EXAMPLE DATA MANAGEMENT AND SHARING PLAN (in compliance with SF-424 Forms H)

EXAMPLE FOR GENOMIC DATA FROM HUMAN RESEARCH PARTICIPANTS

ELEMENT 1: DATA TYPE

A. Types and amount of scientific data expected to be generated in the project:

Our genomic study will be registered with dbGaP, and our raw whole genome sequencing data and derived data will be submitted to the NIMH Data Archive (NDA). Phenotypic and clinical data for all 500 research subjects will be collected and deposited in NDA using the data dictionaries available in NDA (described below).

B. Scientific data that will be preserved and shared, and the rationale for doing so:

All raw and processed genomics files and all clinical and phenotypic data will be shared.

C. Metadata, other relevant data, and associated documentation:

The Institutional Certification will be submitted to NIH during the dbGaP registration process once we have been told that a grant award is likely. Within the first six months following the award, we will submit the Data Submission Agreement to NDA and will create the Data Expected list in our new NDA Collection. A brief study protocol will also be submitted to NDA and will be made freely available.

ELEMENT 2: RELATED TOOLS, SOFTWARE, AND/OR CODE

Genotypic data undergo an extensive automated data cleaning process in the laboratory. Our replication plan for observed associations is outlined in the Research Strategy. While all sequencing data from this proposal will be generated using Illumina pipelines, differences in read depth and primer libraries between studies will require joint re-calling of all genotypes from raw read files to yield the highest possible quality calls and a harmonized dataset for future use in follow-up and unrelated studies. Using the Broad Institute's Genome Analysis Toolkit (GATK), we will apply standard Best Practices workflows for single nucleotide variant (SNV) and Indel discovery from whole genome sequence alignment files (SAM/BAM). These steps should ensure that final association results are representative of "true" genotypes rather than miscalls or confounded genotypes that are unlikely to replicate in independent populations.

ELEMENT 3: STANDARDS

In compliance with NOT-MH-20-067, the following common data elements will be collected to facilitate aggregation of this data set with other data sets:

1) Age (ndar_subject01)
2) Sex at Birth (ndar_subject01)
1) DSM Crosscutting (dsm5crossa0)
2) WHODAS 2.0 (whodas201)
3) PHQ-9 (phq901)
4) GAD-7 (cde_gad701)

As described in the Research Plan, the additional phenotypic and clinical information will be collected using the following data dictionaries obtained from NDA:

1) Genomics Subject (genomics_subject02)

Commented [JB1]: *These example DMS Plans are provided for educational purposes to assist applicants with developing Plans but are not intended to be used as templates and **their use does not guarantee approval by NIH. Do not copy/paste this Plan without modifying it to reflect the types of data that are expected to be generated through your project.***

Note that the example DMS Plans may reflect additional expectations established by NIH or specific NIH Institutes, Centers, or Offices that go beyond the DMS Policy. Applicants will need to ensure that their Plan reflects any additional, applicable expectations (including from NIH policies, ICO policies, or as stated in the FOA).

In addition, these examples may reflect resources or policies that are in place at other institutions but that are not necessarily available at the University of Iowa. If needed, investigators can contact Research Data Services (lib-data@uiowa.edu) if they have questions regarding how to best complete their DMS Plan.

Commented [BJY2]: *Example from NIMH: https://www.nimh.nih.gov/sites/default/files/documents/funding/managing-your-grant/resource-sharing-docs/Human%20Genomic_Template_v2.docx*

2) Genomics Sample (genomics_sample03)
3) Structured Clinical Interview for DSM-V (scidv_dep01)
4) MATRICS Consensus Cognitive Battery (matrics01)

The sequence data will be stored in standard formats FASTQ, SAM/BAM, BED, and VCF. Those data files will all be deposited into NDA. The description of the genomics experiment will be submitted using the NDA genomics_sample03 data structure. Additional experimental protocols will be described in NDA Experiments associated with our NDA Collection.

ELEMENT 4: DATA PRESERVATION, ACCESS, DISTRIBUTION AND ASSOCIATED TIMELINES

A. Repository where scientific data and metadata will be archived:

All data will be deposited to NDA starting 12 months after the award begins and will be deposited every six months thereafter following the usual NDA data submission dates.

B. How scientific data will be findable and identifiable:

Data will be findable for the research community through the NDA collection that will be established when this application is funded. In addition, the dbGaP study, which will point to NDA, will help researchers find the data. For all publications, an NDA study will be created. Each of those studies is assigned a digital object identifier (DOI). This data DOI will be referenced in the publication to allow the research community easy access to the exact data used in the publication.

C. When and how long the scientific data will be made available:
The research community will have access to data at the end of the grant award or when a publication has been submitted. Once the data are submitted to NDA, that archive will control the long-term persistence of the data set. Currently, NDA has no process for deleting or retiring data sets.

ELEMENT 5: ACCESS, DISTRIBUTION, OR REUSE CONSIDERATIONS

A. Factors affecting subsequent access, distribution, or reuse of scientific data:

All research participants will be consented for broad data sharing.

B. Whether access to scientific data will be controlled

To request access of the data, researchers will use the standard processes at NDA, and the NDA Data Access Committee will decide which requests to grant. The standard NDA data access process allows access for one year and is renewable.

C. Protections for privacy, rights, and confidentiality of human research participants:

The NDA GUID tool allows researchers to aggregate data from the same research participant without different laboratories having to share personally identifiable information about that research participant. The NDA data dictionaries do not permit personally identifiable information to be shared. NDA maintains a Certificate of Confidentiality.

ELEMENT 6: OVERSIGHT OF DATA MANAGEMENT AND SHARING

The following individuals will be responsible for data collection, management, storage, retention, and dissemination of project data, including updating and revising the Data Management and Sharing Plan as necessary each year at the time of the Research Performance Progress Report.

[Name, position title, host institution, ORCID, email]

Validation Schedule (this section is required by NIMH)

If funded, within 6 months of the Notice of Award date we will submit a Data Submission Agreement signed by the principal investigators and an institutional business official. We will also define and complete the Data Expected section of this project. Uploads of demographic, clinical, and raw structural MRI, ^{1}H fMRS and fMRI research data will begin at the second submission cycle deadline following the Notice of Award date. Subsequent data uploads will be harmonized, validated, and submitted biannually on the standard January 15th and July 15th submission deadlines.

The NDA Validation and Upload tool will be used for quality control on newly collected phenotypic and clinical data every two weeks.