Quick Intro to the Qualitative Data Repository (QDR)

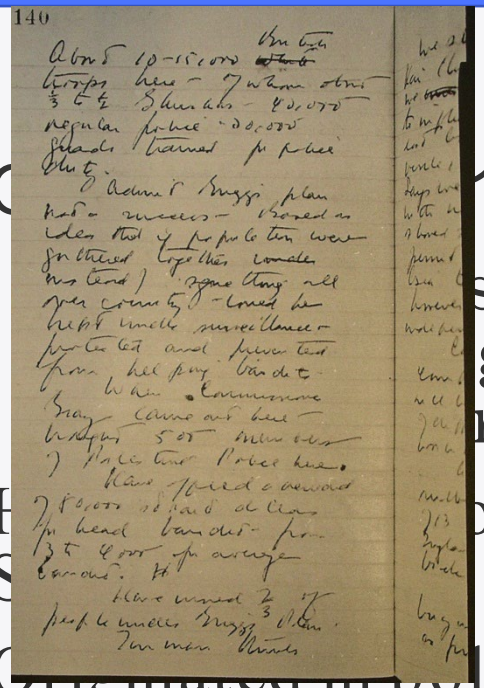
Dessi Kirilova, QDR

Workshop on Open Data and Reuse in Social Science Weather Research

April 11, 2023

QDR
The Qualitative Data Repository

What Is QDR?



- C 14qdr.s stores gital c resear
- H other t S
- Originated in political s
- Currently 154 data projects published
- Over 30 institutional members
- Continuously CoreTrustSeal-certified

A FOCUS GROUP DISCUSSION WITH THE HEALTH WORKERS

DATE: 3rd August, 2016

LOCATION: Federal Teaching Hospital, Abakaliki

DURATION: 74 minutes

I = INTERVIEWER **P = PARTICIPANTS.**

[Names of participants have been omitted. Study team member nam

I: Good morning.

ALL: Good morning.

I: Some of us were not here when we did the introduction. M Mbachu. Here with me are;

I2: Adanna Chukwuma

I3: Eze Nelson

I: What language do you prefer that we use in this discussi combination of both?

ALL: Combination of the two.

es the

Naucalpan de Juárez 2006-2009

GERENCIA DE AGUA POTABLE

HORARIOS DE SERVICIO DE AGUA POTABLE POR COLONIA EN NAUCALPAN

No.	POBLACION	TIEMPO DE SERVICIO	HORARIO DE SERVICIO (HORAS)
	PUEBLOS		
1	LOS REMEDIOS	Diario	8 hrs.
2	SAN ANTONIO ZOMEYUCAN	c/3er día	24 hrs.
3	SAN BARTOLO NAUCALPAN (NAUCALPAN CENTRO)	Diario	24 hrs.
4	SAN ESTEBAN HUITZILACASCO	Diario	24 hrs.
5	SAN FRANCISCO CHIMALPA	No existe infraestructura hidráulica operada por este Organismo	
6	SAN FRANCISCO CUAUTLALPAN	Diario	24 hrs.
7	SAN JOSÉ RÍO HONDO	Diario	24 hrs.
8	SAN JUAN TOTOLTEPEC	Diario	24 hrs.
9	SAN LORENZO TOTOLINGA	c/3er día	24 hrs.
10	SAN LUIS TLATILCO	Diario	14:00 a 6:00 hrs
11	SAN MIGUEL TECAMACHALCO	Diario	24 hrs.
12	SAN RAFAEL CHAMAPA	Diario	6:00 a 13:00 y 19:00 a 6:00
13	SANTA CRUZ ACATLAN	Diario	24 hrs.
14	SANTA CRUZ DEL MONTE	Diario	24 hrs.
15	SANTA MARÍA NATIVITAS	Diario	24 hrs.
16	SANTIAGO OCCIPACO	Diario	24 hrs.
17	SANTIAGO TEPATLAXCO	No existe infraestructura hidráulica operada por este Organismo	

QDR– fills a niche based on existing best principles + innovates in that space

About QDR

Sharing data and its documentation for secondary analysis

Empowering qualitative and multi-method inquiry through guidance and consultation

Providing data and materials to enrich and enliven teaching

Developing innovative approaches for enriching publications with data and analysis

Our Mission

QDR curates, stores, preserves, publishes, and enables the download of digital data generated through qualitative and multi-method research in the social sciences. The repository develops and disseminates guidance for managing, sharing, citing, and reusing qualitative data, and contributes to the generation of common standards for doing so. QDR's overarching goals are to make sharing qualitative data customary in the social sciences, to broaden access to social science data, and to strengthen qualitative and multi-method research.

Learn more about us ▸

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

"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."

https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-sharing-policy/data-management-and-sharing-policy-overview

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
- 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

What Makes Data FAIR?

Findable	Rich metadata, permanent identifiers
Accessible	Data and metadata are machine-readable in standardized ways
Interoperable	Metadata is described in standardized vocabulary
Reusable	Metadata meet community standards and data have clear license

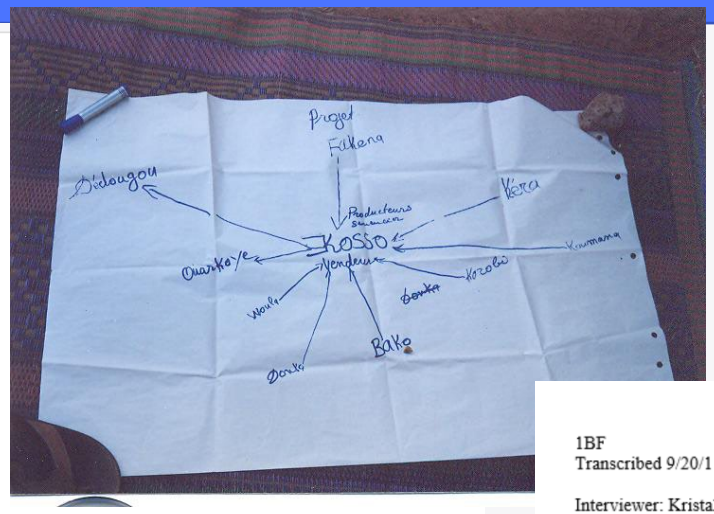
https://www.go-fair.org/fair-principles/

Shared Qualitative Data: Some Examples

Mixed Method Study:
Interviews, Focus Groups, and Quantitative Ecological Data

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)



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?

Filename: Jones_Tabular_RainfallByRegion.tab

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

Download File | Close Preview

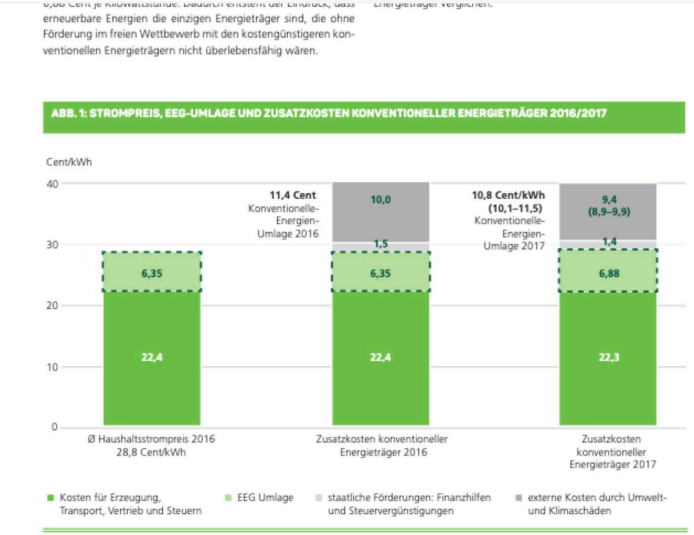
	2010	2011	2012	Farmer ID	Region
1	1104	920	1196	1	Dioila
2	1104	920	1196	2	Dioila
3	1104	920	1196	3	Dioila
4	1104	920	1196	4	Dioila
5	1104	920	1196	5	Dioila
6	1104	920	1196	6	Dioila
7	1104	920	1196	7	Dioila
8	1104	920	1196	8	Dioila
9	1104	920	1196	9	Dioila
10	1104	920	1196	10	Dioila
11	1104	920	1196	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

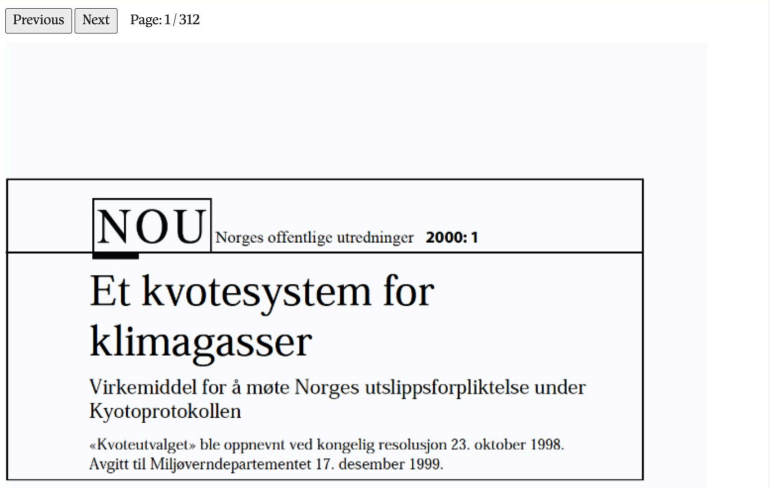
Jones, Kristal. 2020. "Seed systems in West Africa". Qualitative Data Repository. https://doi.org/10.5064/F6URYY1IQDR Main Collection.

Archival Materials:
grey literature, governmental & NGO policy documents, media report

- Mildenberger_CarbonCaptured_TRAX.pdf (384.2 KB)
- Mildenberger_MediaArchive.pdf (282.9 KB)
- README_Mildenberger.txt (19.3 KB)
- Chapter 1
  - Mildenberger_Fn002-003_EPA_Transcript-of-Public-Hearing_20140713.pdf (1.1 MB)
- Chapter 3
- Chapter 4
- Chapter 5
- Chapter 6
- Chapter 7
  - Mildenberger_Fn007_EU-Commission_C(2004)2515-2_20040707.p
  - Mildenberger_Fn008_CDU-SPD_Coalition-Treaty_201402DD.pdf (8
  - Mildenberger_Fn010_GreenpeaceEnergy_Was-Strom-Wirklich-Kost
  - Mildenberger_Fn027_Dion_Climate-Gift-or-Lump-of-Coal_2014091
- Chapter 8
  - Mildenberger_Fn006_WhiteHouse_US-China-Climate-Change_2014
  - Mildenberger_Fn008_NABTU_Letter-to-AFL-CIO_20160914.pdf (2.

Mildenberger, Matto. 2020. "Data for: Carbon captured: How business and labor control climate politics". Qualitative Data Repository. https://doi.org/10.5064/F6GYLSON. QDR Main Collection.





De-identified, Restricted Interviews & Documentation

CFS- 20:19

Interviewer: Well this one's easy. This is the um, exit interview where I get an opportunity to kinda ask you about, um, our sessions and to get ideas from you so that it helps us with our writing of our next grant.

Patient: Oh okay.

Interviewer: So we wanna make sure and see what worked, what didn't work. You know, remember I told you this was a pilot so we're trying it out to see if it worked.

Patient: Right.

Interviewer: So um, in general did you- overall, how was your experience of this Conquer Fear program? Did ya, did it- did it help? Did you like it?

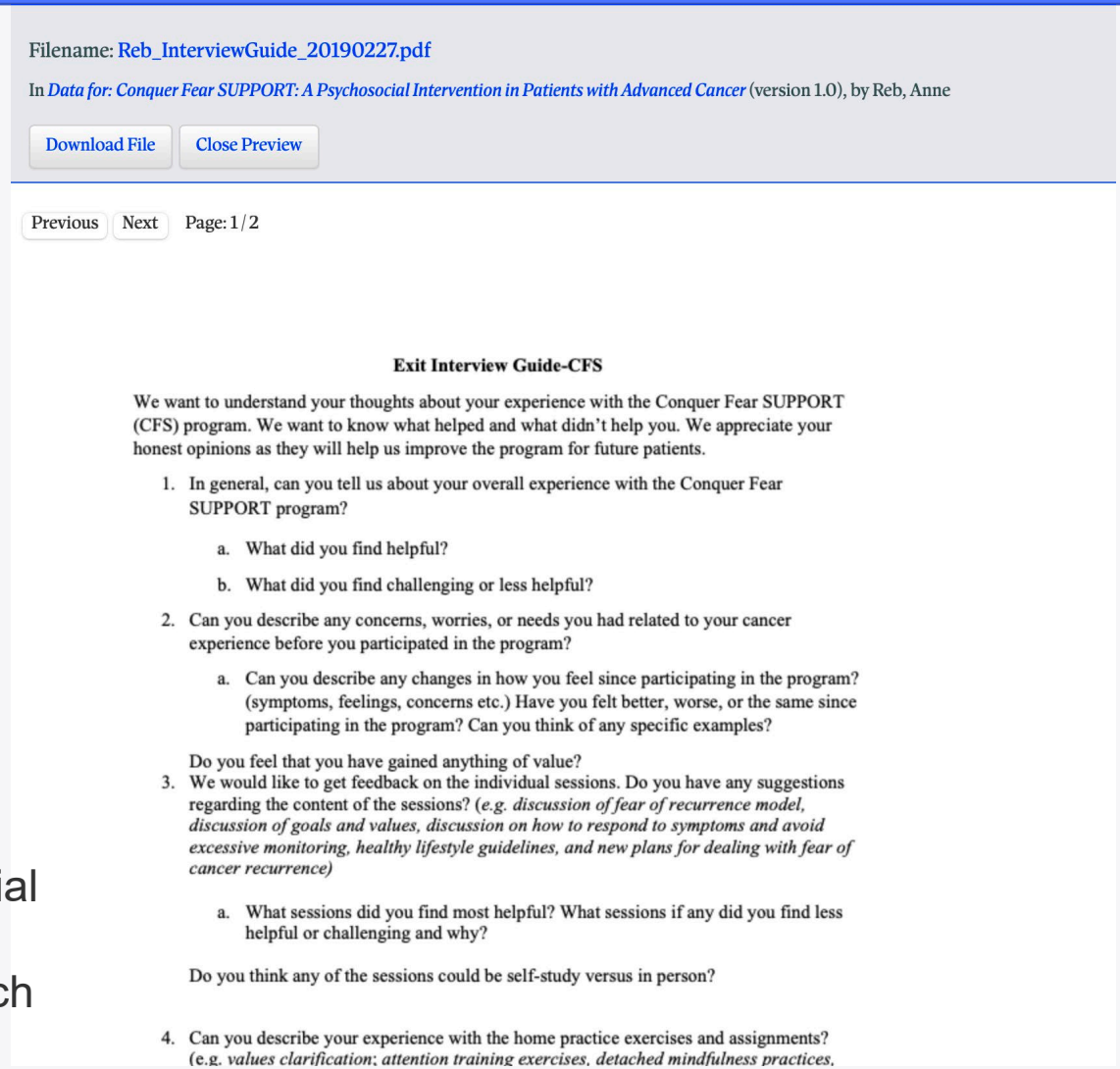
Patient: Yes. It was a positive experience, because I learned things that um, I didn't- I didn't know beforehand. It- it taught me how to um, focus on what I was doing at the time and it taught me like not to um, read into stuff you know, like don't panic right away when I would feel a symptom.

Interviewer: Mhm.

Patient: Uh, you know, and I remember that my last symptom, remember when I went to the hospital?

Interviewer: Yes.

Patient: I didn't panic about that even though it hurt a lot, um, because it happened to me



Filename: Reb_InterviewGuide_20190227.pdf

In *Data for: Conquer Fear SUPPORT: A Psychosocial Intervention in Patients with Advanced Cancer* (version 1.0), by Reb, Anne

Download File Close Preview

Previous Next Page: 1 / 2

Exit Interview Guide-CFS

We want to understand your thoughts about your experience with the Conquer Fear SUPPORT (CFS) program. We want to know what helped and what didn't help you. We appreciate your honest opinions as they will help us improve the program for future patients.

1. In general, can you tell us about your overall experience with the Conquer Fear SUPPORT program?
   a. What did you find helpful?
   b. What did you find challenging or less helpful?
2. Can you describe any concerns, worries, or needs you had related to your cancer experience before you participated in the program?
   a. Can you describe any changes in how you feel since participating in the program? (symptoms, feelings, concerns etc.) Have you felt better, worse, or the same since participating in the program? Can you think of any specific examples?

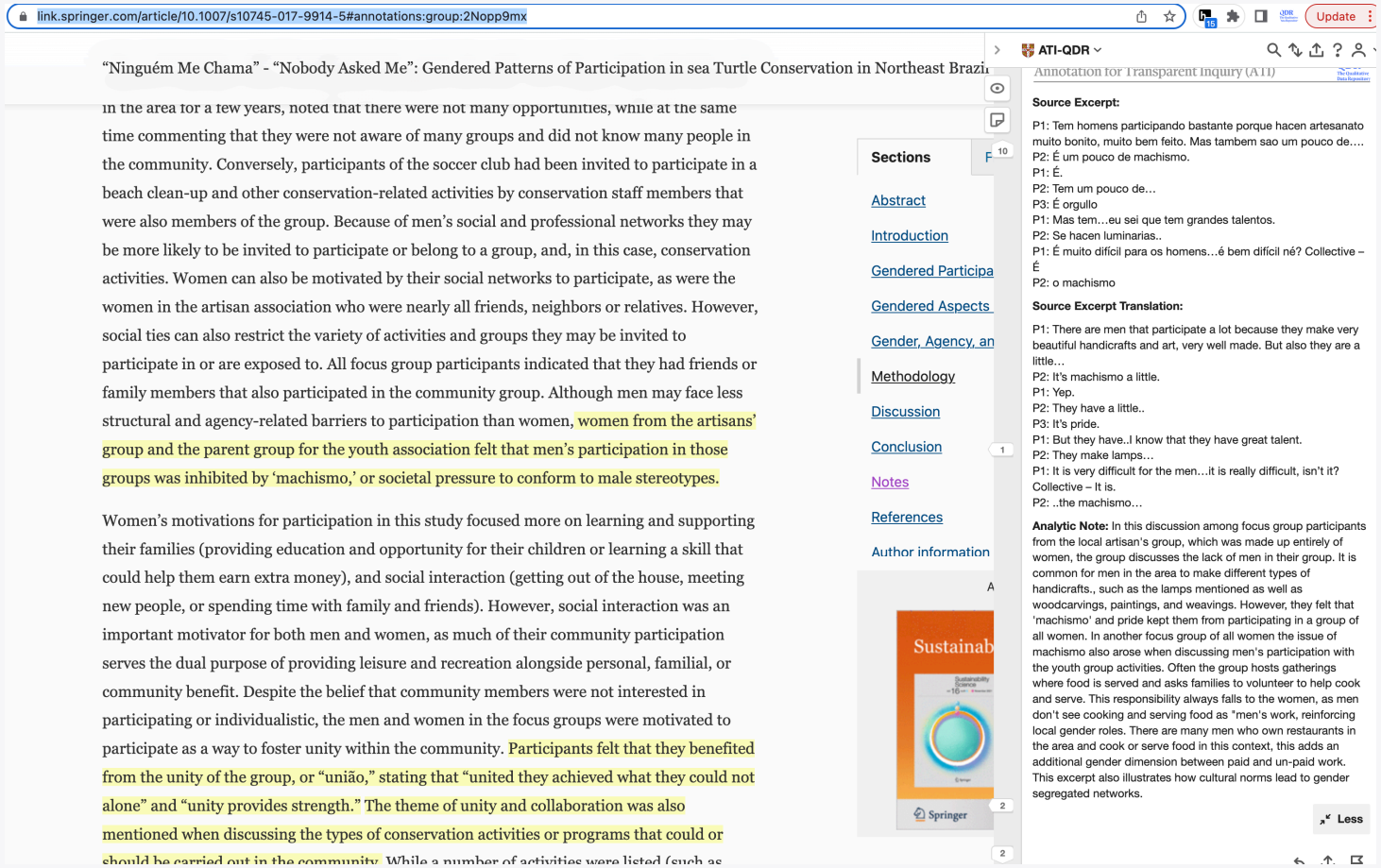
   Do you feel that you have gained anything of value?
3. We would like to get feedback on the individual sessions. Do you have any suggestions regarding the content of the sessions? (*e.g. discussion of fear of recurrence model, discussion of goals and values, discussion on how to respond to symptoms and avoid excessive monitoring, healthy lifestyle guidelines, and new plans for dealing with fear of cancer recurrence)*
   a. What sessions did you find most helpful? What sessions if any did you find less helpful or challenging and why?

   Do you think any of the sessions could be self-study versus in person?

4. Can you describe your experience with the home practice exercises and assignments? (e.g. *values clarification; attention training exercises, detached mindfulness practices,*

Reb, Anne. 2023. "Data for: Conquer Fear SUPPORT: A Psychosocial Intervention in Patients with Advanced Cancer". Qualitative Data Repository.https://doi.org/10.5064/F6FTJ1HUPalliative Care Research Cooperative QDREOLPC . V1

Annotation for Transparent Inquiry (ATI)



Rinkus, Marisa A.. 2019. "Data for: "Ninguém me chama" - "Nobody asked me": Gendered patterns of participation in sea turtle conservation in Northeast Brazil". Qualitative Data Repository. https://doi.org/10.5064/F6SGXPOS. QDR Main Collection.

Sharing Human Participant Qualitative 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*

https://qdr.syr.edu/guidance/human-participants

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 you to document data collection, sharing a 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
- Don’t underestimate participants willingness to help science, not just you
- Opt-in consent for data sharing can be great; IRBs familiar with it as “tiered consent”
  - Careful with quantitative data & opt-in

https://qdr.syr.edu/guidance/templates#informed_consent

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]
- Plan and apply de-identification at time of transcription
- Consistency within research team /project
  - Keep deidentification log of replacements or removals made; keep separate from the processed data files

https://qdr.syr.edu/guidance/human-participants/deidentification

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

Applying Appropriate Access Controls

"As open as possible, as closed as necessary" (European Union)

- Open data (license "CC0" or "CC BY") – often not appropriate for QUAL data
- 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 (typically for data that still has PII)
- Access using enclaves (not currently at QDR, but available for quantitative data at other repositories)

QDR for Weather / Climate / Environmental Social Scientists

Consultations at Different Points in the Project Lifecycle

When planning your research

- DMP / DMSP consultations and review for grant proposals
  - Including help with costing and budgeting
- Informed consent drafting advice and review
- Data management consultations

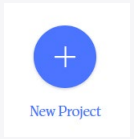
When depositing your data

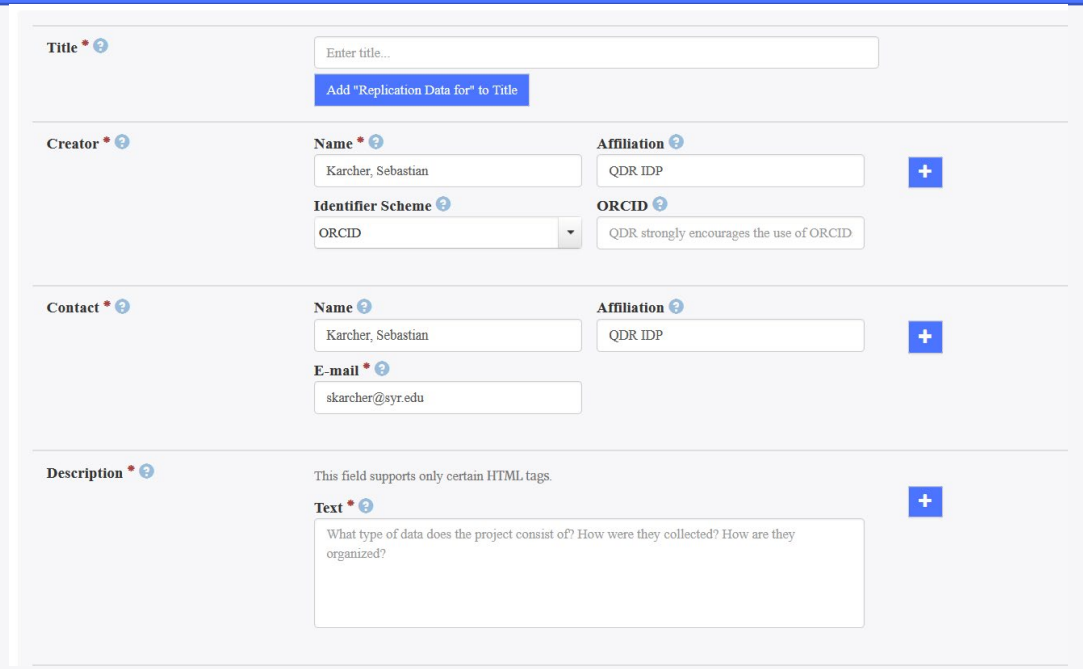
- Disclosure risk review
- Data organization and documentation consultation
- Data/file transformation

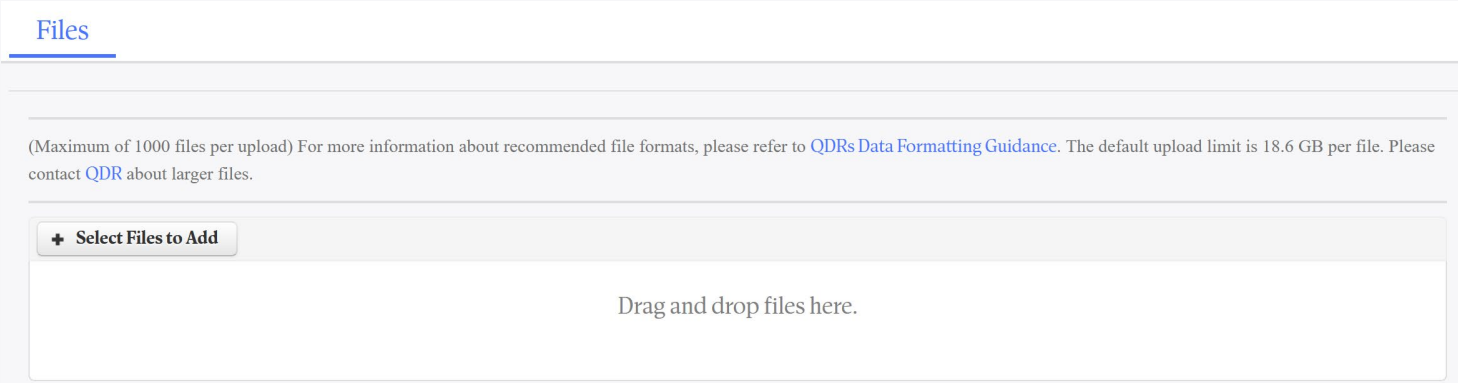
Research Data Lifecycle by LMA Research Data Management Working Group

Depositing *Your*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
- Various data formats acceptable







Questions? Comments? Please Stay in Touch!

Web: https://qdr.syr.edu

@qdr@mstdn.science

@qdrepository

QDR
The Qualitative Data Repository

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

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