



Guide for Sharing Qualitative Data at ICPSR

October 24, 2023

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

The Objective of this Guide

The Inter-university Consortium for Political and Social Research (ICPSR) has created this resource for investigators planning to share qualitative data at ICPSR. The information below provides an overview of elements and considerations for archiving qualitative data, including common sources of qualitative data and its value for reuse. It also provides an overview of how ICPSR supports and partners with the investigator to reduce disclosure risk and optimize usability when qualitative data are archived at ICPSR. The guide also identifies steps for investigators to follow during the research life cycle to ensure that others can share and reuse qualitative data. Finally, information about exemplars of qualitative data available from the ICPSR archive is provided and links to helpful resources external to ICPSR to support the goal of sharing qualitative data.

For more tailored directions based on your specific research project, please contact ICPSR-help@umich.edu.

Table of Contents

Table of Contents	2
Characteristics of Qualitative Data	4
Qualitative Data and the Potential for Reuse	6
Ensuring Qualitative Data Confidentiality	8
Protecting Confidentiality in Qualitative Data	8
Special Considerations for Disclosure Risk in Qualitative Data	10
Disclosure Risk Review by ICPSR	11
Restricted-Use Access Options for Qualitative Data at ICPSR	11
Planning a Project to Collect Qualitative Data	12
<i>Designating a Repository</i>	12
<i>Developing a Data Management (and Sharing) Plan</i>	12
<i>Developing a Proposal Budget</i>	13
<i>IRB and Informed Consent</i>	13
<i>ICPSR Guide to Social Science Data Preparation and Archiving</i>	13
Preparing to Deposit Your Qualitative Data	14
Documentation for Qualitative Data	14
Suggested Data Formats for Qualitative Data	14
Suggested File Organization	14
Submitting Qualitative Data to ICPSR	15
Data Curation at ICPSR	17
ICPSR Curation Steps for Qualitative Data	17
Finding Qualitative Data to Reuse at ICPSR	18
The Qualitative Data Sharing Project Series	18
College and Beyond II	18
More Studies with Qualitative Data at ICPSR	18
Additional Resources External to ICPSR	20
<i>The Qualitative Data Archive</i>	20
<i>Qualitative Data Sharing (QDS) Toolkit</i>	20
<i>UK Data Service</i>	20
References	21

Inter-university Consortium for Political and Social Research

Institute for Social Research
University of Michigan
P.O. Box 1248
Ann Arbor, MI 48106-1248

Voice: +1 (734) 647-2200

Fax: +1 (734) 647-8200

Email: ICPSR-help@umich.edu

Web: <https://www.icpsr.umich.edu/icpsrweb/>

Acknowledgements

This guide has been created with the support of a grant from the National Human Genome Research Institute [R01 HG009351; PI James DuBois] and the institutional members of ICPSR. ICPSR staff who authored this guide include Kyran Reneau and Amy Pienta, and it draws on methods created by the talented curation staff of ICPSR. We thank Sara Britt, Aubrey Garman, Amanda Draft, Julie Eady, Allyson Flaster, Susan Leonard, A.J. Million, Katey Pillars from ICPSR, and Heidi Walsh from Washington University in St. Louis for feedback on earlier drafts of this guide.

Characteristics of Qualitative Data

What is Qualitative Data?

Qualitative research data is a rich and multidimensional source of information within the social sciences and various other fields. Unlike quantitative data, which contains numerical values, qualitative data contains information that does not lend itself well to numerical measurement. This information can reveal nuances about human experiences, behaviors, and perceptions. Qualitative data may be gathered through interviews, observations, surveys, or textual analysis and is typically non-numeric (though, in some cases, numbers can also be considered qualitative). This data type mostly refers to words (including spoken words) and text from many sources.

Sources of Qualitative Data

Examples of types of qualitative data that may be archived for secondary analysis include the following:

- Interview methods, including those captured through notes, audio, and video
 - In-depth and/or unstructured interviews
 - Semi-structured interviews
 - Focus group interviews
- Diary studies that are unstructured or use semi-structured writing prompts
- Observational studies that generate field notes and other text and information
 - Naturalistic observation of real-world environments (e.g., classrooms, workplaces, healthcare facilities, courtrooms, public spaces)
 - Participant observation where the researcher becomes an active part of the setting to collect information (e.g., online gaming, community policing, nightclub culture)
 - Structured observation is where the research has predefined objectives and a systemic approach to collecting information. This would include case studies.
- Text from available sources
 - Meeting minutes
 - Official records
 - Medical records
 - News sources and social media
- Excerpts of copyrighted materials (e.g., literature, film, music)
- Survey methods or questionnaires with substantial open-ended comments

Mixed methods research combines both qualitative and quantitative research methods to gain a more comprehensive understanding of a research problem. Examples of mixed-method studies may include the following:

- **Educational Achievement and Parental Involvement:** A researcher wants to investigate the impact of parental involvement on students' educational achievement. They start with a quantitative survey of a large sample of students to measure academic performance and parental involvement levels. Then, they conduct qualitative interviews with a subset of parents and students to explore the reasons behind these relationships.
- **Employee Satisfaction in the Workplace:** To investigate factors influencing employee satisfaction in a company, researchers begin by administering a quantitative employee satisfaction survey. Then, they conduct qualitative interviews with a smaller group of employees to gain deeper insights into the reasons behind their satisfaction or dissatisfaction, potentially uncovering issues not captured by the survey.

Qualitative Data and the Potential for Reuse

Qualitative research methods have many unique strengths that are important when considering the potential for data reuse. Qualitative research allows researchers to explore complex phenomena in-depth and gain a nuanced understanding of the subject matter. Scholars often highlight the ability of qualitative methods to uncover hidden meanings and deeply explore social and cultural contexts (Creswell & Creswell, 2017). Qualitative methods may be flexible, allowing researchers to adapt their approach during the study based on conditions and initial findings. This may be useful when exploring new topics with less evidence (Merriam & Tisdell, 2016). Qualitative research may also offer context around social phenomena as they have occurred. For example, it is possible to understand the social, cultural, and historical factors that influence lived experiences (Denzin & Lincoln, 2018).

Qualitative data are considered well-suited for reuse due to their richness, depth, and potential for multiple interpretations. Scholars can reanalyze and reinterpret qualitative data to uncover new insights and perspectives, making them valuable resources for ongoing research and interdisciplinary collaboration (Hinds et al., 2019). Additionally, qualitative datasets that are properly documented and shared in repositories contribute to transparency and facilitate data reuse within the research community (Tenopir et al., 2011).

On the other hand, qualitative research may have limited generalizability of findings to broader population groups because it typically involves small, non-random samples (Yin, 2017). The process of data collection (e.g., interviews, fieldwork) and analysis (e.g., coding, and thematic analysis) can be time-consuming (Morse, 2015), making it challenging for researchers to devote effort to planning and preparing their data for sharing. Researchers may have objections to sharing qualitative data for several reasons. First, they may be concerned about ethical issues, such as the potential for identifiable information to be revealed, which could breach participant confidentiality. Second, some researchers may worry about misinterpretation or misuse of their data by others, as qualitative data can be context-dependent and nuanced, and documenting this is difficult. Third, concerns about the time and effort invested in collecting and analyzing qualitative data may lead to a desire to protect their work.

Mixed methods research deserves special mention regarding the potential for reusability. Like qualitative methods, mixed methods research can provide a nuanced understanding of complex social phenomena. Because it is combined with quantitative data, it can also increase generalizability. Those interested in reusing mixed methods data, like the original investigators of the mixed method study, can use the archived data to triangulate their findings, enhancing the validity and depth of their conclusions. The combination of rich, contextual qualitative data with numerical data that is precisely measured is great for helping to triangulate

findings and increase claims of validity but also might increase the chances that a participant could be identified, requiring greater attention to this issue in mixed method data.

Ensuring Qualitative Data Confidentiality

Protecting Confidentiality in Qualitative Data

Before submitting qualitative data to an archive, data depositors should remove information, allowing any of their research subjects to be identified. This de-identification process should be done in a way that is consistent with relevant organizational obligations (i.e., Institutional Review Board, IRB) and the terms participants agreed to in the study's informed consent. However, it is also possible to make arrangements with ICPSR for depositing qualitative data containing personally identifying information (PII) and other direct identifiers (e.g., addresses, organization names) if both the depositor and ICPSR agree to the terms whereby the data can be stored and accessed through ICPSR. This requires a restricted data contributor agreement (RDCA) to be executed between the organization where the researcher collected and produced the data and ICPSR.

The de-identification process, where necessary, can be made less arduous by creating an anonymization protocol before data collection and anonymizing the data as the qualitative files are created. Below in Table 1 are examples of modifications that can be made to qualitative data to address respondent confidentiality:

TABLE 1. REPLACEMENT STRATEGIES FOR IDENTIFYING INFORMATION IN TEXT PASSAGES

Name Replacement	
Anonymization Approach	Example
Replacing actual names with generalized text.	The most basic replacement would be to replace a name with [Name].
Replacing actual names with descriptive text, retaining role or demographic information.	"John" can be changed to [uncle] or "Mrs. Brown" to [teacher]. Or "John" can be changed to [M/W/20] for male, white, 20 years old.

Replacing actual names with generalized text using numeric ordering.	<i>“Mohammed” countered what was said by “Derek” can be changed to... [Subject1] countered what was said by [Subject2]</i>
Replacing actual names with descriptive text, retaining relation, role, or demographic information, and using numeric ordering.	More than one person with the same relationship to the respondent can be subscripted to represent each unique individual — e.g., [friend1], [friend2].
Date Replacement	
Anonymization Approach	Example
Replacing dates	Dates referring to specific events, especially birth dates or events involving the criminal justice system or major medical events, should be replaced with some general marker for the information, e.g., [month], [month/year], or [mm/dd/yy].
Geography Replacement	
Anonymization Approach	Example
Replacing geographic identifiers	Detailed geographic locations should be replaced with some general marker of place information, e.g., [street1], [city1], and so on. Or make a substitution with meaning, e.g., [small city1]
Unique/Publicized Items and Events Replacement	
Anonymization Approach	Example

Removing unique and/or publicized items	If the item cannot be generalized using one of the above options, the entire text may need to be removed and explicitly marked as such, e.g., using either [description of event removed], or [...] as a general indicator.
---	---

Documenting and Identifying Replacements

Data depositors should document the modifications they made to the data in service of masking confidential information. This documentation will allow ICPSR to review the changes to the data easily and communicate to data users what has been replaced or masked. For example, it is commonplace to indicate text modifications by bracketing the changed information that was identifying, e.g., "[Name1] told me that they were also worried about police aggression." Data depositors are asked to indicate how such changes have been designated in accompanying documentation. Appendix A of this guide provides four examples of how to provide ICPSR with documentation of how replacements were made.

Special Considerations for Disclosure Risk in Qualitative Data

As indicated above, direct or near-direct identifiers such as names, dates, geographic locations, and unique items and events pose obvious disclosure risks that must be mitigated. Additionally, ICPSR has accumulated experience detecting more subtle forms of disclosive information in qualitative data. We describe these here:

- Redact (almost) all names: ICPSR recommends redacting all names, including participants, research team members (e.g., [interviewer1] or [notetaker].), or other individuals who may or may not be in the study. *However, it may be acceptable for the names of public figures, such as political officials or celebrities, to remain in the data.*
- Redact places uniquely identifying any masked geography: ICPSR recommends obscuring geographic locations, names of businesses, or places unique to a masked geographic area. For example, "We had [son1] birthday party at Rainforest Cafe in [City1], Tennessee" may identify the city within Tennessee. It so happens that there is only one location in Tennessee, and that is in Nashville.
- Redact acronyms or abbreviations: ICPSR recommends replacing these if they may be used to re-identify masked organizations, schools, clubs, workplaces, etc.
- Redact any other identifying information: ICPSR recommends replacing and obscuring school mascots, nicknames, and so on (e.g., the University of Arkansas uniquely has the razorback, a type of hog, as its mascot).

- Redact identifying information in related quantitative data: For mixed method studies, ICPSR recommends obscuring identifying information in quantitative files that accompany qualitative data when they provide uniquely identifying information about people and places in the qualitative data files.

Disclosure Risk Review by ICPSR

Qualitative data shared through ICPSR undergo a disclosure risk review before release on the ICPSR website or archives (Corden, Eady, Jang, Kirkwood, Reneau & Tang 2022). This includes considering the restriction level for providing users access to the data. ICPSR manages additional conditions on restricted data that may allow different de-identification standards to be applied in the disclosure risk review.

The disclosure risk review process for qualitative data typically includes:

- Reviewing all of the text within each qualitative data file
- Noting any direct or indirect identifiers and contextual information that could result in a disclosure risk
- Implementing advanced strategies to determine the risk of re-identification through indirect identifiers
- Ensuring the released data mitigates the risk of disclosure through masking identifiers and/or designating appropriate restricted-access conditions described below.

Restricted-Use Access Options for Qualitative Data at ICPSR

Sometimes, protective measures taken to reduce disclosure risk reduce the reuse potential of the qualitative data. To retain more information within the files, ICPSR can provide access to restricted-use versions of the data that offer another layer of confidentiality protections by imposing stringent access requirements. ICPSR supports access to restricted-use qualitative data through secure download, Virtual Data Enclave (VDE), Physical Data Enclave (PDE), and delayed dissemination (see Appendix B for more information on these access methods). Compared to purely quantitative studies, which can be easier to remove identifying information from, ICPSR is more likely to present qualitative studies (and mixed method studies) to users as restricted-use data due to the depth of information, details, and contextual information they contain. This provides additional protection against the re-identification of qualitative data.

Planning a Project to Collect Qualitative Data

Designating a Repository

ICPSR suggests that investigators planning to collect qualitative data review potential repositories for their qualitative data when writing their research proposal. It is important to do this before data are collected to ensure data collection decisions and timelines meet any requirements of the repository. It is important to ensure the repository can meet the requirements of your institution regarding the management and protection of human subjects' data and other terms. Sponsors who support the study and data collection may have repository requirements such as guidance from the National Institute of Health for Selecting a Data Repository (NIH, 2023). ICPSR is a CoreTrustSeal repository, meeting most requirements, and accepts qualitative data consistent with the [ICPSR Collection Development Policy](#) (ICPSR, 2023).

Developing a Data Management (and Sharing) Plan

A data management plan (DMP) is a structured document that outlines how your qualitative data will be collected, organized, stored, documented, and ultimately shared or preserved throughout the research project's lifecycle. It is a roadmap to effectively manage your data and ensure its integrity, accessibility, and long-term usability. Key components typically included in a data management plan may encompass:

1. **Data Collection:** Describing the methods and instruments used to collect qualitative and mixed method data, including details about selecting locations and interviewees, data sources, and data formats.
2. **Data Organization:** Outlining how your qualitative data files will be structured and organized, including naming conventions, directory structures, and file formats.
3. **Data Documentation:** Specifying how your qualitative data will be annotated and described, including metadata standards. ICPSR follows the [Data Documentation Initiative metadata standard](#).
4. **Data Storage and Backup:** Detailing where and how your qualitative data will be stored during the research project, emphasizing data security and backup procedures.
5. **Data Sharing and Access:** Identifying how and when your qualitative (and mixed method) data will be shared, specifying access controls, and designating a data repository.
6. **Data Preservation:** Discussing plans for long-term data preservation.
7. **Ethical and Legal Considerations:** Addressing ethical and legal issues related to collecting and providing access to qualitative data, such as informed consent, privacy, and compliance with data protection regulations.
8. **Roles and Responsibilities:** Defining the roles and responsibilities of team members in data management.

Data management plans are considered essential in qualitative research projects, particularly those funded by government agencies or institutions, as they promote transparency, accountability, and the responsible handling of research data. They help researchers ensure that their data is well-maintained, accessible, and compliant with legal and ethical standards. ICPSR provides several data management plan templates for investigators designating ICPSR for their planned data collection (see [ICPSR's general data management plan template](#) and [ICPSR's NIH data management and sharing plan template](#)).

Developing a Proposal Budget

The investigator should outline their plans for preparing the data and documentation for archiving and estimate the cost of that work -- both on their end and to engage repository services for their qualitative data. When designating ICPSR, ideally, these budget costs should be planned in conjunction with ICPSR, which may have costs for non-standard terms of access (i.e., ICPSR members get free access to ICPSR-curated data, whereas all users get free access to data curated by sponsored funds).

IRB and Informed Consent

When planning to share qualitative data in a data repository, researchers must carefully consider the ethical implications and regulatory requirements associated with IRB and informed consent. The IRB ensures that research involving human participants adheres to ethical standards and guidelines. **Investigators must consult with their IRB to plan an informed consent process that adequately covers data sharing and to determine whether specific consent is needed for data repository sharing. Informed consent should inform participants about the potential sharing of their data, including any risks, confidentiality safeguards, and the intended purpose of data sharing.** ICPSR offers [general guidance to investigators when writing informed consent](#) to address plans for data sharing and recommends this paper from Corti, Day, and Backhouse (2000) to qualitative researchers writing informed consent. Investigators should also ensure that any data shared in repositories is appropriately anonymized or de-identified to protect the privacy and confidentiality of participants to be consistent with IRB and legal requirements.

ICPSR Guide to Social Science Data Preparation and Archiving

ICPSR's general purpose [Guide to Social Science Data Preparation and Archiving](#) (now in its 6th edition) is an important companion to this Guide for Sharing Qualitative Data at ICPSR. The Guide to Social Science Data Preparation and Archiving has sections explaining proposal development, how to create, collect, and analyze data, how to protect confidentiality, and steps for depositing data (ICPSR, 2020). Section 5 discusses qualitative data briefly.

Preparing to Deposit Your Qualitative Data

Documentation for Qualitative Data

For qualitative data to be used in secondary analysis, the data must be well documented. Specifically, documentation for qualitative data should include:

- Overview of research methods and protocols (ideally, this is a stand-alone document and/or information entered by the depositor into [ICPSR's deposit form](#))
- A blank copy of the informed consent form with IRB approval number to assist ICPSR in determining if there are any restrictions on data sharing
- Details about the setting of any interviews or observations
- Details on the selection of study participants, observations made, documents, and other textual and visual sources
- Instructions given to interviewers, if applicable
- Copies of all data collection instruments (e.g., interview guides, questionnaires)
- Steps taken to replace direct identifiers in the data (e.g., name, address, etc.) (see Appendix A)
- Any problems that arose during the sample selection and/or interview process and how they were handled
- Interview roster describing participants and their key characteristics, such as demographics and relevant conditions (e.g., this is frequently a spreadsheet listing participants)
- Administrative records, text scraped from the internet, and other types of qualitative data should include any documentation that offers users a complete understanding of how the data were collected and any limitations on its use.

Suggested Data Formats for Qualitative Data

Ideally, depositors are requested to submit qualitative data in the following electronic file formats: plain text (*.txt), rich text (*.rtf), and scanned image of text with OCR (*.pdf), and Microsoft Word (*.doc, *.docx), as these formats easily convert to ICPSR's archival formats. Qualitative data containing annotations and codes may be retained in their original file format (e.g., *.nvp) and enrich the reuse of the raw data. More detailed information on a broader variety of qualitative data formats is also made available by the Library of Congress (Library of Congress, 2023).

Suggested File Organization

ICPSR recommends organizing qualitative data files using a nested structure. Ensuring that the folder names identify what is inside is important. Folder and file names should not include identifying information such as

respondent names. Folders will typically contain similar files (e.g., raw transcripts, transcripts with annotations, interview guides) and may be organized by setting, timepoint, and condition. In cases where data contain annotations or codes, these data should be organized in a folder separate from the raw data. Examples of folder structures can be found in Appendix C.

Submitting Qualitative Data to ICPSR

We provide a general overview of the steps involved in depositing qualitative data into ICPSR:

1. **Preparation and Documentation:** Before depositing your qualitative data, ensure that you have organized and documented your data thoroughly. This includes transcripts, interview guides, field notes, annotations, and other relevant materials. Make sure all sensitive or confidential information is appropriately redacted or anonymized.
2. **Create an ICPSR Account:** If you don't already have one, create an account on the ICPSR website. You will need this account to initiate the data deposit process.
3. **Start a New Data Deposit:** Log in to your ICPSR account and find the option to start a new data deposit. This typically involves filling out an online form with details about your data and research project.
4. **Select Data Type:** Indicate that you are depositing qualitative data. ICPSR accepts various data types, so specifying the data type is helpful.
5. **Describe Your Data in the ICPSR Deposit Form:** Provide detailed information about your qualitative data. This may include information such as the title of your project, a description of the data collection methods, the researchers' names, and funding sources.
6. **Upload Data Files:** Upload your qualitative data files to ICPSR. Ensure that your files are well-organized and labeled appropriately. ICPSR typically accepts data in standard formats, so make sure your data is in a compatible format.
7. **Documentation and Codebooks:** Include comprehensive documentation describing your data collection process, coding schemes, and data preparation or analysis procedures. This documentation is crucial for others who want to use your data.
8. **Access and Sharing Options:** Specify how your data can be accessed and shared. ICPSR offers various options, including restricted access for sensitive data and open access for publicly available data.
9. **Terms and Conditions:** Review and agree to the terms and conditions for data deposition on ICPSR.
10. **Submit Your Deposit:** Once you have completed all the required steps and reviewed your submission, submit your qualitative data deposit to ICPSR.
11. **Review and Approval:** ICPSR staff will review your deposit to ensure it meets their standards and guidelines. They may contact you for clarification or additional information if needed.
12. **Publication and Citation:** If your deposit is approved, ICPSR will publish your data in their repository. You will receive a DOI (Digital Object Identifier) for your dataset, which you can use for your research publications.

13. **Promote Your Data:** Consider promoting your deposited data to the research community to encourage its use and citation.

The details and requirements for depositing qualitative data may evolve, so it's essential to refer to the official ICPSR website or contact ICPSR user support for the most up-to-date information and guidance on the deposit process.

Data Curation at ICPSR

It is difficult to make broad claims about qualitative data curation needs because qualitative data can range from a handful of open-text responses to 1000+ interview transcripts. After an initial review of deposited materials, ICPSR staff will transfer the files to the curation unit. The curation unit is made up of professional data curators and curation supervisors. This staff receives training and supervision to ensure human subject protections remain in place and qualitative data are FAIR (Findable, Accessible, Interoperable, and Reusable; Wilkenson et al., 2016).

ICPSR Curation Steps for Qualitative Data

Curation for qualitative data at ICPSR includes:

- Creating an ICPSR study-level metadata record, including a complete citation to the data for proper attribution, assigning a unique ICPSR study number, overview of the study methodology, and assigning subject terms to aid in data discovery
- Verifying deidentification and/or looking for and masking direct and indirect identifiers
- Ensuring documentation provided by the investigator offers a complete understanding of the study methods and is consistent with the data files
- Entering publications into the ICPSR Data-Related Publications database and linking this to the study record
- Correcting non-ASCII characters and adjusting formatting to improve readability if needed
- Creating a Qualitative Product Suite, a zipped folder containing PDF, RTF, and TXT data files
- Creating an ICPSR README file for each qualitative dataset, which provides contextual information along with relevant processing notes
- Assigning a DOI to provide persistent access to the study archived at ICPSR
- ICPSR quality review before release.

Finding Qualitative Data to Reuse at ICPSR

ICPSR has hundreds of available studies containing qualitative data files, many exclusively qualitative. To find qualitative studies to reuse using the search box on the ICPSR website, we recommend adding the search term [qualitative] along with other keywords into the search box. ICPSR search returns results most relevant to all search terms, so many high-ranking search returns should include qualitative data sources. Searching for keywords of interest and then using the available filters to select “qualitative” for the “Data Type” filter is another strategy for exploring qualitative data. This second option will likely return fewer results but can help you explore part of ICPSR’s archived qualitative data.

ICPSR has some rich qualitative/mixed method studies in the areas of health and higher education, as well as depth in key types of qualitative methodologies that we describe next.

The Qualitative Data Sharing Project Series

The [Qualitative Data Sharing \(QDS\) Project Series](#) was funded by the National Human Genome Research Institute [R01 HG009351; PI James DuBois]. Our QDS team worked to identify and reduce ethical and practical barriers to sharing qualitative research data in health sciences research. The project team developed specific QDS guidelines and pipelines for data de-identification, deposit, and sharing (DuBois).

Additionally, the team developed and tested Qualitative Data Sharing (QuaDS) Software to support qualitative data de-identification. The datasets in this series were de-identified using the QuaDS Software and the project’s QDS guidelines. For more information about the Qualitative Data Sharing Project, visit the [QDS Toolkit](#).

College and Beyond II

College and Beyond II (CBII) is a new data resource for studying the impact and outcomes of higher education. It contains student record and transcript data on bachelor’s-seeking undergraduates enrolled from 2000-2021 at 19 public colleges across seven university systems. In total, CBII contains data on over 1 million students. These data are linked to [qualitative data on course content](#), survey data on long-term life outcomes—including [open-response survey data](#) on meaningful college experiences, National Student Clearinghouse data on educational attainment, and more. College and Beyond II was produced by a team of researchers from the University of Michigan and supported by a grant from The Mellon Foundation.

More Studies with Qualitative Data at ICPSR

Examples of Qualitative Data-Only Studies

- [Identity Formation and Social Problems in Estonia, Ukraine, and Uzbekistan: Focus Group and Oral History Transcripts, 1996-1998 \(ICPSR 36802\)](#)
- [Prostate Cancer Risk in Young Black Men Study, United States, 2015-2016 \(ICPSR 37985\)](#)
- [Generalist-Specialist Palliative Care Social Work Collaboration, Michigan, 2014-2015 \(ICPSR 38000\)](#)
- [Barriers and Facilitators to the Receipt of Treatment for Psychiatric Disturbances following Traumatic Brain Injury, United States, 2015-2016 \(ICPSR 38039\)](#)

Examples of Mixed Methods Studies

- [Teacher Quality Grants Texas, 2012-2014, 2014-2016](#)
- [Young Women Leaders Program \(YWLP\) Longitudinal Follow-up Study, Central Virginia, 2013-2015 \(ICPSR 37360\)](#)
- [Racialized Cues and Support for Justice Reinvestment: A Mixed-Method Study of Public Opinion, Boston, 2016](#)
- [Bullying and Violence on the School Bus: A Mixed-Methods Assessment of Behavioral Management Strategies, United States, 2016-2018](#)

Examples of Ethnographic Studies

- [The Vermont Study on Aid-in-Dying, 2016-2018 \(ICPSR 37209\)](#)
- [Workplace Ethnography \(WE\) Project, 1944-2002](#)
- [The Family Life Project, Phase I, United States, September 2003-January 2008](#)

Additional Resources External to ICPSR

The Qualitative Data Archive

The Qualitative Data Repository (QDR) is a dedicated archive for storing and sharing digital data (and accompanying documentation) generated or collected through qualitative and multi-method research in the social sciences and related disciplines. The QDR website has many resources for qualitative researchers, including a [Guide to Formatting Data](#) and [Organization and Documentation](#).

Qualitative Data Sharing (QDS) Toolkit

The [QDS Toolkit](#) was developed as part of the Qualitative Data Sharing (QDS) project, funded by the NIH National Human Genome Research Institute (R01HG009351). The QDS team developed a toolkit to provide qualitative researchers with resources for preparing and sharing qualitative data and complying with the NIH Data Sharing Policy. The QDS Toolkit offers guidelines and resources for researchers preparing and depositing qualitative data and action ideas about writing data sharing plans, budgets, and making data more visible. Users can also find information about qualitative data sharing software to assist researchers with data de-identification to comply with regulations and maintain context for secondary users.

UK Data Service

The UK Data Service provides access and training to use the UK's largest collection of economic, population, and social research data for teaching, learning, and public benefit. The UK Data Archive offers a learning hub with a [qualitative data subsection](#) and a helpful guide for qualitative researchers on its [Qualitative Data Collection Ingest Processing Procedures](#).

References

- Corden, G., Eady, J., Jang, J., Kirkwood, M., Reneau, K., Tang, S. (2022). ICPSR's Disclosure Risk Guide for Data Depositors. ICPSR <https://dx.doi.org/10.7302/5882>.
- Corti, L., Day, A., & Backhouse, G. (2000). Confidentiality and Informed Consent: Issues for Consideration in the Preservation of and Provision of Access to Qualitative Data Archives. *Forum Qualitative Sozialforschung Forum: Qualitative Social Research*, 1(3). <https://doi.org/10.17169/fqs-1.3.1024>
- Denzin N. K. & Lincoln Y. S. (2018). *The sage handbook of qualitative research* (Fifth). Sage.
- DuBois, J. M., Mozersky, J., Parsons, M., Walsh, H. A., Friedrich, A., & Pienta, A. (2023). Exchanging words: Engaging the challenges of sharing qualitative research data. *Proceedings of the National Academy of Sciences (PNAS)*, 120(43). www.pnas.org/doi/10.1073/pnas.2206981120
- Hinds, J., Murphy, P., & Rivers, D. (2019). A qualitative study of research data sharing and reuse practices in the social sciences. *PLOS ONE*, 14(6), e0219164.
- ICPSR. (2020). *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle* (6th Edition). Inter-university Consortium for Political and Social Research. <https://dx.doi.org/10.7302/3705>
- ICPSR. (2023). *The Collection Development Policy*, ICPSR. <https://www.icpsr.umich.edu/web/pages/datamanagement/policies/colldev.html>
- Library of Congress. 2023. *Sustainability of Digital Formats*. Library of Congress. <https://www.loc.gov/preservation/digital/formats/index.html>
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.
- Morse J. M. (2105). Critical Analysis of Strategies for Determining Rigor in Qualitative Inquiry. *Qual Health Res*, Sep, 25. doi: 10.1177/1049732315588501. Epub 2015 Jul 16. PMID: 26184336.
- NIH. (2023). *Selecting a Data Repository*. NIH. <https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/selecting-a-data-repository#:~:text=Free%20and%20Easy%20Access%3A%20Provides,protection%20of%20other%20sensitive%20data.>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6), e21101.
- Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Yin, R. K. (2017). *Case Study Research and Applications: Design and Methods*. Washington DC: Sage Publications.

Appendix A: How to Document Identifier Replacements in Qualitative Data

ICPSR provides several examples (below) for documenting replacements in qualitative data files.

Example 1:

All direct and indirect identifiers, including text that could identify an individual in combination with other contexts, have been redacted from the data using the following descriptors:

- Participant - redacted as [Participant]
- Interviewer - redacted as [Interviewer]
- Names of individuals - redacted as [Name1], [Name2], [Name3]
- Organizations - redacted as [Org1], [Org2], [Org3]
- Geographic locations - redacted based on the geographic level [City1], [Street1], [Neighborhood1], [State1], [County1]
- Dates/times - redacted as [Day], [Month], [Year], [Time]
- Other information that could re-identify participants - redacted as [Redacted]

Example 2:

Site names have been stripped from the file names and replaced with a letter (Site A, Site B, and Site C) and interviewees have been given a unique ID number. Researchers reviewed all transcripts line by line and removed identifying information that could be used to identify individuals or organizations directly, indirectly, or via Internet search.

Example 3:

Interview respondents' names were replaced with their roles. Role types were collapsed into general groups to avoid re-identification of those in unique roles (e.g., Administrator, Social Services, School Safety). Identifying information within the text was replaced with generic descriptors. For example, school district names were replaced with School Districts. If there were multiple items with the same descriptor in the same interview, they were numbered sequentially in order of appearance (e.g., Student 1, Student 2, Student 3).

Example 4:

The table serves as a key matching identifier in the left column with their replacement text in the right column:

John Doe	NAME 1
Jane Smith	NAME 2

Main Street	LOCATION
Teamsters	ORGANIZATION
etc.	

Appendix B: ICPSR Access Methods for Confidential Data

The vast majority of ICPSR data holdings are public-use files with no restrictions on their access. Sometimes, the protective measures taken to reduce disclosure risk would significantly reduce the research potential of the data. ICPSR provides access to restricted-use versions that protect confidentiality by imposing stringent access requirements in these cases.

Secure Download

For most restricted-use data, ICPSR allows users to request the data via an online application through the ICPSR Data Access Request System (IDARS). Users must sign in to the application system with an ICPSR account or their Facebook or Google passwords. To access ICPSR member-only data, users must be affiliated with a member institution.

The restricted data application requires:

- a. Names, titles, and institutional affiliation of investigators
- b. Description of the proposed research
- c. Information on data formats needed, data storage technology, and data security
- d. Approval for the research project from the Institutional Review Board of the applicant's institution
- e. A signed data use agreement.

Upon completion, requests are reviewed by ICPSR staff. The encrypted data are sent to researchers via a secure link when approved. Please note that ICPSR does not evaluate the scientific merit of the proposed research questions; we merely evaluate the security measures undertaken by the researcher and verify that all the necessary paperwork has been submitted.

Virtual Data Enclave

The virtual data enclave (VDE) provides access to restricted-use data via a virtual machine launched from the researcher's computer but operating on a remote server. The virtual machine is isolated from the user's physical computer, restricting the user from downloading files or parts of files to their physical computer. The virtual machine is also restricted in its external access, preventing users from emailing, copying, or otherwise moving files outside of the secure environment, either accidentally or intentionally. To receive output or other files from the VDE, users must request a disclosure review from ICPSR staff.

Physical Data Enclave

Approximately 50 studies are only accessible for analysis on-site in the physical data enclave at the Perry Building in Ann Arbor, MI. The data in the physical enclave contains highly sensitive personal information collected from, for example, prison inmates, victims of violence, or serious criminal offenders.

Several guidelines are in effect when using the physical enclave:

- a. Investigators cannot bring laptops or other electronic equipment into the enclave.
- b. The enclave has a Windows computer with the Microsoft Office Suite and the SPSS, SAS, and Stata statistical packages. Arrangements must be made in advance for other software.
- c. The computer is not connected to the Internet, and the removable media ports are disabled.
- d. An ICPSR staff member is always present when a researcher uses the enclave. The staff member inspects and approves all material brought into the enclave.
- e. All output, notes, and other material must be submitted for disclosure review before the investigator leaves the enclave.
- f. ICPSR staff will conduct a disclosure review of all files that the investigator wants to use after leaving the enclave.
- g. Approved analysis output will be sent to the researcher electronically.

Delayed Dissemination

In some cases, ICPSR can preserve data under a delayed dissemination model, in which the depositor and ICPSR establish a release date. ICPSR preserves the data until that date and distributes them according to the dissemination plan afterward.

Appendix C: How to Organize Qualitative Files for Deposit

Deposit Materials Organization

Depending on the deposit size and methodological complexity (e.g., multiple sites, samples, methods, time points), depositors are encouraged to organize materials in a manner that denotes the folder's/file's contents to someone outside of the research team. Additional documentation, such as a README that explains file/folder structure, is also encouraged.

Example 1: folder/file structure for a multiple-methods longitudinal study

Deposit1.zip

- Folder: SurveyData
 - Subfolder: Data
 - File: SurveyBaseline.sav
 - File: SurveyFollowup1.sav
 - File: SurveyFollowup2.sav
 - Subfolder: Questionnaires
 - File: QuestionnaireBaseline.pdf
 - File: QuestionnaireFollowup.pdf
 - Subfolder: OtherStudyMaterials
 - File: ChildSurveyAssentForm.pdf
- Folder: InterviewData
 - Subfolder: Data
 - Files: FocusGroup1.txt, FocusGroup2.txt, etc.
 - Files: Teacher1.txt, Teacher2.txt, Teacher3.txt, etc.
 - Files: Student1.txt, Student2.txt, Student3.txt, etc.
 - Subfolder: InterviewGuides
 - File: FocusGroupGuide.pdf
 - File: TeacherGuide.pdf
 - File: StudentGuide.pdf
 - Subfolder: OtherStudyMaterials
 - File: FocusGroupConsentForm.pdf
 - File: TeacherInterviewConsentForm.pdf
 - File: StudentInterviewConsentForm.pdf

Example 2: folder/file structure for a multi-sample, multi-site study

Deposit2.zip

- Folder: Site1
 - Subfolder: SiteVisits
 - Files: Jan2018.docx, Jul2018.docx, etc.
 - Subfolder: StakeholderInterviews
 - Files: Interview1-07012018.docx, Interview2-07142018.docx, etc.
 - Subfolder: ProviderInterviews
 - Files: Interview1-05252018.docx, Interview2-06182018.docx, etc.
 - Subfolder: ClientInterviews
 - Files: Interview1-09092018.docx, Interview2-10112018.docx, etc.
- Folder: Site2
 - Subfolder: SiteVisits
 - Subfolder: StakeholderInterviews
 - Subfolder: ProviderInterviews
 - Subfolder: ClientInterviews
- Folder: Site3
 - Subfolder: SiteVisits
 - Subfolder: StakeholderInterviews
 - Subfolder: ProviderInterviews

Subfolder: ClientInterviews