

Guidance for Preparing and Using Archived Qualitative Data in Research

A Collaboration between Mighty Crow and the Heldrich Center for Workforce Development, Rutgers University

March 2026

Authors¹: Brienne Beaujolais², Victoria Coty³, Andrea Hetling³, Rebecca McCloskey², and Stephanie Walsh³



Introduction

Funders and publishers are increasingly requiring researchers to archive and share data collected and analyzed in their studies. Motivated by the principles of open science, archived data are critical for replication research, as well as new inquiries. Re-using existing data sources saves time and money from the perspective of researchers and funders. Secondary analyses also benefit the research participants and communities being studied by expanding the value of their participation and trust in research beyond the original study. However, because data collection and analyses are intrinsically tied to each other in qualitative research endeavors, secondary analyses of archived qualitative data pose challenges that differ from quantitative data. This brief summarizes methodological challenges and opportunities associated with secondary qualitative analysis, using interview data from the [Health Profession Opportunity Grants \(HPOG\) 2.0 evaluation](#), a federally funded workforce development initiative. Working with the interview data for new analytic purposes required careful attention and extensive time in accessing and analyzing the data, requesting disclosure review, and writing findings, ultimately resulting in research findings that expanded the original purpose of the data.



Drawing on our experiences, this brief shares the challenges we faced, constructive reflection points, and suggestions to further improve best practices in qualitative data archiving and re-use.

Obtaining Access to the Data

As fellow recipients of a federal secondary data research grant, Mighty Crow and the Heldrich Center for Workforce Development, Rutgers University entered into separate Data Use Agreements (DUA) to conduct qualitative analyses of HPOG 2.0 interview data through the ICPSR Virtual Data Enclave (VDE). The VDE is a type of remote access desktop environment that allows access to restricted data in a secure, isolated environment, where users can analyze data without transferring to their local devices. In this context, it was the sole means for research team members to access and analyze the HPOG data. As such, the DUAs included details on data security plans that outlined how each team would work in the VDE. Because DUAs were executed post-award, research proposals were written before data could be accessed and reviewed. Once approvals were in place, the teams embarked on their separate projects. The teams submitted extraction requests to remove data points from the VDE (including participant quotes) for reporting and dissemination. Data could only exit the VDE after analysis completion, requiring a separate approval process and explicit written consent from the data steward. Both research teams successfully completed analyses that furthered the knowledge produced from the HPOG data.

The Mighty Crow project was supported by the Administration for Children and Families (ACF) of the United States (U.S.) Department of Health and Human Services (HHS) as part of a financial assistance award (Award #: 90PE0071-01-00) totaling \$200,000 with 100 percent funded by ACF/HHS. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by ACF/HHS, or the U.S. Government. For more information, please visit the ACF website, <https://www.acf.hhs.gov/administrative-and-national-policy-requirements>.

The Rutgers University project was supported by the Administration for Children and Families (ACF) of the United States (U.S.) Department of Health and Human Services (HHS) as part of a financial assistance award (Award #: 90PE0069-01-00) totaling \$198,400 with 100 percent funded by ACF/HHS. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by ACF/HHS, or the U.S. Government. For more information, please visit the ACF website, <https://www.acf.hhs.gov/administrative-and-national-policy-requirements>.

Our Experience

Data Access and Use Challenges

During implementation, the project teams encountered several data access and operational challenges that affected timelines and workflow.

	Unclear and cumbersome data access processes slowed progress. The requirements for the data security plan led to overlapping and inconsistent requests for information, causing confusion and multiple rounds of communication. Additionally, limited desktops allowed on the server caused system access issues, especially during peak work hours. This disrupted analysis and also prevented team members from accessing audit trails or progress notes, leaving our projects stalled.
	A lack of clarity in communication procedures and data software included in the VDE delayed data access and project timelines. Communication occurred through an online ticketing system, with no designated contact for access questions. After gaining access to the VDE several months post-application, we learned that qualitative analysis was limited to one approved software, requiring the team to learn a new tool.
	Extensive disclosure review process did not align with an already de-identified dataset. Our disclosure requests often required extensive edits in order to have findings extracted from the VDE which created delays and caused confusion about what details could be included in our findings. After each export request, the disclosure review team asked for reorganization, causing delays. Additional quote censoring was needed before export, despite the de-identification. Limited guidance on what to censor complicated the process, causing repeated exchanges with the disclosure review team, review delays, and sometimes the need to reframe findings without the ability to include supporting evidence or quotes.

Analysis Challenges and Missed Opportunities

In qualitative research, data collection and analyses are intrinsically tied to each other. Our teams faced three types of related challenges in analyzing the data, each of which we see as missed opportunities.

	Limited documentation about the research design. Greater transparency and more information on the original research design, including interview procedures and recruitment, interviewer training, interview guide and protocol development, data de-identification, interviewer backgrounds/identities, and site coverage, would help researchers with data analysis and interpretation and increase validity and trust in the findings.
	Incomplete study participant data. Demographic data on the interview participants were very limited, but more details were available in the associated but unlinked quantitative dataset. Linking qualitative data with participant-level survey data or providing additional participant-level demographic information would have strengthened the interpretation of and better contextualized the findings.
	Lack of research process and field notes. Interview transcripts are difficult to analyze without more contextual and process information. The teams could benefit from additional information, including: • Interviewer observation notes/memos on elements such as tone of voice, body language, general affect, and interviewee expression/nonverbal cues. • Interviewer notes/memos on elements such as whether the protocol was followed, the interviewer went off script, they had to skip some questions, etc. • Information on how interview recordings were transcribed, as well as the decisions around what was captured in the transcripts (e.g., participants' body language, pauses, and other speech patterns).

Best Practices for Research Data Management

Our research using the HPOG interviews led to new insights and policy relevant findings related to strengthening health related career training, and we are grateful for the opportunity to have worked with the HPOG interviews. To support and further the re-use of qualitative data, we offer five suggestions for future qualitative project teams.

Prioritize Early Documentation of Research Plan and Design	 Build archiving and de-identification plans into the original study design and project budget. Include input from multiple experts and team members, including interviewers, research analysts, IT staff, and data archivists.
Facilitate Data Access	 Streamline data use agreements. Allow for secure but more flexible storage (including approved cloud-based platforms), potentially eliminating the need for a VDE.
Improve Documentation of Interview Procedures and Metadata	 Include detailed metadata, such as interviewer characteristics, training procedures, site descriptions, and interview timing, to support secondary users' understanding of the dataset.
Link Datasets and/or Provide Demographic Data	 Provide mechanisms to connect qualitative interviews with demographic survey data or site-level implementation data to allow for mixed-methods analysis.
Ensure Coherent Procedures for Confidentiality and Data De-identification	 Align de-identification methods, data access controls, and post-analysis disclosure review so they work together, rather than duplicating efforts. When disclosure review is required, provide explicit guidance on what constitutes sensitive information.

End Notes

¹ All authors contributed equally to this work.

² Mighty Crow – Columbus, OH; for information about the Mighty Crow project, please contact Rebecca McCloskey, rebecca@mightycrow.com

³ Heldrich Center for Workforce Development, Rutgers University – New Brunswick, NJ; for information about the Rutgers project, please contact Andrea Hetling, ahetling@rutgers.edu