

Policy for the NIMH Data Archive (NDA)

Background

The National Institute of Mental Health (NIMH) is the lead federal agency for research on mental illnesses. NIMH is one of the 27 Institutes and Centers that make up the National Institutes of Health (NIH), the largest biomedical research agency in the world. The mission of NIMH is to transform the understanding and treatment of mental illnesses through basic and clinical research, paving the way for prevention, recovery, and cure.

The National Institute of Mental Health (NIMH) Data Archive (NDA) is an NIH-funded collaborative resource that contains partially harmonized data from human subjects that are generally collected as during NIH funded research activities. Multiple data repositories are housed in the NDA infrastructure. Each data repository contains data sets that are useful to a particular research community.

The NIH and NIMH seek to encourage the use of these resources to achieve rapid scientific progress. Moreover, NIMH has made data sharing a requirement for all clinical research it funds (see [NOT-MH-19-033](#)). In order to take full advantage of such resources and maximize their research value, it is important that data are made broadly available, governed by appropriate terms and conditions, to the largest possible number of qualified investigators in a timely manner.

A central function of the NDA is to store and to link together clinical, -omics, neuro-signal recordings, and other data derived from individuals who participate in research studies. Depending on the goals of the research, investigators may collect any combination of data from research participants during a single study. The NDA provides the infrastructure to store, search across, and analyze these varied types of data. In addition, the NDA provides longitudinal storage of a research participant's de-identified information that are generated by one or more research studies, through the use of the NDA Global Unique Identifier (<https://nda.nih.gov/contribute/using-the-nda-guid.html>). The NDA GUID links together a single research participant's data even if those data were collected at different locations or through different studies.

This NDA policy document addresses (1) data submission procedures, (2) data access procedures, and (3) the protection of research participants during the submission of, storage of, and access of data within the NDA.

Principles

The policies contained in this NDA policy document were developed to be consistent with existing NIH policies.

The NIMH believes that the full value of data in the NDA to the public is best realized if de-identified data related to a finding or publication are made broadly available as rapidly as possible to a wide range of scientific investigators (see NIMH notices NOT-MH-19-033, NOT-MH-20-067, NOT-MH-14-015).

Data Submission

Data Deidentification and GUID Generation

Data submitted to the NDA will be de-identified such that the identities of subjects cannot be readily ascertained or otherwise associated with the data by NDA staff or secondary data users (<https://nda.nih.gov/about/standard-operating-procedures.html#sop5>). In addition, de-identified data will be coded using a unique code known as a Global Unique Identifier (GUID). Use of the GUID minimizes risks to study participants because it keeps one individual's information separate from that of another person without using names, addresses, or other identifying information. The unique code also allows the NDA to link together all submitted information on a single participant, giving researchers access to information that may have been collected elsewhere. The GUID is a computer-generated alphanumeric code [example: NDA-1A462BS] that is unique to each research participant (i.e., each person's information in the NDA—or each subject's record—has a different GUID). The NDA has implemented a number of security controls so that GUIDs and the NDA GUID Tool are used appropriately by authorized users, in order to maintain research participant privacy and the confidentiality of their data in the NDA (<https://nda.nih.gov/contribute/using-the-nda-guid.html>).

The NDA's mission is to provide high-quality data to authorized researchers for secondary analysis. The GUID is an integral component to the NDA's strategy for maintaining the integrity and utility of the NDA database. The GUID enables the NDA to de-duplicate research participants within the NDA database, prior to provisioning datasets for secondary analysis. Therefore, all research studies submitting data to NDA are expected to include in their research protocols the collection of information needed to create a GUID. In certain situations, a research study may not be able to collect the information needed to create a GUID. This determination can be made by the study's Institutional Review Board (IRB) upon review of the research protocol. The NDA can generate a unique random identifier called a pseudo-GUID to support data submission from such studies.

Institutional Certification

Tools for analysis of genomic and other complex data types are increasingly able to make inferences about some individual traits (e.g., height; weight; and, skin, hair, and eye color) and to identify predilections for characteristics (e.g., risk of developing some diseases) and behaviors with social stigma. In recognition of these risks, the NDA Policy includes steps to protect the interests and privacy concerns of individuals, families and identifiable groups who participate in research. The NIMH requests that institutions submitting datasets to the NDA certify that an Institutional Review Board (IRB) has considered such risks and that investigators have de-identified the data in accordance before the data

are submitted. Prior to submitting data to the NDA, institutions must complete the NDA Data Submission Agreement - OMB Control Number: 0925-0667. Completed NDA Data Submission Agreements will only be accepted from institutions recognized by the NIH in the eRA Commons and signed by an Authorized Institutional Business Official – commonly called a signing official – as registered in eRA Commons by the institution.

Data Access

Controlled Data Access

The NDA provides basic descriptive and aggregate summary information for general public use. Such summary information may include summary counts and general statistics on completed assessment instruments. Access to subject level datasets submitted and stored in the NDA will only be provided for research purposes through the completion of the NDA Data Use Certification: OMB Control Number: 0925-0667. For the majority of the data available in the NDA, Data Use Certifications will only be accepted from researchers who are sponsored by an institution registered in the NIH's eRA Commons with an active Federal-wide Assurance issued through the Office for Human Research Protections (OHRP). Additionally, the application must include a reason for access related to scientific investigation, scholarship or teaching, or other form of research.

A Data Access Committee (DAC) composed of NIH staff with relevant expertise reviews the Data Use Certification and makes decisions about whether to grant access to subject level data.

The established Data Access Committees or its designees will review data access requests to determine that the proposed research use does not present risks to the research participants (including re-identification, stigmatization, or other violations of confidentiality) and adheres to the consents of the research participants in the data. In the event that requests raise concerns related to privacy and confidentiality, risks to populations or groups, or other concerns, the DAC will consult with other experts, as appropriate.

Pre-Release Data Access

A goal of the NDA is to provide human subjects researchers with the capability to share data and collaborate at the earliest possible opportunity in order to speed scientific discoveries; however, data sharing for secondary or collaborative research may be limited when a study is ongoing. For example, an incomplete dataset may not be appropriate to answer certain research questions and it may engender bias or error. Furthermore, in many cases, new or secondary uses of a dataset will be most effectively undertaken in collaboration with the lab that originated the data. Synergies may be developed to improve both the ongoing study and the new research. The NIMH wishes to encourage collaboration whenever appropriate helping facilitate such sharing.

Study investigators are in the best position to determine whether new uses of data are appropriate during the time that a study is ongoing. As a result, potential collaboration requires communication between the data submitter and the potential data recipient. Therefore, the NIMH recommends that recipients who are seeking access to data from an ongoing study to coordinate with the submitters. This coordination may result in a decision to collaborate, a decision to authorize access absent collaboration,

or a decision to delay access until a study is complete. The NDA is setup to support each of these scenarios, through the permissions feature of the NDA web application.

Protection of Research Participants

Informed Consent

The potential for public benefit through the sharing of research data is significant; however, data related to the presence or risk of developing a disease/disorder and information regarding paternity or ancestry, may be sensitive. Therefore, protecting the privacy of research participants and the confidentiality of their data is critically important. Risks to individuals, groups, or communities should be balanced carefully with potential benefits of the knowledge to be gained. The sensitive nature of information about participants and the broad data distribution goals of the NDA highlight the importance of the informed consent process.

The NDA Policy applies to research data collected both prospectively and retrospectively. For prospective studies, in which data sharing through the NDA is conceived within the study design at the time research participants provide their consent, the NIMH expects specific discussion within the informed consent process and documentation that participant's data will be shared for research purposes through the NDA. For retrospectively collected data, the NIMH anticipates considerable variation in the extent to which data sharing and future research have been addressed within the informed consent documents; therefore, the NIMH expects the submitting institution to determine whether a retrospective study is appropriate for submission to the NDA (including an IRB and/or Privacy Board review of specific study elements, such as participant consent). The NIMH may give programmatic consideration to requests for funds or other resources needed to conduct additional participant consent, when appropriate. Furthermore, the NIMH intends to adapt the NDA Policy to be consistent with best practices for the consideration and risk-benefit analysis of participant research data sharing under this policy.

Removal of Participant's Consent to Share

In the event that participants withdraw consent to share their individual-level data through the NDA, the submitting institution will be responsible for alerting the NDA to request that the specific research participant's data be removed from future data distributions. However, any data that have been distributed to researchers will not be retracted.

Data Security

To ensure the security of the data held by the NDA, the NIMH employs multiple tiers of data security based on the content and level of risk associated with the data. The NIMH has established and will maintain operating policies and procedures to address issues including, but not limited to, the privacy and confidentiality of research participants, the interests of individuals and groups, data access procedures, and data security mechanisms. Procedures safeguarding the NDA and its data will be reviewed periodically. NDA has Authority to Operate from NIMH as a FISMA Moderate system.

Non-research Use of Data

As an agency of the federal government, the NIMH is required to release government records in response to a request under the Freedom of Information Act (FOIA), unless they are exempt from release under one of the FOIA exemptions. Although NDA-held data are de-identified and NIMH does not hold direct identifiers to individuals within the NDA, the agency recognizes the personal and potentially sensitive nature of the data being stored. The NIMH takes the position that technologies available within the public domain today, and technological advances expected over the next few years, make the identification of specific individuals from raw data feasible and increasingly straightforward. Therefore, the agency believes that the release of un-redacted NDA data in response to a FOIA request would constitute an unreasonable invasion of personal privacy under FOIA Exemption 6, 5 U.S.C. § 552 (b)(6). Therefore, among the safeguards that the NIMH foresees using to preserve the privacy of research participants is the redaction of individual-level research data from disclosures made in response to FOIA requests and the denial of requests for un-redacted datasets.

Additionally, the NIMH acknowledges that legitimate requests for access to data may be made by law enforcement offices. The NDA will not possess direct identifiers, nor will the NIMH have access to the link between the code and the identifiable information that may reside with the primary investigator and institution for particular studies. The release of identifiable information may be protected from compelled disclosure by the primary investigator's institution obtained a Certificate of Confidentiality for the original study. Researchers submitting data are encouraged to consider whether a Certificate of Confidentiality might be appropriate for their data as an additional safeguard with regard to involuntary disclosure of research participant identities. For its data, the NDA will hold a Certificate of Confidentiality. Further information about Certificates of Confidentiality is available at the following website: <http://grants.nih.gov/grants/policy/coc/>.

Inquiries

Additional information and detailed implementation guidance related to NDA can be found at <https://nda.nih.gov>.

Specific questions about this policy should be directed to:

Office of the Director

National Institute of Mental Health, National Institutes of Health

6001 Executive Boulevard

Rockville, Maryland 20892

(if overnight delivery): Rockville, Maryland 20852 Phone: (301) 443-3265

NDAHelp@mail.nih.gov