Organizing and Archiving Qualitative Data with the Qualitative Data Repository (QDR)
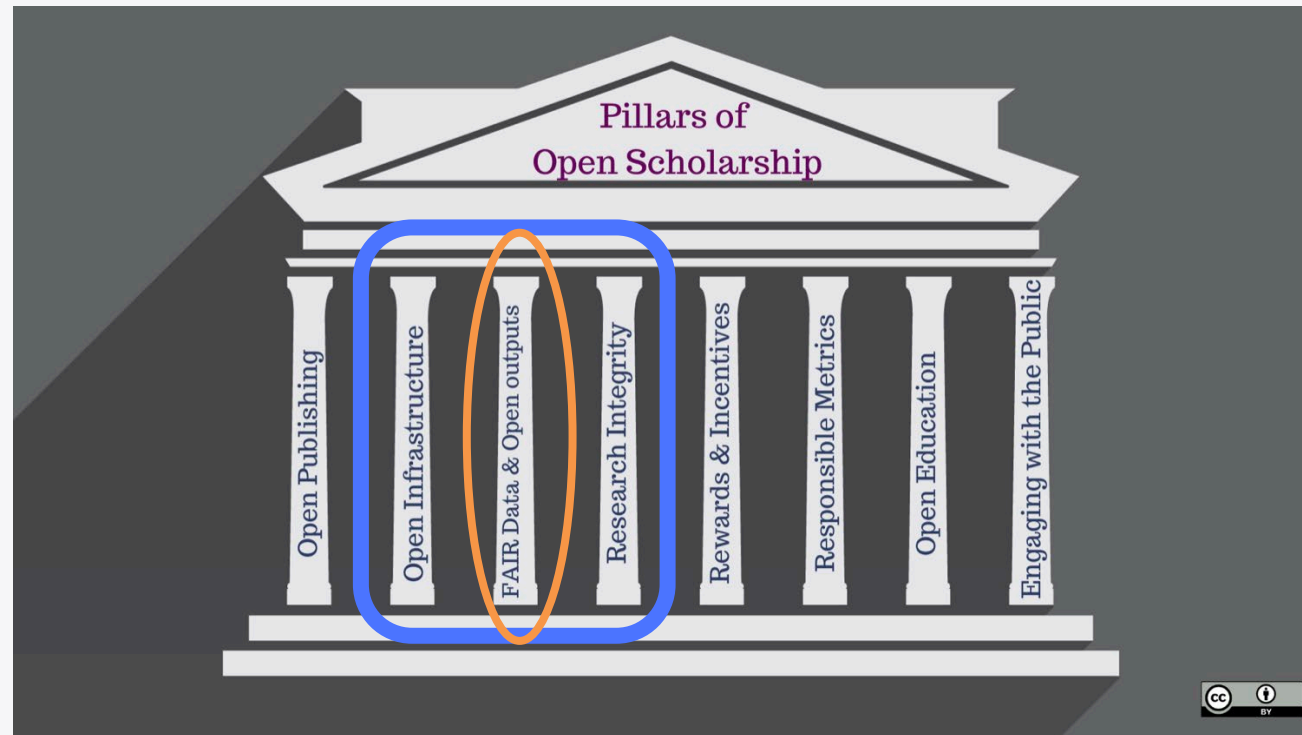
Dessi Kirilova
Senior Curation Specialist
June 16, 2023

Open Scholarship – as a theoretical approach

Open scholarship has the goal of making research / education accessible, reproducible, and freely available
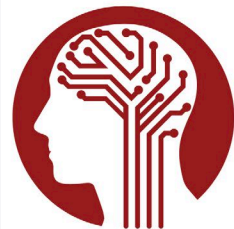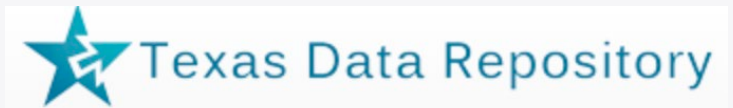
OS practices – a smorgasbord of options

- 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

zenodo

Texas Data Repository

ICPSR 60 years

QDR
The Qualitative Data Repository

DESIGNSAFE

DRYAD

figshare   Browse   Search on figshare...   Log in   Sign up

NIH

need help?   + Follow this search

3,139 results found   sort by: Relevance

DATASET   DATASET   DATASET   DATASET

Biomedical Inequality SQLite database
Dataset posted on 11.07.2017
Yarden Katz

NIH Category Schema
Dataset posted on 22.11.2019
in figshare
Team Figshare

iCite Database Snapshot 2021-07
Dataset posted on 12.08.2021
in The NIH Figshare Archive
iCite

iCite Database Snapshot 2020-05
Dataset posted on 12.06.2020
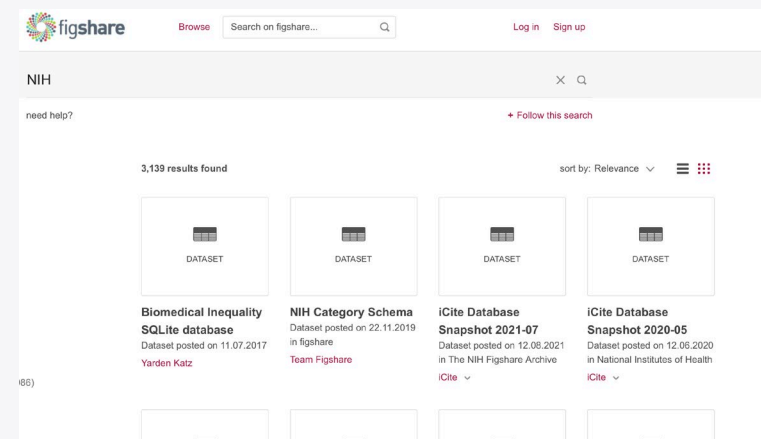in National Institutes of Health
iCite

Welcome to the NIMH Data Archive

The National Institute of Mental Health Data Archive (NDA) makes available human subjects data collected from hundreds of research projects across many scientific domains. NDA provides infrastructure for sharing research data, tools, methods, and analyses enabling collaborative science and discovery. De-identified human subjects data, harmonized to a common standard, are available to qualified researchers. Summary data are available to all.

The NDA mission is to accelerate scientific research and discovery through data sharing, data harmonization, and the reporting of research results.

NIMH common data elements now available: Go to NIMH Common Data Elements

What Is QDR?

- C[...]4qdr.sy[...] funded ([...] curatio[...] (increasingly proposals)
  - [...]tores, [...]gital da[...]esearc[...]
- [...]ther tea[...]
- C[...]tical sci[...] interdisciplinary commu[...]
- 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

Reuse in practice: Graduate thesis

WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Graduate Theses, Dissertations, and Problem Reports

2022

Barriers and Opportunities to Exiting Prostitution: An Analysis of
Prostitution Diversion Programs

Marina Binti Mohd Hamdan
West Virginia University, mbm0031@mix.wvu.edu

https://researchrepository.wvu.edu/etd/11238

"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."

Re-use in practice: Developing & contrasting methodologies

Introduction to Special Issue: Diverse Approaches to Qualitative Data Analysis for Applied Research

Jessica N. Lester
Indiana University, jnlester@indiana.edu

Noah Goodman
EDC's Center for Children & Technology

Michelle O'Reilly
University of Leicester & Leicestershire Partnership NHS Trust, mjo14@leicester.ac.uk

"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 in 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."

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."

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

📄 **README_Jones.txt** (39.1 KB)

▸ 📁 **Documentation**

▾ 📁 **Qualitative Data**

  ▾ 📁 **Farmer interview transcripts**

    ▸ 📁 **Burkina Faso**
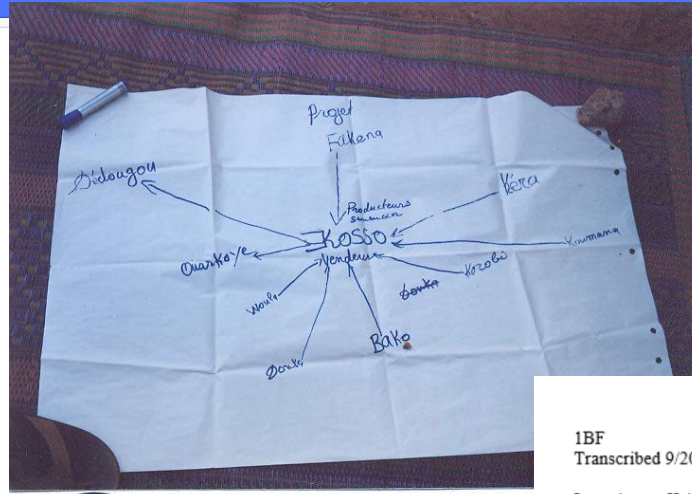
    ▸ 📁 **Mali**

    ▸ 📁 **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 IQDR Main Collection.

Filename: Jones_Tabular_RainfallByRegion.tab

In *Seed systems in West Africa* (version 1.0), by Jones, Kristal

[Download File] [Close Preview]

	2010	2011	2012	Farmer ID	Region
1	1104	920	1196	1	Dioila
2	1104	920	1196	2	Dioila
3	1104	920	1196	3	Dioila
4	1104	920	1196	4	Dioila
5	1104	920	1196	5	Dioila
6	1104	920	1196	6	Dioila
7	1104	920	1196	7	Dioila
8	1104	920	1196	8	Dioila
9	1104	920	1196	9	Dioila
10	1104	920	1196	10	Dioila
11	1104	920	1196	11	Dioila
12	1104	920	1196	12	Dioila
13	1104	920	1196	13	Dioila
14	1104	920	1196	14	Dioila
15	1104	920	1196	15	Dioila
16	1012	828	1196	16	Siby
17	1012	828	1196	17	Siby
18	1012	828	1196	18	Siby
19	1012	828	1196	19	Siby

1BF
Transcribed 9/20/11

Interviewer: Kristal

R: My French isn't very good.

I: No, no, it's fine, it's fine.

R: I didn't go to school. I only did three years.

Background: That's why his French isn't very good.

R: It's mixed up with Bomu. It's not good.

B:...if it's not good, we can translate again.

R: That's already happened.

I: (laughs) Ok, and you are Bomu?

R: Yes, Bobo.

I: Bobo, ok. And you are how old?

R: I was born in 1956.

I: 56. Ah, that's just like my mom! That's 54, no, 55. This year.

B: 55 years old.

I: And you speak Bomu, French?

Professional Community Research–coded excerpts

CODE	SUBCODES	DEFINITION	EXAMPLE 1
Technology	aerial observation	Use of drones, satellites, or other aerial technology to make ground or land surface observations	"Soil moisture monitoring is really complicated. So I think it would make a lot of sense to use more remote sen
	blockchain	Referring to the method of recording transitions on a transparent and decentralized ledger	"You can guarantee that each credit is not double counted, I mean that's the main purpose of blockchain."
	CDSS	The Colorado Decision Support System	"Being able to take CDSS model and putting it into our scenarios and really being able to use that to look at wh
	data visualization	Methods of data visualization include graphics	"We have all of this data, how do we either distill it, or how do we make it available so that people with differe
	forecasting	Related to predicting a quantity of interest at future dates	"The capacity and the ability to predict that pumping on almost a real-time basis, it's important to me."
	in situ	Relating to data collected on site	"Installing the in situ networks, like ground-based towers and stations."
	irrigation	Relating to water used for agricultural or grass	"Deficit irrigation, where you're essentially measuring the amount of water that's going on the field and you're
	metering	Relating to the measurement of water through the use of a gauge or meter	"I want flow meters on everything we've got. So when someone comes to me and says, 'Hey, you don't know v
	modeling	Representation of a complex system, including hydrologic, numerical, or land-surface models	"Any model is only as good as what you put into it."
	Snotel	Relating directly to Snotel sites or snow data from Snotel	"There's like the Snotel network, which is by the NRCS. There's a snow course network. And there's the satellit
	telemetry	Relating to transmission of data, often specifically relating to in situ data	"But put it on a telemetry system where I could go on my computer and pull up and see a running volumetric tc

Holland, Melanie; Demaree, Kat; Thomas, Evan. 2023. "Data for: Investigating technology opportunities toward improved Colorado water monitoring: Insights from key informant interviews and stakeholder surveys". Qualitative Data Repository. https://doi.org/10.5064/F6TFIYHX QDR Main Collection. V2
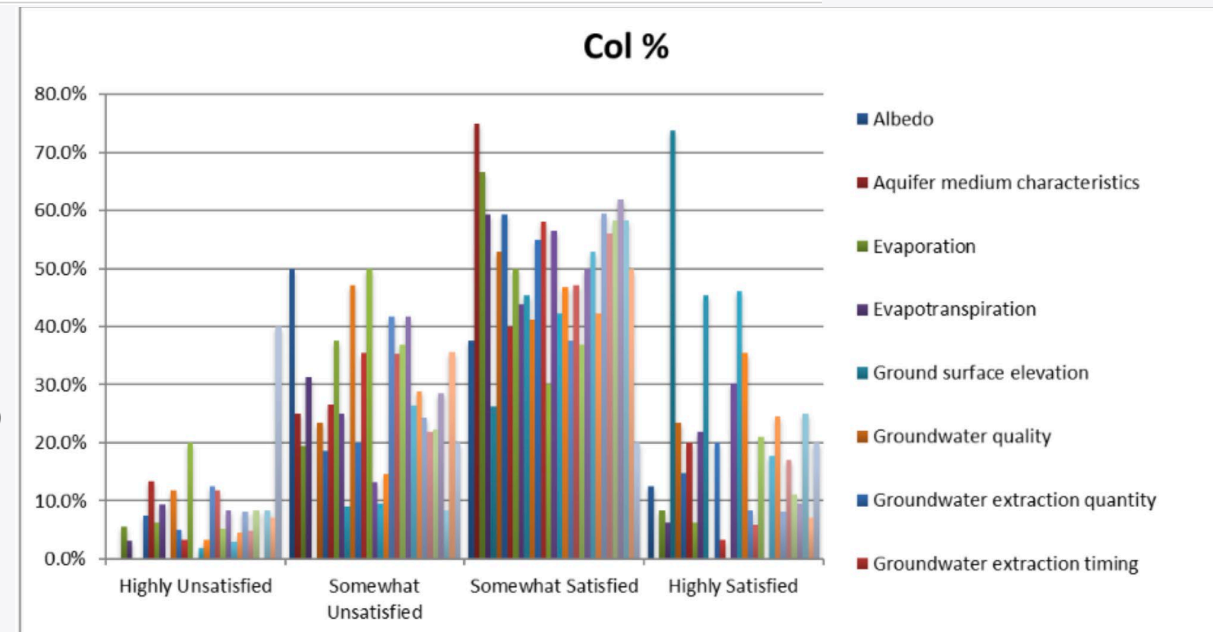
De-identified, Restricted Interviews & Documentation

1	CShdaimah Keisha Interview 1, Copy 2 12-4-12
2	INTERVIEWER: ....cause it would be sad for both of us. Okay. I'm gonna put this
3	just a little closer.
4	RESPONDENT: Okay.
5	INTERVIEWER: But you don't have to worry about speaking in it, it takes a...
6	RESPONDENT: Okay.
7	INTERVIEWER: So thank you again.
8	RESPONDENT: You're welcome.
9	INTERVIEWER: And I'm just gonna start off, I have just like a few demographic
10	questions, and then it'll be more like a conversation.
11	RESPONDENT: Okay.
12	INTERVIEWER: Okay. So how old are you?
13	RESPONDENT: I'm [30-40].

Filename: Shdaimah_Guide_Interview-1.pdf

Description: First Interview with Program Participants

In *Problem-Solving Courts, Street Level Bureaucrats, and Clients as Policy Agents in a Prostitution Diversion Program* (version 1.0), by Shdaimah, Corey

Download File   Close Preview

Previous   Next   Page: 1 / 2

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 Dawn Court. 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?

Shdaimah, Corey. 2020. "Problem-Solving Courts, Street Level Bureaucrats, and Clients as Policy Agents in a Prostitution Diversion Program". Qualitative Data Repository. https://doi.org/10.5064/F6C8VUHP

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 al*

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.

VandeVusse A, Mueller J, Karcher S. "Qualitative Data Sharing: Participant Understanding, Motivation, and Consent." *Qual Health Res.* 2022 Jan;32(1):182-191. doi: 10.1177/10497323211054058

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

De-identification Protocol: Example

Some Guidelines on Making the Transcriptions Anonymous:

We change the names of many particular details (places, people, jobs, etc...) to more general categories. In an interview, the name of a of a particular detail will be assigned the general category in all caps and a number ##. The number will be consistently assigned to this particular detail throughout the transcribed interview. For example, if a broker named Jorge Martinez is mentioned, we will change his name to REFERENTE ##, where ## is a number that is assigned to Jorge throughout the transcribed interview.

Eliminate names of municipalities. Anonymizing the names of the cities provides the best protection for the respondents. Of course, it is not possible to eliminate every detail about a municipality without losing much of the context in the interview. In all of the transcripts MUNICIPIO 1 will be the city, in which the respondent is a broker and/or city council member. Other cities that are named will be assigned a number so that the one city will have the same number throughout the individual transcript. The exception to this rule will be when a respondent refers to Capital or Buenos Aires. In this case, the risk of revealing the identity of the respondent's city or the identity of the respondent is low, and the distinction between Capital and the other cities is an important one. Similarly the names of streets will be termed CALLE and assigned a number. If a particular zone in a city is mentioned, it will be termed ZONA ##. If a neighborhood in a city is named it will be termed BARRIO ##.

All of the names of politicians will be changed to more abstract terms with the exception of the current and preceding presidents and governors. If possible the name of the politician will be changed to the politician's office. For example, if Javier Ramos is a concejal then we would change his name to CONCEJAL ##. If the position of the referenced politician cannot be identified then the name can be replaced with POLITICO ##. Each individual will be assigned a number and this number will be consistent throughout the transcript.

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")
- 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 propo

- Informed consent review and advice
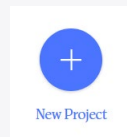
- Data management consultations to plan for project

When depositing data

- Disclosure risk review

- Data organization and documentation consultation

- Data/file transformations

Research Data Lifecycle by LMA Research Data Management Working Group

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!

QDR
The Qualitative
Data Repository

Web: https://qdr.syr.edu

@qdr@mstdn.science

@qdrepository

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

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