychoPy will generate .py files with the script for the experiment from the Builder input as well, which can facilitate learning the language.
Delimited values file	.csv	This comma-separated values file holds conditions for the experiment and is created manually by the experiment creator.
Excel file	.xlsx	Open XML spreadsheet file that holds parameters and conditions for the experiment. Whether one uses a .csv or .xlsx file is a matter of personal preference.

PsychoPy creates a folder after a run called "data" as well as a .py file. Inside of the folder you will find the data for analysis. The files you will see are as follows:

File	Extension	Content
Delimited values file	.csv .tsv	In this context, the delimited values file holds the data collected during the run, for example, onset times, responses and response times.
Log file	.log	A text file containing chronological details about the run. This can be helpful for spotting problems if they arise.
Excel file	.xlsx	The table created in this file contains data much like the delimited values file, only in a different format.
TrialHandler/ Stairhandler object	.psydat	TrialHandler or StairHandler object which may be used by experienced Python users and contains slightly more data than the .csv or .xlsx files.

Analysis

Data analysis and preprocessing for MRI/fMRI typically is performed using a VDI, often accessing files stored in Lagringshotell or within TSD using its own VDI, while accessing files also stored there. The lab engineer can be contacted for more information about how to connect and use VDI.

VDI – Virtual Desktop Infrastructure. A VDI offers users access to a virtual computer with the software and processing power they need. This computer can be used in the same way you use your local computer but can be reached from different devices and operating systems. Which programs that are mounted and can run on the VDI machine is decided together by yourself, your local IT and program managers at the departments. A VDI may offer advantages over using one's office desktop computer for analysis, as the VDI processing capabilities are more powerful.

Software

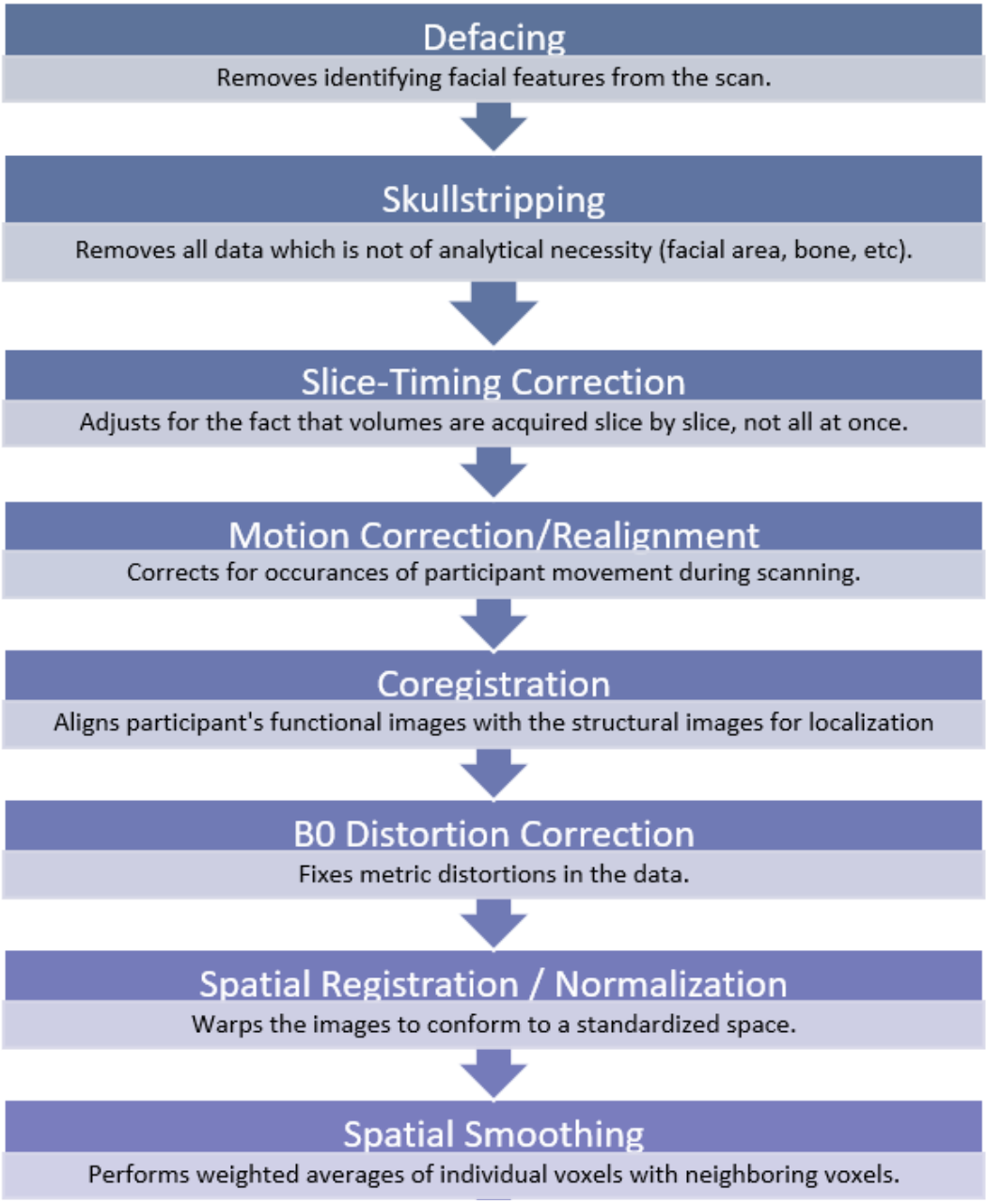
Several programs have been developed by the neuroimaging community to analyze MRI/fMRI data. These different programs and toolboxes often have different goals, strengths and weaknesses that must be considered prior to deciding which one to use. Some of these decisions will be based upon a researcher's personal preference, others will be strategic and based upon the study's specific needs. The open-access nature of some of the programs may also lead researchers to choose one type of software over another. Some researchers additionally opt to run some parts of their preprocessing and analysis using in-house written code.

The following sections will first outline the different steps involved in preprocessing, and then continue with an overview of some of the most commonly used programs employed in preprocessing and analysis, their data outputs and structures.

Preprocessing

fMRI data is inherently noisy. As a result, a number of preprocessing steps must be performed to prepare your data prior to analysis. Preprocessing creates a 4D dataset (the 4th dimension being time) from what begins as a 3D dataset. It also improves the signal to noise ratio. Some steps are also performed to anonymize the data, improve localization within subjects by co-registering the T1 to the structural images and across subjects by warping the images to fit a universal template.

These steps can be carried out in many of the same programs which are used to analyze the data. Each step will create its own output file, which varies slightly depending on the program, but the output is usually a new NIfTI file.



It is important to remember that different researchers, as well as different software, may perform these preprocessing tasks in an order that differs from the visual on the previous page. In some cases, researchers may choose not to employ some preprocessing steps. These decisions are ultimately made based on the study's needs. The following section will provide a brief description of the various steps involved in preprocessing of fMRI data and their purpose.

Defacing and Skull-stripping

One method of ensuring anonymization of fMRI data is by using *defacing* software. Defacing software removes the voxels associated with facial features or makes them unreadable. One concern is that some of the algorithms used in defacing may inadvertently remove data relevant to a study's purpose (Bischoff-Grethe et al., 2007). Some investigators will prefer not to deface the data because of this.

Similarly, *skull-stripping*, or *brain extraction*, removes all voxels that are not necessary for analysis, leaving just the brain, without bone, dura, surrounding air, etc. While this is yet another way of ensuring that the data is anonymized, as well as cutting down on the amount of space the data takes up in storage, not all researchers will choose to perform this step for many of the same reasons as defacing. It is essential to conduct a quality check to ensure the extraction results are precise if they are used.

Artifact-correcting Preprocessing Steps

There are several sources that can contribute to noise and artefacts in fMRI data. Steps can be taken when developing protocols such as making adjustments to TE (time to echo), TR (time to repeat), carefully planning which sequences are used, and adjusting parameters (such as slice thickness or field of view) in order to reduce noise and artefacts (Bell & Yeung et al., 2019). Some noise and artefacts, however, will inevitably need to be dealt with after the images are acquired. This can be accomplished during the preprocessing phase.

Motion correction/realignment. This step corrects for any movement that participants may have made in the scanner by aligning all of the functional images with one reference image (often the first or mid-point image). It is important to check the data carefully after this step.

Slice-timing correction. fMRI analysis assumes that all slices in a volume were taken simultaneously. This, however, is not the case. This step corrects for the fact that each slice from the total volume is taken at a different point time due to the nature of MRI data collection and adjusts for the slight delay.

B0 distortion correction. This step corrects distortions that result due to the B0 magnetic field inhomogeneity.

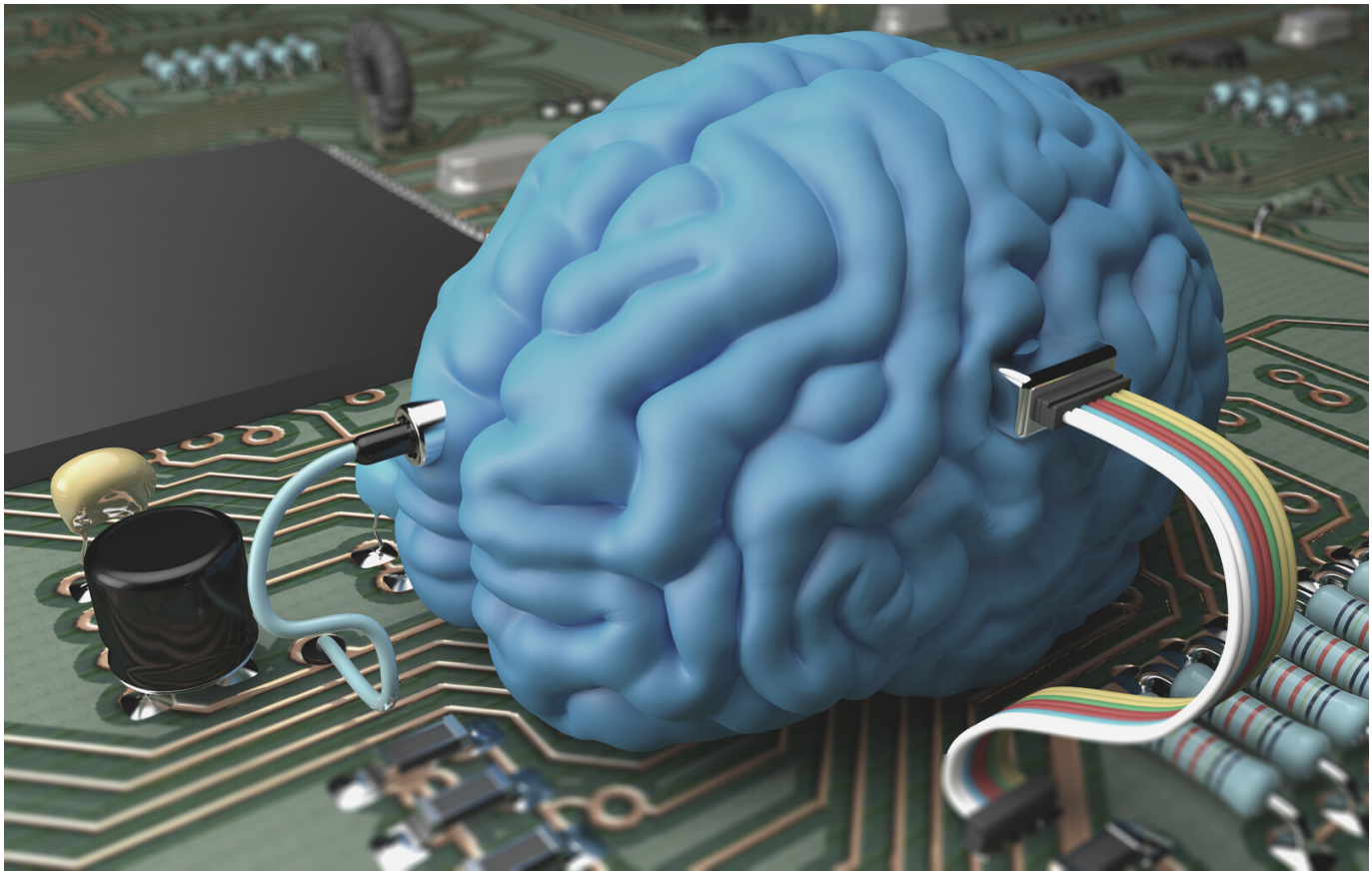
Spatial Preprocessing Steps

There are several sources that can contribute to noise and artefacts in fMRI data. The following steps are used to correct these issues.

Coregistration. In this step, the T1 and/or T2 structural images are used for co-registration of the functional scans so that the functional images align with the anatomical structures/brain regions of the participant.

Spatial Normalization. Human brains can have significant variations from one participant to the next. Thus, when comparing subjects, it is important to ensure that all of the data conforms to the same space so that a voxel in one subject represents the same location compared to another subject. This is achieved by warping the data to a template/atlas brain. This step is necessary to second level analysis in order to compare across participants.

Spatial Smoothing. This step corrects for any limitations of the normalization step by blurring any leftover anatomical variation, improving signal to noise ratio and inter-subject registration. Smoothing may not be performed in studies of only one subject.



Analysis Software & Data Structures

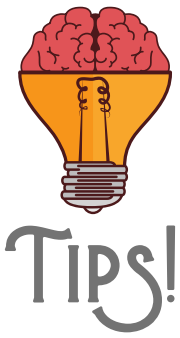
The following section is a basic primer on the different data types produced during analysis by the most commonly used analysis software. It should enable new users who either receive a dataset that is already analyzed or are new to the programs to understand the data structures that are produced by that software and what the different components entail in terms of analysis.

Analysis strategies themselves will vary dramatically depending on what task is being analyzed, and thus, a complete overview of fMRI analysis strategies is beyond the scope of this guide. Several approaches can be used depending on the goals of the study; be they localization of activation in the brain, studying connectivity between regions or to study predictive models. The aims of the study may also dictate which programs and adjacent toolboxes are chosen for analysis.

SPM12

SPM12 is the most widely used neuroimaging analysis tool and can handle various modalities in addition to fMRI. A wide range of open access toolboxes are available for use with SPM12, broadening its capabilities. The program was developed by University College London and is based upon theoretical concepts of Statistical Parametric Mapping. SPM can be a good program to begin learning analysis with due to its easy to use GUI. Although the package itself is free, it must be used with MATLAB or Octavia.

SPM also has several corresponding toolboxes for use in analysis. Like SPM these are open access. For example, MarsBar is used for region of interest (ROI) analysis, Conn is primarily used for functional connectivity analysis, and CAT can be employed for more accurate segmentation and normalization during pre-processing. For a comprehensive list of the different toolboxes available for use with SPM as well as the corresponding SPM versions, see https://www.fil.ion.ucl.ac.uk/spm/ext/.



It is important not to change versions of SPM mid-project, as the files outputted by SPM12 may not properly load in earlier versions. There are also some differences in how preprocessing and other tasks are performed in previous versions of SPM. This is important to be aware of if you are working with an older dataset which may have been preprocessed in an older SPM version.

Preprocessing and File Naming Conventions in SPM12

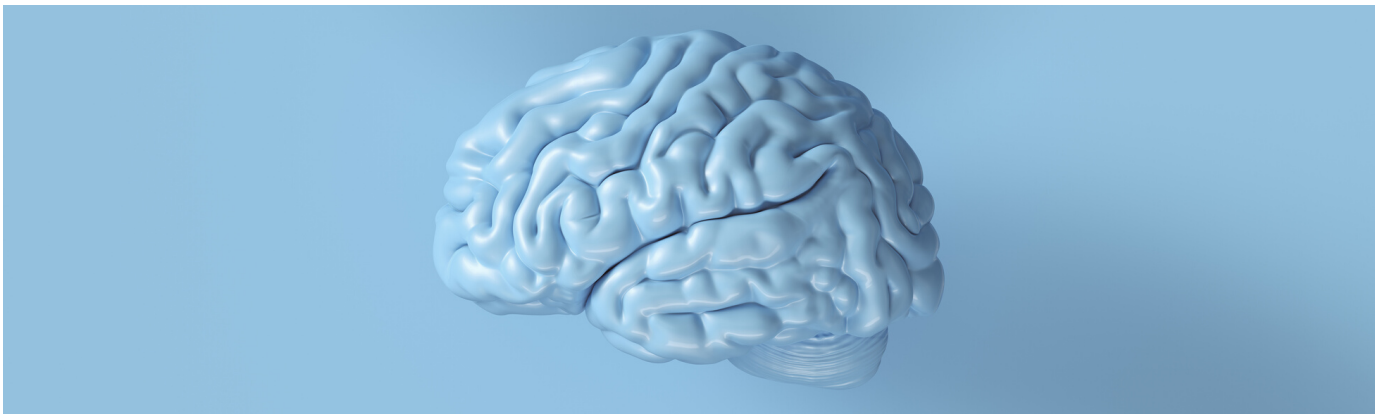
As you are carrying out the different stages of preprocessing, SPM will automatically add (prepend) a prefix letter to the beginning of the file name to prevent overwriting previous files. Although you may designate different prefixes in the batch editor, the defaults are generally well known by SPM users and so maintaining the defaults may help others understand your dataset. The following are the default prefixes:

	Default prepended files in SPM12
r	Realigned images, co-registered bitmap images
a	Slice-timing corrected bitmap images
w	Normalized bitmap images
s	Smoothed bitmap images

SPM12 Data Overview

The table that follows gives a very basic overview of the file types that are produced under analysis with SPM12.

File	Extension	Type	Content
NIfTI-1	.nii .nii.gz .img .hdr	Image	The NIfTI format compiles a whole brain volume into one file. NIfTI can either be represented by one file (.nii) or two files, the image (.img) and a header (.hdr). When a NIfTI file is compressed its extension is .nii.gz. The .nii file contains a bitmap with the data in the image and a transformation matrix which forms the bitmap in a 3D coordinate system.
GIfTI	.gii	Surface	Geometry format made to correspond with NIfTI files. Used for surface-based data in SPM. These files can also be used in FSL and Freesurfer.
ANALYZE	.img .hdr	Image	Older file format that predates the NIfTI-1. The NIfTI format is based upon ANALYZE. SPM will read old ANALYZE files.
M file	.m	Text	MATLAB file containing code in the MATLAB language.
Mat file	.mat	Text	MATLAB specific file storing MATLAB-formatted data.
MEX files	.mex	Exe	MEX stands for MATLAB-executable. These contain compiled functions to perform some of the more complex operations in SPM.
spmT files	spmT*.img	Image	T-score maps that are created to correspond with the con images.
spmF files	spmF*.img	Image	F-score statistical maps
Beta files	Beta_*.img	Image	Beta files are used in first-level analysis to generate the con files used in second-level analysis. You can expect to have one beta file for each column present in your design matrix.
Con files	Con_*.img	Image	Con files contain the T contrast summary statistic images produced by first level (subject level) analysis and show the actual effects when comparing two conditions. The con file is the result of one beta contrast image subtracted from another, i.e. the weighted sum of the related beta files. Con files are used in second level (group/population) analysis. You should have one con file for each contrast specified in first-level analysis. Which contrasts these are is not designated in the file name but can be found by examining them in the contrast manager in SPM.



AFNI

AFNI is another open source program for analysis of fMRI data. It was developed by the National Institutes of Health in the United States. One downside is that it runs only on Unix-based operating systems. Some prefer AFNI for its versatility, more fine-grained options for exploration and visualization of data and for analyzing specific types of data, such as resting state fMRI analysis. AFNI now operates with GUIs for many tasks, you may want to have some familiarity with Unix to truly utilize its features. AFNI runs with C as its primary programming language.

AFNI Data Overview

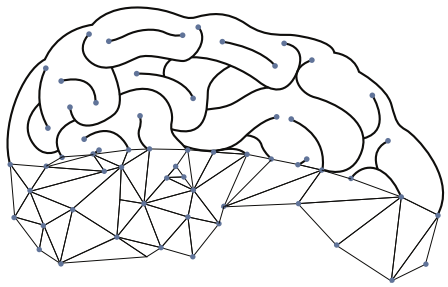
3D arrays in AFNI are called sub-bricks. There is one number per voxel in each sub-brick. Datasets are stored in directories, which are called sessions, because they contain the data from one scanning session with a participant.

AFNI has its own unique set of file systems and extensions that one should be aware of when examining a dataset analyzed with the program:

File	Extension	Type	Content
NIfTI	.nii .nii.gz	Image	Contains the image numeric values. Replaces the Brik and Header files from previous versions. AFNI now writes the Brik and Header into one NIfTI file.
Brik file	.BRIK	Image	Older format for AFNI data sets. Contains the numbers in the sub-bricks.
Header file	.HEAD	Header	Contains metadata/auxiliary info.
Log file	.log	Text	
1D file	.1D	Text	Text file specific to AFNI, indicates tabulated data. For example, you will need to create a 1D file for onsets in your experiment, with one for each condition.

AFNI has a series of strings present either in the file name or descriptors which inform the researcher of what type of image they are viewing and what pre-processing or analysis steps have been performed on them. The following are some examples:

String	Attributes
epan	Functional images. Short for 'echo-planar image'.
anat abuc	Structural images.
volreg	Aligned and co-registered data
+tlrc	Normalized data
blur	Smoothed data
skull	Skull-stripped image
mask	Mask applied to differentiate voxels relevant to the analysis
output.proc	Text file containing output from scripts executed by the GUI. You can check this file for errors, warnings and aborted tasks should problems arise.
stats	Statistical datasets
x	Files related to the design matrix



FREESURFER

Freesurfer is an open source software for Linux and MacOS for the processing and analysis of fMRI data. Freesurfer is commonly used in preprocessing, when a researcher wants to create models from their fMRI data, and to measure the morphometric features of the brain such as regional volumes and cortical thickness. It may also be used to average inter-subject structural and functional data based on cortical folds to produce alignment of different neural substrates. The program creates a 2D surface mesh from the 3D volumes to better locate sources of activation. This is helpful in cases where a particular voxel covers two different areas of the brain or two different tissue types. Boundaries are traced by the program and an inflated brain is produced, much like an brain-shaped balloon. It is also possible to further expand this inflation to the shape of a sphere for comparison across subjects. Rather than thinking about your data in terms of voxels, Freesurfer approaches the data in terms of vertices and edges (Jahn, 2019).

Parcellation is the act of labeling the different brain regions, which Freesurfer performs using two atlases, the Desikan-Killiany atlas and the Destrieux atlas. Freesurfer notoriously uses up quite a lot of processor time and memory, so you will want to run Freesurfer in an environment that supports analysis with a significant amount of processing power. Otherwise reconstruction and analysis of a large dataset can take many days. When using Freesurfer on a very large database, you may find that it is necessary to use a supercomputer. Because of these features, some researchers may find the prospect of using Freesurfer inconvenient. However, it may be worth the hassle depending on what you hope to accomplish with your fMRI data.

Freesurfer uses its own simple command language and jargon set that users must become acquainted with. Some examples of this jargon include:

Term	Meaning
Recon	Shorthand for reconstruction.
Surfaces	Boundaries in a 3D construction of the brain
Vertex	Each point in a surface that gives its 3D shape is called a vertex.
Parcellation	Labelling of the cortex in Freesurfer
Segmentation	Labelling of the subcortical structures in Freesurfer
Registration	Spatial alignment to assist comparison of data within subjects and between subjects.
Pial surface	2D representation in which red areas represent the sulci and green areas are the gyri. Blue areas represent the activation map.

Freesurfer Data Overview

Freesurfer creates its own complex file types for different stages of the processing and analysis. These native file formats include (listed by category):

Surface

File type	File extension	File attributes	Naming and structure
Surface	.white	Refined version of the grey and white matter boundary estimates.	lh.white rh.white
Surface	.pial	Represents the pial surface delineating the grey matter.	lh.pial rh.pial
Surface	.orig	Estimates of the grey and white matter boundaries.	lh.orig rh.orig
Surface	.inflated	Inflated brain created from the .pial file.	lh.inflated rh.inflated

Volume

File type	File extension	File attributes	Naming and structure
COR Volume Format	.COR	Stores structural data in a high-resolution 3D format. Voxel values are stored in coronal slices.	Contains one coronal slice per file, with a standard naming convention of COR-001 to COR256. Because the naming convention is the same across images, there can only be one COR volume in a single directory, due to the risk of overwriting.
COR volume header file	.info	Header file with metadata for the COR volume. Stores the parameters in ASCII text format.	COR-.info is the file name. There can only be one header file in a single directory, due to the risk of overwriting.
Bshort Volume Format	.bshort .hdr	Contains blocks with signed short integers to represent data about the image data. Bshort header files contain info that designates the number of pixels in each of all 3 dimensions (Y, X, and time).	The files can be named as needed but must have the file extension .bshort. The header file should have the same name as the .bshort file but with the .hdr extension.
Bfloat Volume Format	.bfloat .hdr	Stores floating point numbers representing the image data in blocks. Files have a short header containing the number of pixels in each of 3 dimensions (Y, X and time).	Files are named as needed and end in .bfloat. A matching header file that ends in .hdr must accompany the .bfloat.
MGH/MGZ Volume Format	.mgz .mgh.gz .mgh	Stores high resolution structural data. A .mgz (or .mgh.gz) file is a compressed .mgh file (similar to a NIfTI)	In example: files outputted during recon are orig.mgz (original image), T1.mgz, brainmask.mgz (a skull-stripped volume), wm.mgz (white matter mask), aseg.mgz (subcortical segmentation) & filled.mgz (subcortical mass).

Transform

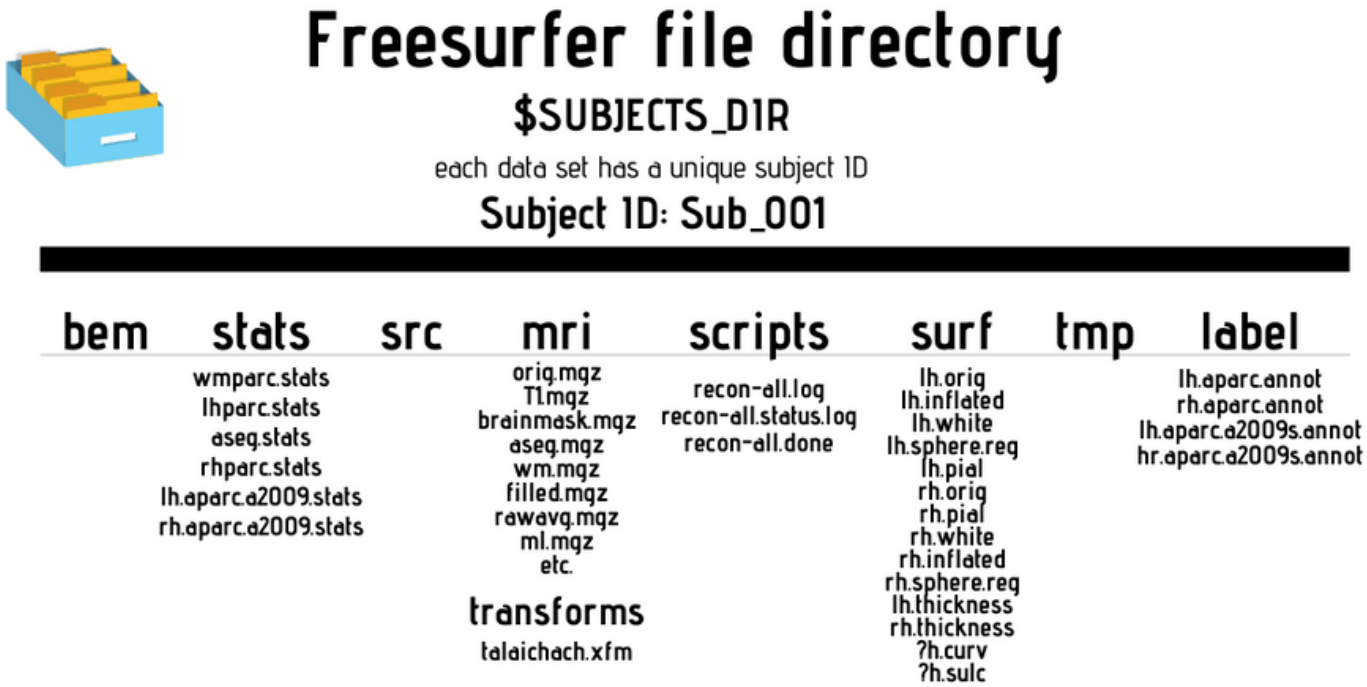
File type	File extension	File attributes	Naming and structure
tkregister-type registration matrix	.dat	The format most Freesurfer programs use to store a 4x4 matrix transformation matrix.	
Talairach transform file	.xfm	Linear Talairach transform file containing a 3x4 matrix found in the <subject name>/mri/transforms directory.	talairach.xfm

Other

File type	File extension	File attributes	Naming and structure
Label file	.label	A text file containing a list of vertices pertaining to a region, including spatial positions(using R,A,S coordinates). A label file only relates to a single label, containing a single list of vertices.	Usually researchers name the label file in a manner that indicates the label intended. Ex: lh.cortex.label
Annotation file	.annot	Contains segmentations for the two different atlases.	lh.aparc.annot (Desikan-Killiany atlas) lh.aparc.a2009s.annot (Destrieux atlas)
Freesurfer Group Descriptor (FSGD)	.txt .fsgd	File that generated to designate covariates/group labels (called 'classes' in Freesurfer).	Can be named after the study. Must be stored in the FSGD directory which the researcher creates.
Contrast file	.mtx	Specifies contrast weights. .Mtx stands for matrix because it designates your contrast matrix. You should have as many .mtx files as you have contrasts. Mtx files are also used to designate covariates.	You should name the file based on the contrasts they represent (for example: RC-RM.mtx; RM-RC.mtx). Must be stored in a CONTRAST directory.
Statistical results	.stats	Contains results from statistical testing.	

Freesurfer additionally uses its own program to view the data and analysis results, Freeview. The program can be used to view standard formats like NIfTI, as well as the Freesurfer-specific formats generated during preprocessing and analysis.

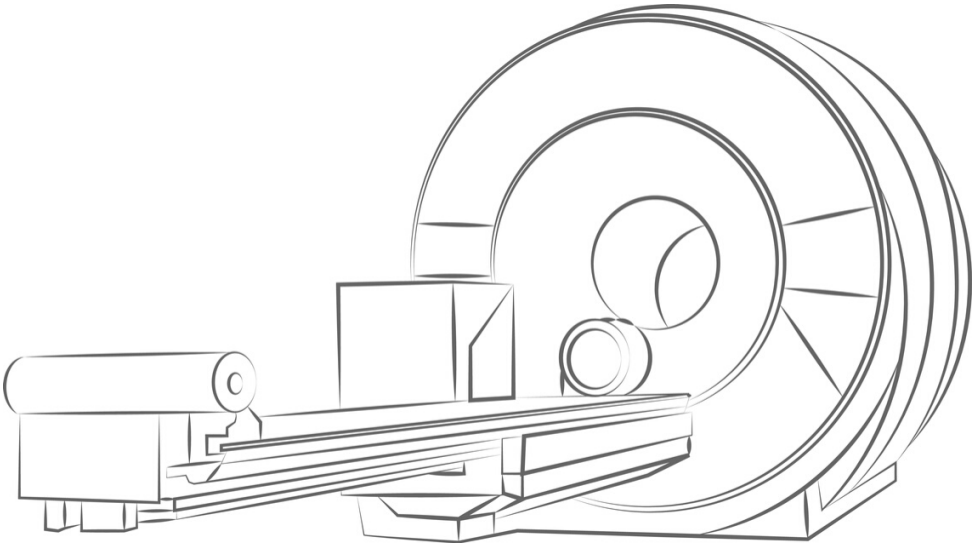
Freesurfer operates with a specific file directory schema. The $SUBJECTS_DIR contains the outputs of recon-all commands. 3D volumes are found in the 'mri' folder, while regions of interests and atlas annotations are found in the 'label' subdirectory. The 'scripts' directory includes log files of events that occurred while running recon-all. The 'stats' directory contains structural measures for the thickness and volume of each parcellation, while the 'surf' directory contains your surfaces, such as pial and inflated surfaces. See the directory mockup on the next page for an overview of the directory structure and the files that are commonly found within each directory.



Parts of the output file names in Freesurfer are separated by periods. Below are the most common file segments/naming conventions created by Freesurfer during analysis and what they mean in terms of the file contents:

File segment	Meaning
lh	Left hemisphere.
rh	Right hemisphere.
aseg	Both hemispheres.
beta	Parametric estimates from the model in group analysis.
sig	Significance at each vertex.
z	Transformed z-scores.
pcc	Correlation coefficients.
gamma	Effect sizes.
cache	Cluster simulation performed using cached simulations recon-all.
th#	Designates the vertex-wise threshold used (i.e. th13 = 0.05).
pos	For positive, designates direction of the test.
neg	For negative, designates direction of the test.
abs	For absolute, designates 2-tailed test.
cluster.summary	Listing of each statistically significant cluster.
cluster.mgh	File used to visualize the results related to the cluster.summary file.
stats	Statistical results.
aseg.stats	Contains segmentations.

After performing second-level analysis and cluster correction (to account for multiple comparisons) the following outputs will be produced:
Cache.th13.pos.pdf.dat
Cache.th13.pos.sig.cluster.mgh
Cache.th13.pos.sig.cluster.summary
Cache.th13.pos.sig.masked.mgh
Cache.th13.pos.sig.ocn.annot
The cluster.summary will list your statistically significant clusters and the cluster.mgh file will allow you to view your results in Freeview. (Jahn, 2019) ROI outputs, when desired, are created by Freesurfer as tab delimited text files (.txt).

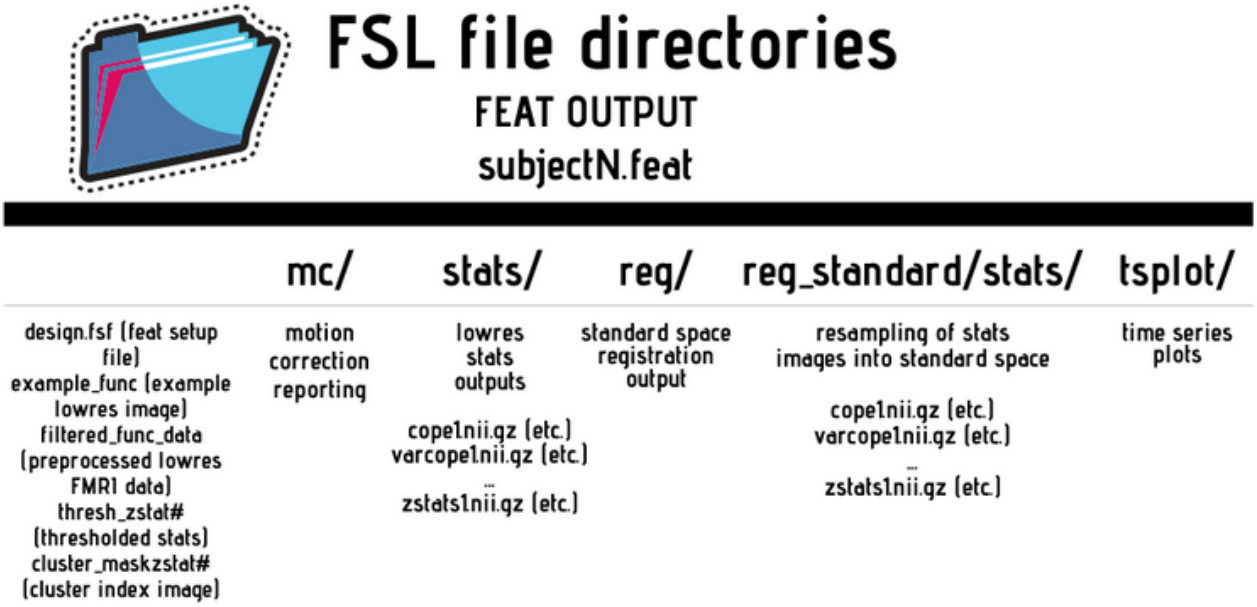


FSL

FSL was created by University of Oxford. It can be used with Windows, Mac and Linux operating systems. Like AFNI, you will need to have some familiarity with Unix and a shell-like bash or tcsh to take full advantage of the program. This can be a drawback for new beginners, but it does have an easy to use GUI to assist new users. Preprocessing and analysis in FSL is performed using the application FEAT.

FSL Data Overview

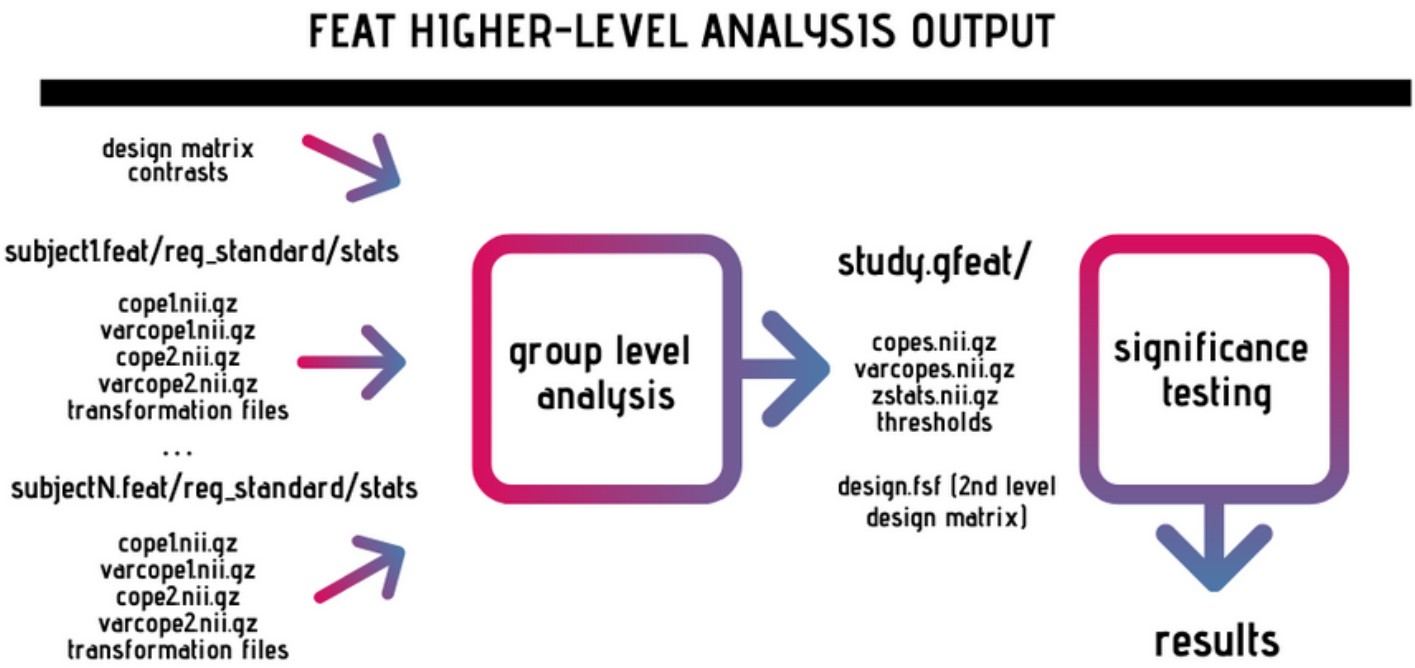
FSL reads and produces the same file types used by SPM, namely, the NIfTI and ANALYZE standards. You will need to create your own timing/onset files in the .txt format for FSL to read. You will also need to create a separate .txt file for each condition and run. FEAT creates its own FEAT output directories for the results. For a comprehensive listing of the file structure and outputs of these directories, see https://poc.vl-e.nl/distribution/manual/fsl-3.2/feat5/output.html.



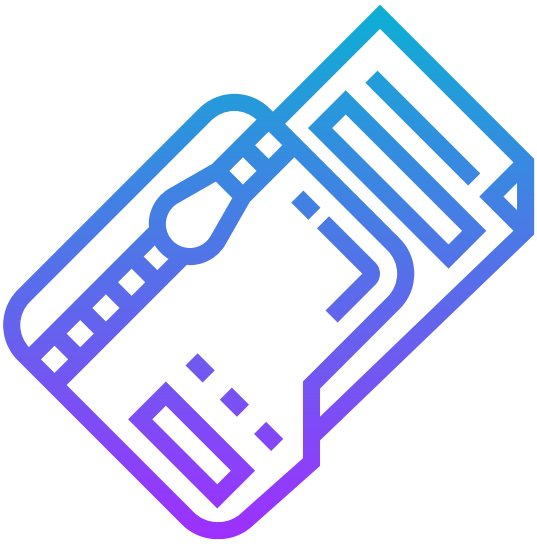
FSL also has a jargon that must be understood for the purposes of analysis and file naming conventions:

Term	Meaning
EV	Explanatory Variables. FSL calls regressors EVs.
COPE	Contrast of Parameter Estimate. Term FSL uses for contrasts.
VARCOPE	The variance of the COPE.
DOF	Degrees-of-freedom. Uncertainty of the variance of the COPE.

Below is a basic visualization of the different inputs/outputs related to second level analysis in FSL:



Metadata



Metadata is data about data. Metadata needed to understand basic traits of the data should be easily readable in file names and structures. Those working with the data should take care to document all steps implemented in data processing. A lack of this information may impede replication. A list of dependencies should be made explicit in the metadata files for your project (Wilson et al., 2017).

Researchers must record exactly what processes were involved in preprocessing of the data, which software was used, which operating system the preprocessing was performed on, and ideally, what order tasks were performed in. Any anomalies like drop-out that could not be corrected for, or a high degree of movement in a certain subject should also be recorded in the metadata pertaining to that subject's scans, as well as if that subject left the scanner during the session (for example, to use the bathroom). If data is compared across scanners (or significant upgrades/maintenance were performed on a scanner during the study), B0 correction may help correct for this, but these challenges should be noted, as these can impact the data in significant ways.

Researchers must be aware that the broad range of methods and software employed in analysis may create problems when it comes to reproducibility. For example, the software, software version or even operating system used can produce sometimes wide variation in results (Bowring, Maumet & Nichols, 2019). Even programming languages and their packages change over time, which can result in difficulty running the same analysis using the same scripts later (Nichols et al., 2017). Thus, documenting which programming language, its version and/or software and its version, as well as which operating system was used is important and should be part of the metadata.

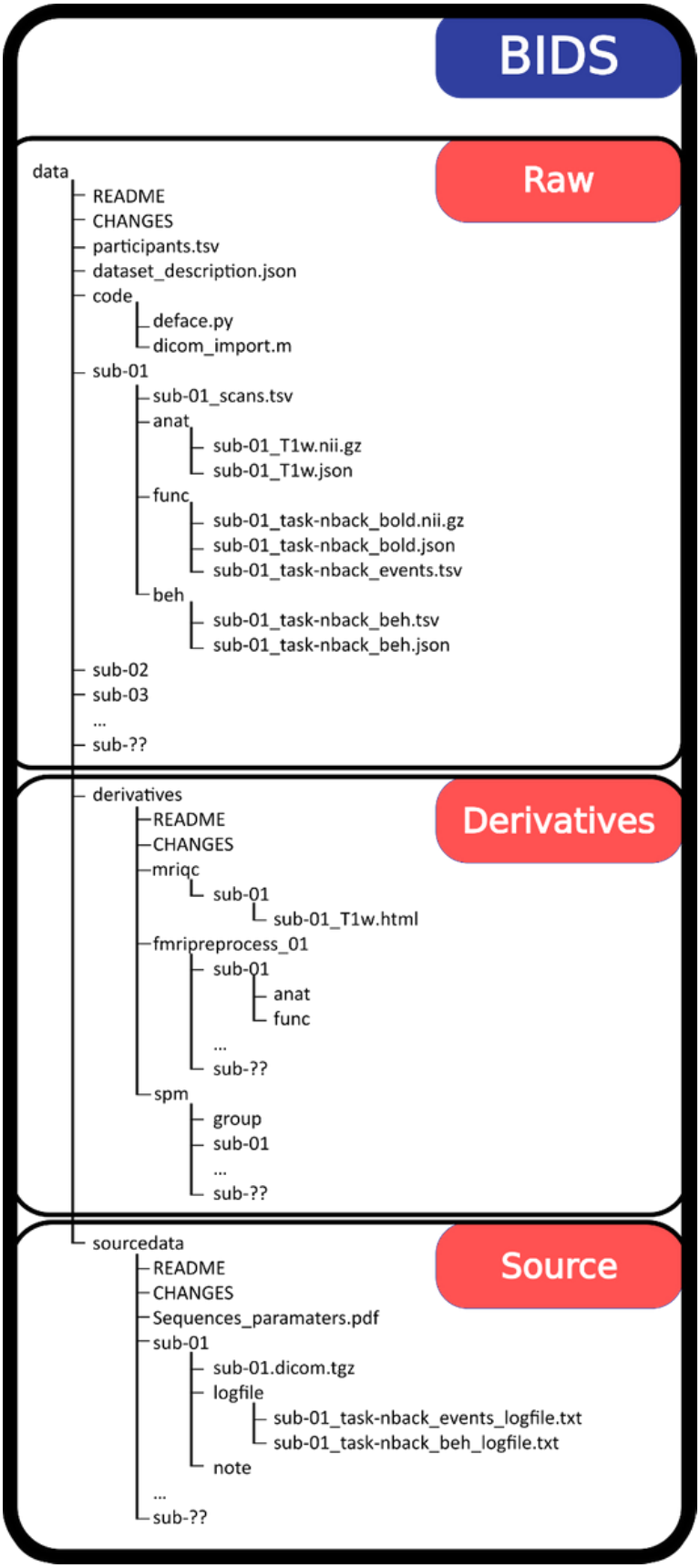
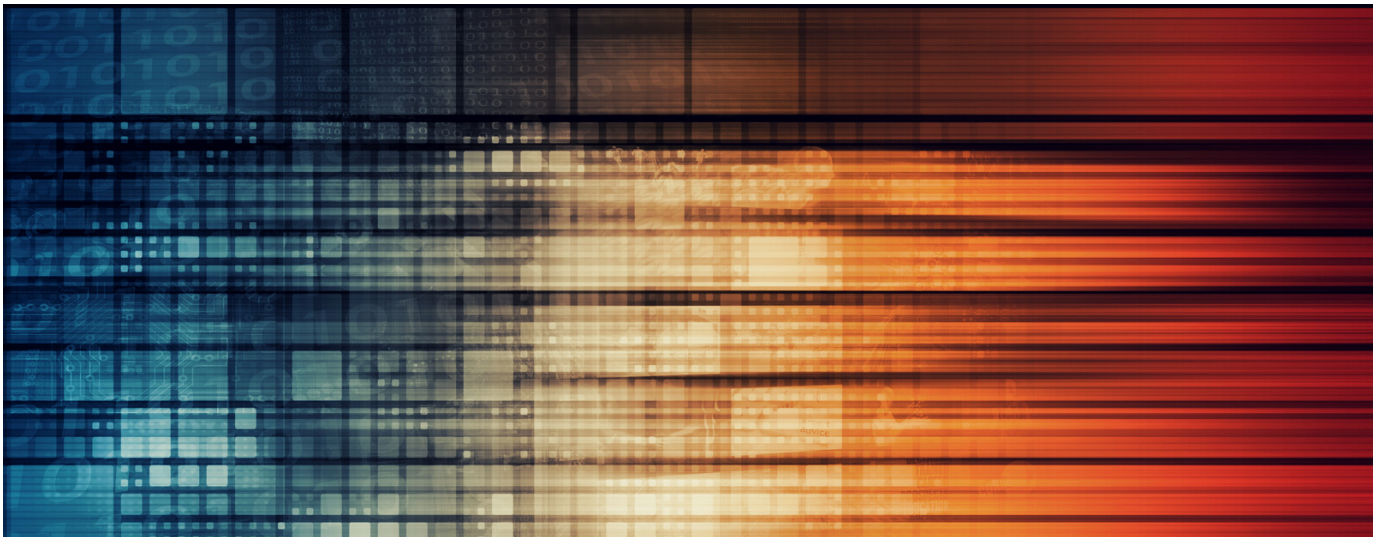


Image source: The Bids Starter-Kit

BIDS

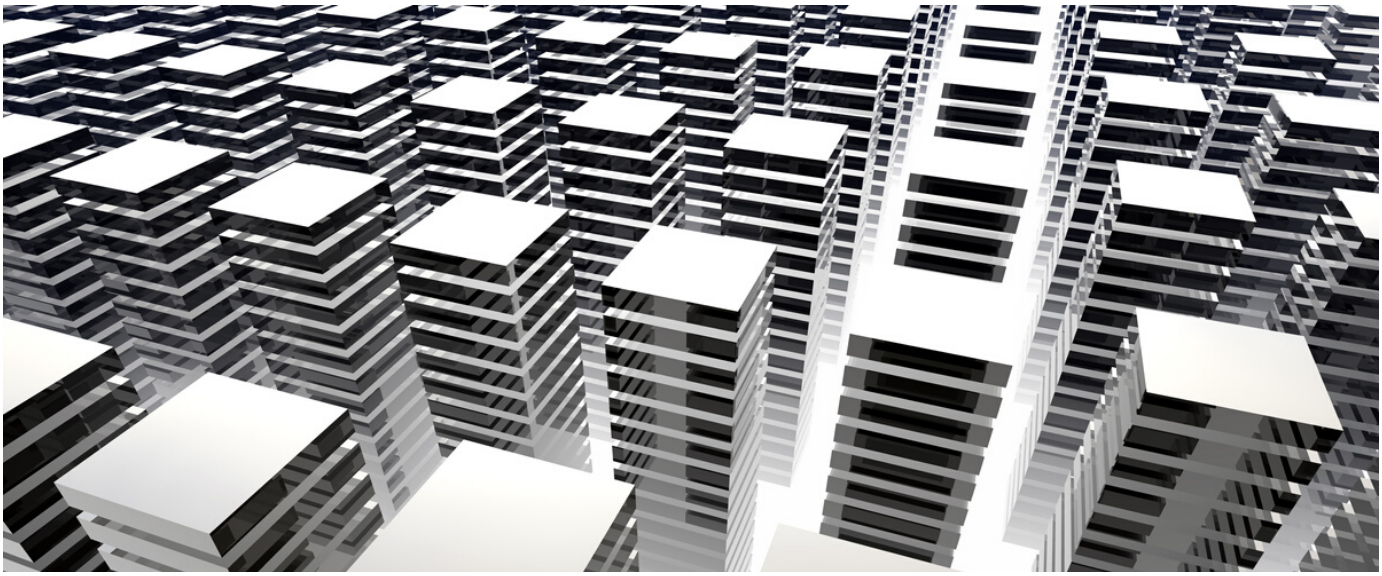
The brain imaging data structure (BIDS) (Pernet et al., 2018) is a format created to facilitate the sharing of neuroimaging data by using agreed-upon standards created by the neuroimaging community itself. BIDS offers a systematic way to organize data into folders using dedicated names, in association with text files, either as tabulated separated value file (.tsv) or JavaScript Object Notation file (.json) to store metadata. We encourage the local neuroimaging research community to share their data using this data structure as it results in greater ease of communication, reproducibility and the development of data analysis pipelines. It also facilitates compliance with the FAIR principles of Findability, Accessibility, Interoperability, and Reusability. See our handbook *Structuring Data with BIDS* for more information on how to use BIDS for fMRI and EEG data.



Anonymization

Neuroimaging data is inherently sensitive and requires special care to ensure safe handling of data. Data is removed from the scanner on password-protected, encrypted hard drives. At the point of transfer of data from the scanner to external encrypted hard drive, one must always remember to check the box for anonymization of the data at the scanner console and designate a participant number or alias instead. This is because DICOMs have headers that contain identifying patient data. If anonymization is not performed at the source, the DICOM headers must be anonymized by hand, which can be time-consuming. Storage of neuroimaging data is permitted only in TSD or Lagringshotellet at UiO. Some patient groups may only be stored in TSD. Prior to data archival and sharing, neuroimaging data should be defaced or skull-stripped. Some may opt to instead make preprocessed data available, however this raises questions as to whether better methods of preprocessing data might be applied in the future (Nichols et al., 2017).

It is important to anonymize data from the start of data collection by providing participants with participant numbers and keeping any identifying information like name, address, phone number, birthday or national identification number separate from the neuroimaging data. Name, contact information and the subject's ID are not stored together. Documents where ID and name are linked are stored on encrypted storage mediums. Those devices should be stored in locked cabinets separate from the data. These steps are of even greater importance when one is working with patient populations.



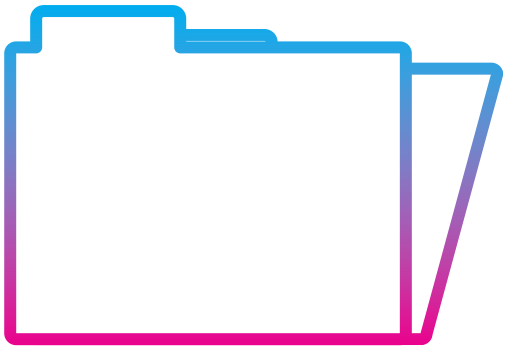
Research data archives

There are national, international, and domain-specific archives that meet international standards for archiving research data and making it accessible. UiO's researchers can choose the archiving solutions that are most appropriate to their discipline and that meet the conditions of applicable legal frameworks. Depositing data resources within a trusted digital archive can ensure that they are curated and handled according to best practices in digital preservation.

Neuroimaging is moving toward open science but there remain several hurdles to this process including policies regarding ownership of the data, individual attitudes toward sharing and ownership, fears that errors will be revealed, the resources involved in curating, storing and sharing data and anonymity (Nichols et al., 2017).

Some archival resources for MRI data are:

- Re3data.ord (a global list of archives)
- Zenodo (EU's archive)
- Open fMRI (domain specific)
- OpenNeuro (domain specific)
- Neurodata (domain specific)
- Neurovault (domain specific)
- The fMRI Data Center (domain specific)
- NIRD/Sigma 2 (national archives)
- DataverseNO (national archives)



For more information, see our guide *Data Management: DMPs & Best Practices.*

Recommended Resources

- Bell, D.J. & Yeung, J. et al. (2019). *Signal to noise ratio.* Radiopaedia. Retrieved 27.06.2019 from https://radiopaedia.org/articles/signal-to-noise-ratio-1.
- Bischoff-Grethe, A. et al. (2007). A Technique for the De-identification of Structural Brain MR Images. *Human Brain Mapping.* 28(9): 892-903.
- Bowring, A., Maumet, C. & Nichols, T.E. (2019). Exploring the impact of analysis software on task fMRI results. *Human Brain Mapping.* 2019; 1-23.
- Flandin, G., & Friston, K. J. J. S. (2008). *Statistical parametric mapping (SPM).* 3(4).
- FSL. (2019). The FSL Wiki. [Website]. https://fsl.fmrib.ox.ac.uk/fsl/fslwiki
- Gorgolewski, K. et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Nature Scientific Data.* 3:160044.
- Gorgolewski, K. J. & Poldrack, R.A. (2016). A Practical Guide for Improving Transparency and Reproducibility in Neuroimaging Research. *PLOS Biology.* 14(7): e1002506.
- Gorgolewski, K. et al. (2017). BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methhods. *PLOS Computational Biology.* 13(3): e1005209.
- Henson, R.N. et al. (2019). Multimodal Integration of M/EEG and f/MRI Data in SPM12. *Frontiers in Neuroscience,* 13: 300.
- Hermans, E. (2016). *SPM12 Starters' Guide.* https://www.ernohermans.com/wp-content/uploads/2016/09/spm12_startersguide.pdf
- Jahn, A. (2019a). Freesurfer Tutorial #9: Cluster Correction. [Website]. *Andy's Brain Book.* https://andysbrainbook.readthedocs.io/en/latest/FreeSurfer/FS_ShortCourse/FS_09_ClusterCorrection.html
- Jahn, A. (2019b). Andy's Brain Book. [Website]. https://andysbrainbook.readthedocs.io/en/latest/

- Mindhive. (2020). Nutshell SPM. [Website]. *Mindhive – A community porthole for MIT brain research.* http://mindhive.mit.edu/node/85
- National Institutes of Health. (2020). *AFNI* [Website]. https://afni.nimh.nih.gov/
- Nichols., T.E. et al. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience.* 20(3): 299-303.
- Pierce, J. et al. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods,* 51: 195–203.
- Psychtoolbox-3. (2020). *Psych Toolbox-3* [Website]. http://psychtoolbox.org/
- Schimke, N., Kuehler, M., & Hale, J. (2011). Preserving Privacy in Structural Neuroimages. *Data and Applications Security and Privacy.* XXV, LNCS 6818: 301-308.
- Spapé, M., Verdonschot, R., & van Steenbergen, H. *The E-Primer: An introduction to creating psychological experiments in E-Prime.* Second Edition. https://tinyurl.com/y729hhby
- The Wellcome Centre for Human Neuroimaging. (2020). *SPM* [Website]. https://www.fil.ion.ucl.ac.uk/spm/software/
- Wagner, T.D. & Lindquist, M.A. (2015). *Principles of fMRI.* LeanPub.
- Wilson, G. et al. (2017). Good enough practices in scientific computing. *PLOS Computational Biology*. 13(6): e1005510.

Elian E. Jentoft & Rene Skukies

Human Time Data 2019

Funded by Fagråd for eInfrastruktur