HCP Users FAQ

This page contains answers to frequently asked questions from users of HCP data and HCP tools. If you don't find your question and answer here, look through the full HCP-Users list archive, join the HCP-Users email list and post questions to hcp-users@humanconnectome.org.

1. What are *dtseries.nii, *.dconn.nii, *dlabel.nii, *.dscalar.nii files?

These are CIFTI-2 format files (see http://www.nitrc.org/projects/cifti/ for specs) that are output from the HCP Pipelines (pipeline code released on GitHub). The extension for CIFTI files is .[type].nii, where type is dconn, dtseries, dscalar, dlabel, pconn, or another such specifier. It ends in .nii because CIFTI uses the NIFTI-2 format, which specifies that the extension must be .nii. The files include a specialized XML header extension to describe the indices, leaving most of the NIFTI-2 header blank/unused (spacing info, etc). Note that CIFTI files are not (and must not be) gzip'ed, as on-disk random access is used for access due to large file sizes. However, we usually gzip our volume files, so the volume files we distribute will generally end in .nii.gz.

CIFTI has the advantage of being able to handle data on surface vertices and subcortical voxels in one file. Collectively, these are termed "grayordinates". Some CIFTI files distributed by HCP (e.g., the rfMRI *dtseries.nii files) do contain data from both the surface and subcortical voxels. The voxels in these files are in MNI space via FNIRT. CIFTI files do not contain the surface coordinates – you must get those from the relevant surface file (.surf.gii in the HCP data). The surfaces used have roughly equivalent spacing to the voxels used, so that the vertices and voxels have a similar "scale". However, two thirds of the data is not in voxels, but in surface vertices.

*.dtseries.nii contain timeseries fMRI data for all CIFTI grayordinates.

*.dconn.nii contain dense connectivity matrix data (correlations between all CIFTI grayordinates).

*.dlabel.nii contain label designations and a label color key for every grayordinate.

*.dscalar.nii contain scalar grayordinate data.

Parcellated CIFTI files (e.g. *.pconn.nii, *.ptseries.nii) also exist that contain data on parcel groups of grayordinates, rather than on individual grayordinates.

Connectome Workbench (http://www.humanconnectome.org/software/connectome-workbench.html) is the tool we use with CIFTI files. You can visualize the data with the wb_view GUI, or perform various command-line operations (including extracting the data into other formats) with wb_command. For more detail on the organizational concepts of CIFTI, see "wb_command -cifti-help", also available online here: https://www.humanconnectome.org/software/workbench-command/-cifti-help.

The full list of CIFTI standard extensions are in the specification, see Appendix A in http://www.nitrc.org/forum/attachment.php?attachid=342&group_id=454&forum_id=1955 for the list, but refer to the main specification in http://www.nitrc.org/forum/forum.php?thread_id=4380&forum_id=1955 for how the special "unknown" type works (its filename still must end in .nii).

2. How do you get CIFTI files into MATLAB?

HCP has developed 2 ways to read and write CIFTI files in MATLAB:

FOR MEG USERS (This tool pads the matrix with NaNs):

A. PLEASE DON'T USE THIS OPTION WITH HCP MRI DATA AT THIS TIME. Using the code developed for the HCP megconnectome data analysis pipelines that are implemented using the FieldTrip toolbox. The code is available in stand-alone format from https://github.com/Washington-University/cifti-matlab, in the ft_cifti folder. This approach loads the complete CIFTI XML header information into a MATLAB structure, including information on what each CIFTI index represents. Where possible it will also load gifti surface files.

FOR MRI USERS (This tool does not introduce NaNs):

B. Using Workbench v1.0 or newer + GIFTI toolbox code (descriptions below). This method has the limitation of not telling you what the CIFTI indices represent (what vertex or voxel, etc) and it creates intermediate files (that are deleted as part of the provided code). For this approach you need a couple of prerequisites:

1) Connectome Workbench (available here) needs to be installed on the system.

2) A matlab set of functions called the GIFTI toolbox by Guillaume Flandin (http://www.artefact.tk/software/matlab/gifti/)

Here are MATLAB functions to download and use for importing CIFTI files (e.g., *.dtseries.nii) using these prerequisite tools:

ciftiopen.m

ciftisave.m

ciftisavereset.m

```
cii = ciftiopen('path/to/file','path/to/wb_command'); CIFTIdata = cii.cdata;
```

Some lines of analysis code:

```
newcii = cii; newcii.cdata = AnalysisOutput; ciftisave(newcii,'path/to/newfile','path/to
/wb_command');
```

Or, if the data matrix has a different number of maps/columns from what you started with:

```
ciftisavereset(newcii,'path/to/newfile','path/to/wb_command');
```

ALPHA TESTING, FOR MRI USERS (This tool is not yet ready for production use):

C. A new CIFTI library has been written for matlab, which is intended to be backwards compatible with ciftiopen, etc from option B, but provides information about what each CIFTI index represents. There are no external dependencies when reading CIFTI-2 files, which should cover all HCP CIFTI files in recent data releases. To read CIFTI-1 files with this code, install Connectome Workbench v1.0 or newer, available here. If you can accept the risk of a very new library generating incorrect output files, the library is available at https://github.com/Washington-University/cifti-matlab.

3. I see that HCP distributes group average dense connectome files. Do you also provide connectivity matrices for individual subjects?

We don't include individual dense connectomes in the HCP releases, because these files are very large (~33GB each). You can generate them by running the following on a subject's rfMRI dtseries file:

```
wb_command -cifti-correlation
```

Because the HCP rfMRI data was collected in 4 runs, you may want to 1) demean and normalize the individual timeseries, then 2) concatenate them, **before** you do the correlation step.

To do that, use these commands in wb_command:

```
wb_command -cifti-reduce <input> MEAN mean.dtseries.nii

wb_command -cifti-reduce <input> STDEV stdev.dtseries.nii

wb_command -cifti-math '(x - mean) / stdev' <output> -fixnan 0 -var x <input> -var mean mean.
dtseries.nii -select 1 1 -repeat -var stdev stdev.dtseries.nii -select 1 1 -repeat

wb_command -cifti-merge out.dtseries.nii -cifti first.dtseries.nii -cifti second.dtseries.nii -
cifti third.dtseries.nii -cifti fourth.dtseries.nii
```

You could do this for each subject (and then z-transform and average the connectomes across subjects).

You could also use -cifti-merge to concatenate all runs for all subjects after the demeaning step and follow with -cifti-correlation to make a group averaged dense connectome file.

We are working on creating a better way to do these things in the future.

4. Where do I find more information about running the HCP Pipelines?

Check out the HCP Pipeline FAQ and documentation on the GitHub HCP Pipelines distribution site.

5. Where do I download the experimental stimuli and presentation protocol software scripts for the HCP task runs?

You can download all of the HCP E-Prime scripts at: https://db.humanconnectome.org/app/action/ChooseDownloadResources?project=HCP_Resources&resource=Scripts&filePath=HCP_TFMRI_scripts.zip

Use your ConnectomeDB account for login. The HCP E-Prime scripts are available under licensing terms in the README document included in the download. You will need to have E-Prime 2.0 Professional and a dual monitor setup to use the scripts as they are. You may need to edit the scripts to suit your own purposes.

6. Where do I find information and definitions of abbreviations for the Behavioral and Demographic data on subjects?

Descriptions of the column headers for both Open access and Restricted HCP behavioral and demographic/individual difference measures are located in the HCP Data Dictionary wiki and in the Data Dictionary within ConnectomeDB (left click on the column name of any data column and select "Data Dictionary").

7. Using gradient unwarping fails, unable to find the file "fullWarp_abs", what do I do?

If you get an error like this:

```
Image Exception : #22 :: ERROR: Could not open image .../T1w/T1w1_GradientDistortionUnwarp
/fullWarp_abs

terminate called after throwing an instance of 'RBD_COMMON::BaseException'

.../scripts/GradientDistortionUnwarp.sh: line 92: 11237 Aborted                (core dumped)
${FSLDIR}/bin/convertwarp --abs --ref=$WD/trilinear.nii.gz --warp1=$WD/fullWarp_abs.nii.gz --
relout --out=$OutputTransform
```

Then what you need is the HCP-modified version of the gradient unwarp code, which is here: https://github.com/Washington-University/gradunwarp

8. There are HCP subjects that are listed as twins with the same parents (Co-Twins), but they have different exact ages or age ranges. Are these subjects not twins?

In the 500 Subjects Release there are **4** sets of twins who do not have the same exact ages in ConnectomeDB. Exact age is restricted data, but it is also used in binning subjects into age ranges, which is open access. The differences in the exact age of these 4 twin pairs is due to the fact that some twin pairs are interviewed and scanned months apart due to their individual schedules. In these 4 pairs, one twin happened to be interviewed in the months before their birthday and the other in the months after their birthday. Because twin status and co-twin IDs are restricted data, we can not be more specific here than to say that 4 twin pairs are affected.

For all HCP participants, we collect the "exact age" information at the time of the SSAGA phone interview that precedes the scanning visit. It should be noted that sometimes months can elapse between the SSAGA and the scan visit, however "exact age" is always the participant's age at the time of the SSAGA interview, not the scan date. Users should be aware of this as they conduct their analyses.

9. How do I map data between FreeSurfer and HCP?

UPDATE May 17, 2017: A new, more accurate atlas-to-atlas registration was computed between fs_LR and fsaverage. If you have resampled data between fs_LR and fsaverage prior to this date, please read Appendix 3 of the new instruction document for details.

Comparisons between HCP-derived data (including the new HCP_MMP1.0 cortical parcellation – Glasser et al., Nature, 2016) and data analyzed in FreeSurfer entail mapping between different surface 'spaces': HCP data are generally on a standard fs_LR mesh (left and right hemispheres aligned), whereas FreeSurfer data are on a native mesh or on the fsaverage mesh (in both cases, no correspondence between hemispheres). Mapping data from one surface mesh to another involves one-step "resample" options within "wb_command", plus preparatory steps that may also be needed.

We have written a document (**Resampling-FreeSurfer-HCP_5_8.pdf**, download here) that details instructions for performing mappings from:

A) fsaverage group data to fs_LR

B) FreeSurfer native individual data to fs_LR

C) fs_LR group data to fsaverage

D) fs_LR individual data to fsaverage

We recommend options (A) or (B) so as to benefit from the correspondences between left and right hemispheres provided by the fs_LR atlas. Option (B) presumes that FreeSurfer was run using mris_register, yielding a "?h.sphere.reg" native mesh sphere registered to fsaverage.

10. What is the order of pipelines for resting-state data?

If you wish to use MSMSulc and MSMAll registration (in conjunction with a recent version of the HCP Pipelines), the MSM binary is available here: https://github.com/ecr05/MSM_HOCR/releases

i) PreFreeSurfer
ii) FreeSurfer
iii) PostFreeSurfer (using MSMSulc registration)
iv) fMRIVolume
v) fMRISurface
vi) ICA+FIX (hcp_fix or hcp_fix_multi_run), followed by PostFix if you want to review the classifications
vii) MSMAll (Do MSM Surface Registration)
viii) DeDriftAndResample (Apply the MSMAll registration) (MSMRemoveGroupDrift is for the uncommon case of generating a new registration template).
ix) *Global Noise Removal* (i.e., temporal ICA; not available publicly yet, but will handle residual global noise in the data without removing the global signal)
x) Resting state analysis

For a resting state analysis, you could do something like dual regression and 'fslnets' to make netmats. You could also use the HCP's multi-modal parcellation to parcellate the MSMAll aligned dense timeseries into a parcellated timeseries, and then use fslnets to estimate functional connectivity (can be either full correlation or partial correlation—which is better if step 10 has not been done). You can also use the MIGP algorithm to combine resting state fMRI data across subjects for group ICA or group dense functional connectivity. You can use wb_command -cifti-correlation to compute functional connectivity as well.

11. How do I find the per subject in-scanner performance data for task fMRI?

We provide this data at three different levels: 1) Averaged reaction time (RT) and accuracy (ACC) measures across task runs, 2) Per run RT and ACC measures, and 3) per trial/event RT and ACC measures within each task run.

1. Average reaction time (RT) and accuracy (ACC) measures across task runs per subject are available in the Behavioral .csv file which you can download from the ConnectomeDB subject dashboard or S1200 project page, accessible via the "Explore Subjects" and "Open Dataset" options, respectively, in the HCP 1200 Subjects section of the datasets page you see upon login. If desired, you can use the filters in the ConnectomeDB dashboard to filter for subjects that have particular values for average RT and/or ACC.

2. Per run (RL or LR) in-scanner reaction time (RT) and accuracy (ACC) measures are available in each subject's preprocessed tfMRI data in a {TASK}_Stats.csv file in the run specific task directory : ${StudyFolder}/${SubjectID}/MNINonLinear/Results/tfMRI_{TASK}_{RUN}/EVs/{TASK}_Stats.csv

3. Per trial/event RT and ACC measures within each run for each task are available in each subject's preprocessed tfMRI data in a *TAB.txt file in the run specific task directory :${StudyFolder}/${SubjectID}/MNINonLinear/Results/tfMRI_{TASK}_{RUN}/{TASK}_run*_TAB.txt

   There's a row for each event (along with rows for other things you'll have to ignore), and columns that contain the reaction times for that event.

   For example:
   ${StudyFolder}/${SubjectID}/MNINonLinear/Results/tfMRI_WM_LR/WM_run2_TAB.txt has a column Stim.RT

   ${StudyFolder}/${SubjectID}/MNINonLinear/Results/tfMRI_GAMBLING_LR/GAMBLING_run2_TAB.txt has a column QuestionMark.RT

   The GLM timing files *.txt files in /EVs/ were all distilled from information in the _TAB.txt file. You can cross-reference these event rows with some of the .txt files in /EVs/ to help clarify event type, accuracy, etc.

   See the 3T Task fMRI scripts and protocol details section on p.45 of http://humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Reference_Manual.pdf for general info on these files.

   The definitions of the key variables used in these data are available in Ref Manual Appendix 6: http://humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Appendix_VI.pdf

12. How do I update my 3T 900 Subjects (S900) imaging data to S1200?

To update to the full set of released S1200 neuroimaging data, you need to do all of the following:

   a. Download the data for the subjects added as part of the S1200 release – there is a subject group for "S1200 New Subjects" in ConnectomeDB.
   b. Download the new reprocessed diffusion data for **all subjects**.
   Note: Previously released diffusion data for subjects 130922 and 150423 was removed in the S1200 release because it was either poor quality or not enough for useful analysis. *S900 data users should ensure that they remove the dMRI data from these two subjects from their local copy.*
   This is noted on the Known Issues wiki page: HCP Data Release Updates: Known Issues and Planned fixes

c. Download the data for subjects 121719 and 169040 that had MR added as part of the S1200 (these subjects were previously released in S900 as "behavior only" subjects).
This is noted on the Known Issues wiki page: HCP Data Release Updates: Known Issues and Planned fixes.

d. Download new FIX-cleaned rfMRI data for the 11 previously released subjects (Quarter Released = S900 and earlier) with Issue Code E (manual reclassification of some FIX components). See https://wiki.humanconnectome.org/pages/viewpage.action?pageId=88901591

List of these 11 subjects:

101107
111716
134728
148436
162733
178647
211720
336841
614439
628248
786569

13. Why are so many DZ twin HCP Subjects same sex twins?

In the released genotyping-verified DZ twins, there is one M/F twin pair, but only one, which is unexpected if our recruiting of twins were matched to the population. Although we did not specifically exclude DZ twins discordant in sex for HCP, we did recruit most of our twin subjects from previous twin studies that specified same sex twins (MZ or DZ) because of potential sex confounds in making comparisons between twins/twin types and to avoid sex being a confound for dizygosity.

14. How do I obtain the 7T Structural data?

We did not collect 7T structural data on the 7T subjects because there would be little benefit beyond the 0.7mm isotropic high resolution 3T T1w and T2w data already collected on these subjects and there are benefits for using a single version of the surfaces and volumes when doing 3T to 7T functional and diffusion data comparisons. Instead, we reprocessed the 3T structural data downsampled to a 1.6mm resolution and 59k mesh for working with the 7T functional data (available as the Structural Preprocessed for 7T (1.6mm/59k mesh) package).

15. How do I obtain "intermediate" files that aren't part of the released data?

Occasionally, users request access to "intermediate" files from the processing that were not included in the released packages (and thus also not available on Amazon S3). These files can be obtained via the REST interface in ConnectomeDB. A Wiki page that discusses this interface is available here: How To Access Subject Data via REST

Additionally, some posts in the HCP User list also contain examples for how to do this:
https://www.mail-archive.com/hcp-users@humanconnectome.org/msg02094.html
https://www.mail-archive.com/hcp-users@humanconnectome.org/msg05458.html
https://www.mail-archive.com/hcp-users@humanconnectome.org/msg05429.html

16. Can I get the fMRI data in a subject's 'native space'?

We have not computed this, because of the increase in storage and computational time necessary to create a 'native space' volumetric time-series with all distortions removed.

If desired, users can produce a 'native space' version of a given 'MNI space' fMRI time-series using spline interpolation and the following code. Note that this particular example resamples the FIX-cleaned rfMRI ("hp2000_clean.nii.gz") data into 'native space':

```
flirt -interp spline \
-in ${StudyFolder}/${Subject}/T1w/T1w_acpc_dc_restore.nii.gz \
-ref ${StudyFolder}/${Subject}/T1w/T1w_acpc_dc_restore.nii.gz \
-applyisoxfm ${fMRIRes} \
${StudyFolder}/${Subject}/T1w/T1w_acpc_dc_restore.${fMRIRes}.nii.gz

applywarp --interp=spline \
-i ${StudyFolder}/${Subject}/MNINonLinear/Results/${fMRIName}/${fMRIName}_hp2000_clean.nii.gz \
-r ${StudyFolder}/${Subject}/T1w/T1w_acpc_dc_restore.${fMRIRes}.nii.gz \
-w ${StudyFolder}/${Subject}/MNINonLinear/xfms/standard2acpc_dc.nii.gz \
-o ${StudyFolder}/${Subject}/T1w/Results/${fMRIName}/${fMRIName}_hp2000_clean.nii.gz
```

where ${fMRIRes} = 2 for 3T data and 1.6 for 7T data.

The use of spline interpolation will induce a minimal amount of blurring (much less than a trilinear interpolation).