

Guidelines for Depositing NIJ and OJJDP Data at NACJD

The National Archive of Criminal Justice Data (NACJD) at the University of Michigan archives databases, codebooks, and other supporting research documentation for the National Institute of Justice (NIJ) and the Office of Juvenile Justice and Delinquency Prevention (OJJDP). NACJD's data-archiving activities are funded by NIJ and OJJDP. Research projects are not charged for our archival services. This document provides instructions for submitting quantitative, qualitative, and geospatial data for archiving at NACJD. To facilitate data reuse, it is important to **fully document and describe the variables in the data files and accompanying documentation**.

The following materials should be submitted to NACJD using our online deposit system. Go to <https://www.icpsr.umich.edu/web/pages/NACJD/index.html> and click the "Share Data" link at the top of the page. Select "NIJ Data Deposits" or "OJJDP Data Deposits". Scroll to the orange deposit form buttons and **select the button that corresponds to your funding source** to access the online deposit system and begin uploading your files.

Questions? Email the NACJD team at nacjd@icpsr.umich.edu. Please include your funding agency (NIJ or OJJDP) in the email so that we can better assist you.

Deposit Materials

Deposits will be reviewed for completeness, accuracy, and disclosure risk within approximately 20 business days. **Data archiving requirements are not fulfilled until all required materials have been submitted. Submissions must abide by the deposit guidelines outlined in this document before NIJ/OJJDP releases the awardee's institution from the archiving requirement at NACJD.** NACJD will work with awardees and assigned Grant Managers to resolve any remaining problems of confidentiality or submission completeness.

The following items are **required for all NIJ and OJJDP deposits**:

1. Data that were collected, or derived from existing data sources, under the grant award, formatted as described in the sections below, along with corresponding data collection instruments/interview protocols.
2. Data archiving plan/strategy approved by NIJ/OJJDP.
3. IRB-approved protocol or determination letter.
4. Privacy Certificate fully executed.
5. Final project report or journal article. Draft form is acceptable and will not be released with the study materials.
6. Completed Submission Checklist and Inventory. All submitted files should be listed in the inventory.
7. IRB-approved informed consent forms or instructions (if applicable).
8. Data use agreements for data obtained from external sources (if applicable).

To assist others in understanding and using the data appropriately, we strongly encourage the submission of a user guide that includes the following information where applicable:

- Descriptions of file formats and structures (i.e., rectangular, hierarchical) and information on how files are related to each other.
- Information about known data anomalies and/or caveats that need consideration when working with the data.
- Modifications made to the data to ensure confidentiality or steps taken to remove identifiers before depositing the data (e.g., use of pseudonyms, recodes, bracketing).
- Glossary of frequently used terms and acronyms.
- Instructions for merging or concatenating files, including the variables required and any syntax used for merge or amend procedures.
- Documentation on important decisions or changes in scope or methods made during the project lifecycle. Include any problems that arose during the selection and/or interview process and how they were handled.
- Guidance for selecting and using weights. Data users often struggle with making decisions about which weight to select when conducting secondary data analysis.
- Bibliographic citations for publications and presentations (current and forthcoming) resulting from the grant award.
- Contact information for a person who can address questions regarding the data or study.

Quantitative Data Deposits

The following materials are required for deposits with quantitative data. Files should be computer-readable and meet the listed specifications:

1. Data files:

- **Format:** SPSS is preferred, Stata and SAS are acceptable, with **embedded** variable and value labels, and missing values assigned. **Contact NACJD to deposit data in non-statistical software packages, including Excel.**
- **Variable labels:** Variable labels should be unique and include either the question number or descriptive information of the variable's content. Avoid duplicate variable labels.
- **Value labels/codes:** Every response should fit into one and only one category, and every category should have a value label assigned. Assign descriptive codes to each variable and use the same codes across variables recording the same type of responses.
- **Missing values:** Assign separate missing codes for refusals, non-response, not applicable and other types of missing data for all variables in submitted data files. Use numeric codes that are above the maximum valid value for the variable (e.g., "97," "98," and "99") because special missing characters are often lost when converting between software packages.
- Submit all data collected under the grant award, particularly variables used to produce analysis, tables, and descriptive information provided in the final report. Include computed, derived, and weight variables (if applicable).
- Document blanks or blank fields in variables if they represent substantive content or codes.
- Include direct identifiers and linking variables in a separate file. See the page "[Data with Confidential Content](#)" located under the "Share Data" tab on the NACJD website for more information.
- Remove character (string) variables or comment/note fields that were not a part of the data analysis. If character variables were used to create new variables, the new variables

should be distinguished from the original variables (e.g., unique variable labels). **If the data contain many character variables, please review the section for depositing qualitative data.**

2. Data collection instruments:

- **Format:** PDF is preferred; Word, Rich Text Format (RTF), or ASCII file formats are acceptable.
- Include questionnaires/interview protocols or other data collection instruments, interviewer guides, and coding instruments where applicable.
- Confirm that the labels in the data files match the data collection instruments (e.g., response categories in the data collection instruments should match the value labels in the data files.)

3. Codebook or data dictionary file:

- **Format:** PDF is preferred; Word, RTF, or ASCII files are acceptable.
- **Codebooks are required if variable and value labels are not embedded in the data files or if data are submitted in a non-statistical software package, including Excel.**
- Codebooks should not simply be data output files. Codebooks should provide details about the variables, values, and formats. Include variable labels, value labels, and/or missing value declarations.
- Include information on linking variables, instances where variables were reverse coded for analysis purposes, and known data anomalies and/or caveats for working with the data.

4. Syntax and setup files and statistical programming code:

- Syntax files containing statistical programming code written to conduct the data analysis, as well as data manipulation code written to transform/recode/create variables for analysis.
- Setup files with code assigning formats and creating weight variables (if applicable).

Qualitative Data Deposits

Qualitative data that have undergone data reduction or content analysis may not need to be archived. Where qualitative data have been coded into quantitative data, awardees are encouraged to submit only the coded quantitative data with the syntax and a description of the coding process. See the specifications listed in the section above for submitting quantitative data. For interview notes that were not transcribed, contact NACJD for guidance on submission requirements.

The following materials are required for deposits with qualitative data. Files should be computer-readable and meet the listed specifications:

1. Data files. Preferred formats include:

- Text stored in plain text (TXT), RTF, Microsoft Word, or PDFs with Optical Character Recognition (OCR), in spreadsheets such as Microsoft Excel, or in databases.
- Microsoft Word extracts from Computer Assisted Qualitative Data Analysis Software (CAQDAS), such as NVivo, ATLAS.ti, MAXQDA, and ANSWR. **We cannot accept the CAQDAS files themselves at this time.**
- Page breaks are not necessary, but headers should be included to clearly separate interviews and transcripts.

- Submit only the transcriptions from video and audio files. If no transcriptions were produced, submit the actual video and/or audio in digital format. Contact NACJD staff to discuss special handling requirements.
- Submit all data used to produce analysis, tables, and descriptive information provided in the final report.

5. Data collection instruments:

- **Format:** PDF is preferred; Word, RTF, or ASCII file formats are acceptable.
- Include questionnaires/interview protocols or other data collection instruments, interviewer guides, and coding instruments where applicable.
- Confirm that the data match the data collection instruments.

6. Data documentation:

- **Format:** PDF is preferred; Word, RTF, or ASCII file formats are acceptable.
- **Coding scheme:** Include the coding scheme or coding tree that identifies the extracted themes, ideas, and patterns used for coding the narrative data. Describe the nature of the themes and any interrelationships between them.
- **Interview “map”:** Provide a document that shows information on any header, speaker tags, and question and answer sequencing.
- **Roster (recommended):** Submit an interview roster or summary listing of all interviews that allow for a more focused review of the data. List respondents by ID number and other key characteristics including interview date, number of pages, file name, interview number, interview summary, notes, or other important information.

Geospatial Data Deposits

The following materials are required for deposits with geospatial data. Files should be computer-readable and meet the listed specifications.

1. Data files:

- **Format:** ESRI shapefile is preferred and must include the main file (.shp), the index file (.shx) and the dBASE table (.dbf). Optional, but helpful, files include the projection file (.prj) and spatial index files (.sbx, .sbn). Other acceptable formats include MapInfo (.tab, .dat, .map, .id) and Geography Markup Language (.gml, .xsd). **Contact NACJD to deposit data in formats not listed here.**
- **Missing data:** Assign unique codes for missing information for all variables in the data. Use numeric codes that are outside the range of valid values for the variable (e.g., 0,0 for unmappable locations or 0000 for street number).

2. Attributes:

- Provide information that can be linked to the spatial data. Attribute information can be tabular or textual and should describe the geographic characteristics of features such as variable name, type, general description, relationship to other features, and other related pertinent information.

3. Analysis variables:

- Submit the final, cleaned geometry file(s) used to produce analysis, figures, and tables in the final report that will permit secondary data users to recreate or extend the original analysis using the same boundaries.

4. Programming code:

- Include the project file used to create the map visualization (e.g., .map), all geographic data (layers) used in constructing maps, and all maps reproduced in project reports or summaries.

5. File list:

- Provide a record layout that lists the Shapefile name, overview, dimension, and attributes as in the example below.

Shapefile Name	Overview	Dimension	Attributes
Campus	Campus structures; buildings; parking lots; recreation areas	Polygon	CAMPUSNAME (main, south, medical) BLDG_ABBRE (abbreviation) TYPE (structure)

6. Copyright:

- Ensure that issues of proprietary visualizations/data have been addressed prior to archiving with the understanding that all archived data will be available for distribution.

Confidentiality

Direct Identifiers

Variables that directly identify research subjects must be removed from the data or the deposit will not be accepted. NACJD will only accept direct identifiers if explicitly permitted through an IRB-approved process for informed consent and submitted in a file separate from the data. Examples of direct identifiers include names, addresses, birthdates, telephone numbers, email addresses, as well as linkable numbers such as social security numbers, driver license numbers, prisoner identification numbers, medical record numbers.

Indirect Identifiers

Indirect identifiers can be used in combination with publicly available data sources or other external information to identify research subjects. Awardees should carefully consider indirect or secondary identifiers and assess their analytic importance before deposit. Examples of indirect identifiers include detailed geography (e.g., state, county, or census tract of residence), organizations to which the respondent belongs, educational institutions from which the respondent graduated (and year of graduation), occupational titles, places where the respondent grew up, detailed income, exact dates of events (such as birth, arrest, graduation), and offices or posts held by the respondent. Such information can make unique cases visible. For example, a ZIP Code variable may not be problematic alone, but when combined with race and income variables, it may allow the identification of unique individuals (e.g., the extremely wealthy).

Confidentiality and Qualitative Data

De-identification is best performed before or during the transcription process. Qualitative researchers are in the best position to find contextual information that can be used to re-identify subjects. Guidance for commonly used methods for de-identifying qualitative data is provided below.

- **Pseudonyms:** Create fictitious names for person(s), organizations, etc., in the data files. Use the same pseudonyms and place names when needed. Keep in mind that directly identifiable names may also be names of facilities, transportation systems, program names, neighborhood names, or other geographic locations. Acronyms or well-known and/or often used nicknames may also be used as direct identifiers.
- **Generalized Text:** Replace actual names with generalized text. For example, “John” can be changed to “uncle” or “Mrs. Briggs” to “teacher.” If more than one person has the same relationship to the respondent, they should be distinguished (e.g., Friend1, Friend2).
- **Remove/mask blocks of text:** If the item cannot be anonymized using pseudonyms or generalized text, the entire text should be removed and explicitly marked as such. For example, replace text with [MASKED] or (“ . . .”) as a general indicator.
- **Replace Dates:** Dates referring to specific events, especially birthdates or events involving the criminal justice system, should be replaced with some general marker for the information, e.g., “month” or “month/year”.

Guidelines for Secondary Analysis Projects using Existing Data

In instances where original data are available to the public or restrictions preclude the deposit of original data at NACJD, the data **may not** need to be deposited; however, **a description of the steps taken to acquire the data from the existing data source is required** so that another analyst will be able to replicate the process. This is also known as a “data road map” and should include the following elements:

- All items listed under Deposit Materials, above, including a user guide and bibliography of publications resulting from the project (current and forthcoming).
- A description of the source(s) of the original data (e.g., the ICPSR study number or a link to a website from which data were downloaded), a description of the original data/study, including sample size, a codebook that lists and describes the variables used in the study, and names of data files acquired (if permitted).
- Any new variables created from the original data. **Original data do not need to be deposited except where new variables were created from the original data.** New, or derived variables, should be submitted with documentation of how the variables were created. Researchers should contact their NIJ/OJJDP Grant Managers in situations where original data producers will not permit any data to be archived.
- Syntax files written to conduct the analysis, including the code to create new variables.
- Contact information and links to webpages for the original data source and/or obtaining necessary materials.