Guidelines for Depositing Geospatial 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

All fields directly identifying 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). Direct identifiers are actual addresses or X,Y coordinates that disclose the position of individuals, organizations, businesses, or location that an incident took place. Data submitted with direct identifiers will be rejected and a resubmission of a repaired dataset will be required. Data should be aggregated to a spatial level that can provide greater subject anonymity (e.g., state, county or census tract).

Deposit Checklist and Guidelines

The following checklist details the items we are looking for with a submission:

- **Data** — Submit the final, sanitized data. Preferred data formats include
  - ESRI Shapefile (essential: .shp, .shx, .dbf; optional: .prj, .sbx, .sbn)
  - Other formats are acceptable as well including
    - MapInfo (.tab, .dat, .map, .id)
    - Geography Markup Language (.gml, .xsd)

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

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

- **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.

- **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.

... checklist continued on reverse

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.
✓ Record counts — Verify that the record counts in the data (e.g., number of incidents) match findings published in the final report. Secondary data analysts often confirm that they are reading the data properly by comparing the documentation with statistics they produce from the data.

✓ File list (recommended) — Provide a record layout that lists the Shapefile name, overview, dimension, and attributes. For example:

<table>
<thead>
<tr>
<th>Shapefile Name</th>
<th>Overview</th>
<th>Dimension</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campus</td>
<td>Campus structures; buildings; parking lots; recreation areas</td>
<td>Polygon</td>
<td>CAMPUSNAME (main, south, medical), BLDG_ABBRE (abbreviation), TYPE (structure)</td>
</tr>
</tbody>
</table>

✓ Missing data — Assign unique code(s) for missing information (e.g., incidents) 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).

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

✓ 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 that describes the data collection as a whole including theoretical or conceptual framework that guided the study, research questions addressed, hypotheses tested, and methods used
- Metadata that captures basic descriptive characteristics about the data including the quality, purpose, spatial extent, and history of the data
- Codebook(s) with Shapefile name and description(s), attributes, variable descriptions, and detailed information describing the files and layers within the data
- Information about known data anomalies and/or caveats for working with the data
- Guidance for what questions the data can and cannot answer
- Glossary of frequently used terms (including any terms created/defined by project staff)
- User guide that provides guidance for using the data well, 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 geospatial data for archiving at NACJD. NACJD offers data deposit guides for quantitative and qualitative data as well.

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