Organizing and Archiving Qualitative Data with the Qualitative Data Repository (QDR)

Dessi Kirilova
Senior Curation Specialist
June 16, 2023
Open Scholarship—has the goal of making research / education accessible, reproducible, and freely available
OS practices— a smorgasbord of options

- Preregistration
- Replication
- Verification
- Sharing of data
- Sharing of code
- Methodological appendices & documentation
- Alternative peer review models
- Open-access publications
- Meta research
- Annotations (e.g., Annotation for Transparent Inquiry)
Data Sharing Factors

- Funder and publisher requirements → *enforcers*
  - Multiplying benefits of resources spent
  - Transparency and better evaluation of publications
- Technological and infrastructure advances → *enablers*
  - A lot of contemporary data are born digital
  - Dedicated research data services and training at university libraries
  - Professional repositories (domain and institutional) develop processes and tools to store, make findable, preserve for the long run
- Researchers → *implementers*
  - Expectations from academia and society (incl. study participants)
  - Internal motivations – show the rigor and richness of their work; do good science; engage with potential collaborators
Myriad Repositories, Multiple Types

- Self-deposit generalist repositories
- Institutional repositories (affiliated with a university)
- Domain / disciplinary repositories
- Hybrid (e.g., Dryad)
- Funder-supported specialized repositories
  - E.g., over 100 NIH-supported ones
What Is QDR?

- Online since 2014: qdr.syr.edu
- Primarily grants-funded (currently: NSF, Sloan; previously RWJF), but increasingly IM and individual curation fees (including commitments in new grant proposals)
- 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 Harvard and UW Seattle
- Originated in political science: today serves an international and interdisciplinary community
- Currently 160 data projects published
- Over 30 institutional members
- Continuously CoreTrustSeal-certified
A Changing Landscape: Funder Expectations for Sharing Data

- NSF: new Open Access & Data Sharing policy just landed
  - 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
- Teaching and training
- Giving back to participants / communities
  - Direct sharing back to individuals
  - Not over-researching same communities
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.”
“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.”
Data Management and Planning - the “how”

- Think about data sharing early on and holistically
- 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
- 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
- Access controls, embargoes, etc.

→ FAIR Data
Shared Qualitative Data: Some Examples
Mixed Methods Data: Interviews, Focus Groups, Ecological Surveys and Seed Maps

Documentation

- Qualitative Data
  - Farmer interview transcripts
    - Burkina Faso
    - Mali
    - Niger
  - Village focus group seed maps
  - Village focus group transcripts

Tabular data

- Jones_Tabular_FarmerCharacteristics.csv (75.1 KB)
- Jones_Tabular_RainfallByRegion.tab (15.0 KB)

Jones, Kristal. 2020. "Seed systems in West Africa". Qualitative Data Repository. [https://doi.org/10.5064/F6URYY1](https://doi.org/10.5064/F6URYY1) QDR Main Collection.
Professional Community Research coded excerpts

<table>
<thead>
<tr>
<th>CODE</th>
<th>SUBCODES</th>
<th>DEFINITION</th>
<th>EXAMPLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>aerial observation</td>
<td>Use of drones, satellites, or other aerial technology to make ground or land surface observations</td>
<td>&quot;Soil moisture monitoring is really complicated. So I think it would make a lot of sense to use more remote sensors.&quot;</td>
</tr>
<tr>
<td></td>
<td>blockchain</td>
<td>Referring to the method of recording transitions on a transparent and decentralized ledger</td>
<td>&quot;You can guarantee that each credit is not double counted, I mean that’s the main purpose of blockchain.&quot;</td>
</tr>
<tr>
<td></td>
<td>CDSS</td>
<td>The Colorado Decision Support System</td>
<td>&quot;I'm being able to take the CDSS model and put it into our scenarios and really being able to use that to look at what our long-term satisfaction would be.&quot;</td>
</tr>
<tr>
<td></td>
<td>forecasting</td>
<td>Related to predicting a quantity of interest at future dates</td>
<td>&quot;We have all of this data, how do we either distill it, or how do we make it available so that people with different needs can use this data in their work.&quot;</td>
</tr>
<tr>
<td></td>
<td>in situ</td>
<td>Related to data collected on site</td>
<td>&quot;The capacity and the ability to predict that pumping on almost a real-time basis, it’s important to me.&quot;</td>
</tr>
<tr>
<td></td>
<td>irrigation</td>
<td>Related to water used for agricultural or grass</td>
<td>&quot;Deficit irrigation, where you’re essentially measuring the amount of water that’s going on the field and you’re trying to get the irrigation efficiency up.&quot;</td>
</tr>
<tr>
<td></td>
<td>metering</td>
<td>Relating to the measurement of water through the use of a gauge or meter</td>
<td>&quot;I want flow meters on everything we’ve got. So when someone comes to me and says, ‘Hey, you don’t know what you want to do’ I say, ‘Well, let’s put it on the meter and see how it does.’&quot;</td>
</tr>
<tr>
<td></td>
<td>modeling</td>
<td>Representation of a complex system, including hydrologic, numerical, or land-surface models</td>
<td>&quot;Any model is only as good as what you put into it.&quot;</td>
</tr>
<tr>
<td></td>
<td>SnoTeL</td>
<td>Relating directly to SnoTeL sites or snow data from SnoTeL</td>
<td>&quot;There’s like the SnoTeL network, which is by the NRCS. There’s a snow course network. And there’s the satellite...&quot;</td>
</tr>
<tr>
<td></td>
<td>telemetry</td>
<td>Relating to transmission of data, often specifically relating to in situ data</td>
<td>&quot;But I don’t put on a telemetry system where I could go on my computer and pull up and see a running volumetric...&quot;</td>
</tr>
</tbody>
</table>

De-identified, Restricted Interviews & Documentation


---

First Interview for Program Participants

A member of the research team will conduct the interview within first week of program acceptance. She will thank the participant for taking time to speak with her and review the letter of explanation. If the participant agrees to continue, the interviewer will stress that there are no right or wrong answers, and that she is interested in the participant's experiences with the Specialized Diversion Program/Project Down Town. The interviewer will remind the participant that she should not provide any identifying information in the course of the interview.

Background

I am going to start with a few demographic questions:
- What is your age?
- What grade did you go to in school?
- What do you consider to be your race or ethnicity?
- What neighborhood or community do you come from in Baltimore?

Motivations

Researchers say that we don’t really know enough about why women engage in prostitution. We think one of the reasons is that not many researchers talk to women. We hope it is okay to ask you these questions, but please feel free to refuse to answer them.
- What was the major reason that you started to engage in prostitution? How old were you?
- Are there any other reasons?
- Do you still engage in prostitution sometimes? If so, is this for the same reasons or other reasons?
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 PI 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 [IDENTIFY REPLACEMENTS / REDACTIONS]
• 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
“As open as possible, as closed as necessary” (European Union)

- Default to open data (license “CC0” or “CC-BY”)
- 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’s Deposit Process and Related Support Services
Consultations at Different Points in a Project Lifecycle

When planning research
• DMP / DMSP consultations and review for grant proposals
• Informed consent review and advice
• Data management consultations to plan for project

When depositing data
• Disclosure risk review
• Data organization and documentation consultation
• Data/file transformations
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
Questions? Comments? Please Stay in Touch!

Web: https://qdr.syr.edu

@qdr@mstdn.science

@qdrepository

Blog: https://qdr.syr.edu/qdr-blog

Email: qdr@syr.edu  Dessi.Kirilova@syr.edu