README

Updated 09/10/2018

Thank you for downloading data from the NIMH Data Archive (NDA). NDA includes the following repositories:

NIMH Repositories

- The National Database for Autism Research (NDAR) http://ndar.nih.gov
- Research Domain Criteria Database (RDoCdb) http://rdocdb.nimh.nih.gov
- The National Database for Clinical Trials Related to Mental Illness (NDCT) http://ndct.nimh.nih.gov
- The NIH MRI Study of Normal Brain Development http://pediatricmri.nih.gov

The ABCD Data Repository: Containing all data generated by the ABCD Study: https://data-archive.nimh.nih.gov/abcd

The Osteoarthritis Initiative (OAI): https://data-archive.nimh.nih.gov/oai

Connectome Coordination Facility (CCF): https://data-archive.nimh.nih.gov/ccf

Each download is identified by a package ID. This ID is displayed in the Download Manager, and on your Packages dashboard: https://ndar.nih.gov/user/dashboard/packages.html. Each package comes with individual data structures delivered as tab-delimited .txt files, labelled with the structure shortname. Within each file, the column names (elements) are in the first row, and the description of each element follows in the second row. The package may also contain subfolders with other associated data files like images, NDA Study results files, eeg data, etc. as well as supporting documentation if requested.

Data definitions for each field, while included in each of the downloaded files, may also be accessed at https://ndar.nih.gov/ndar_data_dictionary.html?type=All&source=All&category=All. The definitions on that page are where you will find coding for values in the data. Common values in all data files include:

- Collection_id: The record's Collection of origin
- (structure shortname)_id: The globally unique row ID of the record in the NDA database
- Dataset_id: The record's dataset of origin
- Subjectkey: The NDA GUID, a subject's globally unique ID across NDA
- src_subject_id: The subject identifier used at the site where the data was collected
- interview_age: Age rounded to whole months
- interview_date: Date data was collected
- Subject sex

If you have downloaded data from a curated release of the **ABCD Data Repository** you should have received a set of supporting documentation labelled Release Notes. These documents provide important information and should be reviewed before making use of the data.

If you have downloaded data from the **OAI** data archive, you may benefit from consulting the documentation on the OAI website: https://ndar.nih.gov/oai/query_data_and_documentation.html.

Non-imaging data are available in the NDA formatted files downloaded via the NDA Download Manager, or as ASCII/SAS archives by data category. Imaging data are available in DICOM format either as .zip files at the visit-level, or as tar.gz files at the series level.

Null Columns: There is an option when building a package to omit null columns from your package. If this was checked, then elements for which your package contained no values in any record will be omitted entirely from the received text files.

Each package also might contain additional files that provide potentially useful information but are not data files requested as part of your query:

Errors.txt:

If you get this file, please send it to NDAHelp@mail.nih.gov along with your package ID. This file will only be generated if a portion of your download package failed to transmit successfully.

QA_Info.txt:

If you downloaded this package through the NDA's feature for identifying and resolving QA/QC issues in your data submission project's Collection, this file will be automatically included in your package and details the potential issues identified and their location (by comparison of row pairs in each structure pair) in your files. You can use this to help identify errors before resolving them and reuploading the structures in question.

GUID_parent_child.txt:

This file shows any GUIDs that have been promoted from pseudo-GUIDS or where more than one Pseudo-GUID was used for the same subject (common with data from biosamples). For example: "NDARUV811UZG" used to be "NDAR_INVKL165TGF", and can still be found in earlier data under that pseudo-GUID. The research group submitting this GUID obtained the appropriate information to generate a full GUID. The Parent is the 'real' GUID, whereas the child is the pseudo-GUID – there may be multiple pseudo-GUIDs associated with a single 'real' GUID.

Omics_experiments.txt:

Contains information related to an omics experiment, including: application, molecule, technology, platform, extraction protocol, processing protocol, analysis software, etc.

Concept_by_GUID.txt:

A list of all concepts as defined in the Query by Concept tool (see https://ndar.nih.gov/query_concept.html) that test true for each subject based on the given rule, including the record information for the relevant rule, the ontological hierarchy of the concept, and child-parent GUID promotion information. As of June 2015, the research field has identified concepts across only a small portion of measures in autism (about 35 out of 1200). Contact us if you are interested in helping extend the concepts currently defined to other concept areas, measures and phenotypes.

Associated Files:

If 'include associated files' was checked when generating the package, associated genomics and imaging files will be placed in sub-directories named for the data structure that the files are referenced in. For example, all files included in the image03 data structure for the package will be included in a sub-directory of that name once the download is complete. When downloading large objects, if errors are encountered the download manager will attempt to retrieve a file three times prior to giving up and generating an errors.txt file. Note that persistence of omics data beyond the time needed for computation is discouraged (see FAQ https://ndar.nih.gov/faq.html#faq64 for more detail).

Additional files specific to an NDA Study may be included (NDA Studies are containers for data specific to a publication, finding, curated release from an ongoing study, etc.):

- Study Attachments are in a directory called associated_files;
- Study Results are in a directory called results
- Study_Aggregate.txt contains information regarding the Study Cohort members, including GUID, age, sex, and phenotype.
- Cohort_Detail.txt contains information regarding the Study Cohorts.

For further assistance, please contact NDAHelp@mail.nih.gov or call (301) 443-3265.