

Accessing Data from the May 2019 HCP Lifespan Data Release

Permissions

Human Connectome Project (HCP) datasets, including Lifespan HCP-Development (HCP-D) and HCP-Aging (HCP-A), are housed in the Adolescent Brain Cognitive Development (ABCD) data repository. Access to HCP data in NDA is governed by the ABCD Data Access Committee (DAC). This DAC has been established to objectively and systematically review data access requests to all datasets in the ABCD repository, which includes data collected and released by ABCD and HCP investigators. The ABCD DAC consists of six federal employees from NIH with expertise in science, policy, or bioinformatics resources. After a request is screened to ensure it satisfies the basic pre-requisites, it is reviewed by the DAC for approval. The ABCD DAC should review all data access requests in 10 business days.

Obtaining HCP Data in NDA – Package Creation

To access NDA data in the cloud or download NDA data, users must add one or multiple filters to the NDA Filter Cart, create data packages from the Filter Cart, and select to download or create a cloud-hosted Oracle database (mindar) from the NDA Packages Dashboard.

The Connectome and NDA teams have set up some shortcuts to facilitate access to frequently requested datasets in the HCP Lifespan Data Release. These shortcuts are presented as Option 1 and Option 2 in the Human Connectome Project tab of NDA's new Query Tool.

Option 1 – Prepackaged Datasets in Packages Dashboard

The full contents of HCP-A and HCP-D data releases, including tabular data, imaging files, and associated documentation have been prepackaged and can be found on the NDA Shared Packages Dashboard, for users with access to the ABCD permission group (see above).

Two frequently requested subsets of the HCP-A and HCP-D data releases - unrelated subjects and preprocessed data - have also been prepackaged and loaded to the NDA Shared Packages Dashboard for authorized users.

Step-by-step Instructions to create a data package:

- Click “Access prepackaged datasets” in the HCP tab of the NDA Query Tool to go to the Shared Packages Dashboard. Alternatively, go to <https://nda.nih.gov/user/dashboard/packages.html> and select “Shared Packages” in the top left drop down menu.
- Select “associate to my account” in the Actions menu for one or multiple packages in the table.
- Select “My Packages” in the top left drop down menu and the prepackaged dataset will appear in the table. This may take a few minutes for the larger prepackaged datasets
- Access data for analysis in the cloud:
 - Select “create mindar” in the Actions menu for one or multiple packages in the table. This will create an Oracle database (mindar) hosted in Amazon Web Services and deploy

the package to the database. Each data structure (table) in the package will translate to a table in this database.

- Create a password to access the mindar and click “submit.”
- Connect to the mindar using these credentials. Additional information and links to tutorials are available at <https://nda.nih.gov/tools/nda-tools.html#cloud>

Option 2 – Predefined Queries

HCP users can use these predefined queries as shortcuts to query frequently-requested subsets of HCP Lifespan datasets in the NDA database.

Step-by-step Instructions to create a data package:

- Select any of the checkboxes in the HCP tab of the NDA Query Tool to launch a predefined query of the NDA database to access the described subset of the HCP-A or HCP-D data releases. Click the info button above each predefined query for detailed information on the dataset.
- These predefined queries will add one or multiple filters to your workspace on the right side of the query tool. Some users may want to further narrow down the predefined queries using the Demographics filters in the NDA query tool.
- From the workspace, click “Add to filter cart.” The Filter Cart will process the request, which may take a few minutes.
- Once the subject numbers are displayed in the Filter Cart (top right of the screen), click the button to “Package/Add to Study,” which takes you to the Landing Page. Here you can view details about the data that will be returned by the selected predefined query. The left panel displays the source Collections of all the data, and the right panel displays a list of all the data structures included. You can check or uncheck structures to remove or include in your data package.
- Click “Create Package” to name and begin creating your data package. Select to “include associated files” if you want imaging files included in the data package.
- Go to your NDA Package Dashboard at <https://nda.nih.gov/user/dashboard/packages.html> and the newly created data package will appear in the table. Package creation time depends on a size of the package.

Obtaining HCP Data in NDA – Download in the cloud or to local machines

If downloading to the cloud, users must provision adequate storage for the download needs and attach that to AWS compute resource. NDA has [computational credits](#) available to assist with this. These tools will download all the objects included in a package to a local directory, preserving the file/directory tree structure.

These are the tools for getting data, and data can be placed on [EBS volumes](#), [FSx for Lustre File System](#), or an [Elastic File System](#) each of which represent a mountable file system to EC2 compute resources; the first two you provision a specific quantity of storage before you begin downloading, EFS you do not have to specify your storage needs up-front.

HCP Lifespan datasets are up to 6TB volume. Depending on network speed, local downloads could take a very long time.

Step-by-step instructions for 3 approaches to access or download data for local or cloud analysis:

- Download Manager Tool (primarily for local downloads):
 - Click on the orange “Download Manager” button in the top right section of the Packages Dashboard (or go to <https://nda.nih.gov/DownloadManager.html>). This opens up a Java tool that will download to your computer and requires a version 8 of Java to run. It allows you to download directly, to a location of your choice, your data package. Recently created packages may be displayed with a "Creating Package" status. You can update the Download Manager's status by using the Refresh Queue button.
 - Click “Browse” to select the download destination. The options at the bottom allow you to stop ongoing downloads, start all selected downloads, and delete or clear packages from the interface. Packages are tied to your account and will persist after you have downloaded them.
 - Select the checkbox next to the package to download and click “Start Downloads”.
 - When prompted to download all associated data, click “Yes”.
- Download with a Command Line Client (local or cloud downloads)
 - Command Line Download Manager
https://nda.nih.gov/jnlps/download_manager_client/downloadmanager.zip
 - Run using `java -jar downloadmanager.jar -h` to see help text, to download a package run `java -jar downloadmanager.jar --username {username} --password {password} {package_id}` from the command line or terminal.
 - Python command line download client - <https://github.com/NDAR/nda-tools>
 - Refer to <https://github.com/NDAR/nda-tools#getting-started> for details on how to configure your system to use the python client, and to <https://github.com/NDAR/nda-tools#downloading-data> for instructions on how to use the client to download data. To download a package execute `downloadcmd -<packageID> -dp` from the command line or terminal.
- Create an Oracle database in the cloud (cloud analysis or download)
 - Select “create mindar” in the Actions menu for one or multiple packages in the table. This will create an Oracle database (mindar) hosted in Amazon Web Services and deploy the package to the database. Each data structure (table) in the package will translate to a table in this database.
 - Create a password to access the mindar and click “submit.”
 - Connect to the mindar using these credentials. Additional information and links to tutorials are available at <https://nda.nih.gov/tools/nda-tools.html#cloud>