Giving structure to the "everything else" box: Creating curation standards for qualitative data

Amanda Draft[1], Annie Beaubien[2], Zachary Bennett[3], Julie Eady[4], Aubrey Garman[5], Matthew Johnston[6]

Abstract

To produce high-quality data for future use, curators must be both proactive in setting FAIR (Findability, Accessibility, Interoperability, and Reuse) standards and flexible in allowing curation standards to evolve alongside methodological and technological advances. This is especially pertinent when working with qualitative and mixed-methods data, which can vary immensely in structure and form depending on the study, even when researchers are willing to share these data. Standards at the Inter-university Consortium for Political and Social Research (ICPSR) for curating qualitative data have evolved over time to account for the varying types of data ingested while building consistency in workflows for curators and in the data sharing experience for end users more familiar with quantitative data curation. We first describe the factors we consider when suggesting curation tasks grouped by intensity ('curation levels') specifically for qualitative data. Building on these factors, we then propose our levels of qualitative data curation. Finally, we briefly discuss how technological advances may impact our curation standards in the future.

Keywords

qualitative data, data curation, workflows, standardization, data archiving

Introduction

The terms 'qualitative' and 'standardization' can feel diametrically opposed for those used to working with qualitative and mixed-methods data, as many forms of qualitative data do not easily fit into a one-size-fits-all package. Researchers may also be reluctant or unable to share qualitative data for secondary use for a multitude of reasons, including participant consent, re-identification risk, misinterpretation of data, and resources required to prepare the data for others to use (Mozersky et al. 2021). To produce high-quality data for reuse, curators must be both proactive in setting FAIR (Findability, Accessibility, Interoperability, and Reuse) standards (Wilkinson et al. 2016) for curating qualitative and mixed-methods data and flexible in allowing these standards to evolve alongside qualitative research and technological advances.

The Inter-university Consortium for Political and Social Research (ICPSR) maintains a repository for social sciences (broadly defined) research data at the University of Michigan. As of April 7, 2025, there

are 11,998 studies with curated data[1] hosted on the ICPSR website, with 1,117 results when searching the term 'qualitative'. A search for qualitative as data type on the ICPSR website returns 213 results dating back to 2006. Depositors can provide information during submission (e.g., study documentation, deposit metadata) defining their data files as qualitative, which is then confirmed by project managers prior to curation. It should be noted that these results do not exclusively reflect studies with qualitative methodology, as this query returns any study that includes files released with the prefix 'qda' (referring to qualitative data) in the back-end file version, including long (greater than 244 characters) open-ended string responses from quantitative datasets that were exported to a text-based file format.[2] Broadening the search results to 'qualitative' anywhere in the metadata includes a study released in 1994 (ICPSR 6148) where a collection of in-depth interviews was publicly released as an ASCII text data file, indicating that qualitative data have a history at ICPSR prior to 2006.

The vast majority of ICPSR's holdings represent quantitative data, but the number and type of qualitative and mixed-methods studies released has been steadily increasing in recent years. ICPSR expects that trend to continue as qualitative and mixed-methods research have gained acceptance in the social sciences, improved computational methods make it possible to analyze large volumes of text in ways that were not possible before, and funder requirements for archiving and data sharing often apply to all data types (A.J. Million, personal communication, October 2, 2025). Out of the total 213 results for qualitative research, approximately half have been released since 2021, peaking to 34 results in a year in 2023. As a result, ICPSR's standards for curating qualitative data have needed to evolve over time to account for the varying types of data ingested while building consistent workflows for curators and a consistent data sharing experience for users more acquainted with quantitative data curation.

The purpose of this paper is twofold: to outline the factors that curators at ICPSR consider when curating and releasing qualitative data, and to propose a set of curation levels for qualitative data as an exemplar for others working in data curation and potential data depositors. We specifically focus on text-based qualitative data, as audiovisual and image-based data have their own challenges and requirements beyond the scope of this paper. First, we provide a history of qualitative data curation practices at ICPSR. Next, we describe the factors we consider when suggesting a curation level specifically for qualitative data: data file type, number/length of files, structure of data collection, file organization and documentation, subject matter, and presence of personally identifiable information (PII) or otherwise disclosive material. Then, building on these factors, we propose our levels of qualitative data curation: least intensive (Fast Release and Level 1), moderately intensive (Level 2), and most intensive (Level 3). Finally, we briefly discuss how technological advances such as de-identification software may impact our curation standards in the future.

---

[1] This query excludes search results from openICPSR, a repository for self-publishing data. Data deposited through openICPSR are not curated by ICPSR Curators.

[2] 'QDA' is a broad internal ICPSR filenaming convention that distinguishes qualitative data files from quantitative data files.

Previous qualitative data curation practices at ICPSR

In 2021, ICPSR began formalizing long-standing practices for curating qualitative data into an internal Guide for Qualitative Data Curation. An external-facing version for depositors was published in 2023 (Pienta & Reneau, 2023). Curators create a package of qualitative data files, which consists of the original data files converted to rich text format (RTF), portable document format (PDF) with optical character recognition (OCR) and embedded fonts, and ASCII (American Standard Code for Information Interchange) text copies with the original file name unless modified due to disclosure risk. These are among the file formats recommended by the Library of Congress for preserving digital text (Library of Congress, 2023). These files are part of the product suite, which at ICPSR refers to the zipped package of downloadable data files available in different formats (in this case, the qualitative suite refers to the TXT, RTF, and PDF listed above. For the quantitative suite, this would be txt, tsv, SPSS, SAS, STATA and R files). Files are grouped by type into subfolders, which are then compressed into a zipped package for dissemination. The zipped package also contains a PDF format ICPSR README with a file inventory, processing notes to describe curation processes or identified discrepancies, and other relevant information for users. The Guide for Sharing Qualitative Data at ICPSR offers multiple suggestions on replacing disclosive information in the data, but any changes curators make to the data are marked with a generic descriptor of the replaced information ('[DESCRIPTOR MASKED BY ICPSR]') formatted to distinguish from de-identification performed by the data depositor. Common descriptors include [ADDRESS MASKED BY ICPSR], [NAME…], [JOB TITLE…], [NEIGHBORHOOD…]. If there are instances of multiple names or other descriptors being used multiple times, curators will number these instances in order to lessen confusion and provide additional information to users. For example, if two names are used in a sentence, these would be masked using [NAME 1 MASKED BY ICPSR] and [NAME 2 MASKED BY ICPSR]. If NAME 1 is used again later in the data, [NAME 1 MASKED BY ICPSR] will be used again so that users understand the same person is being referenced. While curators try to retain depositors' original hierarchical file organization, if there is none present or if it is unclear, curators may organize data files based on other logical criteria, such as sample, study content, or mode of data collection. Currently, ICPSR manually edits qualitative data files to perform remediations and logs changes in an internal processing history file.

A review of ICPSR studies with qualitative files released between 2006 and 2021 shows inconsistencies in released files and approaches to disclosure remediation over the years, yet there are building blocks for what would eventually become the formalized standards. Qualitative releases varied in the presence of internal remediation tracking, release of notes for users, file release structure, available file types, and disclosure remediations. A study released in 2007 (ICPSR 20201) contains an internal processing history file noting what changes were performed by ICPSR staff (none in this case, as the depositor performed all redactions), but the data are only available in a single file format (PDF). Another study released that same year (ICPSR 4278) contains the qualitative product suite as described above. PDF files contained completed document properties fields (Title, Author, Keywords/Tags, and Subject) but with minimal enhancements for accessibility and no matching document properties for RTF files. Other studies in this era also released 1 qualitative data file per dataset (e.g., 1 interview = 1 dataset), rather than grouping multiple files into a zipped package. During this period, anonymization programs such as QualAnon (Nicolai et al., 2025) were used occasionally for disclosure remediations, with the anonymization results and master key files linking identifiers

with their generic descriptor replacement archived as internal files only. Otherwise, curators would manually replace disclosive information with phrases such as '[details removed]' or generic descriptors without marking the change had been performed by ICPSR and not the depositor (e.g., '[Name1]').

Studies released in 2009 introduced ICPSR processing notes for qualitative data and redacting disclosive information by replacing text with '[Masked by ICPSR]'. A further evolution of redaction appears in 2017 with the addition of the nuanced descriptors to better match the replaced information (e.g., '[University - Masked by ICPSR]'). Studies in 2011 began zipping all qualitative data files into a single package, though renaming the zipped files to ICPSR naming conventions (qda12345-0001_interview) rather than retaining the original file names. A study from 2018 ([ICPSR 36802](#)) released a README created by ICPSR in PDF format containing a file inventory with processing notes, as well as releasing data files into separate zipped packages based on file type.

One caveat of these inconsistencies is that ICPSR organizational structure influenced curation practices during this time. Prior to 2017, ICPSR Curators were assigned to specific projects, each of which had their own standards and practices. In 2017, curators from across all projects were reorganized into the Curation Unit, with one of the goals being to increase standardization in all projects across ICPSR. In addition, curators who have experience with qualitative data are more likely to be assigned to curate these studies, so the work is being done by the same small group of curators (as of April 2025, the unit has 28 curators and less than half have experience with qualitative data).

ICPSR curation levels

ICPSR introduced curation levels in 2018 as part of the Curation Unit reorganization effort to facilitate standardization across projects. Each curation level has a series of tasks that curators perform to enhance the data and increase usability. These levels were initially referred to as Beta, Full, and Intensive. In 2019, these were renamed to Level 1, Level 2, and Level 3 respectively, with the addition of the Fast Release (ICPSR, 2019). As the vast majority of ICPSR's holdings are quantitative datasets, the curation levels assume quantitative data as the default.

Before curation begins, a thorough review is completed to determine the appropriate level of curation needed, which ranges from Level 1 (least intensive curation) to Level 3 (most intensive curation). In cases where there is a need to minimize curation effort and expedite release (e.g., available funding, external deadlines), the original files are edited only for disclosure remediations and are released in their deposited form inside a zipped package, otherwise called a Fast Release. Level 1 can be considered a baseline for curation tasks. Each study is released with the ICPSR quantitative product suite (ASCII text, TSV, SAS, SPSS, Stata, R), variable-level metadata, ICPSR-generated codebooks containing frequencies and summary statistics, and study-level metadata containing study information. Curators also perform a full disclosure risk review for each study in which the data are assessed on their risk of identifying research participants and the risk of harm that re-identification could pose to them. This disclosure risk review process involves a thorough examination of the entire qualitative text data, assessing the presence of direct or indirect identifiers and contextual information on their risk of revealing the identity of research participants, and mitigating that risk through masking or redacting information. Based on this assessment, curators will remediate the concerns to protect respondent confidentiality. The exact disclosure remediation is dependent on a number of factors

(e.g., curation level, access restriction), but in general, higher curation levels have more remediation options available. Level 2 includes all tasks for Level 1 plus data enhancements, such as adding variable and/or value labels, designating missing values, and creating online data analysis, and open-ended responses can be reviewed for disclosure risk. Level 3 includes all tasks for Level 2 with options for further enhancements, such as adding question text to variable-level metadata and spot-masking disclosive information from text-based variables.

Any non-standard file types are designated as Level 3 due to the need for more curation resources to handle these. Previously, any study with a qualitative component was typically assigned at Level 3, unless projects requested to move forward with curating only the quantitative components. However, the steadily increasing ingest of qualitative and mixed-methods data has led to a tipping point where these data types are common enough that the all-encompassing 'non-standard' Level 3 curation is no longer adequate and a separate category is needed.

Based on multiple factors, including user demand, available resources and staffing for Level 3 curation, the result in practice has been a backlog of qualitative and mixed-methods studies that have not yet been curated and released. To reduce the backlog, options performed so far have been phased releases and reduced curation effort if the data met certain criteria. The National Archive of Criminal Justice Data has often requested a Level 1 release for mixed-methods studies to release the quantitative data files only, with an ICPSR README that provides a description of the qualitative files in order to help gauge user demand. The Washington University in St. Louis (WUSTL)'s Qualitative Data Series requested Fast Release of ASCII text semi-structured interview data that had been cleaned using a de-identifier tool. To better structurally align with the currently existing levels and adding more options for users, ICPSR began to consider curation levels specific to qualitative data, specifically what factors of the data at the file- and study-levels would define less intensive versus more intensive curation.

Factors considered in curating qualitative data

Data file type
ICPSR considers Microsoft Word Documents, rich text documents, plain text files, and PDF files as standard file types for qualitative text-based data. For these data file types, the text is usually structured, standardized, and easily convertible between alternative qualitative data file types for usability and curation, and the software needed to review and curate these data file types are generally available. The ICPSR qualitative data product suite (with TXT, RTF, and PDF versions of the data) can be generated for data provided in these file formats to increase the accessibility of use.

File types in which the data can be unstructured and difficult to convert to a format that is favored for usability and curation provide challenges, such as NVivo workspaces and Excel spreadsheets. When converting these data file types, the original structure of the text may be lost during the conversion process and would need to be recreated. Additionally, the software needed for these specific data file types (e.g., NVivo) may not be widely available. Thus, data file types that require more work to convert warrant a higher curation level.

Number and length of files

The number of qualitative data files included in the deposit is taken into consideration, as well as the length or size of the data file(s). In general, greater text volume will require more time to review for disclosure risk, therefore requiring a higher level of curation. ICPSR is currently in the process of determining a recommended threshold for the number of files and pages for each curation level.

Structure of data collection

Qualitative data can include a variety of data collection methods resulting in different curation needs. Examples of data collection methods can include field notes, open-ended survey responses, interview transcripts (structured, semi-structured, unstructured), collections of social media posts, observation notes, and more.
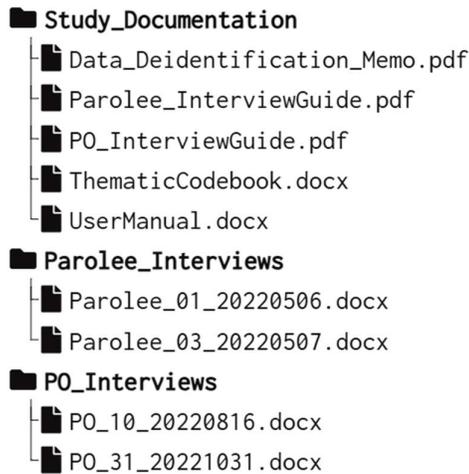
Structured data are often organized and formatted in a standardized file, following a concrete study design and implementation that is consistent in each data file provided, which can aid in the curation process. For example, a collection of interviews in which each question was asked in the same manner and order for every participant or group would qualify as more structured. However, an unstructured qualitative data collection can lead to a higher variance of topics, themes, formats, and a potential for additional disclosure risks. The unstructured format can cause extra complexity in reviewing and demand a higher level of curation.

File organization and documentation

File organization can affect the curation level selection. Curators are better able to review and understand the deposited materials when the files are clearly named and organized appropriately based on the study design. For example, Figure 1 depicts an ideal file organization for a study that includes two study groups, parolees and parole officers (PO). In this example, parolee and PO interview files are organized into two, clearly named subfolders, and each data file follows a consistent and clear file naming structure: study group, followed by respondent ID, and then interview date. If the file organization is not clear, it is possible the study will demand a higher curation level for curators to review the study design and logically reorganize the data for future data users.

The inclusion of detailed study documentation is essential for curation and data re-use. Documents, such as study protocols, user guides, and data collection instruments, allow for a conceptual understanding of the study design and data collection process and can aid in describing how the data files are organized. Other useful documents for understanding the data include codebooks or notes from the data depositor detailing anonymization processes or other important study information. The inclusion of detailed study documentation assists curators in compiling study metadata, conducting the disclosure risk review, and preparing files for dissemination, which make for a more efficient curation process and greater ability to provide a higher level of curation.

Figure 1. Ideal File Organization

```
📁 Study_Documentation
   📄 Data_Deidentification_Memo.pdf
   📄 Parolee_InterviewGuide.pdf
   📄 PO_InterviewGuide.pdf
   📄 ThematicCodebook.docx
   📄 UserManual.docx
📁 Parolee_Interviews
   📄 Parolee_01_20220506.docx
   📄 Parolee_03_20220507.docx
📁 PO_Interviews
   📄 PO_10_20220816.docx
   📄 PO_31_20221031.docx
```

Subject matter

Subject matter of qualitative data is strongly considered during the curation process. If the subject matter is sensitive and re-identification would be potentially harmful to the respondent (e.g., personal political views, criminal history and/or victimization, health history, alcohol/drug use), a higher curation level will be selected to review and address the data for indirect and direct identifying information.

Presence of personally identifiable information (PII) and disclosure risk

The presence of personally identifiable information (PII) is important to resolve in curation. If participants have been promised anonymity or confidentiality, the presence of PII can endanger this. Common PII includes names of study participants, including individuals or organizations, addresses, date of birth, and other identification numbers that can be directly linked to an individual (e.g., Social Security numbers, medical identification numbers, jail booking numbers). However, it is still possible for data to be disclosive without the inclusion of PII due to the presence and combination of indirect information about a respondent (demographic information) and study design (sample size and method, geographic level, etc.). For example, a dataset could include a low sample size, employees of a single company, and job title of each employee, which can have an increased risk of re-identification even without PII. The presence of vulnerable or marginalized populations (e.g., incarcerated individuals, minors, students, individuals with diminished capacity to consent) in the data also warrants closer attention for assessing re-identification and potential harm to participants. Additionally, relationships between individuals directly or indirectly included in the data are often a focal point for collected studies and are considered during curation level selection. Examples of relationships in data can include child-parent, employee-organization, patient-healthcare organization, and parolee-parole officer. Information about other individuals not directly included in data collection are also considered (e.g., teachers answering questions about students).

One of the main factors impacting the time it takes to curate and release qualitative data is the need to carefully review the text for direct and indirect identifiers. Given the rich level of detail often found in qualitative data, reviewing for indirect identifiers that could potentially be used in combination to identify respondents is especially time consuming. If PII is known to be included in the data, a more intensive curation process is needed to provide masking of sensitive information. If the data do not include PII, then a less intensive curation process can be applied. A major factor in curation level selection is if the data depositor can confirm the data were de-identified prior to deposit, either through their manual review and remediation and/or through the use of a de-identification program. If the data were not anonymized by the data depositor, the curation process will be more intensive.

It is important to note that each dataset is unique. The curation level selection process can be heavily influenced by one or several of these factors depending on the state of the deposited data and their accompanying study materials. Since each dataset demands unique curation, ICPSR has created the curation levels to meet these demands.

ICPSR curation levels for qualitative data

Consistent with the current levels of curation for quantitative data, we propose four levels of curation for qualitative data. Each curation level has a series of tasks that curators perform to enhance the data and increase usability. See Table 1 for a summarized version of the different levels.

The goal of Qualitative Fast Release is to provide users with an overview of qualitative data collections by providing access to study-level documentation and a summary of the contents of the qualitative data. Data files would not be included in the fast release. This level of curation provides a mechanism for ICPSR to gauge interest in these data collections and identify which studies should be considered for more intensive curation.[3]

Level 1 and Level 2 qualitative curation are designed to allow access to qualitative data collections through ICPSR's Virtual Data Enclave (VDE), which operates on a remote server and is accessed by researchers via a virtual machine from their own computing equipment. Access to the VDE is limited to vetted secondary researchers and output is reviewed by ICPSR staff before it can be included in a publication. By restricting access to the data through use of the VDE, curators can focus on ensuring that direct identifiers have been removed from the data, while in most cases leaving indirect identifiers as is. This will provide vetted secondary users with quicker access to these data collections. Due to the time-consuming nature of checking for identifying information, Level 1 curation should be considered for smaller collections with less sensitive content and data types such as observation notes or thematic summaries of interviews with minimal direct quotes from individuals. Level 2 curation should be considered for larger collections that may include more sensitive content and data types such as full interview transcripts. For both Level 1 and Level 2 curation, ICPSR will release the files in the same format as they are received.

---

[3] ICPSR has internal tracking mechanisms that can determine the number of unique users downloading specific files. Staff can check these usage statistics on demand and use that information to make decisions on which studies should be considered for more intensive curation.

The goal of Level 3 qualitative curation is to provide the full qualitative product suite, with the caveat that there may be conversion limitations depending on the original file type received. Compared to Levels 1 and 2, which limit access to data via the virtual enclave, Level 3 allows for public or restricted release via secure download if appropriate, which broadens access to the data collection. Curators will carefully review the data for both direct and potentially disclosive indirect identifiers, and remediate any disclosure concerns as appropriate for the requested access level[4].

For all levels of qualitative curation, ICPSR will create study-level metadata and a publicly available overview of the contents of the data collection, including the number and formats of the files. A Digital Object Identifier (DOI) would also be provided for each data collection upon release.

Table 1. ICPSR Qualitative Curation Levels Overview

Curation Level	Ideal for:	Included in ICPSR Release
Fast Release	Providing overview of data, gauging interest level in collection	- Study level documentation only (no data included) - Summary of files in data collection
Level 1	Smaller, fully de-identified collections (e.g., observation notes, summaries)	- Restricted access data - Masked direct identifiers (done by the data producer)
Level 2	Larger collections, partially de-identified data (e.g., interview transcripts)	- Restricted access data - Masked direct identifiers (done by ICPSR)
Level 3	Any type of data, including data on more sensitive topics	- Public or restricted access data - Masked direct or disclosive indirect identifiers (done by ICPSR) - Full product suite

De-identification tools and qualitative data curation

At present, qualitative data curation at ICPSR requires a manual effort to review for and remediate disclosure concerns. Qualitative disclosure risk reviews can be time and cost-intensive as curators

---

[4] ICPSR offers the following access levels for curated studies: public download, restricted secure download, Virtual Data Enclave (VDE), or Physical Data Enclave (PDE).

balance preserving the richness of qualitative data with the need to protect respondent confidentiality. However, evolving technologies in de-identification software aim to meet the increasing demand to curate qualitative data more efficiently. ICPSR curators have explored and tested various tools to aid in qualitative data review, with successes and drawbacks that have the potential to impact qualitative curation standards.

De-identification tools for qualitative data often utilize natural language processing (NLP) to review for a range of text that may pose disclosure risks, such as names of individuals, organizations, or institutions; various geographic indicators; numbers; race, ethnicity, nationality, or indigenous status; or gender and sexual orientation. When disclosive text is identified, some tools can automatically replace it with generalized masking text. For example, replacing individual names found in the data with "[NAME]". Other tools may still rely on human effort to manually approve or reject the suggested disclosure risk remediations. Researchers who utilize such software prior to sending it to ICPSR for curation can save curators time and effort during disclosure risk reviews, which allows for the data to be made available to users more quickly and at a reduced cost. By spending less time conducting disclosure reviews, researchers could have additional curation resources to spend on higher levels of curation that enhances the qualitative data for end users.

Before utilizing these de-identification tools, researchers and curators should be aware of their limitations (Pienta & Redman, 2024). Curators find that these tools miss identifiers beyond what the software is programmed to look for. For example, if a person was referred to by their full name and then a nickname, the software may not flag the nickname as an identifier. Some tools are also programmed specifically with the English language, causing them to overlook non-English language names or text that poses disclosure concerns. Alternatively, these tools may be overzealous in what constitutes a disclosure risk leading to excessive de-identification of important contextual data. These tools may also not allow for nuance in de-identification strategies, such as using the same generalized masking text with all instances of that disclosive category (e.g., all geographic information - country, state, city, street - is replaced with [LOCATION] in some de-identification tools). Privacy concerns should also be considered before using de-identification tools with confidential data to ensure the terms of use will continue to ensure respondent confidentiality. Therefore, while de-identification tools may be an efficient way to begin the curation process, a manual review of the qualitative data by researchers and curators is still strongly recommended to ensure intentional respondent protections are made.

Researchers should continue to inform ICPSR about their process for anonymizing their qualitative data, including the use of any de-identification tools. Using de-identification tools may make higher levels of curation more accessible to researchers. As previously discussed, Level 1 requires that the data types are limited to observation notes or summaries, whereas Levels 2 and 3 can include verbatim transcripts. If researchers can use de-identification software to more efficiently remove identifiers, they may be able to prepare their data for deposit faster or more cost-effectively, allowing curation to prepare and archive more than just documentation. As de-identification tools are developed and improved, ICPSR will continue to assess their utility and effectiveness in the qualitative data curation process.

By describing how ICPSR evaluates qualitative data and proposing these levels of curation effort, our goal is to increase qualitative data sharing and make the depositing process as smooth as possible while maintaining high quality and transparency in our curation standards. As we do with all our workflows, we will continue to evolve as we receive feedback from qualitative data producers and users, or as there are additional breakthroughs or improvements in qualitative data curation technology.

References

ICPSR. (2019, October 17). *ICPSR 101: ICPSR's 3 Levels of Data Curation*. YouTube. https://www.youtube.com/watch?v=i0C26OFYGV8

Library of Congress. *Recommended Formats Statement - Textual Works | Resources (Preservation, Library of Congress)*. (2023). https://www.loc.gov/preservation/resources/rfs/text.html

Mozersky, J., McIntosh, T., Walsh, H. A., Parsons, M. V., Goodman, M., & DuBois, J. M. (2021). *Barriers and facilitators to qualitative data sharing in the United States: A survey of qualitative researchers.* PLOS ONE, 16(12), e0261719. https://doi.org/10.1371/journal.pone.0261719

Nicolai, T., Mozygemba, K., Kretzer, S., Hollstein, B., & Gordeev, E. (2025). QualiAnon - Qualiservice tool for anonymising text data (version 1.5.0). Qualiservice. University of Bremen. https://github.com/pangaea-data-publisher/qualianon

Pienta, A., & Redman, S.M. (2024). *Developing Best Practices for Qualitative Health Data Sharing.* ICPSR. https://dx.doi.org/10.7302/22429

Pienta, A. & Reneau, K. (2023). *Guide for Sharing Qualitative Data at ICPSR.* ICPSR. https://dx.doi.org/10.7302/21539

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., & Gonzalez-Beltran, A. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. Scientific Data, 3(1). https://doi.org/10.1038/sdata.2016.18

[1]**Endnotes**

[1] Amanda Draft is a Lead Data Curator at ICPSR at the University of Michigan. She can be reached by email: adraft@umich.edu.

[2] Annie Beaubien is a Senior Data Curator at ICPSR at the University of Michigan. She can be reached by email: abeaubi@umich.edu.

[3] Zachary Bennett is a Senior Data Curator at ICPSR at the University of Michigan. He can be reached by email: zacharbe@umich.edu.

[4] Julie Eady is a Supervisor at ICPSR at the University of Michigan. She can be reached by email: eadyj@umich.edu.

[5] Aubrey Garman is a Lead Data Curator at ICPSR at the University of Michigan. She can be reached by email: garmana@umich.edu.

[6] Matthew Johnston is a Lead Data Curator at ICPSR at the University of Michigan. He can be reached by email: johnstma@umich.edu.