

# STUDY LEVEL SEARCH USER GUIDE

## HOW DO I DOWNLOAD HARMONIZED DATASETS?

1. At the top of the page, there will be a ribbon with multiple links. Begin by navigating to the study level search. If not yet logged-in, you will be prompted to do so before continuing.
2. Select desired Boolean filters on the left panel. See key terms and definitions in this document, if any of the options are unclear or you would like additional information.
3. After selecting filters, navigate to the Study Panels window to see all the studies that meet the criteria already selected. Review the information about each study and adjust filters accordingly before finalizing study set.
4. At the top of the Studies Panel click “**Generate Harmonized Dataset for Selected Studies.**” This will initiate the dataset download under **Current Working Files Folder** where at the bottom, a progress bar for download will be displayed until it completes.
5. Once the download is complete there will be three separate files. First, a file beginning with the title ‘**object\_alignment**’, **today’s date**, and **unique numerical string**. This is the harmonized data file. Once opened there will be a **STUDY-ID** on the leftmost side at the top of the Excel sheet; this should match the Project #s selected. Each row will correspond to individual subjects within each of those projects. The file will also contain all the variables collected for each study. Because not every study will collect the same information, not every variable will have data for each subject but some common variables will. For example, **CHILD-HUMAN-BIOLOGICAL-SEX** (*which collects the sex of the child under study*).
6. The second file will begin with ‘**codebook.csv**’ in the title. As the title aptly implies, this will be the codebook for harmonized data file. This file will contain the value for each numerically coded categorical variable; *with the codes listed on the first column and the corresponding value on the second column*. The codebook will also include the link to the ontology class on the far right column.
7. The third and last file will begin with ‘**sources.csv**’ in the title. All the DOIs for the projects used in this dataset will be found in this file in order to credit investigators and their work in your project.

## HOW DO I DOWNLOAD DATA FROM A SINGLE STUDY?

1. At the top of the page, there will be a ribbon with multiple links. Begin by navigating to the study level search. If not yet logged-in, you will be prompted to do so before continuing.
2. Select desired Boolean filters on the left panel. See key terms and definitions in this document, if any of the options are unclear or you would like additional information.
3. After selecting filters, navigate to the Study Panels window to see all the studies that meet the criteria already selected. Use the options at the top of this window, to manually review the descriptive information listed for each study, or row, either in table or list form until the desired study is found.
4. Click the button titled '**Access Data and Metadata**' underneath the Project Name of the desired study. This will give the following options: **Access Study-Specific Files**, **Visualize Data**, and **Generate Harmonized Dataset**. To download data and metadata specific to the project select the first option, **Access Study-Specific Files**. The link will open the study-specific webpage.
5. At the top of the study-specific webpage, there will be the Project Description and PIs of the project (this should be identical to the information on the Study Panels window from the Study Level Search). Scroll to find the following categories with options to download file(s): **General Info**, where you will find a file titled, *the last four digits of the Project #\_README.pdf* (**ALWAYS download this first**); **Epidemiologic Data**, where there will be a csv file containing all the epi data specific to the project; one or more **Lab Analytic Data** categories specific to the study, where there will be a csv file(s) containing the laboratory data available for the selected study; and lastly the **Semantic Data Dictionaries (or SDDs)**.
6. It is very important to download the *README.pdf* file under General Info before downloading any Epidemiologic or Data Analytic files. This file will contain important information about the data files available for download, such as how to cite the *CHEAR/HHEAR Data Repository* (note there is also a section on how to do this in this document), the glossary explaining how sample and participant IDs and analyte codes are labeled; a table with all the public files available, their source, a short file description and DOIs; and any important analytic notes to be kept in mind while using the data available.
7. After reviewing the *README.pdf*, download the data files needed to complete any desired analyses.

## HOW DO I ACCESS PREVIOUSLY DOWNLOADED DATASETS?

1. At the top of the page, there will be a ribbon with multiple links. Begin by navigating to the Data Level Search then clicking on the hyperlink, **HHEAR Data Center**. If not yet logged-in, you will be prompted to do so before continuing.
2. The following webpage should be the HHEAR Harmonized Data Portal. Under the **Data Search and Retrieval Options**, select “**Retrieve Datasets/Codebooks**” redirecting to the **Current Working Files Folder**, where all previously and currently downloaded files should be located. All files will be organized by date with the oldest files at the top and recent downloads at the bottom.