Archiving and Sharing Functional MRI data

Dr Cyril Pernet

Centre for Clinical Brain Science, NeuroImaging Science

This document describes how functional MRI data are expected to be stored for archiving and data sharing. It deals with issues of data organization, metadata and de-identification.

Preamble

The data can be shared using the university DataShare service (http://datashare.is.ed.ac.uk/). Three extremely important points have to be considered:

1. Do you have consent from the participants to share the data?
2. All images must be de-identified (no information in the hearder)
3. Structural images (T1, T2, etc) must be defaced

Data Organization

Functional MRI data can have a complex structure and the organization proposed follows the Open Brain Imaging Data-Structure from INCF (http://www.incf.org/activities/our-programs/datasharing/neuroimaging-task-force) and matches closely the schema from OpenfMRI (https://openfmri.org/)

The root folder is the study number, starting from ds000001_Ed. Inside the root folder, each subject has a folder (sub001_Ed, sub002_Ed, etc), with subfolders: anatomy (T1, T2, etc), diffusion, fieldmap, functional. Note that contrary to OpenfMRI, no subject model folder or group folder (statistics) are expected. Instead, we favour sharing analysis workflows and scripts that would re-create such folders.

The Data format is 3D (T1, T2, etc) and 4D (fMRI, DTI) niifti (preferably nii.gz). This allows storing functional tasks, runs, sessions as separate files but within a single folder.

Directory structure for a single session fMRI

This is an overview of the folder and file structure. For details on individual files see descriptions in the next section. Files added to original OBID are highlighted in blue:

- **sub00001**
  - **session #**
    - **anatomy**
      - sub00001_T1w_001.nii.gz
      - sub00001_T2w_001.nii.gz
    - **functional**
      - sub00001_task001_run001_bold.nii.gz
      - sub00001_task001_run001_events.tsv
      - sub00001_task001_run001_physio.tsv
      - sub00001_rest001_run001_bold.nii.gz
      - sub00001_rest001_run001_physio.tsv
    - **diffusion**
      - sub00001_dwi_001.nii.gz
      - sub00001_dwi_001.bval
      - sub00001_dwi_001.bvec
    - **fieldmap**
      - sub00001_fieldmap_001_phase.nii.gz
      - sub00001_fieldmap_001_magnitude.nii.gz
      - sub00001_fieldmap_001.tsv
- task_key.tsv
- scan_key.tsv
- experiment.txt

Directory structure for a multiple session fMRI

- **sub001**
  - **sess001**
    - **anatomy**
      - sub001_sess001_T1w_001.nii.gz
      - sub001_sess001_T2w_001.nii.gz
    - **functional**
      - sub001_sess001_task001_run001_bold.nii.gz
      - sub001_sess001_task001_run001_scan_key.tsv
      - sub001_sess001_task001_run001_events.tsv
    - **diffusion**
      - sub001_sess001_dwi_001.nii.gz
      - sub001_sess001_dwi_001.bval
      - sub001_sess001_dwi_001.bvec
    - **fieldmap**
      - sub001_sess001_fieldmap_001_phase.nii.gz
      - sub001_sess001_fieldmap_001_magnitude.nii.gz
      - sub001_sess001_fieldmap_001_scan_key.tsv
    - sub001_sess001_session_key.tsv
  - **sess002**
    - **anatomy**
      - ….

- ….
  - ○ sub001_participant_key.tsv

- task_key.tsv
- scan_key.tsv
- experiment.txt
- fmri_protocol.zip

Naming

For Structural scans

Name	Suffix	Description
T1 weighted	_T1w_	
T2 weighted	_T2w_	
T1 map	_T1map_	quantitative T1 map (likewise for T2)
FLAIR	_FLAIR_	
Proton density	_PD_	
Combined PD/T2	_PDT2_	
Inplane anatomy	_inplaneT1_	T1-weighted anatomical image matched to functional acquisition (likewise for T2)
Angiography	_angio_	

Meta-Data

1 – General Information and experimental procedure(s)

Experiment.txt

At the folder root is expected to find a document that specifies the data provenance:

- purpose of the study and hypotheses

- people/institutions/funders involved in the data collection

- people involved in the data analysis

- a detailed description of the experimental protocol and, if not provided, a link to a repository to the experiment program/stimuli

fmri_protocol.zip

A simple zip file of the experiment(s) program/code used for the fMRI experiment. Alternatively, this program can be put in an external repository and simply linked into the Experiment.txt document.

2 - Experimental conditions and MRI information

The following files are expected to be .tsv, which stands for Tab Separated Values. This means these are simple text files and each word or value are separated using a tabulation (tab key).

Field Maps

The corresponding **.tsv* file notes the original echo times (in seconds) of the acquisitions.

Diffusion

The sub00X_sess00X_dwi_00X.bval and sub00X_sess00X_dwi_00X.bvec files follow the FSL format. The bvec files contain 3 rows with n space-delimited floating-point numbers (corresponding to the n volumes in the relevant Nifti file). The first row contains the x elements, the second row contains the y elements and third row contains the z elements of a unit vector in the direction of the applied diffusion gradient, where the i-th elements in each row correspond together to the i-th volume with [0,0,0] for non-diffusion-weighted volumes.

```
.bval example:
0 0 0.021828 −0.015425 −0.70918 −0.2465
0 0 0.80242 0.22098 −0.00063106 0.1043
0 0 −0.59636 0.97516 −0.70503 −0.96351
```

The bval file contains the b-values (in s/mm^2) corresponding to the volumes in the relevant Nifti file, with 0 designating non-diffusion-weighted volumes.

```
.bvec example:
0 0 2000 2000 1000 1000
```

fMRI

task_key.tsv and scan_key.tsv

These files are at the root level. The task_key.tsv describes tasks with one line per task. For instance:

```
task001    stop  signal    with  manual      response
task002    stop  signal    with  letter      naming
task003    stop  signal    with  pseudoword  naming
```

The scan_key.tsv describes the MRI acquisition using the following fields. Note slice_order is T for top to bottom, B bottom to top, IT or IB for interleaved from top or bottom – and slice_timing is the time at which each slice was acquired during the acquisition, this is not the slice order - it describes the time (sec) of each slice acquisition in relation to the beginning of volume acquisition. Phase encoding is RL for right-to-left encoding or AP for anterior-to-posterior.

```
repetition_time   3.0
echo_time         0.03
flip_angle  78
slice_order IT
slice_timing  0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8
      3.0
multiband_factor  4
parallel_factor   2
phase_encoding_direction RL
```

sub001_participant_key.tsv

At the subject level, this file describe properties of participants such as age, handedness, gender etc.(same as in demographics.tsv, see below). It is useful especially for multiple sessions and longitudinal data where properties are changing.

```
session_id    acq_time     pre_session_systolic_blood_pressure
sess00001    2009-06-15T13:45:30    120
sess00002    2009-06-16T13:45:30    100
sess00003    2009-06-17T13:45:30    110
```

sub00001_task001_run001_events.tsv & sub00001_task001_run001_physio.tsv

At the functional folder level, these files specify the different events (the log file from an experiment but tabulated) as well as any physiological recordings.

Other meta-data

Other relevant metadata should be provided at the root level – for instance demographics.tsv (also document the content of such file in the Experiment.txt file).

```
ds003  sub001      M      25
ds003  sub002      M      18
ds003  sub003      F      22
ds003  sub004      F      25
ds003  sub005      M      22
ds003  sub006      M      38
ds003  sub007      M      36
ds003  sub008      M      19
ds003  sub009      M      20
```

```
ds003 sub010      F      19
ds003 sub011      F      21
ds003 sub012      M      19
ds003 sub013      F      29
```

3 – Analysis workflow (optional)

Rather than storing analysis results, we encourage archiving analysis workflows and script. Ideally, a single script is called to run the whole analysis (if you have multiple scripts for doing separate task, simple create a new script that call those scripts in the right order). Providing the code is key for replication and will ensure greater reproduction of your results. There is no dedicated format – shell script, Matlab, Python, etc .. we are not proposing a unique way to read your script, simply storing what you have done.

Importantly, we suggest documenting/commenting the scripts to make them easy to read. A README.txt must be provided to explain how to execute the workflow/scripts.

Ed imaging Tools and utilities

Many of the steps presented above can be achieved easily. Knowing the above data structure, simply implement it straight away on your desktop/server which means no extra work. Similarly, all neuroimaging software work with nifti. The only thing remaining is to ensure de-identification. If you are ensure, please contact cyril.pernet@ed.ac.uk

Archiving the data on DataShare

Before starting you will need to register to DataShare see http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/checklist. Most of the other items on that list are already dealt with in the above.

The second aspect is to agree with the fMRI collection administrator (cyril.pernet@ed.ac.uk) on the collection name, community, and subcommunity. Data are necessarily stored under the Sub-community Edinburgh Imaging/fMRI Scans. A given collection can however also appear under our department (e.g. psychology, psychiatry, etc). Once you submit the data, the fMRI collection administrator will have to validate it before being released.

Finally, upload - most of the time, data can be transferred using your browser (Chrome or Firefox). If the data are too big (>10Gb), we will have to use another option.

Data Import

The easiest way is to use a batch import – file size is currently limited to ~2Gb for web submission, and 10Gb in the batch.

For the batch import, we need to submit an xml file which corresponds to different information to fill is using the web-interface, see https://www.wiki.ed.ac.uk/display/datashare/batch_ingest

```
<dublin_core>

  <dcvalue element="title" qualifier="none">YOUR TITLE HERE</dcvalue>

  <dcvalue element="contributor" qualifier="other">University of Edinburgh</dcvalue>

  <dcvalue element="publisher" qualifier="none">FUNDER, University of Edinburgh</dcvalue>

  <dcvalue element="type" qualifier="none">Dataset</dcvalue>

  <dcvalue element="rights" qualifier="none">ODC</dcvalue>

  <dcvalue element="title" qualifier="alternative">Alternative Title</dcvalue>

  <dcvalue element="creator" qualifier="none">AUTHOR</dcvalue>

  <dcvalue element="creator" qualifier="none">AUTHOR</dcvalue>

  <dcvalue element="ABSTRACT</dcvalue>

  <dcvalue element="relation" qualifier="isversionof">OTHER PLACES IT EXISTS</dcvalue>

  <dcvalue element="relation" qualifier="isreferencedby">PMC ID OF THE ARTICLE</dcvalue>

</dublin_core>
```