ARRA Grantees Data Sharing Policy
To be included in Terms and Conditions of Awards
8/17/2009

All data (see Definitions) collected in ARRA autism funded projects that involve human subjects are expected to be made available in the National Database for Autism Research (NDA), along with appropriate supporting documentation to enable efficient use of the data by the broader research community.

The goal of this data sharing policy is to foster collaboration and allow broader use of high quality research data, and doing so at the appropriate times. The terms described below outline a two tiered approach to data submission and sharing of human-subjects data, one for descriptive data and one for experimental data.

1) Descriptive research data that are used to characterize a research subject are expected to be submitted on a semi-annual basis. Included as descriptive data are standard diagnostic assessments, standard clinical measures, family/subject medical history, demographics, raw unprocessed images, and genetic data (see Definitions). Not included as descriptive data are analyzed data, clinical observations, outcome variables, laboratory measures, etc. These are considered experimental data.

   Descriptive research data will be made available for access by the research community within **four (4) months after submission,** allowing time for the research team to complete appropriate quality assurance/quality control (QA/QC) activities.

2) Experimental data (see Definitions) are expected to be submitted within a year of the completion of the research/ primary objective(s) or at the time of publication of the project's objectives, whichever occurs first.

   Experimental research data will be made available for access by the research community within **four (4) months after submission,** which is 16 months after completion of the primary aim(s) or publication.

The objective of this two-tiered approach, detailed below, is to make data available to the research community as soon as possible without compromising the ability of investigators to interpret and communicate formally their findings.

Submission Schedule for Descriptive Data

Cumulative datasets for descriptive data are expected to be submitted to NDA on a semi-annual basis. Included in these semi-annual submissions are data from standard diagnostic assessments, standard clinical measures, family/subject medical history, demographic data, raw unprocessed images, and genetic data (karyotype, gene expression, SNP, exon, microRNA, methylation, etc.) that are being collected in the course of the supported research (see Definitions).

Semi-annual submission cycles conclude in July and January. 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 succeeding July 15 submission cycle, skipping the January submission cycle. Regular semi-annual submissions will continue thereafter. The submission schedule for descriptive data is as follows:

- Data collected through June 1 are to be verified as accurate by the Principal Investigator, and submitted by July 15 or the next business day.
- Data collected through December 1 are to be verified as accurate by the Principal Investigator, and submitted by January 15 or the next business day.

Submission Schedule for Experimental Data

Experimental data are expected to be submitted within 12 months after each primary aim or objective (or set of interdependent aims or objectives) of the research has been completed or at the time of publication of the results of the primary aim(s) of the supported research, whichever occurs first. The Principal Investigator and the NIH program official will determine what constitutes the primary aims of the project, and which are considered to be interdependent, prior to award. Included in this data submission are:

- Experimental data (see Definitions).
- 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).
- 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 expressly 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 human subjects involved in the proposed research is expected to be provided. This includes data from control subjects and related family members. The total number of subjects for which data is 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 for other reasons do not complete the entire protocol.
- 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 Field Integration Tool, allowing those data to be made available for sharing.
- Whenever possible, individual level data rather than summary/aggregate data are expected.

Data Sharing
All submitted data (both descriptive and experimental) will be made available for access by the research community according to the provisions defined in the NDA Data Sharing Policy within **four (4) months** from the time of submission to allow sufficient time for the research team to complete appropriate quality assurance/quality control (QA/QC) activities.

To clarify, an investigator would have 12 months after primary aim(s) was/were met to submit the experimental data to NDA, and then would have an additional 4 months to review the data. Thus, a total of 16 months would elapse between the time the primary aim(s) was/were met and the time the data are shared. For descriptive data, there would be between 5 and 11 months from data collection to data sharing.

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 official before the award is made. If circumstances arise, during the course of the research, that necessitate further 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.

2. Privacy
All data 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.

3. Publication
Investigators are expected to publish the results of their studies using the data they have collected. The data sharing policy is intended to allow investigators sufficient time for data verification, and submission of primary publications.

4. Data access for research
As data become available through NDA, the NIH will provide basic descriptive and aggregate summary information for general research use. 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). Access to the NIH NDA data repository will be provided for research purposes through the NDA Data Access Committee (DAC). The procedures associated with data access are defined at: http://NDA.nih.gov/ndarpublicweb/policies.go#SOP.

DEFINITIONS –
Cumulative Dataset: data set 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: 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.

Genomics: Genomics submissions include a set of templates in which the investigator is required to provide enough information to allow other researchers to repeat the experiment.

Raw unprocessed images: Data acquired from the 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 image data 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 is needed and 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, and a listing of major study publications.

The below procedure will be posted on NDA.nih.gov.

SOP-11 Deviations to Data Sharing Terms

Over the course of research, circumstances may arise that necessitate a change in the data sharing terms associated with NIH funded research. In anticipation of this need, the following procedure has been developed.

Procedure

1. The investigator defines the need to deviate from the established terms of the research, defining the scientific need to deviate from the proposed terms and conditions.
2. The owner of the NDA Collection will upload the file of the concern to their profile in PDF format, defining the document type as "Deviation to Data Sharing Terms" within NDA and associate the request to the appropriate NDA Collection(s).
3. NDA Staff will forward the request to the NDA Data Access Committee and appropriate NIH Program Officer.
4. The NDA Data Access Committee and Program Officer will approve the request or consult with the investigator for clarification/modification.
5. Once approved, the NDA Staff will then set the status of the request in NDA to Approved and this deviation will be honored and included in the terms associated with subsequent awards.