

# Data Management Plan Initiatives at ICPSR

Institute for Social Science, University of Tokyo

March 18, 2024

Presenter: Lynette Hoelter, PhD (ICPSR)

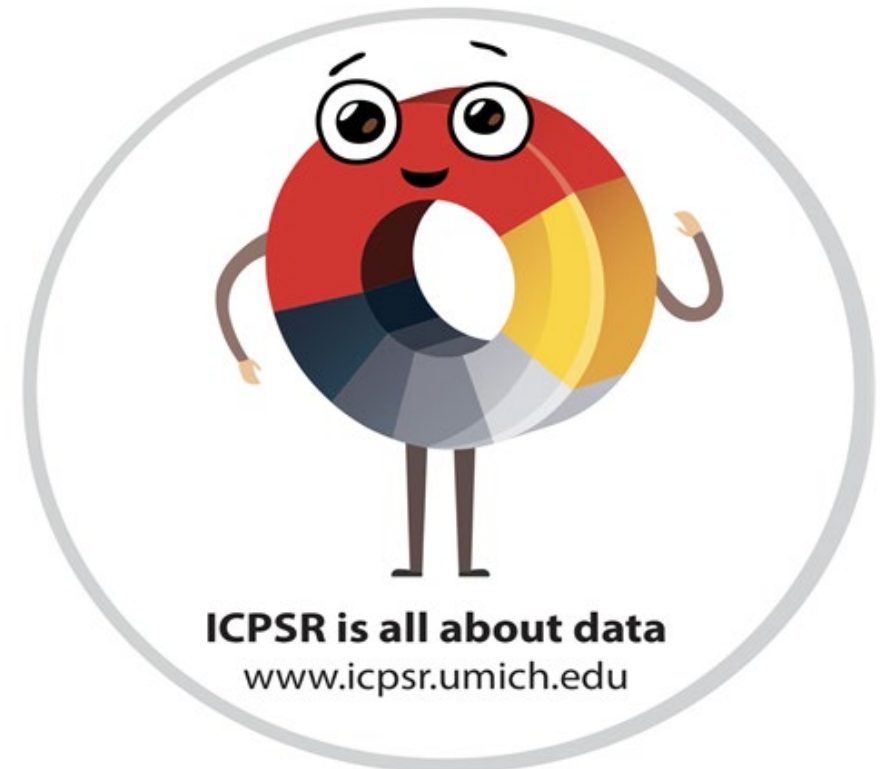


Attribution-NonCommercial 4.0 International

ICPSR

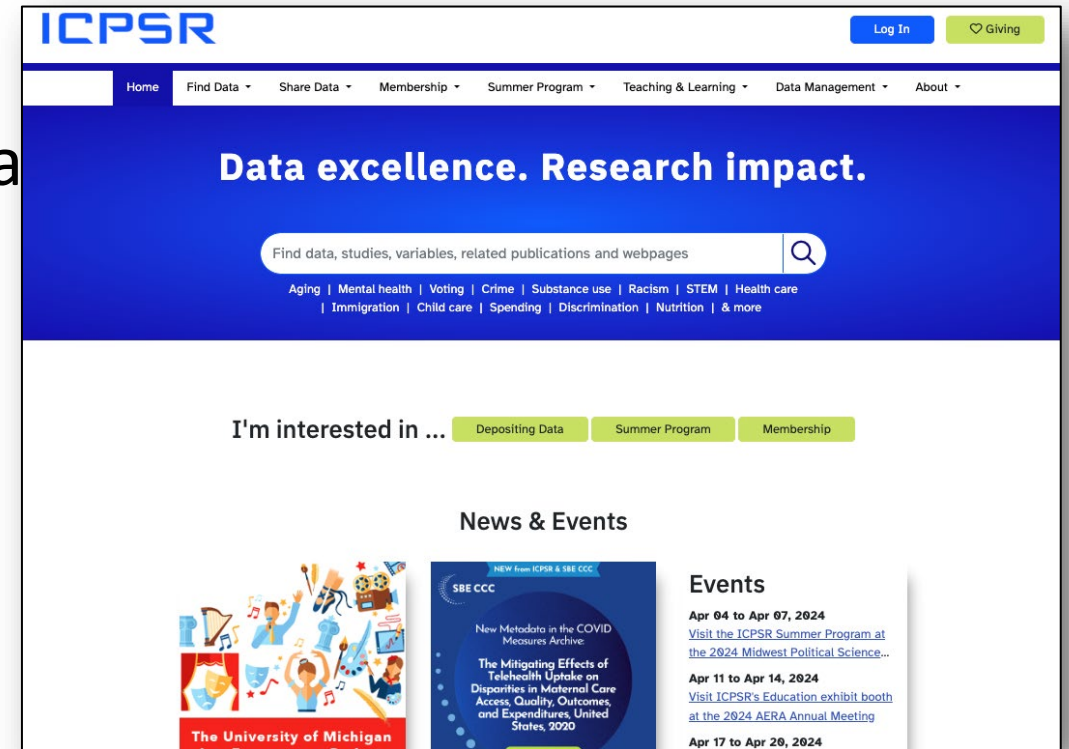
# Today's Presentation

- What is ICPSR?
- U.S. Key Dates in Data Management
- ICPSR's Data Management Philosophy
- Resources for Researchers
- Outreach
- Questions??



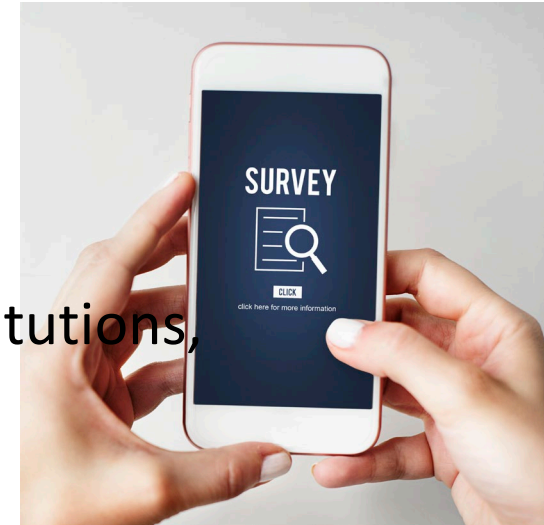
# What is ICPSR?

- [www.icpsr.umich.edu](http://www.icpsr.umich.edu)
- One of the world's oldest and largest data archives for the social and behavioral sciences
  - Founded in 1962
  - 19,700 “studies” (one or more dataset)
  - ~6,425,000 variables
  - 112,500 data-related publications
  - 820 member institutions
- [Summer Program in Quantitative Methods of Social Research](#)



# ICPSR at a Glance

- Social and behavioral sciences, broadly defined
  - Data about individuals (early childhood through elderly), institutions, and events
  - Time periods from ancient history to the present
  - Many disciplines covered
- Public- and restricted-use data
- Tabular/rectangular data files, images (e.g. brain scans), video, narratives, speech transcripts, social media, and more
- Funding model
  - [Funded](#) “collections” (archives) and projects
  - Consortium dues



# Data Management Plans



# Data Sharing: Key Dates

- October 2003, National Institutes of Health (NIH) require data management plans (DMPs) for projects requesting \$500k/year or more
- [January 2011](#), National Science Foundation (NSF) requires DMPs for all proposals
- [January 2013](#), Office of Science and Technology Policy (Executive Office of the President) memo encourages sharing of federally funded data and results
- [January 2023](#), National Institutes of Health issue Data Management and Sharing (DMS) policy

# Data Management Across the Research Lifecycle

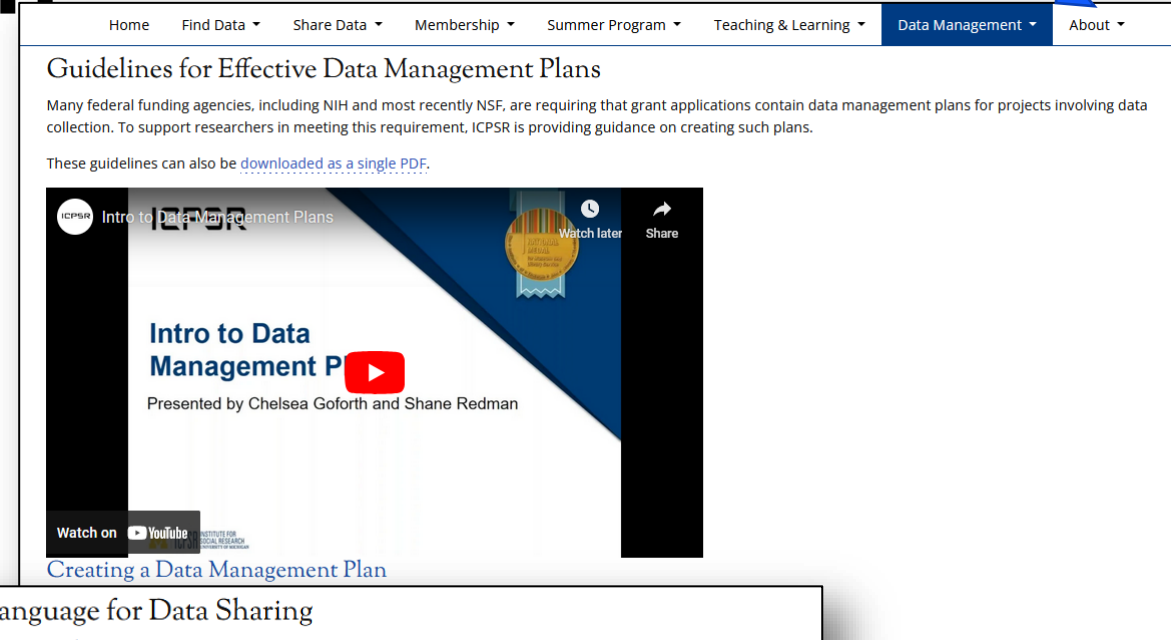


Stage	Data Management
Research idea/proposal	Funding agency requirements, DMP document, letter of support, estimate cost of data sharing
Research design	Informed consent documents to allow sharing, questions asked/concepts measured to minimize disclosure risk
Data collection	Informed consent, safe storage, documentation
Data cleaning and management	Best practices (e.g., remove direct identifiers, “coarsen” indirect identifiers, structure files, document new variables)
Analysis and reporting	Don’t “cherry pick,” think about story your results will tell (especially for vulnerable populations), data/code sharing for journal requirements, cite data
Data sharing	Where will people look? Public vs restricted (or both)? Embargo necessary? Persistent identifiers, strong metadata, preservation <i>and</i> access, demonstrating impact



# Prior to Data Collection

- Create [data management plan](#)
  - ICPSR letter of support and/or budgeting for curation if planning to deposit
- Example [informed consent language](#) that allows sharing



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

## Guidelines for Effective Data Management Plans

Many federal funding agencies, including NIH and most recently NSF, are requiring that grant applications contain data management plans for projects involving data collection. To support researchers in meeting this requirement, ICPSR is providing guidance on creating such plans.

These guidelines can also be [downloaded as a single PDF](#).

ICPSR Intro to Data Management Plans

Watch later Share

### Intro to Data Management Plans

Presented by Chelsea Goforth and Shane Redman

Watch on YouTube

Creating a Data Management Plan

## 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**

# Assistance for Specific Audiences

Home Find Data Share Data Membership Summer Program Teaching & Learning **Data Management**

## Data Management & Curation

ICPSR stores, curates, and provides access to scientific data so others can reuse the data and validate research findings. Curation, from the Latin "to care" uses to add value to data, maximize access, and ensure long-term preservation.

Data curation is akin to work performed by an art or museum curator. Through the curation process, data are organized, described, cleaned, enhanced, and much like the work done on paintings or rare books to make the works accessible to the public now and in the future. With the modern Web, it's increasing. Without curation, however, data can be difficult to find, use, and interpret. Through curation, ICPSR provides meaningful and enduring access to data.

### Quality

Data at ICPSR are enhanced with meaningful information to make it complete, self-explanatory, and usable for future researchers. As a repository, ICPSR adheres to standards that demonstrate it is organizationally, procedurally, and technologically sound as a trustworthy data custodian.

- [Preparing Data for Archiving](#)
- [Guide to Archiving Social Science Data for Institutional Repositories \(pdf\)](#)
- [Curating Data at ICPSR](#)
- [Depositing Data with ICPSR](#)

### Access

ICPSR hosts data in a repository with powerful search capabilities. Indexed by all the major search engines, ICPSR data are easily discoverable and widely accessible to the public.

- [Data Management Plans](#)
- [Guidance for NIH Investigators](#)
- [Guidance for NIDILRR Investigators](#)
- [Guidelines for OSTP Data Access Plan](#)

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

## Share NIH Data

ICPSR is home to more than 1,000 NIH-funded data collections. We can help with the planning and writing of an NIH grant, as well as archiving and disseminating the data collected. ICPSR accepts data that fall within our Collection Development Policy. In addition, ICPSR has developed resources and tools researchers can use when creating an NIH [Data Management and Sharing \(DMS\) Plan](#) and several of our topical archives and projects offer dedicated support for certain NIH-funded data:

- [The National Addiction & HIV Data Archive Program \(NAHDAP\)](#)
- [Data Sharing for Demographic Research \(DSDR\)](#)
- [The National Archive of Computerized Data on Aging \(NACDA\)](#)
- [Social, Behavioral, & Economic COVID Coordinating Center \(SBE CCC\)](#)
- [The National Neighborhood Data Archive \(NaNDA\)](#)

## Resources & Tools for Designating the ICPSR Repository

**ICPSR's Template for the NIH Data Management and Sharing Plan:** [Template NIH DMS Plan from ICPSR](#) (docx): including information on sharing NIH data with ICPSR + copy-paste language

**Budgeting:** Get a cost estimate from ICPSR! Download this [document](#) (docx) and follow the outlined steps to get started on your curation cost estimate. ICPSR will receive notice of your submission and a staff member will contact you directly to provide an estimate.

**Webinars:** [Designating ICPSR in your NIH Data Management and Sharing Plan](#) & [Navigating NIH's Data Management & Sharing Policy with ICPSR's NIH-Funded Projects](#)

## Additional Guidance for HEAL Researchers

To help ICPSR staff identify your project as a HEAL-funded project, please take the following steps when you are working on your data deposit:

1. Include your funding source(s) and grant number(s) in your deposit form in the "Funding Sources" field.
2. Include HEAL as a subject term in your deposit form in the "Subject Terms" field (in the "Scope of Project" section).
3. Upload a copy of your NIH Data Management and Sharing Plan to your deposit.

[Start your deposit](#)

## Contact Us

We are happy to answer any questions you may have about a research project that you are planning or carrying out with NIH funding. Contact us at [ICPSR-help@umich.edu](mailto:ICPSR-help@umich.edu) for any questions or concerns.

## Share Data



### Benefits of Depositing Data

One of the major goals of NACJD and its [sponsors](#) is to support professional researchers test each other's conclusions when professional researchers test each other's conclusions. [National Institute of Justice](#) (NIJ) and the [Office of Juvenile Justice and Delinquency Prevention](#) (OJJDP) data sponsored by these agencies with NACJD.

Sharing data also fosters the development and testing of methods not addressed by the original investigators. Indeed, [deposition](#)

Archiving data affords multiple benefits to the investigator:

- The investigator's study is described on the NACJD website
- NACJD distributes the data and supporting documentation and resources to do so and permitting other researchers to use the data
- Once the data are archived and available, the investigator or she collected.
- The investigator and his or her colleagues can assist in the documentation over a wide variety of computer platforms

### Preparing Research Files for Deposits

NACJD urges researchers to begin their archiving strategy linked above. ICPSR, NACJD's parent organization, has a [page](#)

NACJD encourages researchers to deposit computer-readable science researchers. The following information will assist

- [Guidelines for OJP Analysts, Program Officers, and Other Users](#)
- [Data with Confidential Content](#)
- [Projects Analyzing Existing Data](#)

The screenshot shows the DSDR website interface. At the top right, there is a 'Log In/Create Account' button. Below it is a navigation bar with 'Discover', 'Deposit', 'Resources', and 'About' links. A search bar contains the text 'search studies, publications, variables, webpages'. The main content area features a 'Deposit Data' section with an 'Overview' sub-section. A 'Table of Contents' sidebar lists 'Overview', 'Why Deposit Data With ICPSR?', 'What Should My Deposit Include?', and 'What Happens to My Data After the Deposit?'. A 'Start Deposit' button is prominently displayed. Below this, the 'Why Deposit Data With DSDR and ICPSR?' section lists benefits like Data Curation, Long-Term Preservation, and Widespread availability. At the bottom, it details 'User Support', 'Levels of Access', 'Aggregation of Publications', and 'NIH and NSF data management plans'.

# Sample Plan (General)

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

## Sample Data Management Plan for Depositing Data with ICPSR

This sample plan is provided to assist grant applicants in creating the required Data Management Plans. Researchers should feel free to edit and customize this text before submission. A letter of commitment from ICPSR confirming that it will archive the data should accompany the plan. Please contact ICPSR Acquisitions, [deposit@icpsr.umich.edu](mailto:deposit@icpsr.umich.edu) ✉, to request such a letter. Note that letters of commitment from ICPSR are not provided to researchers applying for National Institute of Justice (NIJ) sponsored research because, in most instances, NIJ requires datasets resulting from funded research to be archived with the [National Archive of Criminal Justice Data \(NACJD\)](#) at ICPSR.

Please review our [guidelines on how to deposit data](#) with ICPSR.

**Data Description** – [Provide a brief description of the information to be gathered – the nature, scope, and scale of the data that will be generated or collected.] These data, which will be submitted to ICPSR, fit within the scope of the [ICPSR Collection Development Policy](#). A letter of support describing ICPSR's commitment to the data as they have been described is provided.

**Responsibility** – The principal investigator will have overall responsibility for data management over the course of the research project and will monitor compliance with the plan. The PI will ultimately transfer responsibility for data management to the Inter-university Consortium for Political and Social Research (ICPSR).

**Designated Archive** – The research data from this project will be deposited with the digital repository of the [Inter-university Consortium for Political and Social Research](#) (ICPSR) to ensure that the research community has long-term access to the data. The integrated data management plan proposed leverages capabilities of ICPSR and its trained archival staff.

**Access and Sharing** – ICPSR will make the research data from this project available to the broader social science research community. *Public-use data files*: These files, in which direct and indirect identifiers have been removed to minimize disclosure risk, may be accessed directly through the ICPSR website. After agreeing to Terms of Use, users with an ICPSR MyData account and an authorized IP address from a member institution may download the data, and non-members may purchase the files. *Restricted-use data files*: These files are distributed in those cases when removing potentially identifying information would significantly impair the analytic potential of the data. Users (and their institutions) must apply for these files, create data security plans, and agree to other access controls. *Timeliness*: The research data from this project will be supplied to ICPSR before the end of the project so that any issues surrounding the usability of the data can be resolved. Delayed dissemination may be possible. The Delayed Dissemination Policy allows for data to be deposited but not disseminated for an agreed-upon period of time (typically one year).

**Selection and Retention** – ICPSR will archive the full dataset and its documentation for the long term, supporting the data through changing technologies, new media, and data formats.

**Metadata** – ICPSR will create substantive metadata in compliance with the most relevant standard for the social, behavioral, and economic sciences—the [Data Documentation Initiative](#) (DDI). This XML standard provides for the tagging of content, which facilitates preservation and enables flexibility in display. These types of metadata will be produced and archived:

- *Study-Level Metadata Record*. A summary DDI-based record will be created for inclusion in the searchable ICPSR online catalog. This record will be indexed with terms from the ICPSR Thesaurus to enhance data discovery.
- *Data Citation with Digital Object Identifier (DOI)*. A standard citation will be provided to facilitate attribution. The DOI provides permanent identification for the data and ensures that they will always be found at the URL specified.
- *Variable-Level Documentation*. ICPSR will tag variable-level information in DDI format for inclusion in ICPSR's Social Science Variables Database (SSVD), which allows users to



# Downloadable Templates/Guides

OMB No. 0925-0001 and 0925-0002 (Rev. 07/2022 Approved Through TBD)

## DATA MANAGEMENT AND SHARING PLAN

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on [sharing.nih.gov](https://www.nih.gov/sharing-nih). The Plan is recommended not to exceed two pages. Text in italics should be deleted. There is no "form page" for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format* shown below.

Public reporting burden for this collection of information is estimated to average 2 hours per response, including the time for reviewing instructions, searching existing data sources, gathering, and maintaining the data needed, and completing and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to: NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-0001 and 0925-0002). Do not return the completed form to this address.

**ICPSR-specific information relevant to each element is highlighted in yellow in the text below. This template assumes you will fill in information about your project and then supplement that information with the ICPSR-specific information where applicable. You may substitute the name of any of ICPSR's NIH-sponsored archives (DSDR, NACDA, or NAHDAP) in place of "ICPSR" in the text below (or combine them, such as "ICPSR/NAHDAP"). Remember to remove this paragraph, the instructions above, and the italic text below from the final version of your DMS Plan.**

### Element 1: Data Type

#### A. Types and amount of scientific data expected to be generated in the project:

*Summarize the types and estimated amount of scientific data expected to be generated in the project.*

#### B. Scientific data that will be preserved and shared, and the rationale for doing so:

## Budgeting for ICPSR Curation Services in a Grant Application

ICPSR accepts data originating in many disciplines that study various dimensions of human health and well-being. If you are writing a grant application, you may designate ICPSR (NIH grantees may also consider designating one of the NIH-funded archives – DSDR, NACDA, and NAHDAP) in your data management plan if the data fall within the [ICPSR Collection Development Policy](#). It is recommended that you include ICPSR curation services in your grant application budget<sup>1</sup>. ICPSR will supply a cost estimate that covers a one-time, final submission for a set of related data and documentation files. When the data are submitted, ICPSR will review the submission and provide an invoice for data it accepts and will curate by validating and standardizing the submission (e.g., ensuring human subject protections are maintained/managed; data are Findable, Accessible, Interoperable, & Reusable).

### Steps to Develop a Budget for ICPSR Curation Services:

1. REVIEW SUBMISSION REQUIREMENTS - A standard data submission must adhere to ICPSR's requirements defined below. If you are planning to collect data that may fall outside these requirements, ICPSR will review information about your plans to help estimate the cost.
2. REGISTER YOUR PLANNED DATA SUBMISSION WITH ICPSR - Anyone wishing to designate ICPSR as the intended repository for data and include an estimate for curation services in a budget **must complete this form to register their plan.** For any submission falling outside the standard data submission requirements, please allow two weeks for an ICPSR staff person to review your plans to determine a cost to include in your budget.
3. DESIGNATE ICPSR IN DATA MANAGEMENT PLAN - You may wish to review and incorporate ICPSR's Data Management Plan (DMP) information into your data management plan. ICPSR offers both a [template for writing a National Institute of Health Data Management & Sharing Plan](#) and a [general DMP Template](#) to help as you write your data management plan.
4. INCLUDE ESTIMATE IN BUDGET - For a standard data submission, ICPSR will provide an estimate after you register your proposed project (see #2 above).

### What are the Requirements for a Standard Data Submission?

**General Guidance** - ICPSR will review all submissions to determine if the data are complete, meet the requirements of a standard submission (below), and determine if ICPSR can curate

<sup>1</sup> Any data submission may ultimately be accepted by an NIH- or membership-supported repository at ICPSR and curated at no cost to the submitter; this is determined at the time of submission.

# Budgeting for Curation

How many quantitative data files do you expect to deposit?

In which format(s) will your data files be? Select all that apply.

- SAS
- SPSS
- Stata
- Excel
- R
- Other (please explain)

How many variables do you expect to deposit?

Please estimate the total number of variables across all files you plan to deposit.

Do any variables contain open-ended responses?

- Yes
- No

What type of qualitative data will you be depositing (text/transcripts)?

How many qualitative data files do you expect to deposit (i.e., number of transcripts, audio files, video files, etc.)?

Please approximate the average size of each expected file type (i.e., average number of pages per transcript, average size of each audio file, etc.).

In what format will your files be deposited (e.g., Word, PDF, MOV, JPEG, etc.)?

Do the qualitative data files include sensitive/confidential information?

- Yes
- No

Can individuals be identified from information in the qualitative data files?

# Data Prep Guide

## Guide to Social Science Data Preparation and Archiving

Best Practice Throughout the Data Life Cycle: 6<sup>th</sup> 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
- Qualitative Data
- Other Data Types
- Best Practice in Creating Metadata



## Table of Contents

<b>About ICPSR</b> .....	1
» Why Should I Archive Data? .....	1
» How Do I Deposit Data? .....	1
» What Services Does ICPSR Offer? .....	2
» Using the Guide .....	3
<b>Importance of Data Sharing and Archiving</b> .....	5
» Planning Ahead for Archiving and Preservation of Data .....	6
» The Data Life Cycle .....	7
<b>Proposal Development and Data Management Plans</b> .....	9
» Recommended Elements for Data Management Plans .....	9
» Other Considerations .....	12
» Summary Table: Elements of a Data Management Plan .....	14
» An Example Data Management Plan for Depositing Data with ICPSR .....	16
<b>Project Start-Up</b> .....	19
» Importance of Good Data Management .....	19
<b>Data Collection and File Creation</b> .....	23
» Best Practice in Creating Research Data .....	23
» Quantitative Data .....	23
» Qualitative Data .....	30
» Other Data Types .....	32
» Best Practice in Creating Metadata .....	32
<b>Data Analysis</b> .....	37
» Master Datasets and Work Files .....	37
<b>Addressing Confidentiality Issues</b> .....	41
» Respondent Confidentiality .....	41

# New Qualitative Guide



## Guide for Sharing Qualitative Data at ICPSR

October 24, 2023

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

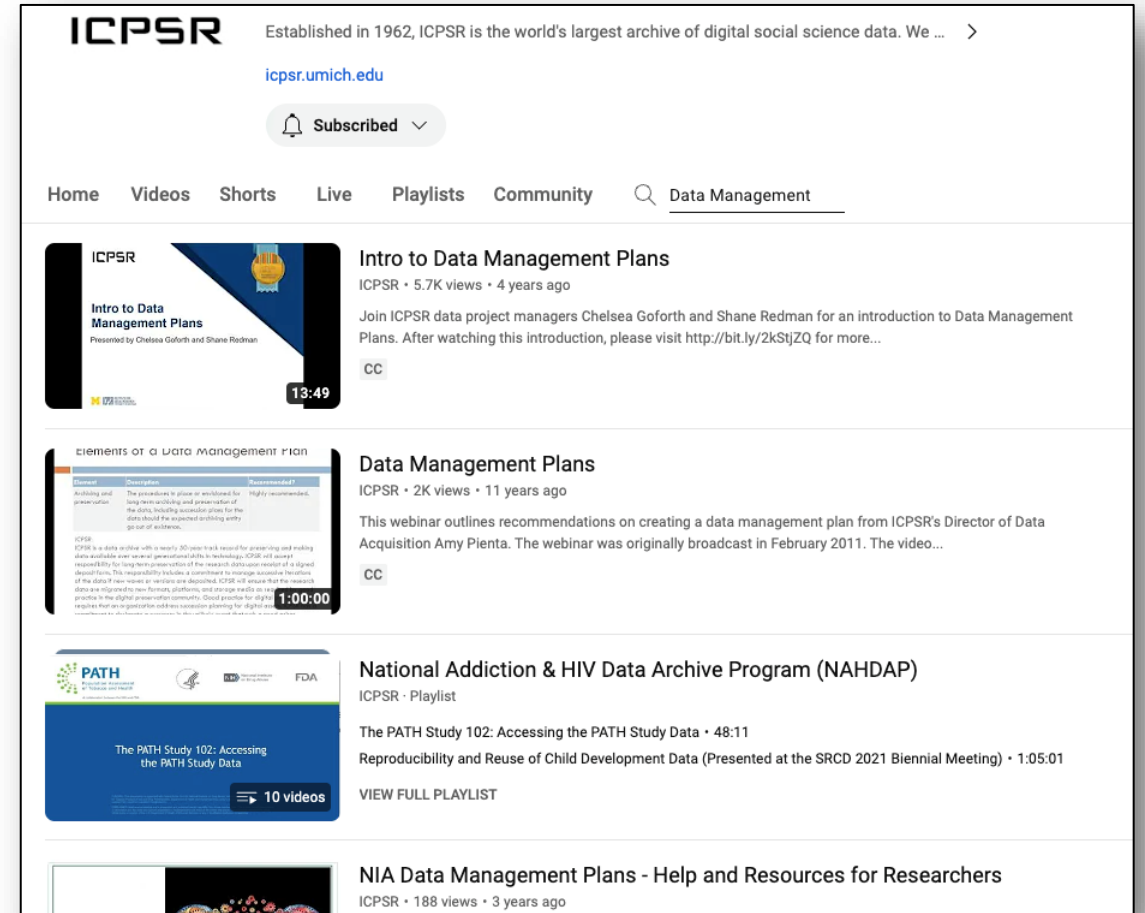


# Restricted-Use Data

- Can be shared!!
- Might create two versions
- Typically restricted because of
  - Potential for disclosure risk
  - Sensitive topics
  - Data provider requirement
- Access requires
  - Description of project
  - Explanation of why the restricted file is necessary
  - List of research team members
  - IRB approval
  - Signed Restricted Data Use Agreement (RDUA)
- Access mechanisms
  - Encrypted download
    - Requires data security plan as part of RDUA
  - Virtual Data Enclave
    - Data never leave ICPSR servers
    - No internet or other network connection
    - Output must be vetted
  - Physical Data Enclave
    - In-house at ICPSR
    - Standalone computers, nothing can be taken in
    - Output must be vetted

# Outreach

- General webinars, saved on [YouTube channel](#)
- Archive-specific webinars and/or included in SP courses
- Workshops, sessions, and posters at conferences
- Inclusion in other webinars (“Introduction to ICPSR”)
- Lots of one-on-one consultations



The screenshot shows the ICPSR YouTube channel page. At the top, the ICPSR logo is displayed next to the text "Established in 1962, ICPSR is the world's largest archive of digital social science data. We ...". Below the logo is the website URL "icpsr.umich.edu" and a "Subscribed" button with a notification bell icon. The navigation menu includes "Home", "Videos", "Shorts", "Live", "Playlists", "Community", and "Data Management".

The main content area features several video recommendations:

- Intro to Data Management Plans**: ICPSR • 5.7K views • 4 years ago. Description: "Join ICPSR data project managers Chelsea Goforth and Shane Redman for an introduction to Data Management Plans. After watching this introduction, please visit <http://bit.ly/2kStjZQ> for more..."
- Data Management Plans**: ICPSR • 2K views • 11 years ago. Description: "This webinar outlines recommendations on creating a data management plan from ICPSR's Director of Data Acquisition Amy Pienta. The webinar was originally broadcast in February 2011. The video..."
- National Addiction & HIV Data Archive Program (NAHDAP)**: ICPSR • Playlist. Description: "The PATH Study 102: Accessing the PATH Study Data • 48:11. Reproducibility and Reuse of Child Development Data (Presented at the SRCD 2021 Biennial Meeting) • 1:05:01. VIEW FULL PLAYLIST"
- NIA Data Management Plans - Help and Resources for Researchers**: ICPSR • 188 views • 3 years ago.

Questions??



# Thank You!

**Phone:** 734-647-2200

**Website:** [www.icpsr.umich.edu](http://www.icpsr.umich.edu)

**Email:** [icpsr-help@umich.edu](mailto:icpsr-help@umich.edu)

Lynette Hoelter (lhoelter@umich.edu)