

The Inter-university Consortium for Political and Social Research in Research: Data Collection, Secondary Data, and Data Sharing

University of Michigan, Ann Arbor, MI USA Amber Bryant, PhD



Who or What is ICPSR?

www.icpsr.umich.edu

 One of the world's oldest and largest data archives for the social and behavioral sciences

- Founded in 1962
- 18,000 "studies" (one or more dataset)
- 6,120,000 variables
- 106,100 data-related publications
- 815 member institutions
- Curated collection and <u>OpenICPSR.org</u>
- Summer Program in Quantitative Methods of Social Research



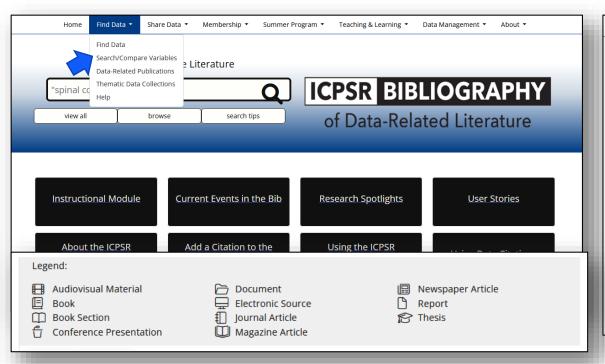


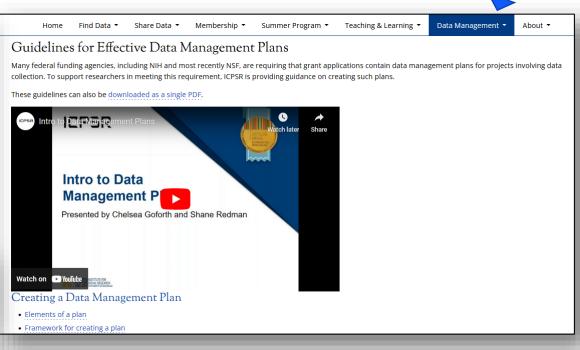
Data Collection Resources

Getting Funding

 Identify gaps in literature and/or view data used to study a given topic via <u>ICPSR Bibliography of Data-Related Publications</u>

Create <u>data management plan</u>







Resources for Data Collection

Recommended Informed Consent Language for Data Sharing

Additional Resources

ICPSR's Approach to Confidentiality

American Statistical Association, Data Access and Personal Privacy: Appropriate Methods of Disclosure Control

The American Statistical Association, Committee on Privacy and Confidentiality - Methods for Reducing Disclosure Risks When Sharing Data

Statistical Policy Working Paper 22 -Report on Statistical Disclosure Limitation Methodology

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.

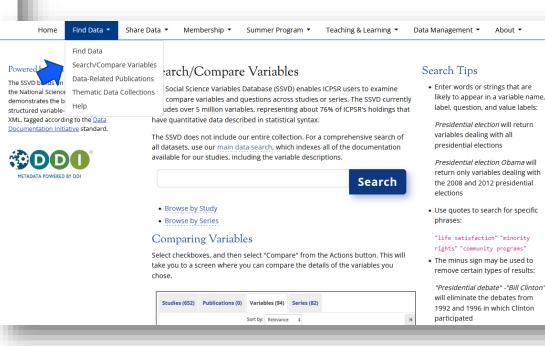
Sample 2. The information in this study will be used only for research purposes and in ways that will not reveal who you are. Federal or state laws may require us to show information to university or government officials (or sponsors) who are responsible for monitoring the safety of this study. You will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

Known Concerns and Recommended Alternatives

Concern 1

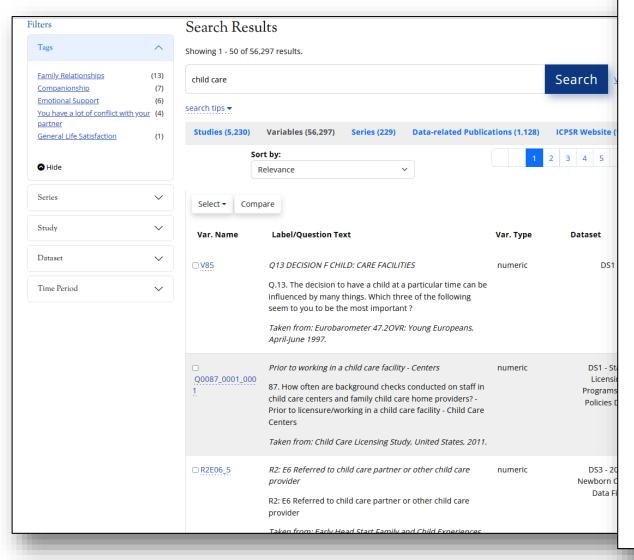
Terms such as "anonymous" and "de-identified" are undefined and left open to interpretation. Some data are collected anonymously as directly identifying information is never obtained. De-identification may involve more than

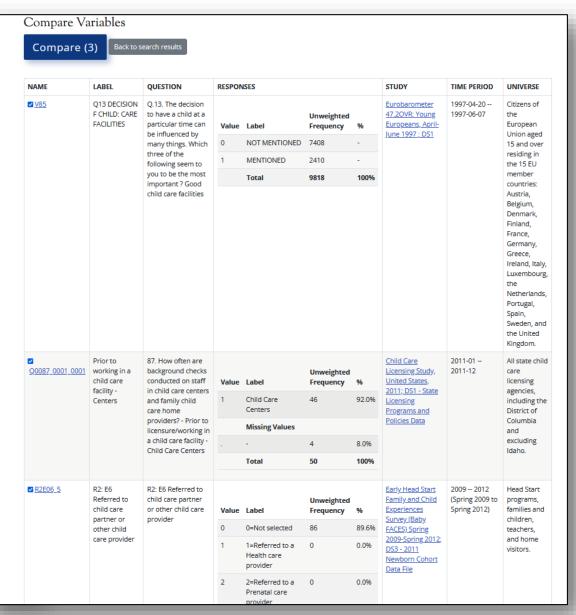
- Example <u>informed consent</u> <u>language</u> that allows sharing
- <u>Social Science Variables Database</u> (question bank)





Question Bank (SSVD)







Data Handling

- "Best practices" guide
 - What information to keep
 - How to structure the file, name variables, etc.
- Also includes information for data management plan
- Applicable even if not depositing with ICPSR

Home Find Data ▼ Share Data ▼ Membership ▼ Summer Program ▼ Teaching & Learning ▼ Data Management ▼ About ▼

Guide to Sockar Science Data Preparation and Archiving
Best Practice Throughout the Data Life Cycle: 6th Edition

Inside the Guide

About ICPSR

- Why Should I Archive Data?
- How Do I Deposit Data?
- · What Services Does ICPSR Offer?
- · Using the Guide

Importance of Data Sharing and Archiving

- Planning Ahead for Archiving and Preservation of Data
- · The Data Life Cycle

Proposal Development and Data Management Plans

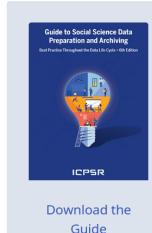
- · Recommended Elements for Data Management Plans
- Other Considerations
- Summary Table. Elements of a Data Management Plan
- An Example Data Management Plan for Depositing Data with ICPSR

Project Start-Up

Importance of Good Data Management

Data Collection and File Creation

- · Best Practice in Creating Research Data
- Quantitative Data





Data Sharing and Acquisitions

This is how data centers contribute to education research in Pursuit of Truth.



Ideas about sharing

Fears/Barriers:

- Takes time/effort researcher does not have
- Worried about protecting research participants
- IRB won't allow sharing
- People might misuse the data
- Fear of being scooped
- Too much work, too little reward



Showing Impact

National Social Life, Health, and Aging Project (NSHAP): Round 3 and States], 2015-2016, 2020-2021 (ICPSR 36873)

Version Date: Nov 17, 2022 ② Cite this study | Share this page ▼

Principal Investigator(s): 2

Linda J. Waite, University of Chicago. Department of Sociology; Kathleen A. Cagney, University of Michigan. Department of Lorent Cancer and Aging. Department of Supportive Care Me Department of Academic Research Centers; Elbert S. Huang, University of Chicago. Department of Medicine; Diane Public Health Sciences; Edward O. Laumann, University of Chicago. Department of Sociology; Martha K. McClintock Department of Comparative Human Development; Colm A. O'Muircheartaigh, University of Chicago. Irving B. Harri Schumm, University of Chicago. Department of Health Studies

Series:

• National Social Life, Health, and Aging Project (NSHAP)

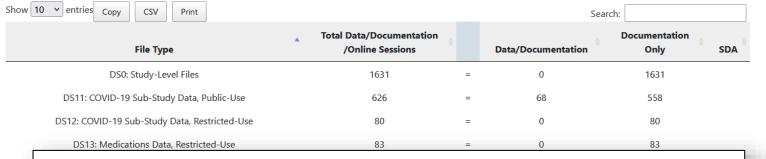
https://doi.org/10.3886/ICPSR36873.v7

Version V7 (see more versions)

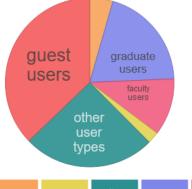


Shov





Unique Users Total Datasets/Files Downloads by Users Total: 1506







Restricted Data at ICPSR

- All studies receive disclosure risk
- Data may be restricted due to increased potential for reidentification, sensitive topics, or both
- Requires an application process
 - Description of proposed project, why restricted data are needed
 - Data security plan
 - Institutional Restricted Data Use Agreement (RDUA)
- Dissemination via secure download, virtual data enclave, or physical data enclave



Data for Secondary Analysis (ICPSR's Bread & Butter)

Why Use Secondary Data?

- Primary data collection is resource intensive
- Good chance someone has already collected data
- Existing data often provide larger, more representative samples, oversampling of small subpopulations, more questions
- Good stewardship of respondents' participation, funders' money
- Get to the fun part of the work sooner!



What Can I Get at ICPSR?

- Social and behavioral sciences, broadly defined
 - Data about individuals (early childhood through elderly), institutions, and even horses
 - Time periods from ancient history (<u>1194-1294</u>) to the present
 - Many disciplines covered
- Survey and other raw data types
- Documentation to help understand the data
- Data easily downloadable into Excel/statistical package
- Publications based on the data
- Public and restricted-use data



Popular Data

- National Longitudinal Study of Adolescent to Adult Health (Add Health)
- India Human Development Survey (IHDS)
- Midlife in the United States (MIDUS)
- Migrant and Seasonal Head Start Study
- Survey of Prison Inmates
- Monitoring the Future: A Continuing Study of American Youth
- National Survey on Drug Use and Health (NSDUH)
- National Incident-Based Reporting System (NIBRS)
- American National Election Study (ANES)
- Law Enforcement Management and Administrative Statistics (LEMAS)
- National Social Life, Health, and Aging Project (NSHAP)



What You Typically Won't Find

- Quick answers to "Google-type" questions
- Summary charts/tables of data
- Identifiable information about individuals or organizations
- Pure genetics, physics, engineering, or other natural science data
- Data to use in GIS or other mapping packages
- Data types we are incorporating:
 - Images
 - Qualitative data (e.g., unstructured interviews, diaries, videos, other text-based data)
 - Social media data



Data vs. Statistics

Statistics

- Numbers (e.g., percentages) you find in written documents or graphics and tables that accompany the reports
- Tell you what is in the data by summarizing patterns and doing calculations
- Are easily "digestible" and tell a story

Data

- The underlying information on which those summaries are based
- Typically come from surveys, administrative records, transactions, or other sources
- Requires some work to use



16 02004184136563568259

16 02004184136563568259

201

2020.1 1602004184136563568259127 160200418413656356825912701 2020.1 1602004184136563568259127 160200418413656356825912702

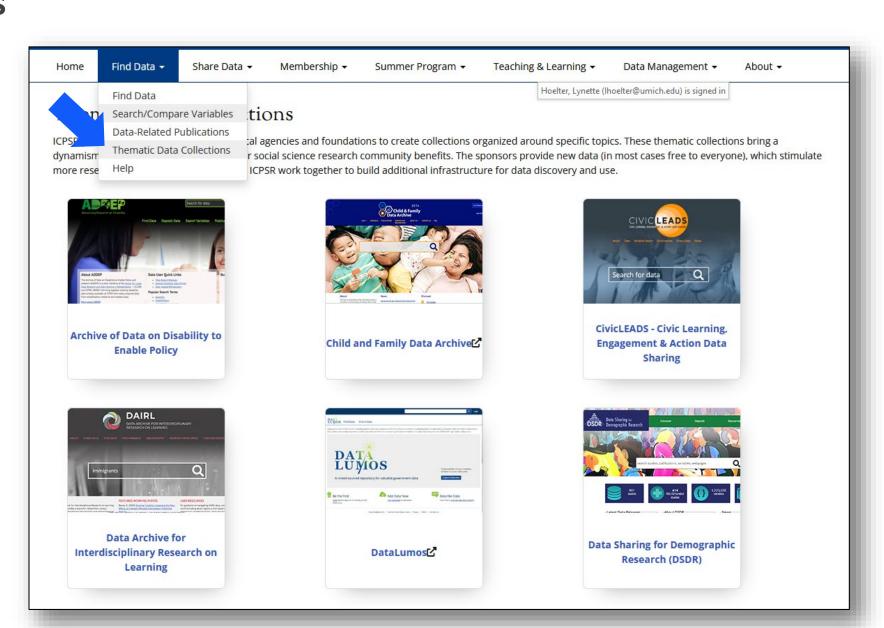


ICPSR Does Data So YOU Can Do Statistics!



Thematic Collections

- Best if you have interest in a broad topic
- Searches within a collection only return results from that collection (may want to back out to ICPSR search)





Collections by Topic

Children/Education

- Child and Family Data Archive
- Civic Learning, Engagement and Action Data Sharing
- Education Research Data Sharing Initiative
- Measures of Effective Teaching
- PEERS Data Hub

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
- Millennium Challenge Corporation
- Resource Center for Minority Data
- US Census Data Repository
- ResearchDataGov

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
- Patient-Centered Outcomes Data Repository
- Social, Behavioral, & Economic COVID Coordinating Center









Summer Program & Connecting with ICPSR

Summer Program in Quantitative Methods of Social Research



- Three-week courses and short workshops
 - Methods-related
 - Substantive
- Networking
- Blalock Lectures
- Scholarship opportunities
- YouTube channel

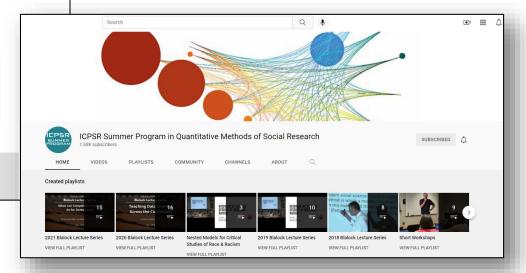
In-person and online courses

Accessible, applied training in a casual learning environment that encourages inquiry and collaboration

Announcing our 2023 dates: Three-Week Sessions: June 19-July 7, 2023 and July 17-August 4, 2023 Short Workshops: May through August 2023



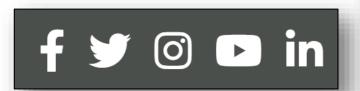
Since 1963, we've trained more than 25,000 researchers.

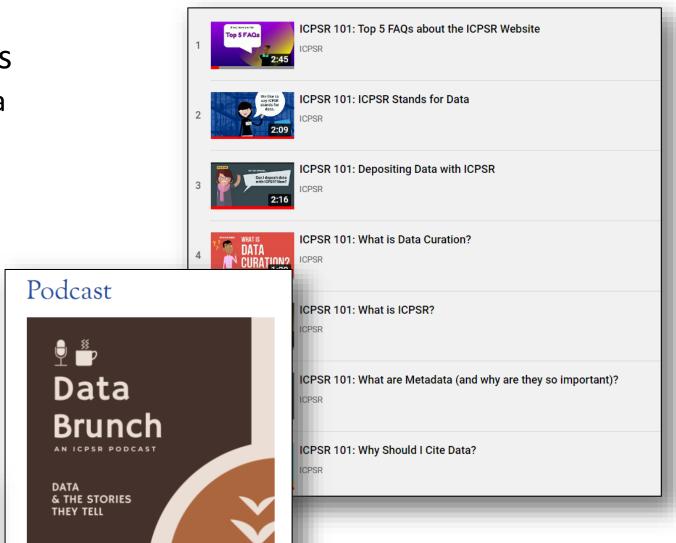




Other Ways to Connect

- YouTube channel full of goodies
 - ICPSR 101 short videos about a variety of topics
- <u>Data Brunch</u> podcast
- Other social media
 - Facebook
 - <u>Twitter</u>
 - <u>Instagram</u>
 - <u>LinkedIn</u>







Thank You!

Phone: 734-647-2200

Website: www.icpsr.umich.edu

Email: icpsr-help@umich.edu

Presenter Contact Info: Amber Bryant,

amberum@umich.edu – happy to

set up video call as well!

