Fall 2019

Sharing and Managing Qualitative Data

Sebastian Karcher
skarcher@syr.edu

Dessi Kirilova
Syracuse University, dpkirilo@maxwell.syr.edu

Christiane Page
Syracuse, cmpage@syr.edu

Sebastian Karcher, Dessi Kirilova, and Christiane Pagé

Published as:

Karcher, Sebastian, Dessi Kirilova, and Christiane Pagé. 2020. "Case Study: Sharing Qualitative Social Science Data." In *Qualitative and Mixed Methods Data Analysis Using Dedoose: A Practical Approach for Research across the Social Sciences*, by Michelle Salmona, Dan Kaczynski, and Eli Lieber, 226–34. Los Angeles: Sage.

Sharing and Managing Qualitative Data

The Qualitative Data Repository (QDR, www.qdr.org) is an archive dedicated to qualitative data and data underlying multi-method inquiry. QDR is the only repository of its kind in the US—dedicated specifically to curating and archiving qualitative data and attuned to the requirements and concerns of qualitative researchers, their materials and sources. Planning for QDR begun in 2007, as discussions about data sharing gained increasing salience in the social sciences. Much of the infrastructure and advice on sharing data, however, focused on quantitative data. QDR was thus founded to fill this gap and provide dedicated guidance and suitable infrastructure for the needs of qualitative and multi-method researchers. With its roots in political science, QDR is now an interdisciplinary and international social science data repository, serving researchers from disciplines as diverse as public health, sociolinguistics, anthropology, and political science.

QDR provides extensive guidance to researchers on key topics such as managing and organizing qualitative data, formatting data for sharing, and working with—and potentially sharing—data based on interaction with human participants, and teaching with and about qualitative data (see qdr.org/guidance). Moreover, the repositories aims to support the particular requirements of qualitative researchers and their data. One example of making data infrastructure work for qualitative research is QDR's Annotation for Transparent Inquiry (ATI) initiative (qdr.org/ati). ATI allows researchers to link specific passages in a published work with source excerpts, notes about their context, and links to the full data sources. In many ways, using annotation and linking to make qualitative research more transparent is rooted in the same practices that underlie REDA software. Such computer assisted analysis of qualitative and multi-method data is rapidly growing among qualitative researchers. Another area of activity for QDR has thus been to develop guidelines and contribute to technological developments that facilitate the sharing of REDA data. This chapter is part of this endeavor, which also includes advice to individual users, a workshop held with developers and practitioners (Karcher & Pagé, 2017) to better understand researchers' needs on both a technical and practical level.

So, why should authors share their qualitative data? And what is it about your REDA project and data that make them particularly good for sharing? Different researchers may share data for different reasons. In many cases, shared data puts research on more solid, transparent foundations and makes it more credible. Data are also a valuable research output by themselves: as other scholars use and cite them, they enhance the authors' profile. Data sharing also supports those teaching methods, allowing students to learn tools and methods in real-world settings. Shared data can help junior scholars or scholar

from low-income economies, who may not have the resources to engage in primary data collection. In some cases, researchers share data because they are required by a funder or a journal they seek to publish in. Finally, where research findings have significant social impact, the dissemination of the data they are built on may be an ethical imperative.

Sharing and exchanging qualitative and multi-methods data projects is independent of epistemological or methodological priors. Research transparency is not about positivist vs. interpretivist approaches, and it is even less about quantitative vs. qualitative methodology. The importance of transparency does not derive from a scientist goal of replication, but from a more fundamental notion that credibility of science if *process dependent:* Its credibility relies on being transparent about the process through which researchers arrive at their conclusions or interpretations (Lupia & Elman, 2014, p. 20). REDA projects lend themselves particularly well to process transparency, because most analysis steps occur in a single environment and can be recorded transparently.

Because data sharing is costly—it can increase cost to your collection, documentation, and data preparation—good research practices often have a significant impact on minimizing those research costs. There is good news for REDA users: REDA already requires or engages multiple levels and types of documentation. REDA analytical attributes make them an excellent match to curation and documentation attributes for data sharing, for metadata creation making your data maximally findable and searchable, and for ensuring their long-term preservation. In effect, data sharing as a REDA users maximizes your return on research documentation and investments which also enabling your data to be citable through the process of assigning them a digital object identifier or DOI. There is a richness and completeness of the research process with REDA, typically beyond what non-computer aided analysis provides, that intrinsically facilitates and readies the data for sharing. The contemporaneous note taking and documentation involved with REDA often leads those data to be more structured than most other qualitative data. The ability to annotate key decision in the research process either via memos or by annotating coding labels is another big plus for data sharing. Such memos can even be time-stamped to document the timing of crucial decision in your research process. In this manner, REDA functionalities and attributes readily map-onto to curation functionality and documentation needs for data sharing.

The additional data visibility and exposure or usage available to REDA researchers often will mean that human participant protections are top of mind. Since the applications add complexity and visibility of data, researchers have often considered additional care by design. Domain repository will benefit from these steps and can help you think through how to legally and ethically share your REDA project. In short, richness and completeness of research process thanks to applications capabilities make REDA a great match for the data sharing. Data sharing accrues benefits, increases and improves collaboration opportunities—in this case championed in the Dedoose design philosophy, while it promotes transparency, evaluability, sharing your lessons learned, all along maximizing the use of your contributed data, information. Moreover, in re-use scenarios, data sharing significantly minimizes the effects of data collection.

Managing your Dedoose project for sharing

You will find it much easier to share your data if you consider the possibility of data sharing from the beginning of your project. To make your data widely usable, we recommend sharing it both as a complete Dedoose project (usable by other Dedoose users) and as individual files, usable by anyone regardless of the software (if any) they use. In the following, we present some detailed, step-by-step advice for sharing your REDA project.

1. File Naming and Organization

Consider what unit of data is going to be most meaningful for organizing. This may be what you think of as a "case" or any other meaningful unit within your research. Often these will be geographic units (a city, region, or country), but they could also be groups (political parties, professions), or historical periods. Organize your data around these main units. An important way to help with this organization is to implement a coherent naming convention for your files from the start. Your file naming convention should be driven by the organization of your project, so there is no one-size-fits-all solution.

Some rules of thumbs for file naming, however, do exist. Typically begin a filename with the name of the main unit you're investigating. We strongly recommend you include the date of when you last edited the file in YYYYMMDD format; this quickly lets you identify the most recent version of a file. To help with organization, also indicate whether a file is data (e.g., an interview transcript) or documentation (e.g., an informed consent form). Other information you may want to consider including is a one-word description of the file's content or an indication that a file needs to remain private or that it is already de-identified. For uniformity, also decide on the use of capital letters in filenames and the delimiter between different components of a filename (hyphen or underscore are the most common). A typical filename following these guidelines would be: `Brazil_InterviewA_20170324_deidentified.`

2. Include File-Level Information

Within individual files, include basic information about their contents, sometimes referred to as metadata or file-level metadata. A common approach is to have a standard header for every file. For example, for interviews, include the date and location of the interview as well as any relevant information about the interview and interviewee following a standard template.

3. Keep Track of Sensitive Information

As you collect your data, keep concerns about private and sensitive data in mind. As you add information that may need redacting, use Dedoose tags to highlight it, so you can quickly identify it later on. Also consider tagging files that you specifically cannot share (e.g., interviews given "off the record" or signed consent forms). If you include file-level metadata, take care to not include strongly identifying information in the file if the data is very sensitive. For sensitive data, keep a separate, encrypted key that links the code or pseudonym used in the transcript in a secure location. You will typically already have described such procedures in your IRB/ethics board application.

4. Keep Memos about Analytic Decisions

As you analyze your data, Dedoose will help you make your analytic process transparent. Using best practices, as recommended throughout this book, will make your own life easier, but it will also make your data more shareable and more valuable. In particular, making any coding and analysis decision explicit in memos will help readers evaluating your conclusions, but also secondary users who can better understand the application of given codes in your data.

5. Create a "Data Narrative"

We recommend that you include a document dedicated entirely to describing the data, the data narrative. The data narrative provides a high-level overview over the data project. It begins with a general outline of your research (akin to an abstract) to contextualize the data. In a second part, it describes the data in more detail:

a) What sort of data are part of your project? (Interviews, photos, videos, scanned documents, etc.)
b) How were the data collected? Describe, e.g., the selection of interviewees and the setting in which interviews were conducted or how particular files in an archive were selected.
c) How were the data processed after they were collected? (e.g. transcription of interviews, text recognition (OCR) on scans)
d) How are the data organized? (see above)
e) How was coding/analysis conducted?

The data narrative will help other readers better understand your data. It can help inform the methods section of your publication. It will also help curators provide better structured metadata that will help to make your data more discoverable for others.

6. Preparing You Data for Sharing

Make a copy of your project and delete any information you do not want to share, such as private notes or sensitive information. If you have followed our advice above, you can now use the tags you have created to redact potentially identifying information from transcripts.

De-identifying qualitative transcripts requires great care as you not only need to remove names and addresses ("direct identifiers"), but also contextual information that can be used to identify participants to whom you have promised confidentiality ("indirect identifiers").

Use pseudonyms for individuals. For indirect identifiers such as dates, specific locations, or institutions, use broader categories such as a date ranges ("between 1980-1985") or a description of the location ("small village in Southern France"). Clearly and consistently mark places where you have redacted data, e.g. using square brackets.

Detailed advice on de-identifying qualitative data can be found on QDR (qdr.org/guidance/human-participants/de-identification) and the UK Data Archive (ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative).

7. Exporting Your Data Sharing

At QDR, we recommend your data shared in two different forms. The first form is the raw full export from Dedoose. Once your data is prepared for sharing, first export the whole project, by clicking on "Export Data" then "Export Project".

Then, create a "human-readable" export, which anyone, regardless of software can use: Export all relevant files in widely used formats (such as RTF, PDF, Excel, as well as widely used video, image, and audio formats). Also export all relevant memos as RTF or PDF files. See the Dedoose documentation for detailed export instructions.

As of this writing, efforts are underway to provide a standardized format for exchange between different REDA software products (qdasoftware.org). As this exchange formats mature and becomes more widely available, we expect it to replace some of these recommendations.

8. Finding a Data Repository

Data repositories provide numerous advantages for sharing data. They provide searchable catalogs that make your data widely findable. They specialize in the long-term storage and preservation of data, so your data will still be accessible 25 years from today. Finally, many data repositories have specialized curators who can help you as you deposit your data.

As you are getting ready to share your data, find a data repository suitable for your purposes. There are several distinctions to consider: some repositories publish data from any scientific discipline and regardless of the origin of the researcher (e.g. Harvard Dataverse, Figshare, Zenodo). These repositories have large holdings, are typically easy to use, but they will provide little assistance and typically no checks on your data. A second group of repositories, "institutional repositories," focuses on data from researchers of a given institution, typically a university. They are interdisciplinary and their expertise given disciplinary data varies considerably, but data librarians at your institution will often be willing to assist you as you deposit data. A third group of repositories, often referred to as "domain repositories," specializes in a given discipline, such as social sciences. Such repositories include the UK Data Archive, ICPSR, and the Qualitative Data Repository (QDR). Depositing with a domain repository lets you benefit from the experience of curators with exactly the sort of data you are depositing. They will curate your data to make it more easily findable, re-usable, and (in the case of sensitive data) safe for sharing.

Deposit all exported files with your repository of choice. For many repositories, especially domain repositories, you will already have communicated with a curation specialist at this point, and they may provide additional guidance.

9. Promoting Your Data

Finally, once your data is published, don't forget to promote it! Data publications are still relatively rare in qualitative research, so this is one way in which you can set yourself apart. Include the publication in your CV and promote it via Twitter, your blog, or any other channels you use to disseminate your research. Most data repositories will assign a

digital object identifier (DOI) to your data. The DOI generates a permanent link and should be used in promoting and in citing your data.

Bibliography:

Karcher, S., & Pagé, C. (2017). Workshop Report: CAQDAS Projects and Digital Repositories' Best Practices. *D-Lib Magazine*, *23*(3/4). https://doi.org/10.1045/march2017-karcher

Lupia, A., & Elman, C. (2014). Openness in political science: data access and research transparency. *PS: Political Science & Politics*, *47*(01), 19–42. https://doi.org/10.1017/S1049096513001716

Pagé, C. (2017). Why and how you should share your CAQDAS project. International Congress of Qualitative Inquiry, Urbana-Champaign, IL, May 19, 2017.