



HCP
AGING



HCP
DEVELOPMENT

Lifespan HCP 2.0 Release
Data Access &
Download Instructions

24 February 2021



Table of Contents

CCF Data Access	3
Requesting Access to NDA	3
Access Eligibility.....	4
Applying for access to CCF data	4
Applying for access as a Group	6
Selecting Data for Download	7
OPTION ONE	7
OPTION TWO	8
Filtering by Subjects	10
Downloading with Java Download Manager	11
Download Speed	13
Download Threshold	13
Downloading via the Command Line	14
Cloud Options	15



CCF Data Access

Lifespan HCP (Aging & Development) and Connectomes Related to Human Disease (CRHD) projects data managed and processed by the Connectome Coordination Facility (CCF) are being released through the [NIMH Data Archive](#) (NDA), a data repository funded by the National Institutes of Health (NIH).

HCP-Aging, HCP-Development, and CRHD imaging and behavioral data are currently only being shared through NDA.

For users familiar with accessing data from HCP-Young Adult from ConnectomeDB or other sources under the HCP Open Access Data Use Terms, the terms for data use for data shared via NDA are more restrictive (see details in the NDA Data Use Certification agreement you sign in [Applying for access to CCF data](#)). There are eligibility requirements, an application process to obtain and maintain annual access to the data, and you must have institutional sponsorship for your application to be approved.

The screenshot shows the NIMH Data Archive website. At the top, there is a navigation bar with links for 'Contribute Data', 'Get Data', 'Tools', 'Webinars & Tutorials', and 'About Us'. A 'MY ACCOUNT' link is circled in pink, with 'Create account/Login' text below it. The main heading is 'Welcome to the NIMH Data Archive'. Below this is a search bar and two columns of menu items. The left column is titled 'CONTRIBUTE DATA' and includes options like 'PREPARE FOR SUBMISSION', 'SUBMIT DATA', 'ACCESS MY NDA COLLECTIONS', 'SEARCH DATA DICTIONARY', 'CREATE QUADS', 'SUBMIT DATA FROM A PAPER', and 'VIEW NDA DATA STANDARDS'. The right column is titled 'GET DATA' and includes 'GET DATA', 'REQUEST DATA ACCESS', 'SEARCH DATA DICTIONARY', 'ACCESS DATA PACKAGES', 'REQUEST CLOUD CREDITS', 'SUBMIT DATA FROM A PAPER', and 'LEARN ABOUT CLOUD ACCESS'. Annotations in pink highlight 'GET DATA' as a 'Query page to select data', 'REQUEST DATA ACCESS' as 'Begin Access Request', and 'ACCESS DATA PACKAGES' as 'Begin Download'.

NDA access for CCF data is being granted by the same NIH permissions committee as for the Adolescent Brain Cognitive Development Study (ABCD) data. Thus, if you already have active NDA access for ABCD data, you also have access to CCF data and vice versa.

If you need help at any step, please do not hesitate to contact the [NDA helpdesk](#). They usually respond within a day on weekdays and often are the only source for information that is difficult to find on the NDA site. Contacting NDA has the added benefit of giving them valuable feedback that can motivate improvements to their process and tools. The more NDA hears from the community about access or download issues, the better for all users.

We also recommend that you [join the HCP-Users Google group](#) where users may ask HCP domain experts and other users about the data and where to find more information. Once you have joined (posts from nonmembers are moderated to control spam), post questions to hcp-users@humanconnectome.org.

Requesting Access to NDA

Researchers can request access to HCP Aging, HCP Development, and other CCF shared data from the [NIMH Data Archive](#) (NDA). Lifespan HCP (Aging & Development) and Connectomes Related to Human Disease projects data are currently only being shared through NDA.

For users familiar with accessing data from HCP-Young Adult from ConnectomeDB or other sources under the HCP Open Access Data Use Terms, the terms for data use for data shared via NDA are more restrictive (see details in the NDA Data Use Certification agreement you sign in [Applying for access to CCF data](#)). There are eligibility requirements, an application process to obtain and maintain access to the data, and you must have institutional sponsorship for your application to be approved.



NDA access for CCF data is being granted by the same NIH permissions committee as for the Adolescent Brain Cognitive Development Study (ABCD) data. Thus, if you already have active NDA access for ABCD data, you also have access to CCF data and vice versa.

Gaining approval for a new request for NDA access is a multistep process that may take some time, possibly a few weeks, to gather the necessary information and signatures, especially if you are at an institution that must establish the eligibility requirements (below). Obtaining data access is more arduous for users than we would like, but our grants have required us to share data through the NDA under this policy.

If you need help at any step, please do not hesitate to contact the [NDA helpdesk](#). They usually respond within a day on weekdays and often are the only source for information that is missing or difficult to find on the NDA site. Contacting NDA has the added benefit of giving them valuable feedback that can motivate improvements to their process and tools. The more NDA hears from the community about access or download issues the better for all users.

Access Eligibility

There are three criteria you must meet to be eligible to request NDA access:

1. You must have a research-related need to access the data.
2. You must be associated with an NIH-recognized research institution, defined as an institution registered in the NIH electronic research administration system (eRA Commons), and have the approval of an authorized signatory official of that institution.
3. Your institution must have an active Federalwide Assurance (FWA).

If you do not know whether or not you meet these, you might need to ask your departmental grants specialist or institutional grants contracts office to determine if you have an active FWA, an institutional eRA Commons account, and to identify the most appropriate signing official at your institution. If your institution is lacking any of these, the following sites have some more info from NDA and the NIH about:

- Who is usually considered an Authorized Institutional signatory official: <https://nda.nih.gov/about/faq.html#ds.3>
- How an institution (legal signatory official) can apply for an eraCommons account: <https://era.nih.gov/erahelp/commons/#cshid=79>.
- Applying for institutional Federalwide Assurance: <https://ohrp.cit.nih.gov/efile/FwaStart.aspx>
- Searching for your institution's FWA status (use advanced options to search by country or state): <https://ohrp.cit.nih.gov/search/fwasearch.aspx?styp=bsc>

Applying for access to CCF data

Steps for getting access:

1. [Create an NDA account](#) for yourself.
2. [Login](#) to your account and navigate to your [Data Permissions dashboard](#).



3. Complete a Draft Data Use Certification (DUC), optionally including other researchers at your institutions who need access to CCF/ABCD data in your DUC.
4. Have your authorized signing official sign your DUC.
5. Upload your signed DUC and submit to the Permissions group for approval (may take up to 2 weeks).

Once you have completed steps 1 & 2, on your [Data Permissions dashboard](#), in the Actions column in the ABCD/CCF Permissions group row, select “Request Access”.

NDA Permission Groups

All NDA Broad Use and Open Access Permission Groups are listed in the table below. Select the Action button to review instructions or initiate a request to access the data in an NDA Permission Group.

Permission Group	Description	Requires Institutional Approval	Status	Actions
NIMH Data Archive	Phenotypic, imaging and genomic data as well as supporting documentation for NIH funded grants. Data Access Requests will be reviewed by the NDA Data Access Committee.	✓		Actions
Osteoarthritis Initiative	The National Institute of Arthritis and Musculoskeletal and Skin Diseases Osteoarthritis Initiative. Data Access Requests do not require Institutional sponsorship and will be approved immediately upon submission of a completed request.			Actions
Adolescent Brain Cognitive Development (ABCD) / Connectome Coordination Facility (CCF)	ABCD is a data repository that houses all data generated by the Adolescent Brain Cognitive Development (ABCD) Study. CCF is a data repository that houses all data generated by the Human Connectome Projects. Data Access Requests to ABCD/CCF will be reviewed by the ABCD Data Access Committee. Approved users will have access to both the ABCD and CCF data repositories.	✓	Access expires in 351 days.	Instructions Request Access

This will open a dialog with instructions and fields to create a draft Data Use Certification (DUC). You can complete this for just yourself, but we recommend applying with a group of any number of people at your institution who want to access the data (see [Applying for Access as a Group](#) below).

To complete the draft DUC you will need:

- A title and brief description of the proposed research project. In 3-4 sentences you should cover your overall research question(s) and how the data you are interested in will be used. This can be modified later in the draft DUC PDF.
- To identify an authorized signing official at your institution (usually someone who signs grants contracts). Contact the [NDA helpdesk](#) if the signing official you want to add is not already listed as an option to select.
- A list of the NDA accounts of any other data users that will be included on your DUC application.

Click “Generate Data Use Certification” to save a copy of your draft DUC locally. Then click “Save” to save it to your account as an Active access request.

Open the draft DUC PDF and make any necessary edits to the Research Data Use Statement.



Sign the document, save and send it to the authorized signing official you selected for their signature.



When you have the signed DUC, go back to your [Data Permissions dashboard](#) and in the Actions column in your Active Request, click “Upload”. Choose your signed DUC file and submit.

Active Requests

Monitor status of current requests for NDA data access.

ID	Type	Permissions	Signing Official	Request Status	Approval Status	Created Date	Actions
7660	Data Access	NIMH Data Archive	Elizabeth Colletta (ecolletta@wustl.edu)	Waiting on Agreement Upload	Processing	Feb 13, 2021	Upload View Delete

NDA Permission Groups

All NDA Broad Use and Open Access Permission Groups are listed in the table below. Select the Action button to review instructions or initiate a request to access the data in an NDA Permission Group.

Once submitted, it can take up to 10 business days to process your request for access. Once approved, access is valid for one full year from approval.

To maintain access, a renewal request should be submitted through the same process as above within 60 days of your access expiration (the UI doesn’t allow you to apply earlier). In your renewal request, submit progress on research related to data access as part of your Research Data Use Statement and any resulting publications, etc. in item 4. Access will be valid one year from Renewal approval date (not one year from original expiration date).

Applying for access as a Group

According to the NDA Data Use Certification (DUC) terms you are required to sign (a legal contract) to obtain data access, users may only share data obtained from NDA with those covered under the same specific DUC agreement created when you apply for access. Thus, the NDA expressly prohibits users from sharing data between individuals or groups, even within an institution, unless groups organize under the same DUC. We understand this data sharing policy is restrictive and unfortunate for users, especially compared to our data use terms for HCP-Young Adult. Our grants have required us to share data through the NDA under this policy.

NDA does allow for any number of investigators at a single institution to apply for access under the same DUC agreement. Therefore, to allow for data sharing within your institutional neuroimaging community and to save repeated annual effort to gather institutional requirements and get necessary signatures, we recommend that groups organize potential users of CCF and ABCD data at their institution to apply for access with a single joint DUC. If organizing your whole institutional neuroimaging community is not feasible, we advise at least including everyone in your lab/neuroimaging group in one DUC access request.

Requirements for a group access DUC:

- Everyone you intend to include in the DUC must have an NDA account.
- You must have a list of all NDA usernames of everyone you intend to include and select them in the Request Access DUC creation process detailed above.
- You should ask all included users how they anticipate using the data (or have used the data, for renewals) and include that in your Research Data Use Statement.

Note: Collaborators at other institutions must meet eligibility requirements through their home institution and apply for NDA access independently by submitting a separate DUC signed by their institutional official.



Selecting Data for Download

To obtain data from the Lifespan 2.0 Release, go to the [HCP Aging & Development Featured Datasets query page](#). You can also get to this page by selecting “Get Data” from the NDA home page, then selecting “Human Connectome Projects > HCP Aging and Development” at the top left of the query page.

On the [HCP Aging & Development Featured Datasets query page](#), the user has two options for accessing the HCP-Aging and HCP-Development data.

The screenshot shows the NDA Query Tool interface. On the left is a navigation menu with categories like 'Featured Datasets', 'Demographics', 'Data from Labs', etc. The main content area is titled 'HCP Aging and Development' and contains detailed information about the datasets. Under the heading 'OPTION ONE', there is a button labeled 'Access Shared Data Packages' which is highlighted with a pink arrow. Below this, 'OPTION TWO' is also visible. At the bottom of the main content area, there is a section titled 'HCP-Aging Release 2.0 Available Datasets' with a list of dataset types.

OPTION ONE

OPTION ONE accesses 8 premade, Lifespan 2.0 Release shared data packages that we recommend as a starting point for download for many users. These are described on the page and in the table below.

Shared Package	Size	Contents
HCPAgingRec	1.7 TB	preprocessed recommended structural, fMRI, and tMRI, plus behavioral data for 725 HCP-A subjects <i>precisely aligned by multimodal cortical surface area registration across subjects and cleaned of spatially specific structured noise</i> <i>fMRI time series in CIFTI format only</i> recommended starting point for most fMRI-based analyses
HCPDevelopmentRec	1.6 TB	preprocessed recommended structural, fMRI, and tMRI, plus behavioral data for 652 HCP-D subjects <i>precisely aligned by multimodal cortical surface area registration across subjects and cleaned of spatially specific structured noise</i> <i>fMRI time series in CIFTI format only</i> recommended starting point for most fMRI-based analyses
HCPAging1Sub	27 GB	all data for a single HCP-A subject with complete data <i>allows exploration of all processed data subsets (Recommended, Legacy, FreeSurfer, Extended, Uncleaned packages) and unprocessed data for all modalities</i>
HCPDevelopment1Sub	32 GB	all data for a single HCP-D subject with complete data <i>allows exploration of all processed data subsets (Recommended, Legacy, FreeSurfer, Extended, Uncleaned packages) and unprocessed data for all modalities</i>
HCPAgingImgManifestBeh	102 kB	behavioral data, plus preprocessed and unprocessed imaging metadata (no imaging data files) for 725 HCP-A subjects <i>includes datastructure_manifest.txt file containing AWS S3 URLs for all released data files useful for command line downloading of specific files of interest (see Downloading via the Command Line below)</i>
HCPDevImgManifestBeh	102 kB	behavioral data, plus preprocessed and unprocessed imaging metadata (no imaging data files) for 652 HCP-D subjects <i>includes datastructure_manifest.txt file containing AWS S3 URLs for all released data files useful for command line downloading of specific files of interest (see Downloading via the Command Line below)</i>
HCPAgingAllFiles	22.8 TB	all 2.0 Release data for all 725 HCP-A subjects (unprocessed and preprocessed imaging data + behavioral data) DO NOT DOWNLOAD in its entirety <i>for exploration/download of subsets of files using NDAs new download tools</i>
HCPDevAllFiles	22.8 TB	all 2.0 Release data for all 652 HCP-D subjects (unprocessed and preprocessed imaging data + behavioral data) DO NOT DOWNLOAD in its entirety <i>for exploration/download of subsets of files using NDAs new download tools</i>

Click on the “Access Shared Data Packages” button.

On the Data Packages page, in the Actions column in the row of the package you are interested in, select “Add to My Data Packages” (pink arrow). It will take some time (seconds to several minutes depending on size) to add the package to your account and there should be a notification at the top of the page when it is complete.

Data Packages (10)

Listed below are data packages and associated mINDARs. There are two data package types available, and you can select from the drop-down menu: My Data Packages, Shared Data Packages, or a combination of both, All Data Packages.

A mINDAR is a cloud-based Oracle database that contains a copy of a data package. You can create a mINDAR from a data package by selecting the Create mINDAR option from the Actions menu. You may also recreate a mINDAR, view connection details, or reset a password from the same menu, for an existing mINDAR.

All Data Packages: A combination of your existing Data Packages, and Shared Data Packages that are available to be added to your Data Packages.

My Data Packages: Data Packages that you have created or added to account. These can be explored and downloaded using the Download Manager, and you may also create a mINDAR from these data packages using the Actions dropdown. Click the Help button for more detail on mINDARs.

Shared Data Packages: Packaged datasets created by NDA or another user that you have permission to add to your account. Adding a Shared Data Package creates a new data package of the same name under My Data Packages.

To see a detailed description of each Shared Data Package, or each data package added to My Data Packages from a Shared Data Package, hover over the data package name.

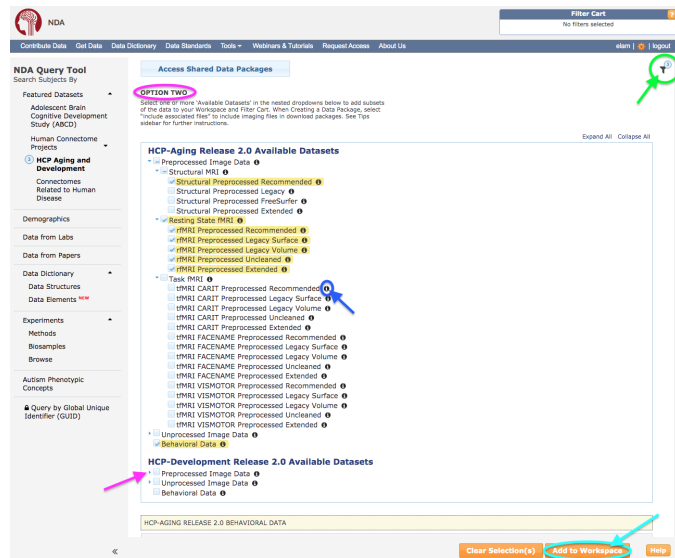
ID	Name	Status	Size	Created Date	mINDAR Last Message	mINDAR Status	mINDAR Created Date	Type	Actions
1183341	HCPDevelopment1Sub	Ready to add to My Data Packages	31 GB	02/06/2021				Shared Packages	Actions
1183340	HCPAging1Sub	Ready to add to My Data Packages	27 GB	02/05/2021				Shared Packages	Actions
1183254	HCPDevelopmentRfc	Ready to add to My Data Packages	1669 KB	02/03/2021				Shared Packages	Actions
1183256	HCPDevelopmentRfc	Ready to add to My Data Packages	102 KB	02/03/2021				Shared Packages	Actions
1183249	HCPDevelopmentRfc	Ready to add to My Data Packages	23789 KB	02/03/2021				Shared Packages	Actions
1183234	HCPAgingRfc	Ready to add to My Data Packages	1742 KB	02/03/2021				Shared Packages	Actions

In the meantime, you can proceed with downloading and setting up the [Java Download Manager](#) or [nda-tools](#) for [downloading on the command line](#).

OPTION TWO

OPTION TWO allows the user to select modality- and processing output-specific “HCP package” filters to access part of the released data by clicking the nested dropdown options under HCP-Aging or HCP-Development Release 2.0 Available Datasets. We refer to these selections as “HCP Packages” because these were curated from the HCP processing pipeline outputs by the data release team to cater to different users’ analysis purposes.

Refer to the table below for descriptions of these filters with data sizes when applied for all release subjects with the specified data available.



NDA Query Tool

Search Subjects By

- Featured Datasets
- Adolescent Brain Cognitive Development Study (ABCD)
- Human Connectome Projects
- HCP Aging and Development**
- Connections Related to Human Disease
- Demographics
- Data from Labs
- Data from Papers
- Data Dictionary
- Data Structures
- Data Elements
- Experiments
- Methods
- Biosamples
- Browse
- Autism Phenotypic Concepts
- Query by Global Unique Identifier (GUID)

Access Shared Data Packages

Select ONE of more Available Datasets in the nested dropdowns below to add subsets of the data to your Workspace and Filter Cart. When Creating a Data Package, select "Include associated files" to include imaging files in download packages. See Tips sidebar for further instructions.

HCP-Aging Release 2.0 Available Datasets

- Preprocessed Image Data
 - Structural MRI
 - Structural Preprocessed Recommended
 - Structural Preprocessed Legacy
 - Structural Preprocessed FreeSurfer
 - Structural Preprocessed Extended
 - Resting State fMRI
 - fMRI Preprocessed Recommended
 - fMRI Preprocessed Legacy Surface
 - fMRI Preprocessed Legacy Volume
 - fMRI Preprocessed Unlabeled
 - fMRI Preprocessed Extended
 - Task fMRI
 - fMRI CART Preprocessed Recommended
 - fMRI CART Preprocessed Legacy Surface
 - fMRI CART Preprocessed Legacy Volume
 - fMRI CART Preprocessed Unlabeled
 - fMRI CART Preprocessed Extended
 - fMRI FACENAME Preprocessed Recommended
 - fMRI FACENAME Preprocessed Legacy Surface
 - fMRI FACENAME Preprocessed Legacy Volume
 - fMRI FACENAME Preprocessed Unlabeled
 - fMRI VISUMOTOR Preprocessed Recommended
 - fMRI VISUMOTOR Preprocessed Legacy Surface
 - fMRI VISUMOTOR Preprocessed Legacy Volume
 - fMRI VISUMOTOR Preprocessed Unlabeled
 - fMRI VISUMOTOR Preprocessed Extended
- Unprocessed Image Data
- Behavioral Data

HCP-Development Release 2.0 Available Datasets

- Preprocessed Image Data
- Unprocessed Image Data
- Behavioral Data

HCP-AGING RELEASE 2.0 BEHAVIORAL DATA

Clear Selection(s) Add to Workspace Help

On the [HCP Aging & Development Featured Datasets query page](#), under OPTION TWO, click

the triangles next to the data types to reveal all the subset options (pink arrow). Click the black “i” information buttons (blue arrow) to see a description of the subset package.

Select the checkboxes of the subsets of the data you are interested in and click the “Add to Workspace” button at the bottom of the page (cyan arrow). **Note:** The OPTION TWO subset package filters filter for data from all released subjects and are additive (if you make more than one selection), so total data sizes may become large. If you are interested in downloading one or a few subjects, see [Filtering by Subjects](#).

Click on the Filter funnel icon at the top right (green arrow) showing the number of filters you added. This will show your Workspace, click “Submit to Filter Cart” at the bottom. It can take several minutes to update the Filter Cart at the top right.

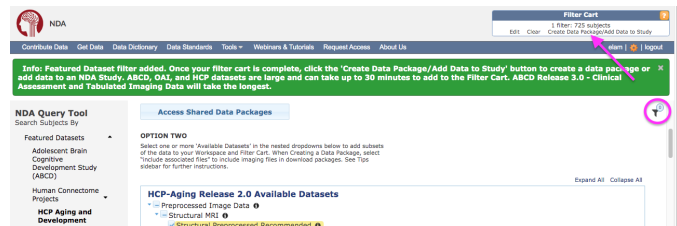


Study	HCP Package/OPTION 2 Filter	Size (All Subjects)	HCP Package Contents
HCA HCD	Structural Preprocessed Recommended	521 GB 476 GB	recommended starting point for structural analyses and contains files precisely aligned across subjects using the MSMAll multi-modal surface registration
HCA HCD	Structural Preprocessed Legacy	526 GB 481 GB	structural files coarsely aligned across subjects using the MSMSulc folding surface registration
HCA HCD	Structural Preprocessed FreeSurfer	585 GB 541 GB	actual outputs from the FreeSurferPipeline stage of the HCP Structural Preprocessing, in FreeSurfer's native file formats and directory structure
HCA HCD	Structural Preprocessed Extended	160 GB 147 GB	additional files related to QC on structural preprocessing outputs and other extra files that may be useful to select users
HCA HCD	rfMRI Preprocessed Recommended	991 GB 883 GB	recommended starting point for rfMRI analyses and contains cleaned files precisely aligned across subjects using the MSMAll multi-modal surface registration
HCA HCD	rfMRI Preprocessed Legacy Surface	989 GB 881 GB	cleaned files coarsely aligned across subjects using the MSMSulc folding surface registration, and hcp_fix_multi_run
HCA HCD	rfMRI Preprocessed Legacy Volume	2.1 TB 2.0 TB	cleaned rfMRI files poorly aligned across subjects using nonlinear volume registration
HCA HCD	rfMRI Preprocessed Uncleaned	3.5 TB 2.4 TB	uncleaned resting state data of all registration types for use in testing alternative data cleanup strategies
HCA HCD	rfMRI Preprocessed Extended	4.9 TB 5.1 TB	additional files related to rfMRI data cleanup and other extra files that may be useful to select users
HCA HCD	tfMRI CARIT Preprocessed Recommended	82 GB 154 GB	recommended starting point for CARIT tfMRI analyses and contains cleaned files precisely aligned across subjects using the MSMAll multi-modal surface registration
HCA HCD	tfMRI CARIT Preprocessed Legacy Surface	82 GB 154 GB	cleaned CARIT tfMRI files coarsely aligned across subjects using the MSMSulc folding surface registration
HCA HCD	tfMRI CARIT Preprocessed Legacy Volume	166 GB 328 GB	cleaned CARIT tfMRI files poorly aligned across subjects using nonlinear volume registration
HCA HCD	tfMRI CARIT Preprocessed Uncleaned	540 GB 1.0 TB	uncleaned tfMRI CARIT data of all registration types for use in testing alternative data cleanup strategies
HCA HCD	tfMRI CARIT Preprocessed Extended	50 GB 93 GB	additional CARIT tfMRI files related to data cleanup and other extra files that may be useful to select users
HCA	tfMRI FACENAME Preprocessed Recommended	93 GB	recommended starting point for FACENAME tfMRI analyses and contains cleaned files precisely aligned across subjects using the MSMAll multi-modal surface registration
HCA	tfMRI FACENAME Preprocessed Legacy Surface	93 GB	cleaned FACENAME tfMRI files coarsely aligned across subjects using the MSMSulc folding surface registration
HCA	tfMRI FACENAME Preprocessed Legacy Volume	189 GB	cleaned FACENAME tfMRI files poorly aligned across subjects using nonlinear volume registration
HCA	tfMRI FACENAME Preprocessed Uncleaned	619 GB	uncleaned tfMRI FACENAME data of all registration types for use in testing alternative data cleanup strategies
HCA	tfMRI FACENAME Preprocessed Extended	50 GB	additional FACENAME tfMRI files related to data cleanup and other extra files that may be useful to select users
HCA	tfMRI VISMOTOR Preprocessed Recommended	56 GB	recommended starting point for VISMOTOR tfMRI analyses and contains cleaned files precisely aligned across subjects using the MSMAll multi-modal surface registration
HCA	tfMRI VISMOTOR Preprocessed Legacy Surface	56 GB	cleaned VISMOTOR tfMRI files coarsely aligned across subjects using the MSMSulc folding surface registration
HCA	tfMRI VISMOTOR Preprocessed Legacy Volume	109 GB	cleaned VISMOTOR tfMRI files poorly aligned across subjects using nonlinear volume registration
HCA	tfMRI VISMOTOR Preprocessed Uncleaned	346 GB	uncleaned tfMRI VISMOTOR data of all registration types for use in testing alternative data cleanup strategies
HCA	tfMRI VISMOTOR Preprocessed Extended	50 GB	additional VISMOTOR tfMRI files related to data cleanup and other extra files that may be useful to select users
HCD	tfMRI EMOTION Preprocessed Recommended	55 GB	recommended starting point for EMOTION tfMRI analyses and contains cleaned files precisely aligned across subjects using the MSMAll multi-modal surface registration
HCD	tfMRI EMOTION Preprocessed Legacy Surface	55 GB	cleaned EMOTION tfMRI files coarsely aligned across subjects using the MSMSulc folding surface registration
HCD	tfMRI EMOTION Preprocessed Legacy Volume	114 GB	cleaned EMOTION tfMRI files poorly aligned across subjects using nonlinear volume registration
HCD	tfMRI EMOTION Preprocessed Uncleaned	308 GB	uncleaned tfMRI EMOTION data of all registration types for use in testing alternative data cleanup strategies
HCD	tfMRI EMOTION Preprocessed Extended	47 GB	additional EMOTION tfMRI files related to data cleanup and other extra files that may be useful to select users
HCD	tfMRI GUESSING Preprocessed Recommended	101 GB	recommended starting point for GUESSING tfMRI analyses and contains cleaned files precisely aligned across subjects using the MSMAll multi-modal surface registration
HCD	tfMRI GUESSING Preprocessed Legacy Surface	145 GB	cleaned GUESSING tfMRI files coarsely aligned across subjects using the MSMSulc folding surface registration

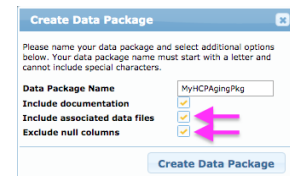


Study	HCP Package/OPTION 2 Filter	Size (All Subjects)	HCP Package Contents
HCD	tfMRI GUESSING Preprocessed Legacy Volume	309 GB	cleaned GUESSING tfMRI files poorly aligned across subjects using nonlinear volume registration
HCD	tfMRI GUESSING Preprocessed Uncleaned	651 GB	uncleaned tfMRI GUESSING data of all registration types for use in testing alternative data cleanup strategies
HCD	tfMRI GUESSING Preprocessed Extended	94 GB	additional GUESSING tfMRI files related to data cleanup and other extra files that may be useful to select users
HCA HCD	Structural Unprocessed	194 GB 203 GB	multi-echo MPRAGE (T1 weighted) and T2-SPACE (T2 weighted) scans (in NIFTI format)
HCA HCD	Diffusion Unprocessed	528 GB 566 GB	dMRI scans (in NIFTI format), bval, and bvec files for the two sets of diffusion sensitizing directions ('dir98' and 'dir99')
HCA HCD	Resting State fMRI Unprocessed	1.3 TB 1.3 TB	both pairs of resting state fMRI scans (in NIFTI format)
HCA HCD	tfMRI CARIT Unprocessed	237 GB 417 GB	fMRI scans for the CARIT task (in NIFTI format; Go/NoGo Conditioned Approach Response Inhibition Task)
HCA	tfMRI FACENAME Unprocessed	271 GB	fMRI scan for the FACENAME task (in NIFTI format; paired-associative memory task)
HCA	tfMRI VISMOTOR Unprocessed	154 GB	fMRI scan for the VISMOTOR task (in NIFTI format; simultaneous motor and visual activation task)
HCD	tfMRI EMOTION Unprocessed	126 GB	fMRI scan for the EMOTION task (in NIFTI format; emotion and face-processing task)
HCD	tfMRI GUESSING Unprocessed	391 GB	fMRI scans for the GUESSING task (in NIFTI format; reward, punishment, anticipatory reactivity task)
HCA HCD	ASL Unprocessed	27 GB 27 GB	mbPCASLhr scan (Arterial Spin Labeling in NIFTI format; multiband 2D EPI pseudo-continuous arterial spin labeling with high spatial resolution)
HCA	HiResHp Structural Unprocessed	10 GB	turbo-spin-echo high spatial resolution hippocampal structural scan (in NIFTI format)

When finished, click on “Create Data Package/Add to Study” (pink arrow) at the bottom of the Filter Cart. On the Data Packaging page, you’ll see the data you selected listed in NDA Data Structure categories (mostly useful for Behavioral data, click on the “i” buttons to see a tabular preview of the data).



Click the “Create Data Package” button to create your custom package with the data you’ve selected. Enter a Package Name, be sure to click the “Include associated data files” checkbox (MRI data are considered associated data files), and “Exclude null columns” so the behavioral data will not have extraneous columns for unused variables.



Click “Create Data Package”. The process of creating the package will take several minutes and can be tracked from your Packages Dashboard, with “My Packages” selected at the top left.

While you are waiting, follow the instructions below to download the Java Download Manager or command line download tools. You can also track package creation within the Java Download Manager GUI by clicking the Refresh Queue button until your package is listed as ready to download.

Filtering by Subjects

If you are interested in downloading subsets of data from one or a few subjects to limit data size, first download the HCPagingImgManifestBeh or HCPDevImgManifestBeh shared package using the OPTION ONE instructions



above. Choose desired subjects based on their behavioral data, completeness (see HCA2.0_Completeness.csv and HCD2.0_Completeness.csv), etc. and make a list of subjectkeys (GUIDs).

On the [HCP Aging & Development Featured Datasets query page](#), **first select filters for the HCP package datasets** you are interested in, add them to the workspace, click on the Filter icon, and submit these filters to the filter cart (green arrow).

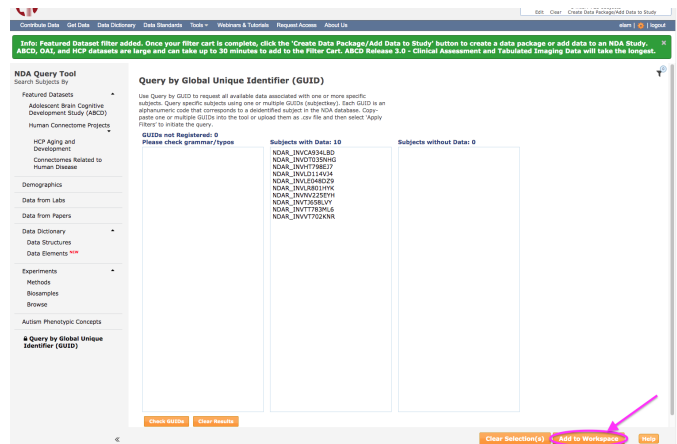
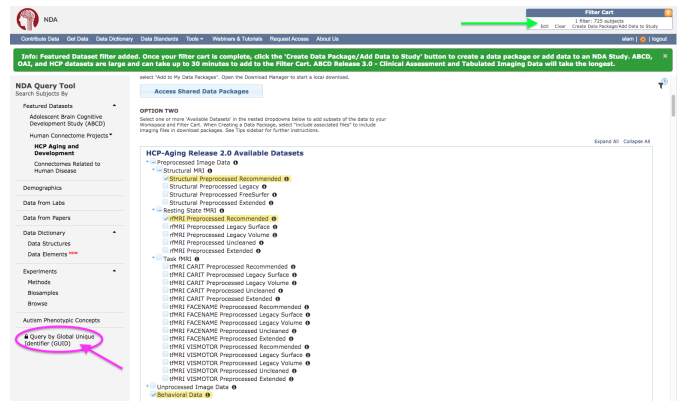
Click on “Query by Global Unique Identifier (GUID)” (pink arrow). Add your list of GUIDs and click “Check GUIDs”. This will check to see if the GUIDs you have listed indeed have data in NDA.

Click on “Add to Workspace”, and click on the Filter icon (pink arrow) and submit the GUID filter to the filter cart. Proceed with creating a package as above.

Note of warning: We found two bugs in testing the GUID filter tool that could affect the results you see using the tool (NDA has these in queue to fix in Spring 2021, we will update as the tool changes).

1) Order matters. Always add all other dataset filters before adding your GUID filter.

2) Extra data. Applying a GUID filter may grab extra data submitted for the subjects in your GUID list. You won’t know this has happened until you download your data. Luckily, with the HCP-A and HCP-D data, this problem is limited, but if you have a lot of GUIDs in your list, the extra data adds up and may significantly add to your download size.



Downloading with Java Download Manager

These instructions for downloading data assume the use of the currently available Java-based NDA Download Manager 5.1 (DM) GUI app. You may also download NDA data using the command line, see [Downloading via Command Line](#). NDA does plan to retire the Java-based DM and some of the tools used for command line downloading in coming months, so we will be updating these instructions as changes occur.

Note: NDA is currently developing a new Electron-based Download Manager and new command line download tools and will be releasing a beta version soon. We have been alpha testing the Electron DM and providing NDA with our feedback, and **we currently cannot recommend using the Electron DM beta version for downloads**, due to its download speed (5-10x slower than the Java DM) and a bug that causes downloads over 1 GB to stall non-systematically.



From the Packages dashboard, click the Download Manager button to download the DownloadManager.jnlp Java web launch application. Click on the [Download Manager Instructions](#) button for operating system-specific instructions on setting up [Java 8](#), which is required for the Download Manager to run. If you are using a Mac OS, you may need to set your System Preferences, Security & Privacy settings to allow the DownloadManager.jnlp file to open. You may need to add an exception for the <https://nda.nih.gov> URL in the JRE configuration to allow the application to run. If you have trouble with setting up Java 8 or running Download Manager on your system, contact the [NDA helpdesk](#) for assistance.

Open the DownloadManager.jnlp file from wherever you downloaded it. If prompted to update Java, DO NOT UPDATE to a newer version of Java (only Java 8 will work). Because the DM is a Java web launch application, part of the app will get downloaded every time you open it and you will get a prompt asking if you want to run the application. Click Run (pink arrow), and enter your NDA username and password when prompted.

A box warning that you are entering a government system, etc. will appear. Click Accept at the bottom.

The DM 5.1 window will open. Click Browse... (pink arrow) to select the directory in which you wish to download the data (make sure it has sufficient space available for the package you're planning to download), click Choose. The directory and path you selected will show up in the Save to: field.

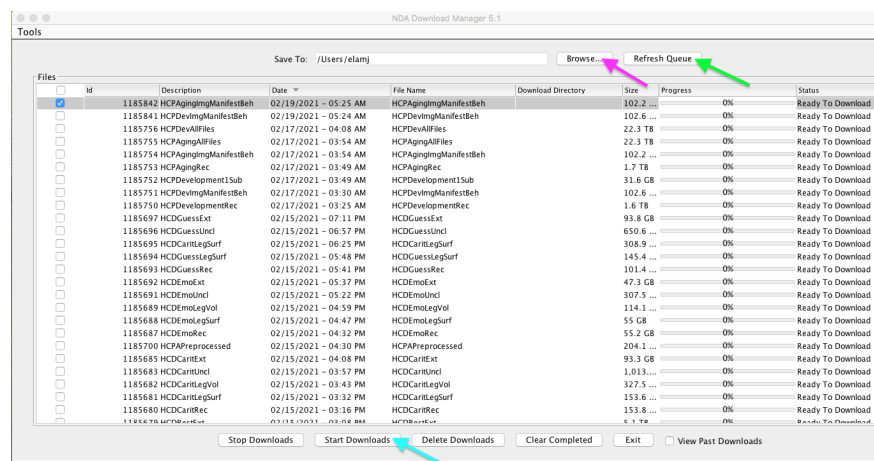
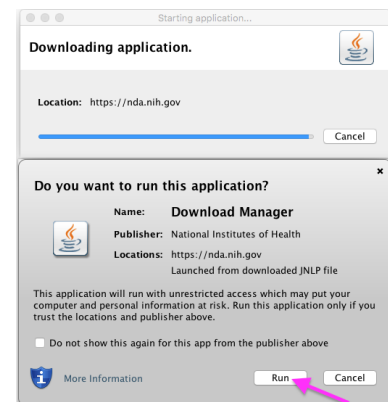
The list of "Files" should show all NDA packages that you have created using [OPTION TWO](#) or "Added to My Data Packages" in [OPTION ONE](#).

If you just created or added a package and you don't see the name of your package listed, click the Refresh Queue button (green arrow) until it appears with the status of "Ready to Download". Click the checkbox to the left of the package you wish to download and click the Start Downloads button (cyan arrow).

You will see the status change to "Downloading" and the progress of the download in the Progress column. You can stop the download, if you'd like, by clicking the "Stop Downloads" button.

If your internet connection is lost, the download will stop and the status will show "Download Incomplete". When you have reestablished your connection, click Start Downloads again and the download should restart.

If there's a problem with the download, the status will also show "Download Incomplete" but an errors.txt file will be written in the top level of your package directory structure. Send a message to the [NDA helpdesk](#) with that





errors.txt file attached and the package ID#. In the meantime, retry adding the shared package to your packages or recreating your custom package as described in [Selecting Data for Download](#). You also may try closing the Download Manager and relaunching it, then trying to restart the download with the Start Downloads button.

When you want to open the Download Manager again at a later time, open the DownloadManager.jnlp file you originally downloaded and follow the prompts. If you click the Download Manager button in NDA, it will download a second copy of the file rather than launch the application.

Download Speed

We have completed extensive testing of downloading the shared packages and custom packages of the Lifespan Release 2.0 data with the Java-based NDA Download Manager 5.1, several alpha versions of the soon to be released Electron-based NDA Download Manager (including the beta release candidate v0.1.29), and the currently available NDA command line tools.

Download speeds depend on 1) the speed of the app itself, 2) RAM available on the computer you are running your download from, and 3) the internet speed of the connection between your download computer and the Amazon Web Services (AWS) location of the NDA’s S3 buckets housing the HCP data.

For the Electron-based DM v0.1.29, the download time is further complicated by a stalling bug that randomly stops the download and requires restarting of the app to restart (NDA is working to address this bug now). To be able to compare the performance of the apps *while they are actually downloading*, we have removed the inconsistent download stalls out of the DM v.0.1.29 download times.

On institutional high speed (10 Gbit) internet, on a OS with ample memory, you can expect download times similar to the following for the Lifespan Release 2.0 shared packages:

Shared Package	Size	Download Time (minutes/hours) to local OS			
		Java DM 5.1	Electron DM v0.1.29	downloadcmd (4 threads)	
				Local storage	AWS EC2
HCPAgingRec	1.7 TB	20-22 h	72-84 h (3-3.5 days)	~17h	7.5 h
HCPDevelopmentRec	1.6 TB	18-20 h	64-72 h (2.5-3 days)	~16h	7.1 h
HCPAging1Sub	27 GB	16-20 min	~70 min	16 min	7 min
HCPDevelopment1Sub	32 GB	18-22 min	~80 min	18 min	8 min
HCPAgingImgManifestBeh	102 kB	1-3 min	3-5 min	<1 min	<1 min
HCPDevImgManifestBeh	102 kB	1-3 min	3-5 min	<1 min	<1 min
HCPAgingAllFiles	22.8 TB	DO NOT DOWNLOAD in their entirety	Options exist for downloading individual files, but loading these large packages is not yet operational	Not tested	Not tested
HCPDevAllFiles	22.8 TB	(only option for Java DM 5.1)			

Download Threshold

The NDA imposes a [threshold of 20 TB on local downloads](#)/user over a 30-day rolling window. Java DM 5.1 on a local OS won’t allow users to start single downloads over 20TB. An indication of your user downloads over the last 30 days is shown at the bottom of [your Data Packages page](#). The Electron DM beta v0.1.29 also keeps track of your download threshold with a “Download Used” bar at the top of the UI. Note: The 20TB threshold does not apply to transfers to AWS IP addresses (EC2 instances).



Downloading via the Command Line

These instructions for downloading data are a simplified version of more detailed instructions/troubleshooting tips (including checking if you have the right prerequisites and setting up your download environment on Linux/Mac/Windows OSs) on our wiki page: [How to get data from the NDA using command line tools](#).

First, you need to find a terminal to a machine with a Linux/MacOS operating system that affords you permission to install software and has access to a filesystem location with enough storage space for the package you wish to download. From this terminal, install `nda-tools` as described on <https://github.com/NDAR/nda-tools> which require `python2` or `python3`, for which you may need to setup a python virtual environment (see [wiki Step 4](#)), and installation of `pip`. Enter these commands in the terminal:

```
> pip install nda-tools # installs nda-tools from https://github.com/NDAR/nda-tools
> downloadcmd --help # shows all the options available for the downloadcmd
> df -h # find out how much space you have for downloading, and decide where to download
```

Locate the ID of the package you want to download from your [Data Packages page](#), with All packages selected. Here is the command to download shared package ID# #1185256 using 8 threads, naming the directory for download 'HCPDevImgManifestBeh' so that it matches the name of the shared package in the NDA. You may be asked to hit 'enter' a couple times if you don't have a token (this method doesn't require a token).

```
> downloadcmd 1185256 -dp -u <your NDA username> -p <your NDA password> -d HCPDevImgManifestBeh -wt 8
```

The download usually takes a minute to start, but since the `-v` (verbose) flag is added, you should start seeing messages about files being downloaded.

If nothing happens, it is likely due to changes that the NDA has made in `downloadcmd` or accessory tools. Look for updates on the NDAR repo pages (<https://github.com/NDAR>). The CCF will monitor NDA's updates and will update our instructions accordingly. If you get errors, first check for typos in the package number, your credentials, or the options used. Send your questions to NDA's helpdesk: NDAHelp@mail.nih.gov with your command (sans password) and the error you are seeing.

If you are successful in downloading the `HCPDevImgManifestBeh` or `HCPAgingImgManifestBeh` package, you will be able to use the `datastructure_manifest.txt` file contained in these packages to find S3 links to download individual files or subsets of files (e.g. all files in the preprocessed recommended rfMRI HCP-style package) and/or of subjects that you can use `grep` to "gather" as described in [Steps 7](#) and [8](#) of the [How to get data from the NDA using command line tools](#) wiki page.

Here's a sample command for downloading a subset of just the Preprocessed Structural Recommended HCP package for a subset of subjects (saved as the text file "PreprocStrucRecommendedSubjectSubsetS3s") created from a list of S3 links from the manifest:

```
> downloadcmd -t PreprocStrucRecommendedSubjectSubsetS3s -u <your username> -p <your password> -d
HCPDevImgManifestBehSUBSET -wt 8 -v
```




You can also just download a single file from the manifest as follows:

```
> downloadcmd s3://NDAR_Central_3/submission_33230/HCD0008117_V1_MR/T1w/T1w_acpc_dc_restore.nii.gz -u  
<your username> -p <your password> -d HCPDevImgManifestBehSingleLink
```

Happy downloading and enjoy the data!

Cloud Options

CCF and the NDA have been putting effort and making plans to support moving data analysis to the cloud *aiming to*: 1) minimize unnecessary data transferring and associated security issues, 2) lower data storage and analysis costs by centralizing a common repository. Currently, the only option available for data use on the cloud through NDA is to download the data as detailed above to an AWS Cloud EC2 instance on which you may setup your analysis tools. The CCF has experimented with this and is [creating wiki tutorials](#) to help our users to get started.

Using the AWS cloud for processing and especially storage can get quite costly over time. To mitigate the cost barrier, NDA has instituted the [NDA Computational Credits \(CC\) pilot program](#) to provide up to \$1000 in AWS credits for successful applicants to begin using the AWS Cloud for research that serves the mission of the NDA.

The NIH [Office of Data Science Strategy](#) has also recently funded a Data and Technology Advancement (DATA) Scholar based at NDA to lead the [Cloud-Native High-Performance Computing for Bioinformatics with Complex Data](#) project over the next 1-2 years with the goal of bringing image processing pipelines into NDA to facilitate computing in the cloud rather than downloading the data to a local server. The CCF is working with the DATA scholar to highlight existing gaps in the infrastructure and to prioritize tools to include in containerized pipelines that may be developed.

Stay tuned, this is a Work in Progress!