Data Management and Sharing Plan

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.

The project will generate longitudinal social, behavioral, and biological data from a large cohort of older adults. Specifically, approximately 13,000 web-surveys and in-person interviews will be conducted, collecting information on health, cognition, psychosocial factors, and demographics. Additionally, around 200 home health exams will be completed to collect biological data, including physical measurements and biospecimen information. The anticipated dataset will include structured survey responses, interview transcripts or coded data, and physiological and laboratory measures, totaling an estimated 13,200 participant records. The total storage size is expected to be approximately 50-100 GB, including data, metadata, and documentation.

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.

All de-identified individual-level data collected from the web-surveys, in-person interviews, and home health exams will be preserved and shared, in accordance with participant consent for broad data sharing. This includes social, behavioral, and biological variables, as well as derived variables relevant to aging research. Sharing these data will facilitate secondary analyses and reproducibility and maximize the scientific value of the study in alignment with NIH and NIA data sharing policies.

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.

The following documentation will be made available to support data interpretation and reuse:

- Data dictionaries/codebooks for all variables

- Survey instruments and interview protocols

- Home health exam protocols and standard operating procedures

- Study protocol and informed consent forms

- Metadata compliant with repository standards, including study design, sampling, and data collection methodology

- User guides and analytic notes describing variable construction and data processing steps

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 shared data will be provided in standard, non-proprietary formats (e.g., CSV, TXT, and Stata/SAS/SPSS files), which can be accessed using commonly available statistical software (e.g., R, Stata, SAS, SPSS) and spreadsheet programs. No specialized or proprietary software is required. If any custom code (e.g., for derived variables or data cleaning) is developed, it will be documented and shared as annotated scripts (e.g., R or Stata do-files) in the repository.

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.

The project will apply established data and metadata standards to promote data interoperability:
- Variable naming and coding will follow conventions established by the Inter-university Consortium for Political and Social Research (ICPSR) and the National Archive of Computerized Data on Aging (NACDA).
- Metadata will be structured using the Data Documentation Initiative (DDI) standard.
- Biological data will be annotated in accordance with relevant NIH Common Data Elements (CDEs) where available.
- Survey and protocol documentation will follow NIA and NIH recommendations for longitudinal aging studies.

If new or domain-specific standards emerge during the project, they will be adopted as appropriate.

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.

De-identified data and associated documentation will be archived at the National Archive of Computerized Data on Aging (NACDA), a domain-specific NIH-supported repository specializing in social and behavioral data on aging. Data may also be deposited with ICPSR as a secondary repository.

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.

All datasets and documentation will be assigned persistent Digital Object Identifiers (DOIs) by the repository. Metadata records will be indexed in repository catalogs and major data discovery portals to facilitate findability.

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 the time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Data will be submitted to the repository and made available for sharing no later than the time of publication of primary results, or at the end of the project performance period, whichever comes first. Data will remain available through the repository for a minimum of 10 years, in accordance with NIH policy and repository guidelines.

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.

All participants will be consented for broad data sharing; however, data will be de-identified to protect privacy and confidentiality. Some sensitive variables (e.g., detailed geographic information, rare health conditions) may be restricted or provided as derived variables to minimize re-identification risk. Users will be required to agree to standard repository terms of use, including restrictions on re-identification and requirements to protect participant privacy.

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).

Access to the de-identified dataset will be controlled via the repository's data access system. Qualified researchers will be required to submit a data use application and agree to conditions of use prior to obtaining access.

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 shared data will be de-identified in accordance with HIPAA and NIH guidelines, with direct identifiers removed and indirect identifiers minimized or masked as appropriate. Data will be reviewed for re-identification risk prior to release. The project holds a Certificate of Confidentiality from NIH to further protect participant data. Access to sensitive data, if any, will be further restricted as necessary to preserve participant 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).

Compliance with this Data Management and Sharing Plan will be overseen by the Principal Investigator, with support from the project's Data Manager and an Institutional Data Steward. Data management activities will be reviewed quarterly during project meetings, and adherence to the plan will be documented in the annual progress reports to NIH. The Institutional Review Board (IRB) and the Office of Research Compliance will also monitor data sharing practices to ensure alignment with approved protocols and regulatory requirements.