

# Policy for the National Database for Autism Research (NDAR)

## Background

The NIH is interested in advancing research to identify common genetic or other factors that influence the prevention, cause, diagnosis, and treatment of autism spectrum disorders. The National Database for Autism Research (NDAR) is a biomedical informatics system and central repository developed by the NIH. NDAR provides a common platform for data collection, retrieval and archiving, while allowing for flexibility in data entry and analysis. The NIH hopes that broad research use of NDAR will accelerate the advancement of research on autism spectrum disorders—a series of related brain disorders that affects a person’s ability to communicate, relate to others, and interact with his or her surroundings.

A central function of NDAR is to store and to link together genetic, phenotypic, images and other data derived from individuals who participate in autism research studies. Depending on the goals of the research, investigators may collect any combination of imaging, genotypic, clinical assessment, or other phenotypic information about research participants during a single study. NDAR provides the infrastructure to store, search across, and analyze these varied types of data. In addition, NDAR provides longitudinal storage of a research participant’s information generated by one or more research studies. In other words, NDAR is able to associate a single research participant’s genetic, imaging, clinical assessment and other information even if the data were collected at different locations or through different studies. By doing so, NDAR gives researchers access to more data than they can collect on their own and provides robust tools to analyze the information, making it easier and faster for researchers to gather, evaluate, and share autism research information from a variety of sources.

Certain autism studies funded by the NIH—the data from which will be deposited into NDAR—will attempt to identify parts of the genetic code that give insight into the prevention, causes, diagnosis, and treatment of autism. These studies may conduct genome-wide association studies. For the purposes of this policy, a genome-wide association study is defined as any study of genetic variation across the entire human genome that is designed to identify genetic associations with observable traits (such as blood pressure or weight), or the presence or absence of a disease or condition, such as autism. Whole genome information, when combined with clinical and other phenotype data, offers the potential for increased understanding of basic biological processes affecting human health, improvement in the prediction of disease and patient care, and ultimately the realization of the promise of personalized medicine. In addition, rapid advances in understanding the patterns of human genetic variation and maturing high-throughput, cost-effective methods for genotyping are providing powerful research tools for identifying genetic variants that contribute to health and disease.

The NIH has given considerable thought to the policy issues surrounding genome-wide association studies and their deposition into NIH-held databases (see <http://grants.nih.gov/grants/gwas/index.htm>). Because NDAR will contain information derived from genome-wide association studies on autism, in addition to other genetic and phenotypic

information, the principles contained in this policy were developed to be consistent with existing NIH policies, principles, and implementation procedures for genome-wide association studies. The policy addresses (1) data sharing procedures, (2) data access principles, and (3) issues regarding the protection of research participants during the submission of, storage of, and access to data within NDAR. The goal of the policy is to advance science for the benefit of the public through the creation of a centralized NIH data repository for autism research information.

### **Protecting Research Participants**

The potential for public benefit to be achieved through sharing autism research data is significant. However, genotype and phenotype information generated about individuals, such as data related to the presence or risk of developing autism and information regarding paternity or ancestry, may be sensitive. Therefore, protecting the privacy of the 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 through NDAR. The sensitive nature of information about participants and the broad data distribution goals of the NIH NDAR data repository highlight the importance of the informed consent process to this research.

The NDAR policy applies to autism research utilizing genetic materials and/or data collected both prospectively and retrospectively. For prospective studies, in which data sharing through NDAR is conceived within the study designs at the time research participants provide their consent, the NIH expects specific discussion within the informed consent process and documentation that participants' genotype and/or phenotype data will be shared for research purposes through the NIH NDAR data repository. For retrospective studies performed using existing genetic materials and previously collected data, the NIH anticipates considerable variation in the extent to which data sharing and future research have been addressed within the informed consent documents. As described in the policy, the submitting institution will determine whether a study is appropriate for submission to the NIH NDAR data repository (including an IRB and/or Privacy Board review of specific study elements, such as participant consent). The NIH anticipates that a number of studies proposing to include pre-existing data or samples may require additional consent of the research participants. The NIH may give programmatic consideration to requests for funds or other resources needed to conduct additional participant consent when appropriate.

As noted elsewhere and reflected in the NDAR oversight structure established to manage implementation of the NDAR policy (see Oversight and Governance section below), the NIH recognizes that the ethical considerations relevant to NDAR data sharing are complex and dynamic. Therefore, NDAR will rely heavily on informational materials previously prepared by the NIH as a resource for IRBs and institutions for their consideration of the issues relevant to reviewing and approving individual studies proposing data submission to the NIH NDAR data repository. The NIH intends to adapt the NDAR policy to be consistent with best practices for the consideration and risk-benefit analysis of genotype and phenotype data sharing under this policy.

In the event that participants withdraw consent for sharing of their individual-level genotype and/or phenotype data through the NIH NDAR data repository, the submitting institution will be responsible for alerting the NIH NDAR data

repository and requesting that the specific record be removed from future data distributions. However, data that have been distributed to researchers will not be retracted.

The NIH recognizes that scientific, ethical and societal issues relevant to this policy are evolving, and the agency has established on-going mechanisms to oversee NDAR policy implementation and data use practices. The NIH will revisit and revise the policy and related practices as appropriate.

### **Non-research Use of Data**

As an agency of the Federal Government, the NIH is required to release Government records in response to a request under the FOIA, unless they are exempt from release under one of the FOIA exemptions. Although the NIH-held data will be coded and the NIH will not hold direct identifiers to individuals within the NIH NDAR data repository, the agency recognizes the personal and potentially sensitive nature of the genotype-phenotype data. Further, the NIH 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 genotype-phenotype data feasible and increasingly straightforward.

The agency believes that release of un-redacted NDAR datasets 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 NIH foresees using to preserve the privacy of research participants and confidentiality of genetic data is the redaction of individual-level genotype, phenotype, and other data from disclosures made in response to FOIA requests and the denial of requests for un-redacted datasets.

In addition, the NIH acknowledges that legitimate requests for access to data made by law enforcement offices to the NIH may be fulfilled. The NIH will not possess direct identifiers within the NIH NDAR data repository, nor will the NIH have access to the link between the data code and the identifiable information that may reside with the primary investigators and institutions for particular studies. The release of identifiable information may be protected from compelled disclosure by the primary investigator's institution if a Certificate of Confidentiality is or was obtained for the original study. The NIH explicitly encourages investigators to consider the potential appropriateness of obtaining a Certificate of Confidentiality as an added measure of protection against future compelled disclosure of identities for studies planning to collect genome-wide association data.

### **Stigmatization**

Tools for analysis of genomic data increasingly are able to make inferences about some individual traits (e.g., height, weight, skin and 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 NDAR policy includes steps to protect the interests and privacy concerns of individuals, families and identifiable groups who participate in autism genetic and other research. The NIH is asking institutions submitting datasets to NDAR to certify that an Institutional Review

Board (IRB) has considered such risks and that investigators have de-identified the data in accordance with 45 CFR 46.102(f) before the data are submitted. In addition, in the event that requests raise questions or concerns related to privacy and confidentiality, risks to populations or groups, or other relevant topics, the NDAR Data Access Committee (DAC) will consult with other experts as appropriate.

### **Oversight and Governance of the NIH NDAR Data Repository, Submission, and Access**

The NIH has developed a governance structure for NDAR that provides oversight tailored to the specific role involved. The Director of the National Institute of Mental Health (NIMH) oversees the NDAR policy and its implementation. In carrying out this responsibility, the NIMH Director participates on a Governing Committee, with several other NIH Institute and Center Directors or their designees, that is responsible for the on-going management and stewardship of NDAR policy and procedures. Reporting to the Governing Committee are several groups and teams charged with the implementation, communication, and development of specific procedures related to the conduct, submission and data release practices for NDAR. One of these groups, the NDAR Implementation Team, is responsible for overseeing NDAR policy and data access to promote consistent and robust participant protections in NDAR.

## **Policy for the National Database for Autism Research (NDAR)**

### **I. Principles**

Consistent with the NIH mission to improve public health through research, the NIH believes that the full value of data in NDAR to the public can be realized only if longitudinal genotype and phenotype data are made available as rapidly as possible to a wide range of scientific investigators. Rapid and broad data access is particularly important for autism research studies because of the significant resources they require, the challenges of analyzing large datasets, and the extraordinary opportunities for longitudinal availability of data across multiple studies.

Protection of research participants is a fundamental principle underlying biomedical research. The NIH is committed to responsible stewardship of data throughout the research process, which is essential to protecting the interests of study participants and to maintaining public trust in biomedical research. In consideration of the evolving scientific, ethical, and societal issues related to this policy, the NIH is establishing a governance structure for NIH NDAR activities that will:

- Ensure ongoing, high-level agency oversight; and
- Obtain regular input from public representatives, including those with expertise in bioethics, privacy, data security, and appropriate scientific and clinical disciplines; and
- Revisit and revise the policy as appropriate.

### **II. Applicability**

This NIH policy applies to:

- Competing grant applications that include NDAR and are submitted to the NIH for the April 15, 2008, and subsequent receipt dates;
- Proposals for contracts that include NDAR and are submitted to the NIH on or after April 15, 2008; and
- NIH intramural research projects that include NDAR and are approved on or after April 15, 2008.

The sharing of autism-related research information through NDAR is a priority to the NIH. As such, the NIH will strongly encourage investigators to address the submission of research information to NDAR as part of their response to autism-related Requests for Applications and Requests for Proposals.

### **III. Data Management**

#### **Data Repository**

To facilitate broad and consistent access to NIH-supported datasets on autism research, the NIH has developed a central NIH NDAR data repository at the NIH Center for Information Technology. The repository will provide a single-point of access to basic information about NIH-supported datasets for autism studies. Although the NIH envisions that access to all NIH-supported NDAR datasets will be possible through this repository, it does not intend the repository to become the exclusive point of data submission for these data, nor does it intend the central database to delimit the structures or tools that may be appropriate for other similar databases. The repository also will accept autism-related datasets contributed from other sources.

To ensure the security of the data held by the repository, the Center for Information Technology will employ multiple tiers of data security based on the content and level of risk associated with the data. The NIH will establish and maintain operating policies and procedures for the repository 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. These will be reviewed periodically by the NDAR oversight bodies.

#### **Data Submission**

All investigators who receive NIH support to conduct autism research are expected to submit to the NIH NDAR data repository descriptive information about their studies for inclusion in the NIH NDAR data repository. All data and information will be submitted to a high security network within the Center for Information Technology through a secure transmission process. Submissions should include the following:

- the protocol,
- questionnaires,
- study manuals,

- variables measured, and
- other supporting documentation.

In addition, the NIH strongly encourages the submission of curated and coded phenotype, exposure, genotype, and pedigree data, as appropriate, to the NIH NDAR data repository as soon as quality control procedures have been completed at the local institution. These detailed data will be made available through a controlled access process according to the NDAR Data Access procedures (described in Data Access section below). Investigators who elect to submit their data to additional data repositories or networks should verify that appropriate data security, confidentiality, and privacy measures are in place for protection of research participants. Irrespective of where the data are submitted, researchers submitting research 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 the research participant identities. Further information about Certificates of Confidentiality is available at the following website: <http://grants.nih.gov/grants/policy/coc/>.

Data submitted to the NIH NDAR data repository will be de-identified such that the identities of data subjects cannot be readily ascertained or otherwise associated with the data by the repository staff or secondary data users (45 CFR 46.102(f)). 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 NDAR 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: 1A462BS] that is unique to each research participant (i.e., each person's information in NDAR—or each subject's record—has a different GUID). Creating the GUID involves several steps.

- The researcher uses his or her computer to enter pre-defined personal identifiers about research participants (e.g., birth name, Social Security number) into a specific computer program provided by the NIH. The program processes these personal identifiers at the researcher's site into several intermediary codes known as "hash codes."
- The hash codes are then sent from the researcher's institution to the NDAR GUID server at the NIH where they are assigned a GUID. The NDAR GUID server also stores the hash code-to-GUID relationship to ensure that the same research participant consistently is assigned the same GUID irrespective of whether he or she participates in different research studies or at different research sites.
- The NDAR GUID server returns the GUID to the researcher, who assigns it to the research participant's health information.
- Once the GUID is assigned to the research information, the combined set may be uploaded into NDAR. NDAR cannot accept data without the GUID.

The process of assigning a GUID keeps direct identifiers from ever being transmitted or stored in the NIH NDAR system. However, although no direct identifiers are transmitted to the NIH for purposes of NDAR, entities covered by

the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule<sup>[1]</sup> may wish to consider whether transmission of the hash codes to the NIH and subsequent disclosure of the GUID with health information constitute a disclosure of protected health information. Researchers should direct questions to their institutions or contact legal counsel about how the Privacy Rule may apply to a specific research project or organization <sup>[2]</sup>. Covered entities may also wish to visit the Office for Civil Rights' Web site for more information on the Privacy Rule at <http://www.hhs.gov/ocr/> and educational materials at <http://privacyruleandresearch.nih.gov>. The Privacy Rule is administered and enforced by the Office for Civil Rights.

Submissions of data to NDAR should be accompanied by a written certification (detailed below) stating that the identities of research participants will not be disclosed to the NIH NDAR data repository. Therefore, the NIH NDAR data repository will be unable to provide individual research results derived from analyses of submitted data to participants. General information regarding known publications analyzing datasets in NDAR will be made available through the repository. In addition, all submissions to the NIH NDAR data repository should be accompanied by a certification by the responsible Institutional Official(s) of the submitting institution that they approve submission to the NIH NDAR data repository.

The certification should assure that:

- The data submission is consistent with all applicable laws and regulations<sup>[3]</sup>, as well as institutional policies;
- The appropriate research uses of the data and the uses that are specifically excluded by the informed consent documents are delineated;
- The identities of research participants will not be disclosed to the NIH NDAR data repository; and
- An IRB and/or Privacy Board, as applicable, reviewed and verified that:
  - The submission of data to the NIH NDAR data repository and subsequent sharing for research purposes are consistent with the informed consent of study participants from whom the data were obtained;
  - The investigator's plan for de-identifying datasets is consistent with the standards outlined above;
  - It has considered the risks to individuals, their families, and groups or populations associated with data submitted to the NIH NDAR data repository; and
  - The genotype and/or phenotype data to be submitted were collected in a manner consistent with 45 CFR Part 46.

---

<sup>[1]</sup> Standards for Privacy of Individually Identifiable Health Information: (*if applicable*) The Department of Health and Human Services (DHHS) issued final modification to the "Standards for Privacy of Individually Identifiable Health Information", the "Privacy Rule," on August 14, 2002. The Privacy Rule is a federal regulation under the Health Insurance Portability and Accountability Act (HIPAA) of 1996 that governs the protection of individually identifiable health information, and is administered and enforced by the DHHS Office for Civil Rights (OCR). Those who must comply with the Privacy Rule (classified under the Rule as "covered entities") began doing so by April 14, 2003 (with the exception of small health plans which had an extra year to comply).

<sup>[2]</sup> Decisions about applicability and implementation of the Privacy Rule reside with the researcher and his/her institution. The OCR website (<http://www.hhs.gov/ocr/>) provides information on the Privacy Rule, including a complete Regulation Text and a set of decision tools on "Am I a covered entity?" Information on the impact of the HIPAA Privacy Rule on NIH processes involving the review, funding, and progress monitoring of grants, cooperative agreements, and research contracts can be found at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-025.html>.

<sup>[3]</sup> Applicable Federal regulations may include HHS human subjects regulations (45 CFR Part 46), FDA human subjects regulations (21 CFR Parts 50 and 56), and the Health Insurance Portability and Accountability Act Privacy Rule (45 CFR Part 160 and Part 164, Subparts A and E).

While the NIH encourages data sharing through this policy, circumstances beyond the control of investigators may preclude submission of autism research data to the NIH NDAR data repository. Applications submitted to the NIH for support of autism research in which the above expectations for data submission cannot be met will be considered for funding on a case-by-case basis by the NIH.

Submitting investigators and their institutions may use the GUID as a means to request removal of data on individual participants from the NIH NDAR data repository in the event that a research participant withdraws his or her consent. However, data that have been distributed for approved research use will not be retrieved.

### **Data Access**

As data become available through NDAR, the NIH will provide basic descriptive and aggregate summary information for general public use. Such summary information may include automated calculations and general statistics on completed assessment instruments, for example. Access to the other datasets submitted and stored in the NIH NDAR data repository, along with appropriate automated calculations (e.g., clinical assessment measures), will be provided for research purposes through the NDAR DAC. Membership of the DAC will include Federal staff with relevant expertise in areas such as the relevant particular scientific disciplines, research participant protection, and privacy. The NIH anticipates that the NDAR DAC may be established based on programmatic areas of interest and the relevant needs for technical and ethics expertise. The NDAR DAC will operate according to common principles and follow similar procedures to ensure the consistency and transparency of the NDAR data access process.

Investigators and institutions seeking data from the NIH NDAR data repository will be expected to meet data security measures (such as physical security, information technology security, and user training) and will be asked to submit a data access request, including a Data Use Certification, that is co-signed by the investigator and the designated Institutional Official(s). Data access requests should include a brief description of the proposed research use of the requested NDAR. Within a Data Use Certification, investigators will agree, among other things, to:

- Use the data only for the approved research;
- Protect data confidentiality;
- Follow appropriate data security protections;
- Follow all applicable laws, regulations and local institutional policies and procedures for handling NDAR data;
- Not attempt to identify individual participants from whom data within a dataset were obtained;
- Not sell any of the data elements from datasets obtained from the NIH NDAR data repository;
- Not share with individuals other than those listed in the request any of the data elements from datasets obtained from the NIH NDAR data repository;
- Agree to the listing of a summary of approved research uses within the NIH NDAR data repository along with his or her name and organizational affiliation;



- Agree to report, in real time, violations of the NDAR policy to the DAC;
- Acknowledge the NDAR policy with regard to publication; and
- Provide annual progress reports on research using NDAR data.

The Data Access Committee or its designees will review requests for access to determine whether the proposed use of the dataset is scientifically and ethically appropriate. 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.

At a future time, the NIH will provide additional details about and instructions that would give autism researchers the capability of working collaboratively on ongoing scientific studies where two or more researchers have common research interests and goals.

### **Data Quality**

The quality of data within NDAR is crucial for ensuring its usefulness and reliability for research. Therefore, the NIH is implementing a two-tiered data control procedure for information and images submitted to NDAR. Such efforts help to ensure that the information submitted has undergone reviews for accuracy, completeness, and availability. The first level of quality control is performed by the researcher who is expected to certify the accuracy of the information prior to submission.

The second level of quality control occurs when data and/or images are submitted to NDAR for broad research access. The NIH provides a period of up to nine months to allow the submitter and the NIH to undertake activities to review the completeness of the submission. Such efforts include verifying that the information received by NDAR is complete (i.e., not missing records intended for submission), contains no identifying information, displays correctly, and that the NDAR toolset functions as expected with the information. During this timeframe, access to data and brain images for research is temporarily suspended to help ensure that NDAR makes available only carefully reviewed information. Should the NIH determine that additional time is necessary to ensure the quality of the submitted information (e.g., time necessary to remedy concerns), the NIH may opt to extend the quality control period as necessary in the interest of science. After quality control measures are satisfied, the submitted information will be certified as accurate by the submitting researcher and will be available for sharing.

### **IV. Publication**

The NIH expects all investigators who access NDAR to acknowledge the Contributing Investigator(s) who conducted the original study, the funding organization(s) that supported the work, and the NIH NDAR data repository in all resulting oral or written presentations, disclosures, or publications of the analyses.

### **Expectations Defined in the Policy for Investigators**

The detailed expectations are enumerated in the individual sections of this policy, and summarized as follows:

**Investigators submitting NDAR data are expected to:**

- Provide descriptive information about their studies;
- Submit coded genotypic and phenotypic data to the NIH NDAR data repository; and
- Submit certification by the Institutional Official(s) of the responsible submitting institution that it has reviewed and approved submission to the NIH, providing assurance that all data are submitted to the NIH in accord with applicable laws and regulations, and that the identities of research participants will not be disclosed to the NIH NDAR data repository.

**Investigators requesting and receiving NDAR data are expected to:**

- Submit a description of the proposed research project;
- Submit a data access request, including a Data Use Certification co-signed by the designated Institutional Official(s) at their sponsoring institution;
- Protect data confidentiality;
- Ensure that data security measures are in place;
- Notify the Data Access Committee of policy violations; and
- Submit annual progress reports detailing significant research findings.

**Inquiries**

Additional information and detailed implementation guidance related to NDAR can be found at <http://ndar.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, Room 8252, MSC 9649  
Rockville, Maryland 20892-9649  
(if overnight delivery): Rockville, Maryland 20852  
Phone: (301) 443-3879  
[NDAR@mail.nih.gov](mailto:NDAR@mail.nih.gov)