NACDA
National Archive of Computerized Data on Aging

NACDA OAR

NACDA Data Sharing Best Practices for NIA Funded Research

nacda-aging.org

icpsr-nacda@umich.edu

1. **Address the risk of respondent confidentiality**
    a. Remove direct identifiers; Public-use data must not contain direct identifiers such as name, address, email, phone number, SSN, etc.
    b. Review the data for indirect identifiers; Indirect identifiers in combination with each other could result in accidental respondent identification. Indirect identifiers include geographic locations, dates and times of interview, other date and time variables, exact income, open ended responses, etc. Some indirect identifiers may be location specific - such as school and business names. The data may need to be aggregated or recoded to eliminate identification.

Email us if you have questions about data access and protecting respondent confidentiality for your specific datasets or to learn more about restricted-use access level options.

2. **Provide complete collection metadata (data about the data)**
    a. Use a clear and consistent title across all files (the ICPSR format is as follows: description, place (for instance "[United States]"), timeframe).
    b. Use a file and folder structure with unique naming, descriptive titles; if providing multiple datasets, include dataset to documentation legend or name correspondingly
        i. File-naming should be considered with respect to long term use and have little to no changes over time (one small change can mean a large amount of effort across multiple files and web pages)
    c. Provide a description of the project; goals, main topics, etc.
    d. Provide PI full names and organizational affiliations (consistent with the Virtual International Authority File)
    e. Include data collection dates or timeframes
    f. Be explicit about the intended unit of analysis (a description of who or what is being studied)
    g. Provide information about the sample
    h. Provide information about the universe
    i. Does the project/study have a website? If so, please provide it.
    j. Include the funding source and grant number

This collection level metadata can be submitted directly within the deposit form fields or in a documentation file; the clearer the information is, the more discoverable the data will be.

3. **Provide complete variable metadata (variable and value labels) – directly in the data file(s) or by providing supporting syntax file(s)**

a. Ensure variable names DO NOT contain characters that are not acceptable across stats packages, such as periods
b. Variable names should reflect a direct correlation to survey questions, where applicable
c. Variable names in the data files should match variable references in provided documentation
d. Include complete, descriptive, and unique variable labels
e. Include complete value labels (for SAS, a formats catalog is required)
f. Designate missing or ensure missing and special missing are labeled clearly; if assigned in SAS or Stata, provide numeric codes that match the letter missing values for ease of conversion across stats packages

We accept multiple data file formats, such as SPSS, SAS, Stata, R, Excel, csv, text, DDI Codebook (these examples are mainly quantitative, although we do accept other data types – email us if you are interested in depositing other types of data).

4. **Provide documentation (codebook, questionnaire, user guide)**
    a. Provide a copy of the original questionnaire, if available
    b. Include a codebook; with unweighted frequencies is best for data comparison
    c. Identify and describe computed and derived variables
    d. Include interviewer instructions, especially if they indicate any question specific universe information
    e. Create and include a user guide; this may also be known as a "read me" in some cases
    f. A final report is not required, but may include relevant methodology
    g. Provide a list of citations of related publications, if available
    h. If available, include the question text, DDI, or another text-source for the question
    i. If available, provide variable groupings (especially for large datasets, it is useful to categorize variables into conceptual groupings)

We accept Word docs, PDFs, plain text files, Stata codebooks, Excel files just to name a few solid examples – we can accept almost any format, the main priorities are that the documentation is complete, and the text is recognizable and searchable

5. **Is your data longitudinal? If yes...**
    a. Ensure variable names are consistent across waves or follow a consistent pattern across waves so it is clear to secondary data users that those data were collected and measured at each time point (for instance, T1_income, T2_income, etc.)
    b. Provide a merged data file

c. If you cannot provide a merged file, explain why/how the data are intended to be used together. Also, highlight variables users should use to merge, or include syntax which may be used to replicate merging
d. Provide weighting information and specify weight variable name(s)
e. Be explicit about longitudinal weights; provide longitudinal weight variable names and syntax, if needed for replication
f. Identify all merging variable name(s)
g. Provide necessary syntax for merging (based on origin stats package, there is no need to translate it into multiple formats prior to depositing)
h. Provide a variable crosswalk – this could be in any file format, such as Excel, Word, PDF, etc. – the priorities here again are that the information is complete and that the text is recognizable and searchable

Keep in mind - We have a variety of data users - independent of their skills and experience, **they will all be new to your data**. By providing complete information, you are telling them how to use the weights, how sampling occurred, providing the questionnaire and any assistive materials used to facilitate the data collection so that they can replicate analysis.

More Resources:
- ICPSR Guide to Social Science Data Preparation and Archiving
- Download a checklist based on these best practices
- NIH website – Selecting a Repository