The Realities of Data Sharing

Research Standards in Criminology seminar series

Margaret Levenstein
August 20, 2020
Agenda

Introduction to ICPSR
NACJD, the National Archive of Criminal Justice Data
Resources for journals
Restricted data
  Access modalities
  Curation
Other resources to improve criminal justice research
  Bibliography
  Data Management Plans
  Consent Statement Guidance
Founded in 1962 by 22 universities, now consortium of ~800 institutions world-wide

Focus on social and behavioral science data, broadly defined

Current holdings
- 15,000 studies, quarter million files
- 1700 are restricted studies, almost always to protect confidentiality
- Bibliography of Data-related Literature with 92,000 citations

Data Stewardship: high quality, curated data in user-friendly formats

Data Protection: multiple strategies for safe research using sensitive data
Thematic Collections

Community
- Nat’l Archive of Data on Arts and Culture
- Open Data Flint

Criminal Justice
- Nat’l Archive of Criminal Justice Data
- Firearm Safety Among Children and Teens

Demography
- Nat’l Archive of Computerized Data on Aging
- Data Sharing for Demographic Research
- Resource Center for Minority Data
- US Census Data Repository
- ResearchDataGov

Education
- Child and Family Data Archive
- Civic Learning, Engagement and Action Data Sharing
- Education Research Data Sharing Initiative
- Measures of Effective Teaching
- NCAA Student-Athlete Experiences

Health/Wellness
- Archive of Data on Disability to Enable Policy
- COVID-19 Data Repository
- Health and Medical Care Archive
- Nat’l Addiction and HIV Data Archive Program

At-Risk Government Data Data Lumos
The National Archive of Criminal Justice Data
About NACJD

**Mission**
To support research on crime and justice by facilitating the discovery and use of data

**Audience**
Criminal justice students and researchers, news organizations, other non-academic research institutions

**Funding**
Bureau of Justice Statistics (BJS)
National Institute of Justice (NIJ)
Office of Juvenile Justice and Delinquency Prevention (OJJDP)
NACJD by the Numbers

Over 10,322 research datasets available for analysis

From January 1, 2020 - June 30, 2020:

NACJD released 22 NIJ, 10 BJS, and 5 OJJDP studies
16,944 unique NACJD data users downloaded 252,640 data or document files
About Our Data

- Official U.S. crime data
- Grantee data collections
  - NIJ and OJJDP mandate to share data with NACJD
- Topical areas:
  - Policing
  - Corrections
  - Court case processing
  - Juvenile delinquency
  - Guns and other weapons
  - Recidivism
  - Violence against women
  - Human trafficking and sex crimes
  - Drugs and alcohol
  - And more
Popular Data Collections

- National Crime Victimization Survey (NCVS)
- Uniform Crime Reporting Program (UCR) and National Incident-Based Reporting System (NIBRS)
- Federal Justice Statistics Program (FJSP)
- National Corrections Reporting Program (NCRP)
- National Intimate Partner and Sexual Violence Survey (NISVS)
- Project on Human Development in Chicago Neighborhoods (PHDCN)
- Pittsburgh Youth Study (PYS)
Accessing NACJD Data

- NACJD data are available without charge
- Many NACJD studies (over 1,100) are restricted access
- Restricted access is offered via secure download, Virtual Data Enclave (VDE), Physical Data Enclave (PDE)
- Prospective data users can navigate to the application by visiting the study home page of the data they are interested in
Research Resources

Learning Guides
• Replicating Results from the National Crime Victimization Survey
• Law Enforcement Agency Identifiers Crosswalk

Resource Guides
• Available for several popular data collections, including UCR, NIBRS, NCRP, NCVS, and FJSP
• Includes information needed to properly use and understand data

(Data) sharing is caring

- Reproducibility
- Respect for study participants
- Minimize burden
- Increase impact
- Incremental knowledge building
- Trust and credibility
Sharing data

The 7 Habits of Data Deposit
Resources for Journals
Any researcher can deposit data and receive DOI immediately

- Available to share with journal prior to publication
- Data are not curated, but they are checked for confidentiality
- Data can be restricted or embargoed

Dedicated journal repositories

- Journal-branded
- Journal-specific criteria
- Work flow allows journal to approve deposit prior to data publication
Welcome to JEDI

The "Journal Editors Discussion Interface" (JEDI) mailing list offers the editors and editorial staff of social science journals, and personnel from the data management community, a shared forum in which to ask and answer questions, pool information and expertise, and build a fund of collective knowledge.

Join JEDI  Post to JEDI

Data-PASS JEDI

Bringing together journal editors and data professionals

https://dpjedi.org/
Depositing data for curation: Data for re-use
Data quality matters
I can’t share. My data’s private

Why is the data private?

- Privately owned by third party, governed by NDA and/or RDUA
- Researcher promised privacy to respondent/participant and/or IRB
- Data cover sensitive topic

Solutions

- Consent statements and RDUAs that permit sharing for reproducibility and research
- Share code, process for obtaining access
- Mask the data
- Restrict access
  - IRBs will often allow data sharing with restricted access, even when that was not explicitly permitted in research protocol
Recommended Informed Consent Language for Data Sharing

Language to Avoid

Promises in the informed consent can appear to limit an investigator’s ability to share data with the research community. In reality, investigators can inform study participants that they are scientists with an obligation to protect confidentiality and still share the study data with the broad scientific community. Many effective means exist to create public-use data files or share restricted-use data files under controlled conditions. That is, data can be modified to reduce the risk of disclosure or shared with additional safeguards while preserving their value for science.

Model Language

Here are two model statements investigators may use in informed consents to describe protection of confidentiality that also allows data sharing:

Sample 1. Study staff will protect your personal information closely so no one will be able to connect your responses and any other information that identifies you. Federal or state laws may require us to show information to university or government officials (or sponsors), who are responsible for monitoring the safety of this study. Directly identifying information (e.g., names, addresses) will be safeguarded and maintained under controlled conditions. You will not be identified in any publication from this study.

Sample 2. The information in this study will be used only for research purposes and in ways that will not reveal who you are. Federal or state laws may require us to show information to university or government officials (or sponsors) who are responsible for monitoring the safety of this study. You will not be identified in any publication from this study.

Known Concerns and Recommended Alternatives

https://www.icpsr.umich.edu/icpsrweb/content/data-management/confidentiality/conf-language.html
Temptation is to promise that no one else will see the data
- Or even that the data will be destroyed
  - This is the direction GDPR has taken

Promise instead to create the most scientific impact while protecting confidentiality
- Separate and encrypt Personally Identifiable Information (PII)
- Restrict use to scientific and evidence-building purposes
- Never reveal information about individual or share with those who try to use to re-identify individuals
Safe data

Curation includes steps to create safe data
  ➢ Consistent with fitness for use

Criteria for evaluating data safety
  ➢ Living persons
  ➢ Vulnerable populations
  ➢ Expectations of privacy
  ➢ Data type and level
  ➢ Unit of analysis
  ➢ Sampling
  ➢ Longitudinal
  ➢ Availability
  ➢ Social relationships
  ➢ Geography
  ➢ Date specificity
  ➢ Sensitivity
  ➢ Small or distinct populations

Methods for “anonymization”
  ➢ Aggregation
  ➢ Suppression
  ➢ Swapping
  ➢ Perturbation/noise infusion
Safe places

Secure online analysis

- Expensive to setup, but easy to access
- Requires confidence in automatic disclosure checks

Encrypted downloads

- Secure, local computing environments
- Training
- Restricted data use agreements
- Researchers do own disclosure review
Safe places

Virtual Data Enclave
- Secure, controlled computing environment
- Training
- Restricted data use agreements
- Third party disclosure review

Physical data enclave
- Secure, controlled physical and computing environment
- Training
- Restricted data use agreements
- Third party disclosure review
Sharing data safely

Is possible
Multiple solutions
Increases impact of research
Increases credibility of research
New resource!

A special place for students to deposit their data for use by other students!
Data Jeff is happy to take your questions