NIMH Data Archive: Data Sharing Terms and Conditions

Effective Date: January 25, 2023

Contents

Overview

This Data Sharing Terms and Conditions document defines the expectations and procedures for data submission and sharing with the NIMH Data Archive (NDA). Investigators submitting data to NDA should review this document to ensure they understand how the data they submit will be made available.

Related Notices

NOT-MH-23-100, Notice of Data Sharing Policy for the National Institute of Mental Health
NOT-OD-14-124, NIH Genomic Data Sharing Policy

NIMH Data Archive (NDA)

The National Institutes of Health (NIH) and NIMH have developed a data infrastructure to store diverse types of data from participants in research studies, regardless of the source of funding. The extensive information collected by these studies is harmonized and subsequently stored in one of several data repositories within NDA data infrastructure, providing a rare and valuable scientific resource. NDA data repositories include the NDA, the Adolescent Brain Cognitive Development (ABCD) Study, the Connectome Coordination Facility (CCF), the Osteoarthritis Initiative (OAI), and the NIAAA Data Archive ($NIAAA_{DA}$). A current list of all NDA data repositories and links to their websites is maintained at https://nda.nih.gov/nda/about-us.html.

Data Sharing Overview

Researchers who are funded by NIMH are required to deposit all raw and analyzed data (including, but not limited to, clinical, genomic, imaging, and phenotypic data) from experiments involving human subjects into this infrastructure. All data should be de-identified and should be collected from subjects who have broadly consented to share their data for research use or who have consented to share their data for research, with data use limitations.

Non-NIMH funded researchers with relevant data may deposit their data if they are willing to adhere to the data sharing terms and conditions outlined in this document and the NDA Data Submission Agreement (DSA).

All deidentified data from research involving human subjects (https://grants.nih.gov/policy/humansubjects/research.htm) must be submitted to NDA using data structures that are defined by the research community. Researchers are encouraged to reuse existing data structures to allow data from different laboratories to be analyzed together. This includes clinical trials, epidemiological surveys, human laboratory investigations, and other types of studies involving human subjects.

Researchers who are collecting human genomic data should explicitly seek consent for broad sharing of subject data for research. They should submit with their DSA an Institutional Certification that has been approved by their Institutional Review Board. Genomics researchers will register their studies with dbGaP and add a link to the NDA collection to the dbGaP registration. After registration, all data (including but not limited to clinical, genomic, imaging, and phenotypic data) will be deposited in NDA.

Exceptions to Data Submission

Data collection date (interview_date) and age in months (interview_age) are required data fields in all NDA data structures. NDA expects this information in order to help investigators identify errors since their last submission. However, institutional IRBs may restrict submission of date and age variables into NDA collections. In this case, collections should follow NDA data masking methodology, where date or age are modified by a persistent "offset" number that is generated for each subject, securely stored by the collection administrators, and not shared outside of the study. Dates may also be masked by using a common month or calendar day for all entries in a dataset. This common methodology will support appropriate secondary use of the data.

Electronic medical records will not be submitted to NDA. De-identified data derived from electronic medical records may be submitted.

Physical biospecimens will not be submitted to NDA, but reference numbers to biosamples in other repositories will be submitted.

Videos and pictures of faces or other external body parts (i.e. parts that could potentially compromise a participant's identity) will not be submitted to NDA.

Participants who do not consent to share data through NDA are not excluded from the study, but their data cannot be uploaded to NDA.

Data Submission and Sharing Timeline

Pre-Award Expectations

Per NOT-MH-23-100, all applicable grant applications are expected to include a Resource Sharing Plan. The portion of that plan dealing with data must include: 1) a summary of the data that will be shared 2) a description of the standard(s) and/or data dictionaries that will be used to describe the data set, and 3) the proposed schedule to validate that the data are compliant with the data dictionary that is being used.

Other steps:

- Plan to collect the personally identifiable information from research subjects in order to generate a secure Global Unique Identifier (GUID) for data submission to NDA.

- Use the Cost Estimation tool on the NDA website to calculate an estimated cost to include in the grant application's budget.

- Include appropriate language in subject informed consent documents to allow for the broad sharing of data through NDA. An Informed Consent template is available on the NDA website.

- Provide the Institutional Certification (for sharing human genomics data) prior to award, along with any other Just-in-Time information requested.

- Review the NDA Data Dictionary and identify data structures (see Data Dictionary section below for a definition of data structures) that can be used to collect data for the proposed research. These data structures should be listed in the Resource Sharing Plan.

Post-Award Milestones and Timeline

Submit an NDA Data Submission Agreement (DSA) within 6 months of grant award as defined by the start date on the Notice of Award. The DSA is signed by the PI and their institutional signing official and submitted to NDA through the web interface. PIs must submit the signed DSA in order to gain access to their NDA Collection page, a virtual container where data and metadata are submitted and shared.

Define the collection's Data Expected section within 6 months of grant award. See Data Dictionary Expectations below for more details.

If the research is a clinical trial, report the NCT# in the Data Submission Agreement or as soon as it is available.

Communicate data submission timelines and expectations to appropriate research staff to ensure the timely submission of data.

Submit subject-level data (individual and derived variables) biannually. Standard NDA data submission deadlines are January 15th and July 15th each year. Alternative submission dates may be approved by NDA and the NIH Program Officer. Subject-level data should be submitted in the data submission cycle following data collection (for individual variables) or data generation (for derived variables).

- The first data submission date is the second cycle after the grant is awarded. For example, if the grant is awarded in February 2020, the first data submission is expected by January 15, 2021.

- If there are no data available to submit for the first data submission cycle, delay the first data submission by one cycle by updating submission dates on the Data Expected tab of their Collection to the next submission cycle date. No approval is necessary.

- To delay the first submission by more than one cycle, request a submission exemption and include justification language. The request is initiated from the Collection page, in the Data Expected tab. The PO will be notified and must approve or deny the request.

- To delay a submission after data have already been submitted to a collection, request a submission exemption, as described above. The PO will be notified and must approve or deny the request.

Clinical and phenotypic data submissions are cumulative (i.e., submitted in full upon every submission), for QA purposes and to capture data updates over time. Cumulative submissions increase data quality and integrity. Previous submissions are archived following successful QA.

Neurosignal recording and omics data submissions are not cumulative. They are submitted as they are collected or generated and are appended together in the NDA database.

All Collection data are shared automatically one year after the grant end date specified on the first Notice of Award.

Any subject-level data and the associated analyzed data used in a journal publication will be shared at the time of publication, if the publication occurs before the one-year automatic share date. PIs may request to share their data before the one-year automatic share date by updating the sharing dates in their collection's Data Expected tab.

Sharing Data used in Publications and Tool Development

Upon acceptance of a manuscript for publication**, authors are required to** create an NDA Study that links analytical output to the underlying collection data (individual and derived variables). The NDA website contains detailed instructions for creating and populating an NDA Study. Creation of the study prior to publication allows the DOI for the Study to be referenced in the publication.

Original data already submitted to the Collection does not need to be re-submitted. Submit to NDA additional data used in the publication that is of sufficient detail so as to allow future analysts to replicate and expand upon the results, including but not limited to:

- a description of the analytic methodology
- derived variables and algorithms used in the derivation
- statistical/analytical models and their output

NDA will generate a persistent digital object identifier (DOI) for the Study. The DOI should be included in the publication so that other researchers can find the data and reanalyze it or reproduce the original findings.

Investigators should appropriately acknowledge NDA in any research publication where they analyze data submitted to NDA. Current acknowledgment language is available at https://nda.nih.gov/nda/manuscript-preparation.html.

Data Standardization Expectations

Global Unique Identifier

Investigators will use NDA's global unique identifier (GUID) to identify all subjects in a Collection. The GUID is a common human subject identifier across NDA Collections and is a secure approach to increasing the quality of cross-study analyses.

If the subjects' information needed to generate a GUID is not collected, investigators can generate pseudo GUIDs for those subjects.

To access the NDA GUID tool, select "GUID Tool Access" in the NDA User Dashboard to enable access. Download the GUID tool software to a local machine and enter personally identifiable information (PII) on each research subject. The PII entered into the GUID tool is kept on the investigator's local machine only and is not transmitted to NDA. The software processes the identifiers into several intermediary codes using a one-way hash function (a cryptographic algorithm) and transmits the codes (hashes) in a secure manner to a secure NDA data enclave, where an alphanumeric GUID is linked to the hash codes. NDA returns the GUID to the investigator. The investigator is expected to maintain a link between internal study identifiers and the GUID in a secure file. The crosswalk between hash codes and GUIDs never leaves the secure NDA data enclave and is only accessible by authorized NDA staff.

Data Dictionary (Data Expected)

The NDA Data Dictionary contains over 2,500 data collection instruments (structures) and is extended as new data structures are added by NDA investigators. The NDA data harmonization approach is described at https://nda.nih.gov/nda/nda-data-harmonization-approach.html.

Create a data dictionary request for each individual grant by selecting existing or new data structures in the Data Expected tab of an NDA Collection. The NDA curation team will extend existing NDA Data Dictionary data structures (add variables, extend value ranges, or add aliases) or create new NDA Data Dictionary data structures, to ensure that all new data is harmonized to the NDA Data Dictionary.

Input targeted enrollment numbers for each data structure. This is the total number of subjects for whom data will be collected and submitted in that structure. When data is collected from a subject across multiple visits, that subject is counted once in the 'targeted enrollment' field.

Initial Submission and Initial Share dates should be populated according to these NDA Data Sharing Terms and Conditions. Any modifications to these will go through the approval processes outlined above.

Data Expected will be defined within 6 months of grant award.

Data Harmonization Standards

All collections must include several standard NDA data structures and supporting documents, depending on the type of collection. Current NDA data standards are maintained at https://nda.nih.gov/nda/harmonization-standards.html.

All collections

- Collections must include the required core data elements pertaining to the individual subject data in a "Research Subject and Pedigree" category of data structure. All collections except those submitting genomics data should use the 'Research Subject' structure.

Collections submitting genomics data

- Investigators submitting genomics data must submit the 'Genomics Subject' type of "Research Subject and Pedigree" data structure instead of the 'Research Subject' structure.
- Investigators must create an NDA Experiment that describes the parameters for each genomics assay and it will be assigned an ID number.
- The 'Genomics Sample' structure should be used to submit sample information, including repository identifiers for banked biosamples and the Experiment ID. This structure includes a field for specifying the file path of the associated genomics data files, which is then uploaded to NDA.

Collections generating imaging and/or other neurosignal recording data

- Collections must contain methods-specific structures for neurosignal recording data:
  - fMRI/DTI/PET: Image
  - EEG: EEG Subject Files
  - Eye Tracking: Eye Tracking subject experiment
  - MEG: Magnetoencephalography
  - EGG: Electroglottography form

- Investigators must create an NDA Experiment that describes the data collection protocol for each imaging or neurosignal recording experiment. The Experiment ID is then referenced in the related data structure (above).

Clinical Trials collections (including human laboratory studies)

- Clinical trials collections will submit a 'Treatment Assignment' data structure, which maps subjects to treatment groups/study arms over the course of the trial.
- Investigators will submit the clinical trial protocol, in accordance with the Good Clinical Practice guidelines, must be uploaded as supporting documentation. The protocol must include the assessment schedule.
- Manual of Procedures and Case Report Forms must be uploaded as supporting documentation.

Studies using survey instruments must upload survey instruments and supporting documentation.

Privacy

All data (see Definitions) made available for public use via NDA will be de-identified data, such that the identities of participants cannot be readily ascertained or otherwise associated with the data by NDA staff or secondary data users.

Exceptions to accommodate submission of sensitive data will be reviewed on a case-by-case basis.

Data Access for Research Purposes

Access to data for research purposes will be provided through an NDA Data Access Committee (DAC). Investigators and institutions seeking data from NDA will be expected to meet data security measures and will be asked to submit a Data Use Certification, which is cosigned by the investigator and the designated Institutional Official(s) at the NIH-recognized sponsoring institution with a current Federal Wide Assurance (FWA). The procedures associated with data access are described at https://nda.nih.gov/nda/standard-operating-procedures.html#sop4a.

NDA has an active Certificate of Confidentiality to protect the identity of research subjects in the Archive. The Certificate expires in 12/31/2032.

Definitions

Data Structure – A table in the NDA Data Dictionary that represents a single measure, data collection instrument, or assessment. These are updated regularly as new projects add variables, create aliases, and update descriptions; all changes are recorded in the change history record.

Users can download data collection templates to simplify data collection and submission to NDA.

NDA Data Dictionary – A database of over 2,500 tables, each of which is a single data structure that has been harmonized to a measure, instrument, or assessment. All data submitted to NDA is submitted to one of these structures and allows researchers to easily query across the entire NDA database.

Data Expected - The list of all data structures in which data will be collected and submitted for a given collection as well as targeted enrollment numbers, initial submission, and initial sharing dates for each structure.

NDA Collection – A virtual container and organization structure for data and associated documentation from one grant or one large project/consortium. It contains tools for tracking data submission and allows investigators to define a wide array of other elements that provide context for the data, including all general information regarding the data and source project, experimental parameters used to collect any event-based data contained in the Collection, methods, and other supporting documentation. They also allow investigators to link underlying data to an NDA Study, defining populations and subpopulations specific to research aims.

Global Unique Identifier (GUID) – A computer generated alphanumeric code that serves as a common subject identifier across all NIMH Data Archive studies and is unique to each individual participant.

Human Subjects Research – Research conducted on a living individual about whom an investigator (whether professional or student) conducting research obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens.

NCT# – A unique identification code given to each clinical study record registered on ClinicalTrials.gov. The format is "**NCT**" followed by an 8-digit number (for example, NCT00000419). Also called the ClinicalTrials.gov identifier.

Data Sharing Plan (DSP) – A template-based amendment to the grant application, completed by the PI and approved by the PO, that contains standard language for how the PI plans to share data with the NDA and thus comply with the NDA's data sharing policy (NOT-MH-23-100).

NDA Data Sharing Terms & Conditions – A document that outlines the NDA data submission and sharing policies and procedures, include key milestones throughout the grant lifecycle.

NDA Data Submission Agreement (DSA) – A document that outlines the NDA Data Submission Terms and Conditions and must be signed by the Principal Investigator and their authorized institutional business official prior to providing access to their collection.

NDA's Guide Notice (NOT-MH-23-100**)** – a Federal Guide Notice issued June 17, 2019 describing NIMH's policy whereby investigators and their institutions are expected to submit grant-related human subjects data to the NIMH Data Archive.

Subject-level data – the information collected and maintained on individual human research subjects.

For questions contact:

NDA Help Desk: NDAhelp@mail.nih.gov