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 proposed project involves collecting social, behavioral, and biological data through web-surveys, in-person interviews, and home health exams. Approximately 13,000 web-surveys and interviews and 200 home health exams will be completed, generating a substantial amount of longitudinal data on aging.

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 data collected from web-surveys, interviews, and home health exams will be preserved and shared. The rationale for sharing these data is to contribute to the broader research community's understanding of aging processes, facilitate replication studies, and enable meta-analyses that can inform policies and interventions aimed at improving outcomes for older adults.

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.

Metadata will include participant demographics, survey and interview protocols, home health exam procedures, and data collection timelines. Other relevant data include linkage to previous waves of the longitudinal study to enable analysis of changes over time. Associated documentation will comprise study protocols, informed consent forms, data dictionaries, and detailed descriptions of data cleaning and processing methods.

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.

Specialized tools and software needed for data access and manipulation include statistical analysis packages like R or SAS, and possibly data visualization tools. These are widely available and commonly used in the research community, facilitating broad accessibility to the shared scientific data.

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 common data standards for social sciences and health research, including those recommended by the NIH for data sharing, such as using standardized variable names and formats (e.g.,

SDTM for clinical trials data). For longitudinal studies, we will follow guidelines that enable integration with existing datasets in the field of aging, ensuring interoperability and facilitating combined analyses.

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.

The National Archive of Computerized Data on Aging (NACDA) at the University of Michigan's Inter-university Consortium for Political and Social Research (ICPSR) has been selected as the repository for archiving the scientific data and metadata.

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.

Each dataset will be assigned a unique DOI (Digital Object Identifier), enabling easy location and citation. Additionally, detailed metadata records will be created to facilitate discovery through search engines and the NACDA/ICPSR catalog.

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.

The scientific data will be made available no later than the time of publication of the main findings or within 12 months of the completion of data collection, whichever comes first. Data will remain accessible for a minimum of 10 years after the initial release to ensure their utility for secondary analyses and replication studies.

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.

Given that all participants are consented for broad data sharing, the primary considerations will be related to protecting participant privacy through de-identification methods and adhering to HIPAA guidelines where applicable.

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 scientific data will not require additional approval beyond registration with the repository, promoting open access while still allowing for tracking of data usage.

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).

To protect participant confidentiality, all shared data will undergo rigorous de-identification procedures, including but not limited to removing direct identifiers (e.g., names, addresses) and applying statistical methods to prevent re-identification. Additionally, the project has obtained a Certificate of Confidentiality from the NIH to further safeguard sensitive information.

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 Plan will be monitored by the study's Principal Investigator in conjunction with the institution's Research Integrity Office. Regular checks (every 6 months) will ensure that data management practices align with the proposed plan, and adjustments will be made as necessary to maintain compliance with NIH policies on data sharing.