Grantees Data Sharing Policy
To be included in Terms and Conditions of Awards
03/01/2012

All data resulting from this autism-related NIH-funded research involving human subjects are expected to be submitted to the National Database for Autism Research (NDA), along with appropriate supporting documentation to enable efficient use of the data. The goal of this data sharing policy is to facilitate autism spectrum disorder research and foster collaboration by giving the broader research community access to publicly available high-quality data.

Outlined below is the two-tiered approach for data submission to, and sharing through, NDA. The first tier is for descriptive data, and the second for experimental data (see Definitions). The objective of this two-tiered approach is to make data available to the research community as soon as possible without compromising the ability of the authors to interpret and communicate formally their findings.

Submission Schedule for Descriptive Data

Descriptive data are data used to characterize a research subject (see Definitions), including data from standard diagnostic assessments, standard clinical measures, family/subject medical history, demographic data, raw unprocessed images, genetic data, and genetic test results (karyotype, Fragile X, MeCP2, etc.) that are being collected in the course of the supported research. Not included as descriptive data are analyzed data, clinical observations, outcome variables, laboratory measures, etc. These are considered experimental data.

Descriptive data are expected to be submitted to NDA on a semi-annual basis. Semi-annual submission cycles conclude July 15 and January 15. The first submission of descriptive data will be expected during the second semi-annual submission cycle after the award is made. For example, for an award made in October, the first submission of descriptive data would be expected by the following July 15, skipping the January submission cycle. Regular semi-annual submissions will continue thereafter. For clinical data, cumulative data are expected (see Definitions)

The submission schedule for descriptive data is as follows:

- Data collected through June 1 are to be deemed reasonably accurate by the Principal Investigator, and submitted by July 15.
- Data collected through December 1 are to be deemed reasonably accurate by the Principal Investigator, and submitted by January 15.

Submission Schedule for Experimental Data

Experimental data (see Definitions) are expected to be submitted within twelve (12) months after accomplishment of each primary aim or objective (or set of interdependent aims or objectives) of the supported research, or at the time of publication of the results of the primary aim(s), whichever occurs first. Prior to award, the Principal Investigator and the NIH program official will determine what constitutes the primary aims of the project, and which primary aims are considered to be interdependent. Included as experimental data are:

- Experimental results.
- Data from custom or proprietary clinical assessments/measures that support the aims of the proposed research or are otherwise not included in the semi-annual submissions.
- Final data and/or images derived from processed images (see Definitions).
- Sufficient supporting documentation to enable efficient and appropriate use of the data by the broader research community (see Definitions).
- All other de-identified research data acquired through the supported research but not explicitly listed here.

Provisions for Data Submission

- All human subject data provided must include an NDA Global Unique Identifier (GUID) and must not include personally identifiable information (PII).
- All data collected on all human subjects involved in the NIH-supported research are expected to be provided. These include data from control subjects and related family members. The total number of subjects for which data are provided should be consistent with the total number of subjects reported on the 2590 Inclusion Enrollment Report. It is understood that gaps in data will exist in the event that not all participants agree to share their data, or do not complete the entire protocol for other reasons.
- Custom or proprietary measures not currently defined in the NDA Data Dictionary will require the investigator to define the data measures, data structures, and discrete data elements using the NDA Data Dictionary Tool, allowing those data to be made available for sharing.
- Individual subject-level data rather than summary/aggregate data are expected.
- Item-level detail on core autism measures such as ADOS, ADI, Vineland, and a research participant's medical/family history are expected.
- Due to the challenges inherent in de-identifying video footage, video material should not be submitted.

Data Sharing

All submitted data (both descriptive and experimental) will be made available for access by members of the research community according to the provisions defined in the NDA

Data Sharing Policy. The data sharing policy is intended to allow investigators sufficient time for data verification, and for submission of primary publications based on the collected data.

Descriptive research data are made available for access to other researchers within **four (4) months after submission,** allowing the Principal Investigator and their team sufficient time to complete appropriate quality assurance/quality control (QA/QC) procedures. Thus, there would be between five (5) and eleven (11) months from collection to sharing of descriptive data.

Experimental research data are made available for access to other researchers within **four (4) months after submission,** which is sixteen (16) months after completion of the primary aims or publication. To clarify, an investigator would have twelve (12) months after the primary aims were met to submit the experimental data to NDA, and then would have an additional four (4) months to review the data. Thus, a total of 16 months would elapse between the time the primary aims were met and the time the data are shared.

It is expected that any deviations from the above in terms of timelines or types of data to be shared may be negotiated with the NIH program officer for the grant (or other award mechanism) before the award is made. If circumstances arise during the course of the research that might cause deviations from these terms, such deviations must receive approval as defined in NDA SOP-10 Request Time Extension for Sharing or NDA SOP-11 Deviations in Data Sharing Terms.

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 the repository staff or secondary data users. Submissions of data to NDA must be accompanied by the NDA Data Submission Agreement.

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 access request, including 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 http://NDA.nih.gov/ndarpublicweb/policies.go#SOP.

DEFINITIONS

Cumulative data: A dataset that includes all data collected from the beginning of the study to designated time point; each submission replaces previously submitted data sets in order to avoid the challenges of tracking interim changes or corrections in the database. Data containing references to large files (e.g., genomic, imaging, and other rich data types), may be provided incrementally for efficiency reasons.

Data: For human subjects, data include all research and clinical assessments and information obtained via interviews, direct observations, laboratory tasks and procedures, records reviews, genetic and genomic data, neuroimaging data, psychophysiological assessments, data from physical examinations, etc. The following are not included as data: laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as gels or laboratory specimens.

Descriptive data: Descriptive data include family/medical history, demographic data, data from standard diagnostic instruments, or custom measures supporting a categorization of a subject's phenotype. Examples include but are not limited to ADOS, ADI-R, IQ, Vineland, M-CHAT, Medical History, etc. Additionally, raw unprocessed images and genomic submissions are also categorized as descriptive data. For longitudinal neuroimaging studies, where images at different time points are considered outcome measures, only baseline raw images are expected as descriptive data.

Genomic data:
Descriptive genomic data are defined as the raw or primary data specific to the technology platform used for the research study. If a microarray technology is used, an example of descriptive data is the intensity data such as an Affymetrix CEL file. Descriptive data submissions from research studies using the next generation of sequencing technology should include the read data, the second most frequent base and the quality data. Formats for these submissions include fastq, AB SOLiD Native, AB SOLiD SRF, Illumina Native, Illumina SRF, and Roche 454 SFF.

Experimental genomic data are defined as data derived from the primary or raw data. For the example of the next generation of sequencing technology, experimental data would be alignments or mapped data in the BAM (Binary Alignment/Map) format or the Sequence Alignment/Map (SAM) Format. Examples of experimental data from the SNP microarray technology would include copy number and/or genotype. For the gene expression microarray technology, an example of experimental data would be normalized gene expression levels.

The investigator is required to provide enough information to allow other researchers to repeat the experiment. Information provided using NDAR's Experiment Definition Tool includes the experimental molecule, used technology and experimental platform,

protocols used for molecule and experiment preparation and kits used for these purposes, as well as names of analysis software, experimental equipment and description of analysis protocols.

Raw unprocessed images: Data acquired from a scanner in a standard medical imaging format (e.g. DICOM, NIFTI).

Processed images: Derived data generated as the final result of image analysis applications in any standard medical research format (e.g. NIFTI, Analyze, DICOM, MINC, MIPAV, AFNI, SPM, etc.). Intermediate image datasets should not be submitted unless the investigator feels that they are pertinent.

Experimental Data: Data specific to the primary aims of the research being conducted (e.g. outcome measures, other dependent variables, observations, laboratory results, analyzed images, volumetric data, etc.)

Supporting documentation: Clear documentation expected in order to enable an investigator unfamiliar with the dataset to understand and use the data. For example, supporting documentation may include non copyrighted data collection forms, study procedures and protocols, data dictionary rationale, exclusion criteria, website references, a listing of major study publications, and the definition of a genomic experiment using the NDA Experiment Definition Tool.