

# NIAAA Data Archive (NIAAA<sub>DA</sub>): Data Sharing Terms and Conditions

---

**Effective Date: 01/13/2023**

**Version 1.2**

## Contents

Overview .....	1
Related Notices.....	1
NIAAA Data Archive (NIAAA <sub>DA</sub> ).....	1
NIMH Data Archive (NDA).....	2
Data Sharing Overview .....	2
Data Submission and Sharing Timeline .....	3
Data Standardization Expectations .....	5
Privacy.....	7
Data Access for Research Purposes.....	7
Definitions .....	7

## Overview

This Data Sharing Terms and Conditions document defines the expectations and procedures for data submission and sharing with the National Institute on Alcohol Abuse and Alcoholism Data Archive (NIAAA<sub>DA</sub>). Investigators submitting data to NIAAA<sub>DA</sub> should review this document prior to submitting a Data Management Sharing Plan for a grant application. Investigators who sign a Data Submission Agreement agree to the Data Sharing Terms and Conditions outlined in this document.

## Related Notices

[NOT-AA-23-002, Data Sharing Expectations for Human Research funded by NIAAA](#)

## NIAAA Data Archive (NIAAA<sub>DA</sub>)

The National Institute on Alcohol Abuse and Alcoholism Data Archive (NIAAA<sub>DA</sub>) is an NIAAA-supported data repository housed within the NIMH Data Archive. NIAAA-funded

investigators<sup>1</sup> conducting human subjects research are expected to submit de-identified, individual-level data to this data archive. Non-NIAAA funded investigators with alcohol-related data are welcome to deposit their data to the NIAAADA as well. The data in the NIAAADA will be catalogued and made available to the general research community at the time of an associated publication, or the end of the award/support period, whichever comes first. Researchers are strongly encouraged to visit the [NIAAADA website](#) for more information.

## **NIMH Data Archive (NDA)**

The National Institutes of Health (NIH) and NIMH have developed a data infrastructure to store the collection 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 the NIMH Data Archive (NDA) data infrastructure, providing a rare and valuable scientific resource. NIMH Data Archive 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 (NIAAADA). 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**

All de-identified data resulting from this NIH-funded award involving human subjects are expected to be submitted to the NIAAA Data Archive (NIAAADA) at the item level and subject level along with appropriate supporting documentation to enable efficient use of the data. This includes clinical trials, epidemiological surveys, human laboratory investigations, and other types of studies involving human subjects.

Genomic data will be submitted to NIAAADA unless NIAAA agrees to a different data archive during the negotiation of the terms and conditions of the grant award. NIAAA expects that studies with GWAS data will also submit these data to the Psychiatric Genomics Consortium. Awardees who are collecting human genomic data will register with [dbGaP](#). After registration, all data (including but not limited to clinical, genomic, imaging, and phenotypic data) will be deposited in the NIAAADA. A link to the NIAAADA collection will be added to the dbGaP registration.

## **Exceptions to Data Submission**

Data collection date (interview\_date) and age in months (interview\_age) are required data fields in all data structures. However, institutional IRBs may restrict submission of date and age variables into NDA collections. In this case, collections should follow the 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. This common methodology will support appropriate secondary use of the data.

---

<sup>1</sup> Fellowship (F), Training (T), Small Business (SBIR/STTR), and Education (R25) grants are not expected to share data with NIAAADA

Electronic medical records will not be submitted to NIAAAD<sub>A</sub>.

Physical biospecimens will not be submitted to NIAAAD<sub>A</sub>, but reference numbers to biosamples 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 NIAAAD<sub>A</sub>.

Participants who do not consent to share data through NDA are not excluded from the NIAAA-funded study, but their data cannot be uploaded to NIAAAD<sub>A</sub>.

## **Data Submission and Sharing Timeline**

### **Pre-Award Expectations**

Applications submitted to or assigned to NIAAA on or after January 25, 2023 must follow the Final NIH Policy for Data Management and Sharing ([NOT-OD-21-013](#)) and provide a Data Management and Sharing (DMS) Plan. Per [NOT-AA-23-002](#), for grants involving human subject data, NIAAA expects the data to be submitted to the NIAAA-sponsored data repository, the NIAAA Data Archive (NIAAAD<sub>A</sub>). Per [NOT-OD-22-195](#), the DMS Plan should be included in the "Other Plan(s)" attachment field of the PHS 398 Research Plan Form. Applicants to NIAAA are strongly encouraged to provide specific information described in Notice of Guidance for Data Management and Sharing Plans for the National Institute of Alcohol Abuse and Alcoholism (NIAAA) ([NOT-AA-23-001](#)).

Prior to award, the Principal Investigator(s) and the authorized business official must sign a NIAAA Data Archive Award Agreement indicating that the study will submit data to the NIAAAD<sub>A</sub> and will fulfill the following summary of tasks and expectations:

- Obtain [Informed Consent](#) that allows for broad sharing of the research subjects' de-identified data.
- Collect Personally Identifiable Information (PII) from research subjects that will allow for the creation of the [NDA GUID](#) (a de-identified subject ID).
- [Create a NDA Account](#) with access to the GUID Tool.
- Complete and submit a [Data Submission Agreement \(DSA\)](#) within 6 months of the Notice of Award (NoA) issue date.
- Create a list of data items to be collected in the project in the [Data Expected](#) tab of the NDA Collection within 6 months of NoA issue date.
- Create a [GUID](#) for each research subject using the NDA GUID tool.
- Format and submit data according to the [NDA Data Dictionary](#) and/or work with NDA staff to define an appropriate data structure, if one does not already exist.
- Submit data on or before the NDA submission due dates (April 1 and October 1 each year) in accordance with the applicable [Data Sharing Terms and Conditions](#) of award.
- [Perform QA/QC](#) checks on data within 4 months after the submission due dates and address any issues identified by the NDA.
- Share data in accordance with the default NIAAAD<sub>A</sub> data sharing schedule in accordance with the applicable [Data Sharing Terms and Conditions](#) of award.

Effective Date: 01/13/2023

- Upload supporting documents that allow future analysts to effectively use the data, including: study protocols, assessment schedules, operating procedures, original assessment instruments (if not proprietary), and analytic/statistical algorithms used to derive variables in publications.
- Create an [NDA Study](#), linking relevant publications to submitted data at the time of publication, and use the appropriate Digital Object Identifier (DOI) and [Acknowledgement template](#) in these publications.
- Submit a statement of progress on data sharing in non-competing renewals and progress reports. Further instructions are available in the [RPPR Instruction guide](#) under section C.5.b. Resource Sharing. Refer to specific [Tutorials](#) about data submission and sharing via the NDA or refer to the data sharing planning section where the data will be shared.

### **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 NIAAADA collection page, a virtual container where data and metadata and submitted and shared.

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

If the research is a clinical trial, report the NCT# as soon as it is available.

*Submit subject-level data (individual and derived variables) biannually (April 1<sup>st</sup> and October 1<sup>st</sup> each year). 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 October 1, 2020.
- If there are no data available to submit for the first data submission cycle, delay the first data submission by request a submission exemption and include justification language on the Data Expected tab of their Collection to the next submission cycle date.
- 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 two years after the grant end date specified on the first Notice of Award. PI requests for exceptions will be considered for only extreme and unusual circumstances (e.g., unanticipated life event [death of family member] or other force majeure).

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

Communicate this data sharing plan to appropriate research staff to ensure the timely submission of data.

### **Sharing Data used in Publications and Tool Development**

Upon acceptance of a manuscript for publication, submit to NIAAAD<sub>A</sub> data used in that publication that is of sufficient detail so as to allow future analysts to replicate and expand upon the results, including:

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

Create an NDA Study that links analytical output to the underlying collection data (individual and derived variables). The NIAAAD<sub>A</sub> website contains detailed instructions for creating and populating an NDA Study.

NDA will generate a persistent digital object identifier (doi) for the Study. The doi must 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 NIAAAD<sub>A</sub> in any research publication where they analyze data submitted to NDA. Current Acknowledgement language is available at <https://nda.nih.gov/niaaa/manuscript-preparation.html>.

## **Data Standardization Expectations**

### **Global Unique Identifier**

Collections will use NDA's [global unique identifier \(GUID\)](#) to identify all subjects in a collection. The GUID is a common identifier across NDA studies 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

Only investigators on an active DSA can 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. This PII entered into the GUID tool is kept on the investigator's local machine only and 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.

***Investigators should never submit PII to NDA for any reason.***

### **Data Dictionary**

The NDA data dictionary contains over 2,000 data collection instruments (structures) and is extended as new data structures are added to all repositories.

Create a data dictionary for each individual grant by selecting existing or new data structures in the Data Expected collection tab. 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 will be automatically populated according to these NIAAADA Data Sharing Terms and Conditions. Any modifications to these will go through the approval processes outlined above.

## Data Standards

All collections must include several standard NDA data structures and supporting documents, depending on the type of collection.

### 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 filepath of the associated genomics data files, which is then uploaded to NIAAAD<sub>A</sub>.

### Collections generating imaging and/or other neurosignal recording data

- Collections must contain methods-specific structures for neurosignal recording data:
  - EEG: [EEG Subject Files](#)
  - fMRI/DTI/PET: [Image](#)
  - 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)

- ‘Treatment Assignment’ data structure, which maps subjects to treatment groups/study arms over the course of the trial
- 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.

### **Data Access for Research Purposes**

Access to data for research purposes will be provided through the 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 co-signed 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>.

### **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 2,000 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.

**NIAAA Data Archive (NIAAA<sub>DA</sub>)** – NIAAA data repository housed in the NIMH Data Archive.

**NIAAA<sub>DA</sub> Award Agreement** – an agreement signed prior to grant award by the Principal Investigator and their authorized institutional business official indicating that the study will comply with NOT-AA-23-002; and thus, study data will be submitted to the NIAAA<sub>DA</sub> by fulfilling the required tasks and expectations

**NIAAA<sub>DA</sub> Data Sharing Terms & Conditions** – A document that outlines the NIAAA<sub>DA</sub> 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.

**NIMH Data Archive (NDA)** – NIMH Data repository system which houses the NIAAA Data Archive

**NIAAA's Guide Notice ([NOT-AA-23-002](#))** – a Federal Guide Notice issued January 9, 2023 describing NIAAA's policy whereby investigators and their institutions are expected to submit grant-related human subjects data to the NIAAA 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](mailto:ndahelp@mail.nih.gov)

NIAAA's NIAAADA Coordinator: Daniel Falk, Ph.D., [falkde@mail.nih.gov](mailto:falkde@mail.nih.gov)