



Best Practices for Protecting and Sharing Data

Margaret C. Levenstein

Director, Inter-university Consortium for Political and Social Research

Promoting transparency and confidentiality

- What is ICPSR?
- Preserving and accessing shared data and data-related content
- Journal repositories and journal-related deposits
- LinkageLibrary
- Researcher passport

ICPSR



- Founded in 1962 by 22 universities, now consortium of over 760 institutions world-wide
- Focus on social and behavioral science data, broadly defined
- Current holdings
 - 10,000 studies, quarter million files
 - 1500 are *restricted studies*, almost always to protect confidentiality
 - Bibliography of Data-related Literature with 70,000 citations
- Approximately 60,000 active MyData (“shopping cart”) accounts
- Thematic collections of data about addiction and HIV, aging, arts and culture, child care and early education, criminal justice, demography, health and medical care, and minorities

Preserving and accessing shared data and data-related content

- Make data sharing feasible
 - ICPSR's General Archive
 - Anyone can deposit
 - Curated and preserved
 - Guidance over data life cycle
 - Templates for consent, IRB, DMP consistent with transparent and reproducible access
- Incentivize data sharing
 - Standard citation
 - Bibliography
 - Usage statistics

Publication-related sharing

- Journal repositories for article-specific materials
 - Data extracts (exactly as used in article)
 - Maintaining provenance: code to reproduce data extract & point to original data source
 - Code for reproducing analytical results
 - Integrated work flow with publication
- Data-specific collections
 - Extracts and re-codes used in publications
- Moving toward capsules
 - Jupyter notebooks with code, data, and results
 - Articles are the advertisement
 - Code and data are the research
 - Credit: Victoria Stodden, University of Illinois

ICPSR LinkageLibrary

- Linked data present challenges for both confidentiality and reproducibility
 - Linkage more accurate with more detailed information
 - Need standards for safe, ethical ways to enhance data with new linkages
 - Linked data easier to re-identify, even after removing unique identifiers
 - Need safe places to analyze linked data
 - Linkage strategies introduce differences in datasets that are often not well documented

ICPSR LinkageLibrary

➤ LinkageLibrary

- Encourage researchers to share linked (or linkable) data, and linkage strategies

 - Algorithms, code

- Compare approaches across projects, datasets, disciplines

 - Improve linkage practices

 - Improve transparency

➤ Coming this fall

- Looking for beta testers now!

Protecting Confidential Data

Safe data: Modify data to reduce risk of re-identification

Safe places: Physical isolation and secure technologies

Safe people: Training and data use agreements

Safe outputs: Results reviewed before being released

Safe data and safe places are complements, not substitutes

The safer the data, the more we need safe places

Safe places

- Require technological *and* social solutions
- Transparency via restricted computing environment protects confidentiality
 - Intentionally limits access
 - Need to make process data simpler, more efficient, limiting as much as necessary but not more

Alternative safe places

- On-line analysis with automatic disclosure control
 - Researcher cannot see underlying data
 - Limited use in teaching, as we want students to look at data!
 - Limited use for non-designed data with minimal metadata
- Secure download to local computing environment
 - With data use agreements
 - Requires institutional credibility and enforcement
- Virtual data enclaves
- Physical enclaves

Safe people

- The Researcher Passport: Improving Data Access and Confidentiality Protection
 - ICPSR's Strategy for a Community-normed System of Digital Identities of Access
 - <https://deepblue.lib.umich.edu/handle/2027.42/143808>
 - Identifies inconsistent language and policies that impede access
- Passports for safe people
 - Verified identities, institutional affiliation
 - Training
 - Experience (good and bad)
- Visas to control access
 - Permission to "enter" (access) specific data specifying
 - Passport holder
 - Project, Place, Period

Researcher passport

- Implementation plan
 - Issue passports as digital identifiers associated with ICPSR MyData Accounts
 - Other passport issuers?
 - Blockchain
 - Issue visas to access restricted data
 - At ICPSR, visa controls access to secure download or VDE
 - Visa issued after signed DUA
 - Establishing content of training requirements
 - Determining which existing training include that content
 - Governance system to provide due process for allegations of misconduct

Questions

- Can reach ICPSR at help@icpsr.umich.edu
- Questions or interest in Researcher Credentialing to me at MaggieL@umich.edu