U.S. Department of Justice
Office of Justice Programs
National Institute of Justice

NATIONAL INSTITUTE OF JUSTICE
Issues and Practices

Depositing Data With the Data Resources Program of the National Institute of Justice:

A Handbook

About the National Institute of Justice

The National Institute of Justice is the research and development agency of the U.S. Department of Justice established to improve the criminal justice system and to prevent and reduce crime.

Specific mandates established by Congress in the Omnibus Crime Control and Safe Streets Act of 1968, as amended, and the Anti-Drug Abuse Act of 1988 (Public Law 100-690) direct the National Institute of Justice to:

• Sponsor special projects and research and development programs that will improve and strengthen the criminal justice system and reduce or prevent crime;

• Conduct national demonstration projects that employ innovative or promising approaches for improving criminal justice;

• Develop new technologies to fight crime and improve criminal justice;

• Evaluate the effectiveness of criminal justice programs, identify programs that promise to be successful if continued or repeated, and recommend actions that can be taken by Federal, State, and local governments, and private organizations and individuals to improve criminal justice;

• Develop new methods for the prevention and reduction of crime and delinquency, and test and demonstrate new and improved approaches to strengthen the justice system;

• Provide to the Nation's justice agencies information from research, demonstrations, evaluations, and special projects;

• Serve as a national and international clearinghouse of justice information for Federal, State, and local government; and

• Deliver training and technical assistance to justice officials about new information and innovations developed as a result of Institute programs.

The Director of the Institute is appointed by the President and confirmed by the Senate. The Director establishes the objectives of the Institute, guided by the priorities of the Department of Justice and the needs of the criminal justice field. The Institute actively solicits the views of criminal justice professionals to identify the most critical problems confronting them and to develop projects that can help resolve them. Through research and development, the National Institute of Justice will search for answers to what works and why in the Nation's war on drugs and crime.

134178

U.S. Department of Justice
Office of Justice Programs
National Institute of Justice

Depositing Data With the Data Resources Program of the National Institute of Justice:

A Handbook

Prepared by

James L. Peterson, Ph.D.
Ugo A. Colella

Sociometrics Corporation
170 State Street, Suite 260
Los Altos, CA 94022
415-949-3282

October 1991

Foreword

Secondary data analysis is a vital aspect of contemporary public policy research. Original data collection is expensive and time consuming. Yet rarely can the original investigators explore all the policy questions their data permit. Re-examination of these data resources by other researchers permits a relatively economical and scientifically productive exploration of important policy issues. It brings the analytical talents of a larger number of researchers to bear on questions of concern to criminal justice professionals. Secondary data analysis not only permits original findings to be validated by independent investigators, but also allows alternative explanations and approaches to be tested quickly at a fraction of the original data collection costs.

It is for these reasons that the National Institute of Justice has led the way among Federal agencies by establishing the Data Resources Program. This program promotes secondary data analysis by requiring that original data gathered under grants from the Institute be documented and delivered to the Institute at the time the projects are completed. Such data sets are then archived and made publicly available.

This handbook for investigators is written for National Institute of Justice grantees and contractors who are collecting and preparing data for deposit with the Institute at the conclusion of their projects. The handbook has been created to promote efficiency and convenience for the original investigators depositing the data in the Data Resources Program. The handbook advises investigators about the nature, content, and format of the data and documentation to be deposited in the Data Resources Program. In so doing, we hope to avoid incomplete documentation, unnecessary data preparation steps by the original investigators, or unnecessary delays in preparing the data for public distribution.

Charles B. DeWitt
Director, National Institute of Justice

I. The Data Resources Program of the National Institute of Justice

In 1984, the National Institute of Justice (NIJ) established the Data Resources Program (DRP). This program is designed to assure that data collected with NIJ funding are available to other researchers. The archiving of data is a cooperative effort between the original investigators and the DRP staff. Machine-readable data and documentation are provided by the original investigators. Staff of the DRP examine, edit, and archive the machine-readable data and documentation. In this process, staff identify technical difficulties with machine-readable data or the documentation and, if problems are discovered, work with the original investigators to clarify discrepancies. A well-documented data product that will be easily accessible to future users is the goal of this effort. Currently, the processing and archiving of data are performed under contract to NIJ.

All archived data sets and their documentation are deposited with the National Archive of Criminal Justice Data, housed at the Inter-University Consortium for Political and Social Research at the University of Michigan. NIJ periodically publishes *Data Resources of the National Institute of Justice*, a resource directory listing the files and a brief description of each in the archive. Copies of this directory may be obtained by contacting:

National Criminal Justice Reference Service
Box 6000
Rockville, MD 20850
Telephone: 800–851–3420
(from Maryland or Metropolitan Washington, D.C.): 301–251–5500

or

Pamela K. Lattimore, Ph.D.
Manager, Data Resources Program
National Institute of Justice
633 Indiana Avenue NW.
Washington, DC 20531
Telephone: 202–307–0144

II. Preparing Data for Archiving: Principal Investigators' Responsibilities

Principal investigators are responsible for:

(1) Submitting data and supporting materials.

(2) Responding to requests by DRP staff for additional or clarifying information.

(3) Reviewing and correcting draft materials prepared by the DRP staff.

Submitting Materials

Principal investigators must submit the following materials to their NIJ grant or contract monitor at the conclusion of their grant or contract period:

(1) Data file(s).

(2) Data collection instrument(s).

(3) Codebook or data dictionary.

(4) Final project report, project summary, or other description of the project.

(5) Bibliography of publications pertaining to the data.

(6) Printout of the first and last data records.

If available, summary statistics (frequency distributions, means, etc.) of all variables should also be submitted.

Data files must be submitted in machine-readable format. A variety of transmittal media and formats are acceptable. The most common and currently accepted media and formats are listed below. To determine if others are acceptable, please call the NIJ Data Resources Program manager. (The name and telephone number are listed on page 1.)

The following media are currently acceptable:

Mainframe tape (reel)
1600 or 6250 bytes per inch
EBCDIC (preferred) or ASCII
9-track (preferred) or 7-track

Microcomputer
5 1/2-inch floppy disk (high or low density)
3 1/4-inch disk (high or low density)
1/4-inch tape cartridge
8-inch bernoulli cartridge

The investigator should indicate on the tape, disk, or cartridge what operating system (VMS, DOS, UNIX, Apple/Mac, etc.) was used to write the data file.

Data files can be submitted as raw data or as system files. Raw data must be accompanied by a listing of the record and byte positions of variables on the data file. SPSS and SAS system files should be created as transportable files by using the SPSS EXPORT command or SAS PROC XCOPY with export option. (Mainframe and personal-computer versions of these programs create portable files that are equivalent; either may be used.) Using portable files preserves the variable names and variable and value labels that the investigator has already created. Common microcomputer spreadsheet files, such as those generated by dBASE, Lotus, Framework, etc., are also acceptable.

Investigators collecting qualitative information through interviews or case studies also may be required to submit these data for archiving. Text files generated by commonly used word processing formats should be transmitted as unformatted ASCII (DOS) files. Investigators collecting qualitative information are encouraged to confer early in their grant/contract period with their NIJ monitor or the NIJ DRP manager to determine responsibilities to submit data.

When preparing the data for submittal, the following points should be kept in mind:

(1) Information that could identify an individual (e.g., name, address, or Social Security number) should not be included on the data file. If identification variables are needed to link multiple files, these should be created variables and not names or numbers that could be traced to individuals.

(2) If the data set consists of two or more related files, variables that link the files should be included on each file. The nature of the relationship between files and the variables to be used to link them should be described clearly in accompanying documentation.

(3) In general, rectangular files with rows corresponding to cases and columns corresponding to variables are preferred over hierarchical files. Complex hierarchical files should, if at all possible, be converted to a series of linkable rectangular files.

(4) Each variable should have a set of exhaustive, mutually exclusive codes. These codes should be thoroughly documented in the codebook.

(5) Data should be reduced to numeric codes whenever possible. If alphabetic codes are used, they should be identified clearly in the codebook as alphabetic so that users of the data will know to use alphabetic formats to read such data.

(6) Missing data should be coded differently than inapplicable data. As a rule, blanks should not be used for coding variables since statistical programs generally do not provide for labeling of blank values.

(7) Variables created by transforming other variables may be included. The transformations and the transformed variables' values should be thoroughly documented. Variables that the investigator has determined are not of sufficient quality to be used in any analysis should be removed from the file before submittal.

(8) If weights are developed that adjust the data, they should be included in the data file. Their locations in the file and the appropriate ways to use the weights should be documented in the codebook.

(9) Investigators are encouraged to check for out-of-range codes that are inconsistent with skip patterns and codes that are internally inconsistent. Errors identified through these checks should be corrected on the data file.

Data collection instruments and instructions to interviewers, if appropriate, should be submitted. One clean, unused copy of each instrument, including interview schedules, self-administered questionnaires, data collection forms for transcribing information from records, paper tests and scales, screening forms, and call-report forms, should be submitted. If different forms of the instrument were used, each form should be included along with a description of the circumstances in which each was used (study populations, time periods, etc.).

Codebooks may be submitted either in machine-readable or paper format. For example, a machine-readable version of an SPSS dictionary file is acceptable, providing the dictionary describes the variables and their values completely. The codebook should specify the data position for each variable, describe the contents of each variable, and identify the range of possible codes and their meanings for each variable.

The **project description** should include the theoretical or conceptual framework that informs the study, research questions addressed by the study, any specific hypotheses to be tested, and methods used to collect the data. The final report prepared for NIJ will generally meet this requirement. The description of methods should address the following points that are applicable:

(1) Unit(s) of analysis.

(2) Universe from which the study population is drawn.

(3) Sampling method (e.g., probability, purposive, convenience) used to select elements of the universe.

(4) Use of strata or quota (if any).

(5) Time period and geographic area covered by the data collection.

(6) Source(s) of data.

(7) Response rate (the proportion of the sample about which data were obtained).

(8) Retention rate (in a multiwave study, the proportion of cases successfully followed up).

(9) Other elements of the study design (for example, experimental treatments, assignment to groups, timing of followup data collection).

A **bibliography** of research reports, journal articles, and other publications pertaining to the data set should be provided. This material is included in the *User's Guide.*

Summary statistics, if generated by the investigator, should be submitted in either machine-readable or paper form. This information is useful to the archivist as a standard for comparing data. The form of the statistics should be determined by the characteristics of the data; namely, frequencies for categorical variables, means, and standard deviations for continuous variables.

Responding to Requests for Additional Information

Archivists at the DRP will contact principal investigators during the archiving process to obtain necessary information not available in the data and documents submitted.

Reviewing Draft Materials

User's guides and codebooks drafted by the DRP staff will be submitted to the principal investigator for review. This provides investigators an opportunity to clarify, amplify, or correct any information before it is released to the public. Generally, these materials will be sent to investigators within 6 months of the original receipt of the data by NIJ. Investigators are asked to complete their reviews within 3 weeks of receipt.

III. Preparing Data for Archiving: DRP's Responsibilities

DRP staff are responsible for:

(1) Preparing the data files in a format suitable for transmittal to the National Archive of Criminal Justice Data.

(2) Preparing a user's guide that describes the project and data.

(3) Reviewing the codebook for completeness and accuracy and augmenting the codebook as necessary.

(4) Making copies of data sets archived since October 1989 available (for a small fee) to the research community and providing technical support for these data.

Preparing Data Files

Machine-readable data files are inspected for quality, usability, and correspondence with documentation. Checks are made for out-of-range values, missing values, and data inconsistencies. Data files may be restructured to promote usability. However, no changes will be made in the data values unless authorized by the original investigator.

Regardless of the format in which the data are obtained, the DRP staff will produce a raw ASCII data file for transmittal to the National Archive of Criminal Justice Data. In cases where system files have been submitted, a portable version of the system file may also be archived and disseminated by the DRP.

While the DRP staff are preparing the data set for public use, they may contact the investigator directly if more information, documentation, or clarification is needed.

Preparing a *User's Guide*

A user's guide is prepared as part of the archiving process. The guide provides the potential user a general overview of the study to help determine the suitability of the data set for the potential user's purpose. The guide, associated instruments, and codebooks are intended to provide enough information about the study to enable the user to work with the data without recourse to the original investigator. The table of contents for a typical guide is as follows:

TABLE OF CONTENTS

Reviewing the Codebook

The codebook is reviewed for completeness and accuracy. If necessary, the codebook is reorganized or augmented to ensure it links the data collection instruments to the data file; matches the organization of the data file; and provides byte positions, variable definitions, codes, and code definitions for each variable.

Making Data Sets Available

Researchers can obtain copies of data sets from the National Archive of Criminal Justice Data. Researchers who work at institutions that are members of the

Consortium can obtain the data at no cost through their institutional representative. Others can order the data directly from the National Archive. Information about prices and formats for the data can be obtained by contacting:

Victoria W. Schneider, ICPSR
Institute for Social Research
P.O. Box 1248
Ann Arbor, MI 48106
Telephone: 313–763–5010 or 800–999–0960

Data files archived since October 1989 are also available for a small fee directly from the Data Resources Program. These data sets are available in several media, including mainframe tape and low- or high-density 31/2- or 51/4-inch diskettes. Each data set and its documentation will be sent by first class mail within 3 days of an order. A catalog can be obtained by contacting:

Pamela K. Lattimore, Ph.D.
Manager, Data Resources Program
National Institute of Justice
633 Indiana Avenue NW.
Washington, DC 20531
202–307–0144