DATA MANAGEMENT AND SHARING PLAN
An example from an application proposing to collect clinical and MRI data from human subjects.

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on sharing.nih.gov. The Plan is recommended not to exceed two pages. Text in italics should be deleted (but this has not been done in the sample below). There is no "form page" for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format* shown below.

Element 1: Data Type
   A. Types and amount of scientific data expected to be generated in the project:
   Summarize the types and estimated amount of scientific data expected to be generated in the project.

Demographic, clinical, and MRI, ^{1}H fMRS and fMRI imaging data will be acquired from 110 affected youth and 110 matched healthy controls (described in detail in sections C.3 and C.4 of this application). All data will be de-identified prior to receipt by the repository, but the information needed to generate a global unique identifier for the NIMH Data Archive (NDA) will be collected for each subject.

   B. Scientific data that will be preserved and shared, and the rationale for doing so:
   Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.

Sufficient data from this project will be preserved to enable sharing via NDA data of sufficient quality to validate and replicate research findings described in the Aims. NIMH requires data measured from human subjects to be shared using the NDA.

   C. Metadata, other relevant data, and associated documentation:
   Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

In addition to the subject level data described above, all ^{1}H fMRS and fMRI task related paradigm designs and experiment definitions will be deposited in the NDA.

Element 2: Related Tools, Software and/or Code:
   State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

The clinical data will be analyzed with custom Python code written using the statsmodels, numpy, and pandas packages, all of which are freely available. ^{1}H fMRS spectra will be analyzed with LCModel 6.3 software using LCMgui, which is freely available from http://s-provencher.com/lcm-test.shtml. fMRI images will be analyzed using the SPM8 toolbox (https://www.fil.ion.ucl.ac.uk/spm/software/spm8/) for MATLAB (http://www.mathworks.com/products/matlab/). While MATLAB is commercial software, most universities have site licenses available and the SPM8 toolbox is free. It is also possible that the toolbox might run in Octave, an open-source alternative to MATLAB (https://www.gnu.org/software/octave/index), but we have not tried it. All code will be shared on our lab website at https://github.com/labname the main readme.md file for the project will also include instructions and parameter choices for the GUI-based analyses.

Element 3: Standards:
   State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist.

Participant age, sex, ethnicity, height, weight, socioeconomic status, and other demographic data will be collected using the following instruments as defined in NDA:

1) Research Subject and Pedigree (ndar_subject01)
2) Demographics Short Form (demsf01)
3) Ethnic Group Questionnaire (ethgrp01)
4) Height and Weight (height_weight01)
5) Hollingshead Socioeconomic Rating Scale (ses01)
6) Pubertal Development Scale (pds01)
7) Edinburgh Handedness Inventory (edinburgh_hand01)
8) WASI-2 (wasi201).

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

1) DSM Crosscutting for Youth (dsm5crossch01)
2) RCADS-25 (rcads2501)

The clinical assessments we plan to collect for this study include:

1) Kiddie-SADS-Present and Lifetime Version (ksads_pl01)
2) Children's Yale-Brown Obsessive Compulsive Scale (cybocs01)
3) Schedule for Obsessive-Compulsive and Other Behavioral Syndromes (Hanna. Schedule for Obsessive-Compulsive and Other Behavioral Syndromes, Ann Arbor: University of Michigan, 2010, new data dictionary will be defined in NDA)
4) Dimensional Obsessive Compulsive Scale (docs01)
5) Yale Global Tic Severity Scale (yale01)
6) Child Behavior Checklist (cbcl01)
7) Multidimensional Anxiety Scale for Child Parent and Self (masc_p01)
8) Conners 3 (conners3_ps01)
9) Adolescent Depression Rating Scale (doi:10.1186/1471-244X-7-2, new data dictionary will be defined in NDA)

^{1}H fMRS and fMRI data will be shared with the Image (image03), Imaging Work Flow (iwf01), and Imaging Collection (imagingcollection01) data dictionaries as defined in NDA.

Element 4: Data Preservation, Access, and Associated Timelines
   A. Repository where scientific data and metadata will be archived:
   Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see Selecting a Data Repository).

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:
   Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

Data will be findable for the research community through the NDA Collection that will be established when this application is funded. 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:
   Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

The research community will have access to data when the award ends. As required by NDA, studies will also be created that contain the data used for every publication. Those studies will be shared when the pre-print is available. NDA studies have digital object identifiers (DOI) to aid in findability. We will include that DOI in relevant publications. NDA will make decisions about how long to preserve the data, but that data archive has not deleted any deposited data up to now.

Element 5: Access, Distribution, or Reuse Considerations
    A. Factors affecting subsequent access, distribution, or reuse of scientific data:
        NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See [Frequently Asked Questions](#) for examples of justifiable reasons for limiting sharing of data.

All research participants will be consented for broad data sharing.

    B. Whether access to scientific data will be controlled:
        State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

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:
        If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

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:
    Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

The Office of Sponsored Programs at University X that will be administering this award has created a data management and sharing plan compliance system as part of their process for submitting the annual NIH progress report. That Office is collecting information related to the number of research participants that are deposited each reporting year. The Office of Sponsored Programs will also look for the NDA data DOIs from NDA Studies and will include that information in the annual progress report.

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, as well as define and complete the Data Expected section of this project. Uploads of all initial demographic, clinical, and raw structural MRI, ^{1}H fMRS and fMRI research data will be completed using 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.

We also plan to use the NDA validation tool as a quality control measure in the laboratory. The data manager in charge of submitting data to NDA will help researchers in the group validate their data once every month.