

**National Archive of Criminal Justice Data  
National Institute of Justice  
Data Submission Checklist & Secondary Data Analyst User's Guide**

---

Project Title \_\_\_\_\_

NIJ Grant Number \_\_\_\_\_

Grant Manager Name \_\_\_\_\_

Principal Investigator \_\_\_\_\_

Grant Manager Email \_\_\_\_\_

-----

Please use the following checklist to guide and assist you in submitting project data and documentation to the National Archive of Criminal Justice Data (NACJD) fulfillment of the submission requirement for National Institute of Justice sponsored research.

Project deliverables are to be submitted to NACJD via the secure, online deposit form. This checklist must be completed prior to uploading your data deposit and other materials on the NACJD website. Once completed, it will assist you in responding to fields in the online deposit form.

To submit files for archiving, go to the NACJD website (<http://www.icpsr.umich.edu/icpsrweb/content/NACJD/archiving/index.html>), click on the *NIJ Deposit Form* button and upload your files. **The checklist must be uploaded with all project deliverables.**

**Helpful Resources**

Before submitting materials, you may want to review video from a recent Webinar on archiving with NACJD:

<http://www.icpsr.umich.edu/icpsrweb/content/NACJD/archiving/index.html>

and

The Guide to Data Preparation: <http://www.icpsr.umich.edu/files/deposit/dataprep.pdf>

For additional information and details on archiving data:

<http://www.icpsr.umich.edu/icpsrweb/content/NACJD/archiving/deposit-nij-data.html>

NACJD staff can provide guidance at any point in the archiving preparation process. Send questions to [NIJdeposits@umich.edu](mailto:NIJdeposits@umich.edu).

## NACJD DATA SUBMISSION CHECKLIST

---

- 1. Documentation:** Provide all documentation needed for others to understand the archived data including:
  - Project title
  - NIJ Grant Number
  - Principal Investigator
  - Name and email address of your Grant Manager
  - Project Abstract
  - The Project report which describes the study and processing procedures including research design, study rationale, sampling and recruitment, time frame, geography, data collection procedures and dates of data collection, descriptions of weighting, imputation methods, recodes and other transformations
  - Project findings including frequency distributions and descriptive statistics for all variables used in the analysis.
  - Questionnaires or other data collection instruments
  - Codebook(s) to provide details about the variables, values, and formats including **all** applicable labels (i.e. variable, value, missing values)
  
- 2. Data:** Submit cleaned, final, and de-identified data meeting the following criteria:
  - All **direct identifiers\*** must be removed from the data set unless consent was explicitly provided for the purpose of archiving.
  - **Indirect identifiers\*\*** should be recoded to prohibit re-identification. Various methods exist for use in de-identification (such as suppression, generalization, and perturbation). NACJD staff can provide guidance at any point in the preparation process pertaining to de-identification in preparation for submitting data for archiving. See also *Treating Indirect Identifiers*, page 37 in the [Guide to Data Preparation](#).
    - Submit only data used in the analysis eliminating all unused and unnecessary content in comment fields before submission. Make certain that all variables names, value labels, and missing value codes used in the analysis of data are provided in the syntax and codebook. Include all syntax used to carry out data transformations and analyses and for purposes of replicating all results.
    - NACJD staff will perform an independent confidentiality review of datasets submitted to the archive and will work with the Principal Investigators via the Grant Manager to resolve any remaining problems of confidentiality or other issues.

**\*Direct identifiers are names, residential addresses, birthdates, telephone numbers, fax numbers, electronic mail addresses, social security numbers, medical record numbers, account numbers, certificate/license numbers, vehicle identifiers, biometric identifiers, full face photographic images, any other unique identifying number, characteristic, or code.**

**\*\*Indirect identifiers are items that could be used to re-identify study participants. Variable identifiers include items such as gender, race/ethnicity, age, zip code, detailed geography, respondent member organizations, educational institutions from which the respondent graduated (and year of graduation), exact occupation, places where the respondent grew up, exact dates of events or incidents, detailed income, and offices or posts held by the respondent.**

## NACJD DATA SUBMISSION CHECKLIST

---

**Please Note: Secondary analysis of existing data.** Data does not need to be deposited for projects involving only the analysis of existing data already publicly available where the product of the project is the analysis alone. For these projects, only the computer code or syntax that produced the analysis should be deposited. With the submission, include the source(s) of the original data (e.g. the ICPSR study number or a link to a website from which data were downloaded).

- **Analysis variables:** Submit final, cleaned data file(s) used to produce the analyses and/or tables in the project report including transformed, computed, or recoded variables constructed or derived from variables collected using the data collection instrument(s). Be sure to provide all transformation/compute/recode code or syntax along with the data files. Remove from the data file(s) all string variables that were not used for analysis.
  
- **Data Formats:** Data files may be submitted in any of the three following data formats.
  - **Quantitative:** SPSS system files with embedded variable elements are preferred but other formats are acceptable (e.g., SAS, Stata). Be sure to provide all syntax including format definitions, and describe the source of any transformed variables and method used for deriving them by depositing the recode syntax statements and/or providing a more extensive explanation in the documentation. Include all variable labels and values, missing data declarations, and other syntax/computer code used in data manipulation. Please do not submit variables comprising comments or note fields especially if not used in data analysis.  
(See: <http://www.icpsr.umich.edu/files/NACJD/pdf/quantitative-data-nij.pdf>).
  
  - **Geospatial:** ESRI Shape file(s) are preferred but other formats are acceptable including MapInfo and Geography Markup Language. All geospatial data should be aggregated to the block level or higher unless informed consent is provided. Please submit use agreements along with data acquired from government and law enforcement agencies. (See: <http://www.icpsr.umich.edu/files/NACJD/pdf/geospatial-data-nij.pdf>).
  
  - **Qualitative:** Qualitative or text narrative formatted data should be submitted when not captured in quantitative or geospatial formats. Data in this format needs to be completely de-identified unless informed consent was secured from participant sources. Preferred qualitative data formats include:
    - Text files – Plain text (.txt), Rich Text (.rtf), Microsoft Word, or machine readable PDF files with *OCR*
    - Text stored in spreadsheets – Microsoft Excel) should be converted to text format.
    - Make sure that all primary and secondary identifiers are redacted unless consent is provided for archiving.  
(See: <http://www.icpsr.umich.edu/files/NACJD/pdf/qualitative-data-nij.pdf>).

## NACJD DATA SUBMISSION CHECKLIST

---

- 3. Other Project Documents:** Submit copies of all applicable project documentation listed below
  - The project approved Data Archiving Plan
  - IRB approved protocol
  - Privacy Certificate
  - IRB approved Informed Consent forms or instructions
  - Data use agreements (for data obtained from external sources)
  - This Data Submission Checklist
  
- 4. Secondary Data Analyst's User's Guide:** The following information should be provided in a separate document (e.g. Microsoft Word) and will be used on the project data web page to provide users with important guidance.
  - Description of file format(s) and structure(s) (i.e. rectangular, hierarchical, etc.)
  - Information on if and how files are associated or related to each other (subject ID Numbers, variables representing relationships between files)
  - Instructions for merging or concatenating files (i.e. list variables required, merging syntax). Be sure to provide any syntax used for merge or amend procedures
  - Documentation on important decisions or changes in scope or methods made during the project lifecycle
  - Guidance for selecting and using weights (*if applicable*)
  - Information about known data anomalies and/or caveats that need consideration when working with the data
  - Guidance on data limitations including what questions the data can/cannot address in further analysis
  - Methods used to ensure confidentiality for direct/indirect identifiers (e.g. use of pseudonyms, recodes, bracketing).
  - Bibliographic citations for publications and presentations resulting from the grant
  - Glossary of frequently used terms and acronyms
  
- 5. The Data Archiving File Inventory:** Using the form on the next page, list the files being uploaded with brief descriptions, number of variables, case counts, etc. Please list each file and a brief description and submit this list as an inventory of the files being uploaded along with the completed checklist. Add additional pages if needed.



