Organizing and Sharing Qualitative Data via the Qualitative Data Repository (QDR)

Dessi Kirilova
Johns Hopkins Love Data Week
February 14, 2024
What Is QDR?

- Online since 2014: qdr.syr.edu, NSF funded
- QDR curates, stores, preserves, publishes, and enables the download of digital data generated through qualitative and multi-method research in the social sciences.
- HQ at Syracuse; other team members at UW Seattle
- International and interdisciplinary
- Currently over 180 data projects published
- Over 30 institutional members
- Continuously CoreTrustSeal-certified
Special Collections

The Trachtenberg Papers

This data collection - The Trachtenberg Papers - broadly concerns Cold War policy from the end of WWII to 1968. The data was accumulated in order to write several books and articles relating to Cold War relations during this period, most notably A Constructed Peace: The Making of the Korean Settlement, 1951-1953 (Princeton: Princeton University Press, February 1999).

It includes five data projects. The three Cold War Documents projects contain archival data, whereas the remainder contain the appendices to A Constructed Peace and a broader collection of pedagogical materials.

Search this collection...  
Advanced Search

Security Studies

Collection for supplementary data for the Security Studies journal. By agreement between Security Studies and the Qualitative Data Repository, data deposited to supplement articles accepted for publication in Security Studies do not accrue deposits fees.

Search this collection...  
Advanced Search
A Changing Landscape: Funder Expectations for Sharing Data

• NSF: 2023 update to **Open Access & Data Sharing policy**
  • Existing Data Management Plan (DMP) requirement since 2011; renamed to Data Management and Sharing Plans (DMSP) to better align with intent and other funders’ nomenclature
  • DMP reviewed by grant panel
  • Anecdotally, program officers increasingly insistent to address data sharing
  • Little accountability for DMP content, but new policy suggests change

• NIH: **New DMSP requirement** for *all* funded research data began in 2023
  • Strong data sharing requirements
  • Updating of DMSP possible with interim and final grant reports
  • Implementation of DMSP – a consideration for future support applications
Uses of Shared Data – the “why”

- Secondary analysis and collaboration
- Transparency for specific publications
  - Replication
  - Reproducibility
  - Verification
  - Traceability of methodological choices
- Giving back to participants / communities
  - Direct sharing back to individuals
  - Not over-researching same communities
- Methodological innovation
- Teaching and training

“Sharing scientific data accelerates biomedical research discovery, enhances research rigor and reproducibility, provides accessibility to high-value datasets, and promotes data reuse for future research studies.

Ultimately, the sharing of scientific data expedites the translation of research results into knowledge, products, and procedures to improve human health.”

“This study is a critical program evaluation of two court-affiliated diversion programs with data collected by Corey Shdaimah from the University of Maryland School of Social Work. Shdaimah conducted a longitudinal study examining the perspectives of women exiting prostitution (...) through open-ended interviews with the participants as well as significant stakeholders (Shdaimah 2020).

I will be utilizing secondary data, her interview transcripts, of program participants from Specialized Prostitution Diversion program (SPD) in Baltimore City and Project Dawn Court program (PDC) in Philadelphia.”
“We suggest that the field can benefit from new and practical examples that allow researchers to compare different qualitative approaches, how they can be used in applied settings, and the unique lenses they bring. In order to achieve this goal, we have asked each author in this special issue to analyze one shared data set from a study exploring “postnatal care referral behavior by traditional birth attendants (TBAs) in Nigeria” (Chukwuma, Mbachu, Cohen, Bossert, et al., 2017). The shared data set, which was secured from Syracuse University’s Qualitative Data Repository, consists of transcripts from three focus groups—one with hospital health care workers, another with traditional birth attendants, and a third with TBA delivery clients (Chukwuma, Mbachu, Cohen, McConnell, et al., 2017). Contributing authors—who work within a range of disciplines—were encouraged to focus less on the substantive findings of factors affecting TBAs’ postnatal referral behavior, and instead aim to unearth the analytical contributions and insights that their approach might contribute to a similar data set.”
Re-use in Practice: Teaching

From Alastair Tomlinson, Cardiff Metropolitan University, UK:

“I teach a postgraduate research methods course for approximately 130 students per year, across a range of health-related degree subjects. We have used the QDR to enable students to develop their understanding of qualitative methods, and to demonstrate their qualitative data analysis skills. Students register with the QDR and access a specified research study. As part of their course assessment, students have to critique the qualitative design and methods used in the research study (relevance and appropriateness of design methodology to research question, evidence of rigour in sampling and data collection methods). They then access a subset of interview transcripts from the study and conduct a qualitative thematic analysis using the approach recommended by Braun & Clarke (2013). They produce a short research report summarising their evaluation and analysis, presenting the key themes that they have developed from their data analysis, and draw a brief conclusion. The QDR has been invaluable in enabling our students to work with real-world qualitative materials relevant to their subject area – and raising their awareness of the broader issues about making qualitative research data available to other researchers, and the appropriate access controls that need to be applied.”
Why Share (Qualitative) Data?

• **Be a good colleague:** Allow others to scrutinize and build on your work
• **Be a good teacher:** Students benefit from working with “real” data
• **Be a good citizen:** Open science is more impactful science
• **Get funded and published:** Scientific funders and publishers increasingly expect data sharing and research transparency
• **Get cited:** Lots of empirical evidence across different disciplines that publications whose data are shared, do get more visibility and attention
Data Management and Planning – the “how”

- Think about data sharing early on and holistically – fine-grained choices
- Consider data security and access: Both during project and after
- Plan & organize early: Folder names and structure, file names
- Create documentation throughout the project
Benefits of Sharing Data in a Repository – the “where”

- Stable links (Digital Object Identifiers – DOIs)
- Long-term digital preservation (incl. keeping formats up to date)
- Meeting institutional requirements
- Data management planning and curation assistance
  - can help you with sharing data well
- Makes data more visible/easier for others to discover, access, cite
- Interoperability across disciplines
- Custom solutions for sensitive data (access controls, use agreements, embargoes, etc.)

→ FAIR Data
Shared Qualitative Data: Some Examples
<table>
<thead>
<tr>
<th>Date of Interview: 22Aug14</th>
<th>Start Time: 12:10 pm</th>
<th>End Time: 12:50 pm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewer’s Name: TJ</td>
<td>Location: An NGO clinic, Dushanba, Dushanbada (2)</td>
<td></td>
</tr>
<tr>
<td>Respondent’s Name: Dr. Numinal Islam (Pseudonym)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of Respondent: Private (NGO) Sector</td>
<td>Number: 2</td>
<td></td>
</tr>
<tr>
<td>Gender of Respondent: Male</td>
<td>Year of Graduation: [1993-1998]</td>
<td></td>
</tr>
<tr>
<td>Position/Rank: MO: [Health Clinic]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Partial Sharing of Full Transcripts

Joarder, Taufique. 2017. "Understanding and measuring responsiveness of physicians in rural Bangladesh", Qualitative Data Repository. [https://doi.org/10.5064/F6P55KNC](https://doi.org/10.5064/F6P55KNC). QDR Main Collection. V2
Hitt, Matthew. 2019. "Replication Data for: Inconsistency and Indecision in the United States Supreme Court". Qualitative Data Repository. https://doi.org/10.5064/F6W7QRSX
Leonard, Don. 2019. "Data for: Exposure to trade and postcolonial divergence on the island of Hispaniola". Qualitative Data Repository.
https://doi.org/10.5064/F6MFN8D4. QDR Main Collection. V1
De-identified, Restricted Interviews & Documentation

CShaimah Keisha Interview 1, Copy 2 12-4-12

INTERVIEWER: ...cause it would be sad for both of us. Okay. I'm gonna put this just a little closer.

RESPONDENT: Okay.

INTERVIEWER: But you don't have to worry about speaking in it, it takes a...

RESPONDENT: Okay.

INTERVIEWER: So thank you again.

RESPONDENT: You're welcome.

INTERVIEWER: And I'm just gonna start off, I have just like a few demographic questions, and then it'll be more like a conversation.

RESPONDENT: Okay.

INTERVIEWER: Okay. So how old are you?

RESPONDENT: I'm [30-40].

QDR’s Take on Responsible Sharing of Human Participant Data
Data Sharing and Ethical Concerns

- Research must be designed, reviewed, and undertaken to ensure integrity, quality, and transparency
- Any promises for confidentiality and/or anonymity of respondents must be respected
- Harm to participants must be avoided in all instances
- Any conflicts of interest must be explicit
- Participation is voluntarily, free of any coercion, and informed
DMP-IRB Nexus

• IRB: required for human subject research based on federal regulations; typically based at an institution

• DMP: a document required by funders; typically written by PI and not closely monitored by institution

• Both require researchers to document data collection, sharing and security details
  • It is critical for PIs to ensure that the two documents align
Informed Consent Considerations

- Balance: being transparent about data use, but remain intelligible
- Participants are often willing to help science broadly, not just individual researcher
- Opt-in consent for data sharing can be great; IRBs familiar with it as “tiered consent”
  - Careful with quantitative data & opt-in
**Data Sharing in Informed Consent: Example**

Alicia VandeVusse and Jennifer Mueller, Guttmacher Institute

**Potential for Data Sharing:** If you agree, the transcript of your interview may be shared with researchers at other organizations in the future. We will take out or change any information that could identify you before sharing. You can be in the study whether you agree to data sharing or not (see *Optional Consent* below).

Then after the consent to participate, optional data-sharing consent included:

Do you agree to allow a written copy of your interview to be shared with other researchers in the future?

- Yes
- No

In a qualitative study on abortion using this consent script, 92% of respondents opted into data sharing.

De-identifying Qualitative Data

• Removing / replacing information in text can distort data, make them unusable, unreliable or misleading: A balance to preserve context
• Remove direct identifiers, or replace with pseudonyms – often not essential research info
• Avoid blanking out; use pseudonyms or replacements [FLAG REPLACEMENTS]
• Plan and apply de-identification at time of transcription
• Consistency within research team / project
  • Keep de-identification log of replacements or removals made; keep separate from the processed data files
Entrevistador: ¿Y en qué barrio pensás, digamos, cuando, vos en qué barrio..?

Entrevistada: Bueno, yo soy BARRIO 1 y BARRIO 2. BARRIO 2 y BARRIO 1.

Entrevistador: Y, digamos, si tuvieras que, de esos referentes que conocés de diferentes partidos, dar un número, ¿te animás a dar un número, cuántos son?

Entrevistada: Y, son unos cuantos, son muchos eh, yo entiendo que son más de cincuenta

Entrevistador: Bueno, cincuenta me parece...

Entrevistada: Te digo, por ejemplo, en MUNICIPIO 1 hay [detalla el número] bibliotecas populares, reconocidas por la CONABIP, que es una institución nacional que las agrupa y ....

Dunning, Thad; Camp, Edwin. 2015. “Brokers, voters, and clientelism: The puzzle of distributive politics”. Qualitative Data Repository. https://doi.org/10.5064/F6Z60KZB. QDR Main Collection. V1
Controlling Access

“As open as possible, as closed as necessary” (European Union)
- Default to open data (license “CC0” or “CC-BY”) for documentation
- Public-use data
- Timed Embargo (in 1, 3, 10, 100 years) – NB: not a publication-based embargo
- Access by application
  - Identity and affiliation checks
  - Specific training requirements (e.g., CITI)
  - Signed DAUAs
  - Secondary IRB (appropriate for data that still has PII)
- Access using enclaves (not currently at QDR, but available for quantitative data at other repositories)
QDR for JHU’s Researchers and Students
Johns Hopkins University's Institutional Membership with QDR

**QDR**
The Qualitative Data Repository

- Free curation & storage for typical qualitative and multi-method projects
- Priority DMP consultations
- Recommended by and compliant with federal and private funders
- Strategies and tools for sensitive data
- Virtual class visits (methods courses; graduate research seminars)
Johns Hopkins University’s Local Data Service Resources

JHU’s Dataverse instance (for open-access data): https://archive.data.jhu.edu/

JHU Libguide on “Protecting Human Subjects Identifiers”: https://guides.library.jhu.edu/protecting_identifiers

Guidance on data sharing from 3 IRBs: General JHU IRB Expectations for Sharing of Individual Level Research Data

Trainings on de-identifying data with Dave Fearon (for JHU affiliates only)

- PROTECTING HUMAN SUBJECT DATA PRIVACY: AN INTRODUCTION: https://jhu.libcal.com/event/11787246
- PROTECTING HUMAN SUBJECT DATA PRIVACY PART II: DE-IDENTIFICATION: https://jhu.libcal.com/event/11787268

https://dataservices.library.jhu.edu/ dataservices@jhu.edu
Consultations at Different Points in a Project Lifecycle

When planning research
• DMP / DMSP consultations and review for grant proposal
• Informed consent review and advice
• Data management consultations during project

When depositing data
• Disclosure risk review
• Data organization and documentation consultation
• Data/file transformation

Research Data Lifecycle by [LMA Research Data Management Working Group](link)
Depositing Data in QDR

- Initial Consultation: Possible anytime, best at project start
- Initial Deposit & Curation
- Final Review
- Publication & Promotion
- Can be qualitative only or mixed methods