Automated Scan Downloading and Project Organisation

23042021

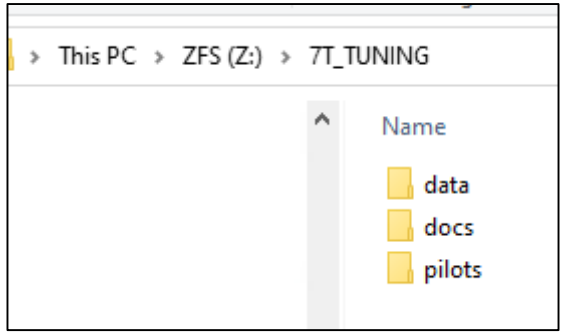
Overview:

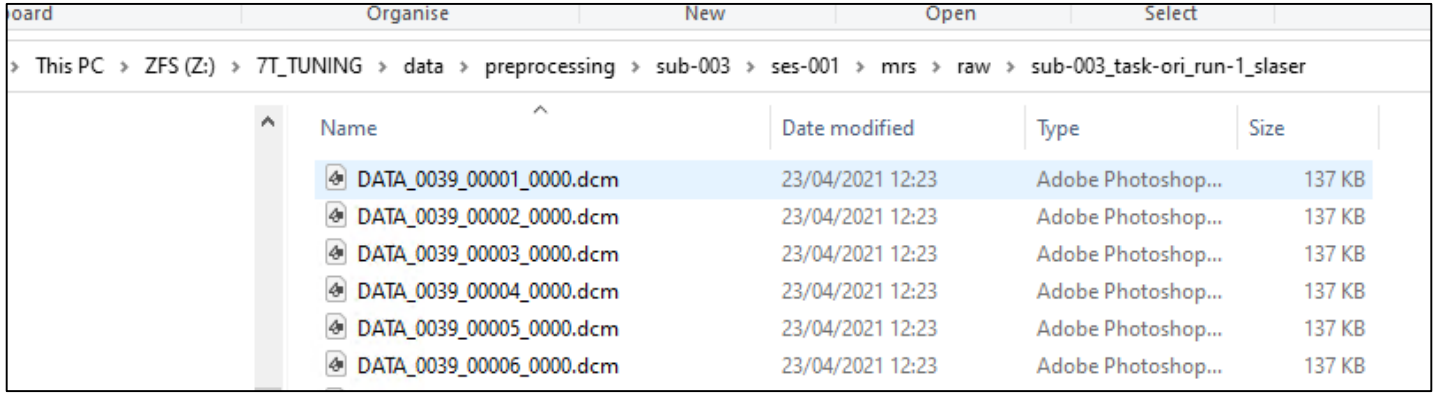
- Initialise project in pre-determined BIDS folder structure
- Download scans from WBIC, organise with Scan ID (potential for automatic backups)
- Copy user-specified raw scans to pre-processing folder
- Users only need to specify options in a configuration file – no coding required
- Object-orientated approach means easy to duplicate for new projects and customise
- Full logging and testing

Current Features:

- Download scans from WBIC to HPC then from HPC to Hivemind
- Deletes data from HPC to avoid data buildup
- Reorganise downloaded data to ABL format
- Copy raw data to preprocessing with flexible user-specified search strings.
  Sessions are automatically added and text files including all important subject IDs, scan date, time etc. written to session folder
- Optional flags for skipping runs (e.g. if a run fails)
- Every command is logged (e.g copy file, move file) for convenient checking
- All user input is tested to ensure no input errors
- Scan / Session IDs are tested against scan time / date to ensure order is corrected
- Maintains BIDS organisation throughout project

Project / BIDS structure





/mnt/zfs/PROJECT_NAME/ – docs
– logs
data
– pre-processing
– sub-00X
– ses-00X
– mrs
– raw
– sub-00X_task_XXX_run-X_slaser
– .dcms
– func …
– anat …
– mpm …
– ses-00x_info.txt (WBIC ID, ABL ID, sub, ses ID, scan zk ID, scan date, scan time

– raw_scans - zk…..001
- zk-….002

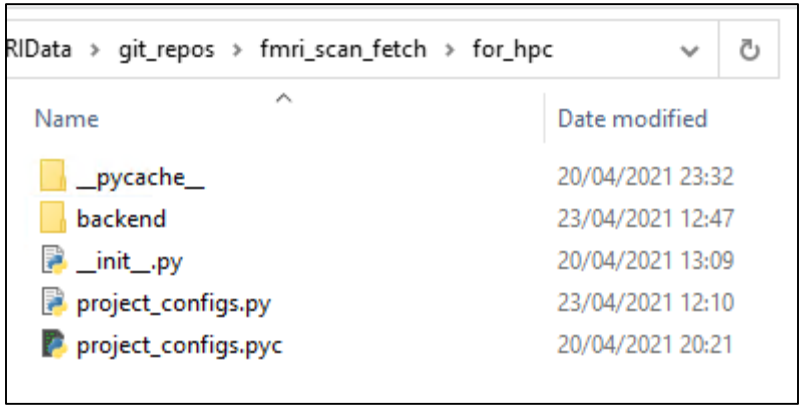
https://github.com/bids-standard/bids-specification

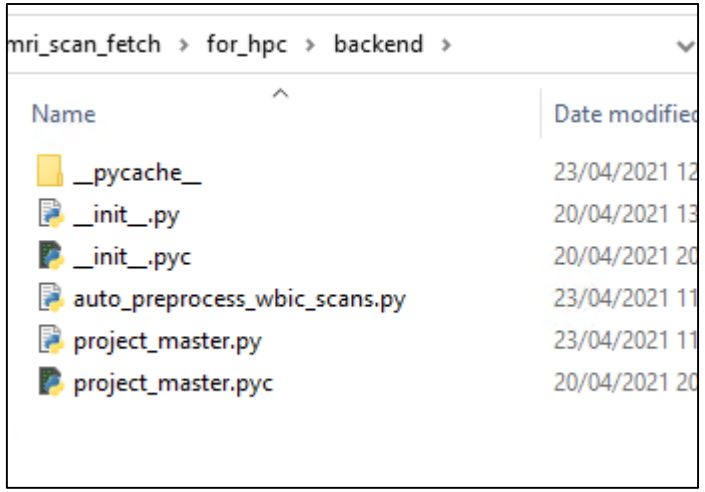
Code Organisation

/git-repos/frmi_scan_fetch/

- backend
  - all backend code
- project_configs.py (only thing user needs to change)

ProjectMaster superclass
Project subclass that inherets ProjectMaster methods and overwrites config attributes.
This means full generalisation across projects and easy customisation by method subclassing





Config File – Project Details

```
#       Project Settings ------------------------------------------------------------------------------------------

        self.base_path = "/mnt/zfs/7T_TUNING"                                   # Base directory for the folder
        self.scripts_path = "/mnt/Data/jjz33/hpc/backend"                       #
        self.project_code = "P00461"                                            # WBIC project code for the project
        self.account = "jjz33"                                                  # User ID. This should be the same for the Hivemind / HPC / WBIC
        self.scanner_format = ".dcm"

        self.mrs_scan_info = {"slaser":
                                {"search_str": "*_sLaser_WS128_OCC_w_Pad_LongTE",   # search_str used with glob.glob, see docs for it's format options (its different to regexp)
                                 "task_name": "ori"
                                 }
                              }

        self.func_scan_info = {"bold_ptx":
                                {"search_str": "*rslh_ep3d_vaso_ma4a_*TReff4s_30sl_Run?_E00_M",
                                 "task_name": "ori",
                                 },
                               "bold_old":
                                {"search_str": "*rslh_ep3d_vaso_ma4a_TReff4s_30sl_R?_E00_M",
                                 "task_name": "ori",
                                 }
        }

        self.anat_scan_info = {"mp2rage":
                                {"search_str": "*mp2rage_sag_p2_0.65mm_UNI_Images",
                                 "task_name": "ori"
                                 },
                               }

        self.mpm_scan_info = None
```

Project details

< Flexible string searching for files to copy from raw scans to preprocesing

< handles different tasks / runs / scan types

Input expected number of volumes for testing

```
        self.num_expected_func_files = 292
        self.num_expected_anat_files = 240
        self.num_expected_mrs_files = 136
        self.num_expected_mpm_files = None
```

Config File – Participants

```
#       Participants ----------------------------------------------------------------------------------------------------

        self._participant_log = {

        "31970": {"sub_id": "sub-001",
                  "lab_id": "3470",
                  "scans": {"scan_1": {"date": "20210420",
                                       "zk_id": "zk21w7_006",
                                       "time_start": "09:30",
                                       "flags": ["ignore_mrs_1"]},
                            }
                  },

        "31974": {"sub_id": "sub-002",
                  "lab_id": "3471",
                  "scans": {"scan_1": {"date": "20210420",
                                       "zk_id": "zk21w7_007",
                                       "time_start": "12:45",
                                       "flags": ["ignore_mrs_1"],
                                       },
                            }
                  },

        "31879": {"sub_id": "sub-003",
                  "lab_id": "3472",
                  "scans": {"scan_1": {"date": "20210421",
                                       "zk_id": "zk21w7_008",
                                       "time_start": "14:00"}
                            }
                  }

        }
```

All sub IDs in one place

Supports multiple scans per participant

Indicate bad runs

Logging

Logs all SSH connection information for easy troubleshooting

```
Logger initialised. Pulling scans from HPC...
starting thread (client mode): 0x176c6ba8
Local version/idstring: SSH-2.0-paramiko_2.7.2
Remote version/idstring: SSH-2.0-OpenSSH_7.4
Connected (version 2.0, client OpenSSH_7.4)
kex algos:['curve25519-sha256', 'curve25519-sha256@libssh.org', 'ecdh-sha2-nistp256', 'ecdh-sha2
ssh.com', 'umac-128-etm@openssh.com', 'hmac-sha2-256-etm@openssh.com', 'hmac-sha2-512-etm@openss
Kex agreed: curve25519-sha256@libssh.org
HostKey agreed: ssh-ed25519
Cipher agreed: aes128-ctr
MAC agreed: hmac-sha2-256
Compression agreed: none
kex engine KexCurve25519 specified hash_algo <built-in function openssl_sha256>
Switch to new keys ...
Adding ssh-ed25519 host key for login.hpc.cam.ac.uk: b'ebe3a1f06468cf9c63da84db2eee1583'
Trying SSH key b'6767daab4c9f4b757b6f4f4e6ebb2c6a'
userauth is OK
Auth banner: b'\n        <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>\n        <
Authentication continues...
Methods: ['keyboard-interactive', 'hostbased']
userauth is OK
Authentication (keyboard-interactive) successful!
[chan 0] Max packet in: 32768 bytes
Received global request "hostkeys-00@openssh.com"
Rejecting "hostkeys-00@openssh.com" global request from server.
[chan 0] Max packet out: 32768 bytes
Secsh channel 0 opened.
[chan 0] Sesch channel 0 request ok
[chan 0] EOF received (0)
[chan 0] EOF sent (0)
Dropping user packet because connection is dead.

 Fri Apr 23 00:00:00 2021 pulled scans from wbic to hpc  ------------------------------------------

project: P00461, wbic_id 31970, date: jjz33
      Patient ID Date      Study Description    Acc. Number
1    31970      20210420 WBIC Protocols Ready U-ID51976       a: 1   AAHead_Scout_32ch-head-c
                                                              b: 2   AAHead_Scout_32ch-head-c
                                                              c: 3   AAHead_Scout_32ch-head-c
                                                              d: 4   AAHead_Scout_32ch-head-c
                                                              e: 5   b1_mapping_2mm
                                                              f: 6   b1_mapping_2mm
                                                              g: 7   mp2rage_sag_p2_0.65mm_IN
                                                              h: 8   mp2rage_sag_p2_0.65mm_T1
                                                              i: 9   mp2rage_sag_p2_0.65mm_UN
                                                              j: 10  mp2rage_sag_p2_0.65mm_IN
                                                              k: 11  t2_tse_tra_fast for repo
                                                              l: 12  gre_b0map_more_slices
                                                              m: 13  gre_b0map_more_slices
                                                              n: 14  rslh_ep3d_vaso_ma4a_TRef
                                                              o: 15  AAHead_Scout_32ch-head-c
                                                              p: 16  rslh_ep3d_vaso_ma4a_TRef
                                                              q: 17  rslh_ep3d_vaso_ma4a_TRef
```

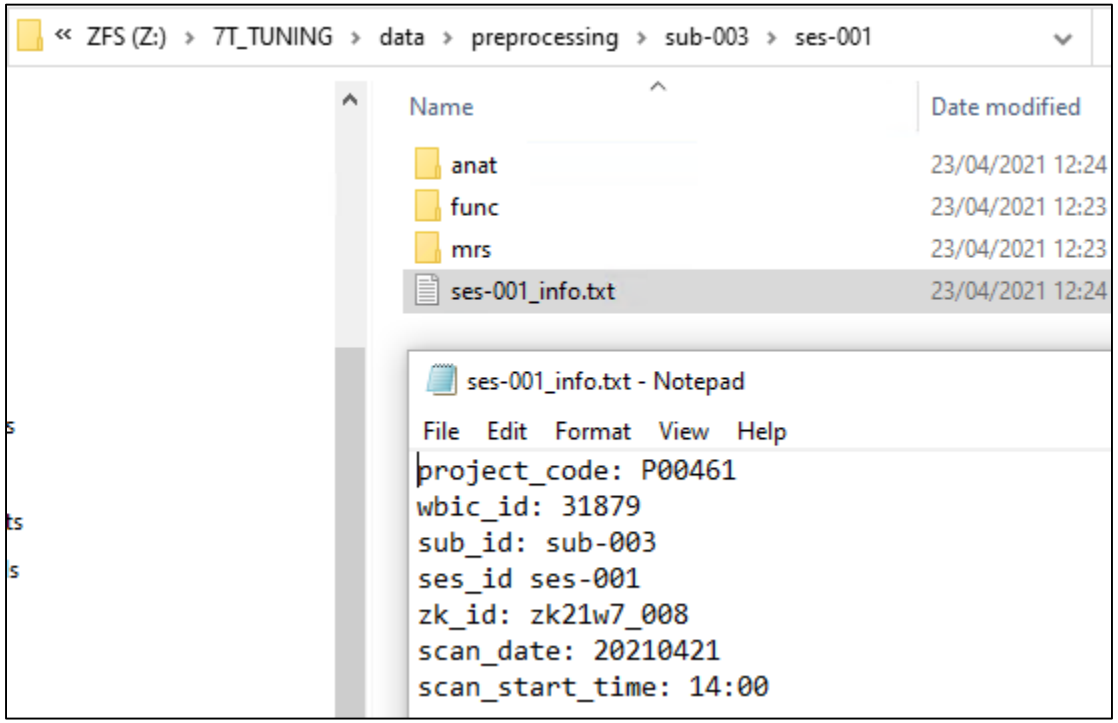
Every file operation is logged, and tests reported

```
 Fri Apr 23 00:00:00 2021 Processed a new session  -------------------------------------------------------------------------------------

sub: sub-003,
existing ses files: []
new_ses_id: ses-001

 Fri Apr 23 00:00:00 2021 copying raw data to preprocessing folder ---------------------------------------------------------------------

data type: mrs, zk_id: zk21w7_008, sub_id: sub-003

copied from: /mnt/zfs/7T_TUNING/data/raw_scans/zk21w7_008/Series_039_sLaser_WS128_OCC_w_Pad_LongTE/*
copied to: /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/mrs/raw/sub-003_task-ori_run-1_slaser
scan has the expected number of files: 136


 Fri Apr 23 00:00:00 2021 copying raw data to preprocessing folder ---------------------------------------------------------------------

data type: func, zk_id: zk21w7_008, sub_id: sub-003

copied from: /mnt/zfs/7T_TUNING/data/raw_scans/zk21w7_008/Series_015_rslh_ep3d_vaso_ma4a_TReff4s_30sl_Run1_E00_M/*
copied to: /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/func/raw/sub-003_task-ori_run-1_bold_ptx
WARNING: Only 28 scans in /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/func/raw/sub-003_task-ori_run-1_bold_ptx but expecting 292

copied from: /mnt/zfs/7T_TUNING/data/raw_scans/zk21w7_008/Series_017_rslh_ep3d_vaso_ma4a_TReff4s_30sl_Run1_E00_M/*
copied to: /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/func/raw/sub-003_task-ori_run-2_bold_ptx
scan has the expected number of files: 292

copied from: /mnt/zfs/7T_TUNING/data/raw_scans/zk21w7_008/Series_018_rslh_ep3d_vaso_ma4a_TReff4s_30sl_Run2_E00_M/*
copied to: /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/func/raw/sub-003_task-ori_run-3_bold_ptx
scan has the expected number of files: 292

copied from: /mnt/zfs/7T_TUNING/data/raw_scans/zk21w7_008/Series_019_rslh_ep3d_vaso_ma4a_TReff4s_30sl_Run3_E00_M/*
copied to: /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/func/raw/sub-003_task-ori_run-4_bold_ptx
scan has the expected number of files: 292

copied from: /mnt/zfs/7T_TUNING/data/raw_scans/zk21w7_008/Series_020_rslh_ep3d_vaso_ma4a_TReff4s_30sl_Run4_E00_M/*
copied to: /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/func/raw/sub-003_task-ori_run-5_bold_ptx
scan has the expected number of files: 292

copied from: /mnt/zfs/7T_TUNING/data/raw_scans/zk21w7_008/Series_021_rslh_ep3d_vaso_ma4a_TReff4s_30sl_Run5_E00_M/*
copied to: /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/func/raw/sub-003_task-ori_run-6_bold_ptx
scan has the expected number of files: 292

copied from: /mnt/zfs/7T_TUNING/data/raw_scans/zk21w7_008/Series_043_rslh_ep3d_vaso_ma4a_TReff4s_30sl_Run5_E00_M/*
copied to: /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/func/raw/sub-003_task-ori_run-7_bold_ptx
WARNING: Only 10 scans in /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/func/raw/sub-003_task-ori_run-7_bold_ptx but expecting 292
```

Warnings when something goes wrong

```
copied from: /mnt/zfs/7T_TUNING/data/raw_scans/zk21w7_008/Series_015_rslh_ep3d_vaso_ma4a_TReff4s_30sl_Run1_E00_M/*
copied to: /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/func/raw/sub-003_task-ori_run-1_bold_ptx
WARNING: Only 28 scans in /mnt/zfs/7T_TUNING/data/preprocessing/sub-003/ses-001/func/raw/sub-003_task-ori_run-1_bold_ptx but expecting 292
```

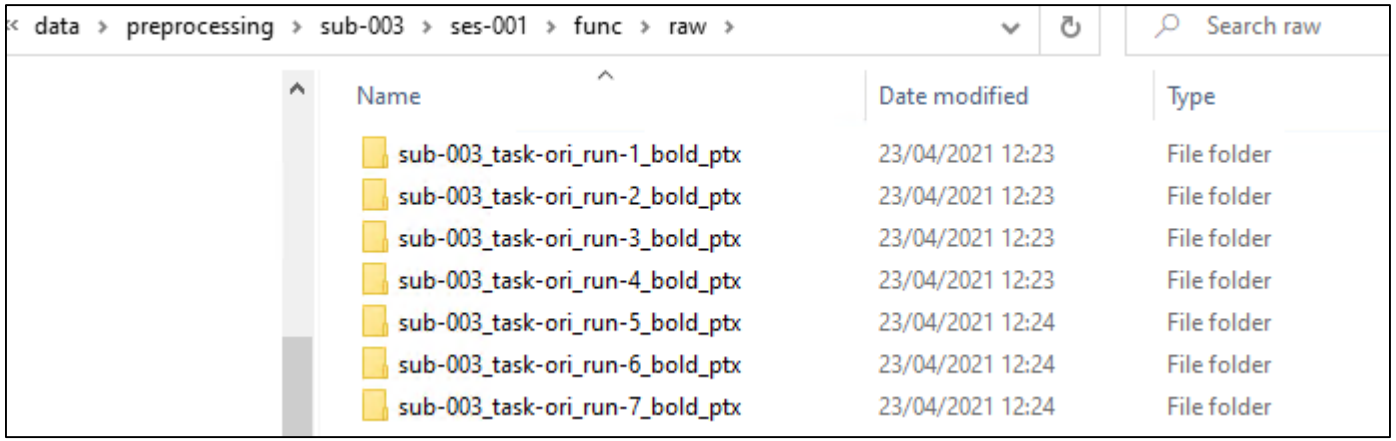
Folder Organisation and Session info file



BIDS folder structure is imposed

< all session details collated in single file per session

< scan date / times used to test subject order



BIDS file structure

Next: Automated early pre-processing steps in AFNI, FSL, SPM