

Managing Qualitative Data Safely and Securely

Webinar, November 17, 2016

Sebastian Karcher and Dessi Kirilova (Qualitative Data Repository)

Attribution: Parts of this presentation are based on slides used in a course co-taught by personnel from QDR and the <u>UK Data Service</u>.

All materials are provided under a <u>CC-BY license</u>

Overview

Managing Qualitative Data Safely and Securely

- 1) The value of planning and DMPs
- 2) Transparency and data documentation
- 3) Ethical, legal and logistical challenges to sharing and best practices to address them
- 4) Keeping your data safe and secure



Qualitative Data Repository

- Social science data (any relevant discipline or topic)
- From qualitative or multi-method research projects
- Online since 2014
- Have been conducting in-person data management training in various venues
- Recently posted data management guidance documents on the website (<u>www.qdr.org</u>)
- First webinar based on user request

1) What is research data management?

Research data management is caring for, facilitating access to, preserving and adding value to research data throughout its lifecycle.

Source: University of Edinburgh Information Services

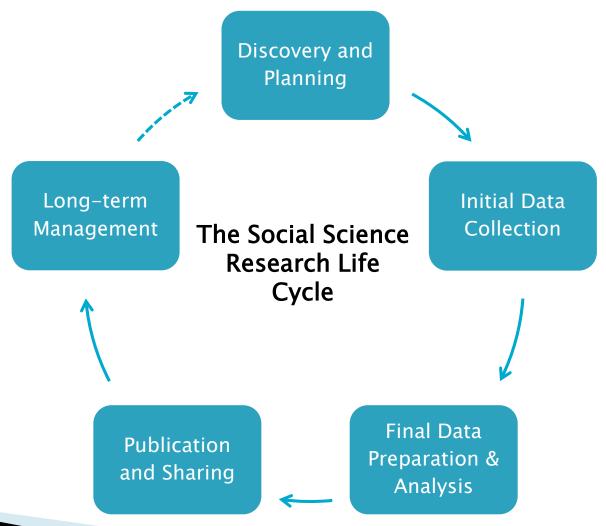


Why manage research data well?

- Creating data is expensive
- Data are basis for your findings
- Protect yourself against data loss
- Keep your participants safe
- Compliance with norms, rules, and laws
- To benefit your future self
- Transferable skill in Data Management Planning
- To help advance science!



Data Management for Sharing



Based on Green and Gutmann, 2007



How to write a DMP: Resources

- Funder template/guidance and sample DMPs
 - NSF-SBE DMP content guidance
 - ESRC <u>research data policy</u> and <u>DMP guidance</u>
- Generic resources:
 - DMPTool
 - DMPOnline tool
 - DCC guidance on DMP creation



DMP Tool

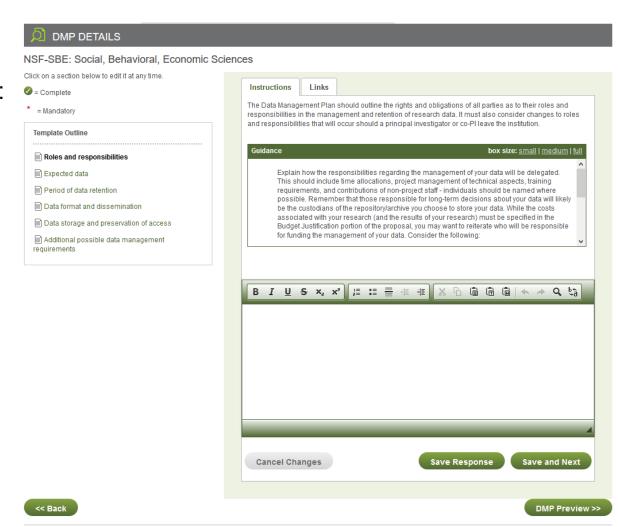
https://dmptool.org/

(Used, among others, by: DoD, DoJ, DoE, IMLS, Sloan Foundation)

Example output:

https://dmptool.org/plans/19207.pdf

"A Political Ecology of Value: A Cohort-Based Ethnography of the Environmental Turn in Nicaraguan Urban Social Policy"





Topics of a DMP

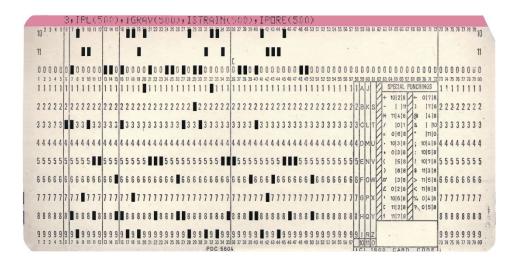
- Kinds of data that are being created
- Any applicable data sharing policies
- File formats
- Data descriptions, standards & metadata
- Data storage
- Access and use, incl. appropriate restrictions
- Intellectual property ownership / copyright
- Human participant constraints
- Roles and responsibilities in a team
- Budget for data activities

Key planning issues

- Know your legal, ethical and other obligations
- Know your institution's policies and services
- Assign roles and responsibilities
- Implement and review management of data during project meetings



Can you understand/use these data?



SrvMthdDraft.doc

SrvMthdFinal.doc

SrvMthdLastOne.doc

SrvMthdRealVersion.doc





2) Why document your data and processes?

- ▶ Enables you to understand data when you return to them
- Makes data and research reusable and verifiable
- Helps avoid incorrect use/misinterpretation
- Allow for archiving and creation of metadata

Guiding Question: What would a new user need to know to make sense of the data?

What should be captured?

Contextual information about project and data

- background, project history, aims, objectives, hypotheses
- publications based on data collection
- final reports, working paper, lab books
- user guides
- anything useful...

What should be captured?

- Data collection methodology and processes
 - data collection process and sampling choices
 - instruments used questionnaires, show-cards, interview schedules, topic guide
 - temporal/geographic coverage
 - data validation steps: cleaning, error-checking
 - compilation of derived variables (QUAL EX: codes you develop as part of content analysis)
 - citations for any secondary data sources used

What should be captured?

- Information on data files structure
 - Inventory of files (smart naming conventions help)
 - relationships between those files
 - records, cases...
- Variable-level documentation
 - labels, codes, classifications
 - missing values
 - derivations and aggregations
- Data confidentiality, access and use conditions
 - anonymisation carried out (anonymization protocol)
 - consent or copyright conditions/forms/procedures
 - access or use conditions of data



Consider documentation early on

- Good documentation and metadata depends on what you can provide
- What you can provide depends on what you can remember
- Start gathering meaningful information from as early on in the research process as possible

Data-level documentation suggestions

Embed documentation in your data files

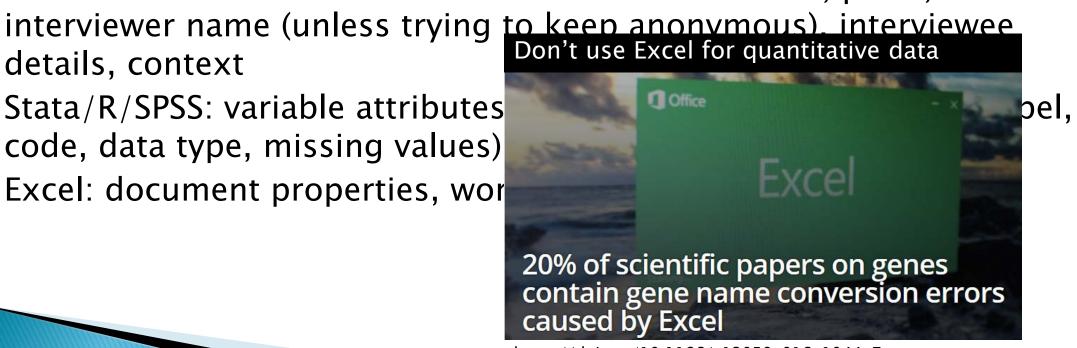
Interview transcript speech demarcation (speaker tags)

Document header with brief details of interview date, place,

details, context

Stata/R/SPSS: variable attributes code, data type, missing values)

Excel: document properties, wor



https://doi.org/10.1186/s13059-016-1044-7

In practice: Documentation in transcript

Raúl L. Madrid Interview with Felipe Quispe, Leader of Movimiento Indigena Pachacuti (MIP) La Paz, Bolivia, July 29, 2004

Quienes crearon el MIP?

Un poco de introducción. Ese movimiento indígena viene desde los anos 70s. Luego hemos estado en los Ayllus Rojos. Se ha transformado en el Ejercito Tupak Katari. Por dos años actuamos en acciones revolucionarias. Nos capturaron. Nos encarcelaron. Estuve 5 años en la cárcel. En 1997 salí. Todavía tengo proceso.

Source: https://doi.org/10.5064/F6MS3QNV



Metadata: Data about data

- Highly structured documentation
- Data collection metadata examples:
 - Components of a bibliographic reference
 - Core information that a search engine indexes to make your data findable
- International standards/schemes
 - Data Documentation Initiative (DDI)
 - DataCite Metadata
 - Dublin Core

Sample QDR catalog record metadata

PRESIDENTIAL CAMPAIGN ADVERTISING IN CHILE, BRAZIL, AND PERU

Description

These videos of presidential campaign advertising in Chile, Brazil and Peru were collected in order to characterize the evolution of campaign strategies in new democracies.

Show More

Suggested Citation

Boas, Taylor. (2015) "Presidential Campaign Advertising in Chile, Brazil, and Peru." Data Collection, QDR:10053. Syracuse, NY: Qualitative Data Repository [distributor]. http://doi.org/10.5064/F62Z13FQ

PID	QDR:10053 PROJECT
Geography	Brazil, Chile, Peru, Latin America
Topic	Brazil, campaign advertising, Chile, Latin America, Peru, presidential elections, television
Creator	Taylor C. Boas, Department of Political Science, Boston University (tboas@bu.edu)
Collection Type	Data Collection
Version	Original (Pilot Project)
Date	2005-01-06 to 2011-12-31
Temporal Coverage	1985-01-01 to 2011-12-31
Language	Spanish, Portuguese
Rights	Only a sub-set of the full data collection of over a thousand videos is shared here. A decision about the online accessibility of the full set of project files is pending, awaiting further investigation of their provenance (originally broadcast publicly on television). Interested registered users can contact QDR to arrange viewing all videos on site.
Publisher	Qualitative Data Repository
DOI	10.5064/F62Z13FQ



3) 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 respondents must be respected
- Participation is voluntarily, free of any coercion
- Harm to participants must be avoided in all instances
- Any conflicts of interest must be explicit

Planning is key! (again)

- Collecting identifying information
 - Avoid collecting unless necessary
 - Where confidential: Keep directly identifying info separate
- Informed consent an active process
 - Careful with restrictions in consent script
 - Oral vs. written consent
 - Cultural context
 - Ask for permission for data sharing explicitly (& include in IRB)



Publishing Sensitive Data: Quantitative vs. Qualitative Techniques

- Quantitative: Technical solutions (differential privacy, masking, synthetic data, multi-party computing)
- Manual anonymization: researcher knowledge

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



Publishing Sensitive Data: Access Controls

Safe Qualitative Data



Safe Quantitative Data

4) Stuff happens or how to keep your data secure

"I'm sorry but we had to blow up your laptop."



"What....all my client case notes and testimony, writing, pictures, music and applications. Years of work. NO!!!!"

Source: lilysussman.wordpress.com

Backing-up data

- It's not a case of *if* you will lose data, but *when* you will lose data!
- Digital media are fallible
- Protect against: software failure, hardware failure, malicious attacks, natural disasters, YOURSELF!

Rule 1: Three versions in two locations

Rule 2: Regular, automatic, and incremental



sticky notes never crash.'



Cloud storage services





- Cloud services increasingly popular
 - Very convenient
 - Background syncing
 - Mobile apps available
- Consider if appropriate, as services can be hosted outside your country (personal data laws)
- Encrypt anything sensitive (ex: via Pretty Good Privacy; BitLocker) or avoid services altogether
- Look for services with end-to-end encryption aka "zero knowledge"



Data destruction

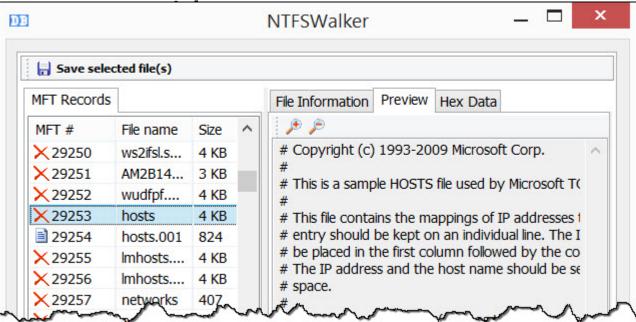
Beware of mandates to destroy the data but, if required, keep the following in mind:

 When you delete a file from a hard drive, it's still retrievable – even after emptying the recycle bin

Files need to be overwritten (ideally multiple times) with

random data to ensure they ar m

Free file and folder-shredding





How to securely destroy data

Disintegrate, Pulverize, Melt, and Incinerate. These sanitization methods are designed to completely Destroy the media. They are typically carried out at an outsourced metal Destruction or licensed incineration facility with the specific capabilities to perform these activities effectively, securely, and safely.

NIST Special Publication 800-88: Guidelines for Media Sanitization



https://youtu.be/4WIMSP_IMGs *