Guidelines for Depositing Qualitative Data at NACJD for National Institute of Justice Sponsored Research

To facilitate secondary use, it is important to fully document and describe variables in the context of the data as well as in the documentation. This document provides guidance and a checklist to assist in preparation for archiving. Accordingly, NACJD staff review deposits for completeness and accuracy and send a report to NIJ. Any outstanding issues will need to be resolved before the submission can be finalized and assigned to a data processor. Please note the following:

Confidentiality

Direct Identifiers

Information that directly identifies research subjects must be removed prior to deposit or transfer to NACJD (unless explicitly permitted and provided for through an IRB approved process for informed consent). Examples of direct identifiers include names, addresses, and linkable numbers such as social security numbers, driver license number, prisoner identification number, etc. Data submitted with direct identifiers will be rejected and a resubmission of a repaired dataset will be required.

Indirect Identifiers

Researchers should also carefully consider indirect or secondary identifiers and assess their analytic importance. Indirect identifiers can be used in combination with publically available data sources or other external information and can lead to identification of an individual. 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, exact dates of events (such as birth, arrest, graduation), and offices or posts held by the respondent. Such information can make unique cases visible. De-identification is best performed before or during the transcription process. Guidance for some commonly used methods for de-identifying is provided below:

- **Pseudonyms**: Create fictitious names for person(s), organizations, etc., in the data files. Use the same pseudonyms and place names used in any publication. Keep in mind that names that can be direct identifiers may also be store names, names of facilities, transportation systems, program names, neighborhood names, or other geographic location. 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.” More than one person with the same relationship to the respondent can be subscripted to represent each unique individual — 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 ellipses (“…”) 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.”

Guide to Archiving

The *Guide to Social Science Data Preparation and Archiving* (http://tinyurl.com/d7ufjd6) is provided as a resource to assist you in the preparation of your data for use by your own project team and the research community. Additional guidance for addressing confidentiality can be found in Phase 5: Preparing Data for Sharing of *The Guide*. 
Deposit Checklist and Guidelines

Prior to deposit, researchers should consider whether it is necessary to archive the qualitative data. Qualitative data often undergo data reduction or content analysis whereby the information is coded into quantitative data. In those instances, depositors are encouraged to submit only the coded quantitative data. The following checklist details the items we are looking for with your submission:

- **Data** — Submit the final, sanitized data. Preferred data formats include:
  - Text files — Plain text (.txt), Rich Text (.rtf), Microsoft Word, or PDF files with OCR
  - Text stored in spreadsheets — Microsoft Excel
  - Text stored in databases — Microsoft Access
  - Computer Assisted Qualitative Data Analysis Software (CAQDAS) — NVIVO, ATLAS.ti, MaxQDA, AnSWR.
  
  Export the final coding tree and any useful memos to the raw data.

- **Analysis information** — Submit all data (i.e., interviews, text files, etc.) used to produce analysis and tables and descriptive information provided in the final report.

- **Project documents** — Submit copies of the final, approved Data Archiving Strategy, IRB approval, Privacy Certificate and Consent forms.

- **Coding scheme** — Include the coding scheme or coding tree (where applicable) that identifies the themes, ideas and patterns used in the data.

- **Record counts** — Verify that the record count in the data file(s) matches findings published in the final report.

- **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.

- **Video/audio** — Submit only the transcriptions from the video and audio. If no transcriptions were produced, submit the actual video and/or audio in digital format. Contact NACJD staff to discuss special handling requirements.

- **Documentation** — Provide all documentation needed for others to sufficiently understand the data. NACJD often distributes the original project documentation with the data. Documentation can include:
  - Project report describing the entire data collection including study design, rationale, research methods and practices
  - Data collection instruments such as interview questionnaires
  - Guidance for what questions the data can and cannot answer
  - Glossary of frequently used terms (including any terms created/defined by project staff)
  - Descriptive information about the data including the extracted themes, the nature of those themes, and possible interrelationships between the themes
  - Details on interview scheduling and instructions
  - Details on selection of interview subjects
  - Steps taken to remove identifiers from the data
  - Any problems that arose during the selection and/or interview process and how they were handled
  - User guide that provides guidance for using the data appropriately which can include information to cover the points listed above.

**About Us**

The National Archive of Criminal Justice Data (NACJD) archives the databases, codebooks, and other supporting research documentation for the National Institute of Justice Data Resources Program. A crucial part of making data user-friendly, shareable, and usable over the long term is to ensure they can be understood and interpreted by any user. This requires clear data description, annotation, contextual information, and documentation. This document provides guidance and a checklist for submitting qualitative data for archiving at NACJD. NACJD offers data deposit guides for quantitative and geospatial data as well.

**Need more information?**
Email: nacjd@icpsr.umich.edu
Phone: (800) 999-0960
Web: www.icpsr.umich.edu/NACJD