

## National Database for Clinical Trials Related to Mental Illness (NDCT) Data Sharing Plan

09/15/2014

### i. Data Sharing Overview

As indicated in the NIH Guide Notice “Data Sharing Expectations for NIMH-funded Clinical Trials ([NOT-MH-14-015](#)),” all de-identified data resulting from this NIMH-funded award are expected to be submitted to the National Database for Clinical Trials Related to Mental Illness (NDCT, <http://ndct.nimh.nih.gov>), along with appropriate supporting documentation to enable efficient and appropriate use of the data. The goal of this data sharing policy is to enable rapid, widespread sharing of high-quality, human-subjects data to the research community, and thereby maximize the value and utility of the data and the research.

NIH has established a two-tiered approach for data submission to, and sharing through, NDCT. The first tier is for the submission of descriptive/raw data while the study is ongoing, while the second tier is for the submission of analyzed data at the publication of results or after the completion of the award period, whichever comes first (see Definitions). Regardless of the timing of submission, neither descriptive/raw data nor analyzed data will be made available for access until publication of findings or completion of the award period, whichever comes first—this delayed access includes investigators conducting ongoing, blinded studies. This tiered approach enables summary-level reporting on the progress of the study and prevents the premature release of data, while also enabling rapid availability of data at the time of publication or study conclusion.

### ii. Submission Schedule for Descriptive/Raw Data

Descriptive/raw 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, -omics (e.g. proteomics, genomics, metabolomics) data, raw neurosignaling recordings, and genetic test results that are being collected in the course of the clinical trial. Analyzed data, outcome variables, processed neurosignal recordings, etc., are not considered descriptive/raw data.

Descriptive/raw data are expected to be submitted to NDCT on a semi-annual basis (on or before January 15 and July 15, beginning six months after the award budget period has begun). NDCT support staff will contact the Principal Investigator following award to plan an appropriate data submission schedule. NIH expects cumulative submission of descriptive/raw data during each submission cycle, which will enable data corrections

throughout the duration of the award. Raw -omic, EEG, and neuroimaging data are expected to be submitted incrementally as new data are acquired.

### iii. Submission Schedule for Analyzed Data

Analyzed data (see Definitions) are expected to be submitted prior to publication/public dissemination (whether the findings are positive or negative). Even if a publication focuses on only part of an analyzed dataset, the entire analyzed dataset should be submitted when the first paper/finding is published or communicated. The data that are not part of the paper will not be immediately shared, but rather along the timeline described in the Data Sharing section below.

Analyzed data include:

- 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 award but not explicitly listed here.

Additionally, Principal Investigators are expected to associate the data deposited in NDCT with their publications/findings—both positive and negative—using the NDCT Study feature (see: <http://ndct.nimh.nih.gov/results/>).

#### Provisions for Data Submission into NDCT

- All human-subjects data provided must include a Global Unique Identifier (GUID) and must not include personally identifiable information (PII).
- Submission of data into and sharing of data via NDCT should be mentioned in the informed consent process of the study.
- The awarded institution and Principal Investigator must ensure that submission and sharing via NDCT are consistent with the informed consent of study participants from whom the data are obtained.
- All data collected on all human subjects involved in this NIH-supported clinical trial are expected, including data from control subjects. The total number of subjects for which data are provided should be consistent with the total number of subjects reported on the annual progress report.
- In the event that the research involves custom or proprietary measures not currently defined in the NDCT Data Dictionary, the Principal Investigator will ensure the definition of the data by defining the specific data elements and sending these definitions to NDCT for curation. Once these measures have been defined, the associated data can then be submitted to NDCT.
- NIH expects individual subject-level data, rather than summary/aggregate data.

- Video recordings of research participants are expected when necessary to demonstrate a specific clinical trial result, and only if the recordings can be effectively de-identified. Otherwise, video recordings are not expected.
- The Principal Investigator is expected to communicate this data sharing plan to appropriate research staff to ensure the timely submission of data.

#### **iv. Data Sharing Schedule**

All submitted data (both descriptive/raw and analyzed data) will be made available for access by qualified members of the research community according to the provisions defined in the NIMH Data Repositories Data Access Agreement and Use Certification. These procedures are intended to allow investigators sufficient time for data verification, and for submission of primary publications based on the collected data. Descriptive/raw research data are held until the finding is communicated or published.

Analyzed data are expected at the time a manuscript is accepted for publication. These data will then be shared when the publication is released, along with the associated descriptive data. Data that remain unpublished are expected prior to project completion and will be shared within one year after the original project completion date, allowing the Principal Investigator and his/her team sufficient time to complete appropriate quality assurance/quality control procedures.

In the event circumstances arise during the course of the award which the Principal Investigator believes necessitate deviations from this data schedule, the Principal Investigator must receive approval from the NIH Program Official overseeing the award.

#### **v. Privacy**

All data (see Definitions) made available for public use via NDCT 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 NDCT must be accompanied by the NIMH Data Repositories Data Submission Agreement, which is expected within 3 months of award.

#### **vi. Data Access for Research Purposes**

Access to data for research purposes will be provided through the NDCT Data Access Committee (DAC). Investigators and institutions seeking data from NDCT 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).

## vii. Definitions

**Analyzed 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.)

**Cumulative data:** A dataset that includes all data collected from the beginning of the study to designated time point; each submission replaces previously submitted datasets 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 (i.e., not cumulatively) 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, EEG, eye tracking, 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/raw data:** Descriptive/raw data include family/medical history, demographic data, data from standard diagnostic instruments, or custom measures supporting a categorization of a subject's phenotype. Additionally, raw unprocessed images and genomic submissions are also categorized as descriptive/raw data. For longitudinal neuroimaging studies, where images at different time points are considered outcome measures, only baseline raw images are expected as descriptive/raw data.

**Experiment definition:** The Principal Investigator is expected to use the NDCT Experimental Definition Tool, an online resource, to provide enough information to allow other researchers to repeat the experiment (see <http://ndct.nimh.nih.gov/submit/#tab-3>). For -omics data, experiment definition information includes the experimental molecule, the 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. For neurosignal recordings, experiment definition includes timing sequences, event definition, and acquisition hardware/software specification.

### **-Omics data:**

Descriptive/raw 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/raw data is the intensity data such as an Affymetrix CEL file. Descriptive/raw 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.

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

**Processed images:** Derived data generated as the final result of image analysis applications in any standard medical research format (e.g. NIFTI, AFNI, etc.). If applicable, supporting de-identified video and imaging materials that define the experiment (e.g., timing sequences in fMRI) should accompany processed images. Intermediate image datasets should not be submitted unless the investigator feels that they are pertinent.

**Raw unprocessed images:** Data acquired from a scanner in a standard medical imaging format. DICOM format is preferred.

**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 NDCT Experiment Definition Tool. Definition related to a specific finding or publication is to be defined and documented through the NDCT Study feature.