The National Institute on Aging (NIA) Division of Social and Behavioral Research (BSR) provides the following sample Data Management and Sharing Plan for a project involving collection of electronic health record data (EHR) for a new study. Click here for more sample plans from NIA.

DATA MANAGEMENT AND SHARING PLAN

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 are instructions and should be deleted before submitting the plan. 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.

Clinical data will be obtained from electronic health records (EHR) for ~15,000 patients in a primary care clinic. Clinical data include demographic data, medical history, laboratory data, medications, physical exams, and cognitive exams. Cognitive assessment data for a subset of individuals (~40) will be collected via validated assessment instruments. This project proposes to use clinical data to generate cognitive risk scores.

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.

De-identified summary measures such as risk scores and assessment outcomes will be made openly available. De-identified individual level data will be available through a controlled access plan requiring a confidentiality agreement and institutional approval as described in Element 5 of this DMS Plan.

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.

Operational definitions, data dictionaries, and programming code will be made available with data at the [Repository Name Here].

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.

Data will be analyzed and data which can be publicly shared will be formatted for widely available statistical packages such as R, SPSS, STATA, and SAS.

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.

Clinical and laboratory data will be standardized using LOINC. ICD-10 code lists will be shared to facilitate implementation of shared codes.

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).

Data and metadata will be deposited with [Repository Name Here].

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.

Clinical data will be findable using a Digital Object Identifier (DOI) assigned by [Repository Name Here].

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.

Data and metadata will be deposited into the repository at the time of publication or by the end of the project period, whichever comes first. Data will be preserved for at least five years following the end of the grant period. Researchers may request controlled-access data using standard processes at [Repository Name Here].

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.

Individual-level data will be shared with controlled access as allowed by informed consent agreements approved by the Institutional Review Board (IRB).

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).

Prior to release of individual-level data, researchers will be required to complete data sharing and confidentiality agreements as established by [Repository Name Here]. Researchers must commit to 1) not attempting to re-identify participants, 2) secure the data with appropriate technology and enforce strict access rules, 3) destroy data after analyses are completed, 4) meet any requirements as stipulated by the health system involved in the study, and 5) if research is conducted within a covered entity, HIPAA requirements must be met prior to use.

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).

All accessible data will be de-identified and in the case of HIPAA conflicts, a limited data set will be created. The institution will determine whether the de-identified individual-level data are subject to the NIH Certificate of Confidentiality policy, and if so, will protect accordingly and ensure that recipients of the data are aware of the protection.

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 PI will oversee archiving and sharing of the analysis codes generated from this project upon publication or at the end of the project period, whichever comes first. The final project report will summarize adherence to this data management and sharing plan.