How to download the Healthy Brain Network (HBN) imaging data from AWS?

There are two options for downloading neuroimaging data from the Healthy Brain Network (HBN) dataset. Option 1 uses a Bash script, and option 2 uses Python. While option 1 is simpler to use (since it does not depend on Python), option 2 is helpful if you need to download a large amount of data. We recommend reading through both methods before selecting the option that best suits you.

You will need to install the latest version of AWS’s Command Line Interface. You can follow the step-by-step instructions on their website here.

Note: to use our download script, you do not need to perform any setup, such as configuring the AWS Access Key ID or AWS Secret Access Key.

### Option 1: Download HBN data using a bash script

1. Download the “HBN\_download\_links.sh” script. You can download it here.
2. Download the full list of links to the files on AWS. You can download the “HBN\_aws\_paths.csv” file here.
3. Modify the “HBN\_aws\_filepaths.csv” file only to have the rows of files that you would like to download. If you do not modify this file, the script will download the entire dataset.
4. Open a terminal and go to the location where you saved the “HBN\_download\_links.sh” script.
5. Run the command:

bash HBN\_download\_links.sh -i HBN\_aws\_filepaths.csv -o path/to/outputs

Note: Before running the script, ensure you have created the output folder and that the “HBN\_aws\_filepaths.csv” file is either in the same location as the script or includes the full path to it.

### Option 2: Download HBN data using a Python script

The Python script enables you to specify various options for downloading a customized subset of the sample from the S3 bucket. You can specify:

* Age ranges
* Sex
* Site
* Scan type
* Task

If none of these items are specified, the script will assume that you would like to download all available data from HBN.

The script requires Python 3.9 or higher, pandas, and boto3.

#### Basic Usage

Run the script from the command line to download files. For example:

python download\_from\_AWS.py --out\_dir /data/output

This line will download all the files from HBN and store them in the /data/output directory.

#### Command-Line Arguments

|  |  |  |  |
| --- | --- | --- | --- |
| Argument | Required | Type | Description |
| --out\_dir | Yes | str | Output directory where files will be saved |
| --aws\_links | No | str | Path to hbn\_aws\_links.csv |
| --age\_min | No | int | Minimum participant age |
| --age\_max | No | int | Maximum participant age |
| --sex | No | str | Download only files from participants of this sex (M or F) |
| --site | No | str (space-separated) | List of acquisition sites (e.g., Site-CUNY) |
| --scans | No | str (space-separated) | List of scan types to download (e.g., anat) |
| --tasks | No | str (space-separated) | List of functional tasks to download (e.g., REST1) |
| --dry\_run | No | str | Dry run to have a count of files that would be downloaded |
| --help | No | str | Display a complete list of arguments |

#### Valid Input Options

* Sex (--sex): M or F
* Site (--site): Site-SI, Site-RU, Site-CBIC, Site-CUNY
* Scans (--scans): anat, dwi, fmap, func
* Tasks (--tasks): REST, REST1, REST2, PEER1, PEER2, PEER3, MOVIEDM, MOVIETP

#### Advanced Examples

Download only functional scans from Site-CBIC for male participants who are more than 25 years old:

python download\_from\_AWS.py --out\_dir ./data --site Site-CBIC --age\_min 26 --tasks func

Download the anatomical data and the functional data from the movie The Present for all sites:

python download\_from\_AWS.py --out\_dir ./data --scans anat func --tasks MOVIETP

Download all anatomical and resting-state series for female participants between the ages of 6 and 15:

python download\_from\_AWS.py --out\_dir ./data --scans anat func --tasks REST REST1 REST2 --age\_min 6 --age\_max 15