The Qualitative Report 2021 Volume 26, Number 6, 1996-2010
https://doi.org/10.46743/2160-3715/2021.5012

How Data Curation Enables Epistemically Responsible Reuse of Qualitative Data

Sebastian Karcher
Syracuse University, New York, USA

Dessislava (Dessi) Kirilova
Syracuse University, New York, USA

Christiane Pagé
Syracuse University, New York, USA

Nic Weber
University of Washington, Seattle, Washington, USA

Data sharing and reuse are becoming the norm in quantitative research. At the same time, significant skepticism still accompanies the sharing and reuse of *qualitative* research data on both ethical and epistemological grounds. Nevertheless, there is growing interest in the reuse of qualitative data, as demonstrated by the range of contributions in this special issue. In this research note, we address epistemological critiques of reusing qualitative data and argue that careful curation of data can enable what we term "epistemologically responsible reuse" of qualitative data. We begin by briefly defining qualitative data and summarizing common epistemological objections to their shareability or usefulness for secondary analysis. We then introduce the concept of curation as enabling epistemologically responsible reuse and a potential way to address such objections. We discuss three recent trends that we believe are enhancing curatorial practices and thus expand the opportunities for responsible reuse: improvements in data management practices among researchers, the development of collaborative curation practices at repositories focused on qualitative data and technological advances that support sharing rich qualitative data. Using three examples of successful reuse of qualitative data, we illustrate the potential of these three trends to further improve the availability of reusable data projects.

Keywords: data sharing, qualitative data, Qualitative Data Repository

Contemporary social science research has benefited greatly from infrastructure innovations that make both the collection and exchange of data more efficient. Along with data-sharing policies promoting greater access to materials that are used to produce a research finding, these innovations enable social scientists to rapidly build upon previous studies in producing new knowledge. In this environment, data sharing and reuse are quickly becoming the norm in quantitative research. At the same time, significant skepticism still accompanies both the sharing and reuse of *qualitative* research data. This skepticism is, at least partially, based on epistemological and ethical grounds: Can any researcher who was not part of the original team ever properly understand qualitative data? Is it possible to separate qualitative

data from the contextually rich settings of researchers originally involved in their collection, and further, are qualitative data capable of being reliably analyzed without this original context (e.g., Feldman & Shaw, 2019; Mauthner et al., 1998)? Despite practical and conceptual barriers to doing so, there is growing interest in and support for the reuse of qualitative data for producing new scholarship (Bishop & Kuula-Lummi, 2017). This special issue of *The Qualitative Report* is a salient example of the emerging practices and challenges that researchers face when attempting to produce new findings through the secondary analysis of shared qualitative data.

The authors of this research note are research staff at the Qualitative Data Repository (QDR), the publisher of the data analyzed by participants in the current issue. Collectively, we believe that this perspective allows for valuable reflection on the affordances and limitations to effectively reusing qualitative data. More specifically, our goal is to provide readers of this special issue with some background on the history, current state, and potential future for sharing and reuse of qualitative data. We begin with a definition of data and a review of epistemic critiques of the sharing and reuse of qualitative data in particular. We then introduce the concept of "epistemically responsible reuse" of research data and describe how data curation can be designed to facilitate responsible reuse. We further describe how three recent advances make us hopeful about researchers' ability to overcome challenges to effective sharing and reusing of qualitative data and repositories' ability to enable epistemically responsible reuse:

1. **The increasing sophistication of data management**: Data management helps researchers with the storage and documentation of their data throughout the research process. As more funding organizations require data management plans and provide funds for their implementation, and as more universities—typically through their libraries—offer training and support for data management, researchers plan for sharing from the beginning of their projects. Thus, data are better organized and better documented, providing more context for reusers.

2. **Collaborative curation**: We describe the critical role of curation at QDR, which we approach as a collaborative interaction between researchers and data archivists through which qualitative data and meaningful documentation are made available in secure long-term preservation environments. Collaborative curation ensures—among other things—that data are properly contextualized and that sharing occurs ethically, respecting the informed consent and wellbeing of human participants.

3. **Technological advances which facilitate the sharing of richer qualitative data**: Improvements in repository technology, as well as improvements in available standards, further enable reuse. For example, the recent introduction of a common exchange format (REFI-QDA Standard, qdasoftware.org) between major qualitative data analysis software (QDAS) tools allows researchers to share not just data but also include their memos and original codings, facilitating a richer understanding of the original research and more informed reuse.

We conclude by briefly discussing three examples of successful reuse of qualitative data. These examples, spanning a wide range of methodologies, disciplines, and reuse contexts, demonstrate that responsibly reusing qualitative data is possible.

Epistemic Critiques of Reusing Qualitative Data[1]

Qualitative researchers have voiced concerns about the epistemological basis for reusing qualitative data going back at least to the early days of the United Kingdom's Qualibank archive in the 1990s (Corti & Thompson, 1997; Mauthner et al., 1998). Data sharing, in this view, is appropriate for quantitative research based in positivist epistemologies that view data as factual representations of an objective reality. As such, data can easily be understood and reused by other researchers. Much of qualitative research, however, rests on different epistemological principles. Here, "because interpretive researchers assume the relationality and constitutive character of data collection and analysis, they typically generate data in which they play a key role" (Feldman & Shaw, 2019, p. 704). In other words, the original researchers and their relationship to participants and the environment, as well as their tacit knowledge, are deeply embedded in data collection and analysis.

An aspect of this debate might be resolved if researchers of various methodological and epistemological commitments can nevertheless subscribe to a common definition of data. The definition we start with in this essay—and which informs our daily curatorial operations at QDR—sees data as "any representations of the social world relevant to a particular type of inquiry and rendered in a form suited to the analysis to be undertaken" (Kapiszewski & Karcher, 2020, p. 199). This definition also relates to information scholar Christine Borgman's description of data as "entities used as evidence of phenomena for the purposes of research or scholarship" (Borgman, 2015, p. 29). Both statements recognize that data are selected and thus "constructed" with a particular scholarly purpose in mind and thus acknowledge the critical role of the researcher in this process. In the case of qualitative materials in particular,[2] what makes them data is their selection, organization, and use for scholarly inquiry and analysis. Documentation plays a key role in contextualizing those processes on the project and file levels. Our argument for epistemically responsible reuse thus applies to any form of shared qualitative data and their accompanying documentation.

While such all-embracing definitions of data above might be accepted by a wide variety of social scientists, they do not necessarily result in an agreement on the possibility or appropriateness of qualitative data reuse. From the perspective of skeptics, it is impossible for any other researcher to consult *the same* data; arguably, even the researchers themselves cannot revisit the same data (Mauthner et al., 1998, pp. 737-743). But could this problem not be overcome by providing detailed documentation (Mannheimer et al., 2019)? Wouldn't it be possible to provide enough contextual information so that it enables the reuser to understand the data in the same way the original researcher did? The interpretivist answer to this question is a resounding no: "We do not agree that data can be repaired by supplying relevant contextual information" (Mauthner et al., 1998, p. 736). The problem, interpretivists argue, is that "'findings are not in the data but, rather, are created through the interaction of particular (either primary or secondary) researchers with particular respondents in particular locations and at particular historical junctures" (Feldman & Shaw, 2019, p. 705; Mauthner et al., 1998, p. 735). As a result, many qualitative researchers working in the interpretivist tradition remain skeptical

[1] Critiques of sharing qualitative data typically focus on three areas: ethics, in particular participant confidentiality; practical obstacles, including the workload imposed on researchers; and epistemic objections to the idea of reusing qualitative data. As our engagement in this note is with the practices of curating and reuse, we focus solely on the last issue. For a sense of the debates on the ethics and practicality of sharing qualitative data (e.g., Alexander et al., 2020; Bishop, 2005, p. 2005; DuBois et al., 2018; Parry & Mauthner, 2005; Tsai et al., 2016).

[2] In practice, qualitative data can comprise a wide range of products, including recordings and texts or transcripts representing participants' words, both in full or de-identified/redacted; researcher notes from observation of activities; artefacts such as drawings, maps, blogs, etc. originally produced for non-research purposes; as well as archival-type materials such as official documents, meeting minutes, letters, party platforms, campaign commercials, etc. See https://qdr.syr.edu/deposit/data for a non-exclusive list.

about secondary use of qualitative data and believe it is only possible, at most, where direct collaboration between primary and secondary researchers is involved (Feldman & Shaw, 2019, p. 713f).

Other qualitative researchers remain more optimistic. While not denying the epistemological challenges of decontextualized data, they believe that they can, at least in part, be addressed. Fielding and Fielding (2000), for example, argue that qualitative researchers are already constantly engaged in trying to understand the effect of relationships on their data. Extending the same type of reflexivity to secondary data is feasible, given the right information and a continued commitment to reflexivity. In a similar vein, Mannheimer et al. (2019) suggest that data repositories and librarians can facilitate such reuse, both by helping depositors to create documentation that provides sufficient context and by advising reusers in identifying data projects that are suitable given their epistemological commitments.

Enabling Epistemologically Responsible Reuse of Qualitative Data

Even the more optimistic voices agree, however, that the epistemological challenges to reusing qualitative data are substantial. These challenges may be exacerbated by different epistemological commitments among qualitative researchers. Qualitative researchers disagree about the goals of their endeavor, the right tools to reach such goals, and even about basic ontological assumptions such as the existence of an underlying "truth" that research could try to discern. This diversity is arguably one of the strengths of qualitative research, but how can data be shared among researchers of such different outlooks?

To inform our approach to enabling the reuse of qualitative data, we draw on the concept of "epistemological responsibility" as used in the philosophy of knowledge (especially Kornblith, 1983). As opposed to attempts to describe rules for ideal reasoning, epistemic responsibility asks, "whether a subject was reasoning 'as best he [sic] could'" (Kornblith, 1983, p. 33). As Kornblith argues, this epistemic responsibility is in some respects less demanding than ideal reasoning, but in other respects more so. Most importantly for our purposes, "[b]eing justified requires more than simply reasoning properly; it requires that one gather evidence properly as well" (p. 35; see Hall & Johnson, 1998 for an even stronger version of this proposition).

We believe the role of curation of qualitative data for reuse is to enable secondary users of data to act epistemically responsibly, that is, to allow them to learn additional evidence about the data they are reusing, which in turn allows them to shape justifiable beliefs based on those data. We call this function of curation "enabling epistemically responsible reuse," and define epistemically responsible reuse (ERR) as secondary use of data that actively aims to understand as much of the original data and context as possible and does not make claims beyond what can be justifiably inferred from the data.

We contend that ERR is not just an *epistemic*, but also an *ethical* imperative. This does not follow from a standard procedural approach to research ethics: a secondary analysis of qualitative data that completely "misses the point" could easily pass procedural requirements regulating research ethics—if de-identified data are analyzed, it may not even require formal approval by an IRB or other ethics committee. Qualitative researchers have long been skeptical of this purely procedural approach and its application to their work and have emphasized the need to engage in ethics as a process (e.g., Guillemin & Gillam, 2004; Pollock, 2012). Researchers often build trust with their participants, and this trust includes participants' beliefs that their story will be told "right." In one of the few empirical studies on participants' views on sharing and reusing qualitative data, a participant warns: "the further away from the interview you are, and whether it's the sound or it's just transcripts, as [sic] researchers going to get an inaccurate measure of what I actually meant" (Yardley et al., 2014, p. 108).

What does enabling epistemically responsible reuse mean in practice? We identify three aspects of curation that are essential.

1. *Provide context.* A successful curation process produces detailed documentation that helps secondary users properly understand the data. Such documentation includes fairly standardized information—when were the data collected and by whom, using what methods—but can also contain information more specific to a project. The role of the curator is to request such information from the data creator, to ask for missing details, and to help the data creator turn that information into documentation intelligible to secondary users of a data project. This function of data curation is widely accepted, including for curating quantitative data. As noted above, it is also the most common response by proponents of reusing qualitative data to its skeptics. We believe that it is a necessary component of curation, but we do agree that it may not, by itself, be sufficient to allow for responsible reuse.
2. *Give a sense of what is missing.* Good documentation commonly describes how data were generated and what they contain. To enable epistemically responsible reuse, documentation should also give a clear sense of what they do not contain, to help secondary reusers frankly assess how the information available to them differs from that available to the original researchers. There are many reasons for "missing" information. In some cases, specific interviews might have been "off the record," or participants did not consent to data sharing. Other information may not be shareable because it remains under copyright or because the holding archive did not grant permission to share (as is common, for example, for many archives in Europe). In many ethnographic traditions, field notes are considered highly private and not suitable for sharing (Cramer, 2015). Finally, some information, such as broad contextual knowledge absorbed during fieldwork, may just not lend itself well to sharing. Both good data management practices and conversations during the curation process can help clarify, and document, what is missing in a data project.[3]
3. *Help the data reflect the researcher(s).* Part of good curation involves standardization, such as using standardized metadata fields and vocabularies. But curation that enables epistemically responsible reuse of qualitative data also avoids over-standardization. Given the link between original researchers and data, it seeks to display their perspective of the data, their epistemology, and, in some cases, their positionality. As a result, data and documentation may contain stronger reflexive and subjective considerations for interpretivist work (e.g., Johnson et al., 2017), whereas data produced by researchers with a more positivist approach, such as those by Chukwuma et al. (2017) used in this special edition, reflects the depositors' epistemological outlook by using a more objective voice. As we will show, technology can play an important role in closing the gap between the researchers' experience and the way data are made available to potential reusers.

In the following sections, we describe how responsible reuse of qualitative data can be supported through data management, curation, and technical innovations. In particular, we draw attention to how researchers facing epistemically based challenges of qualitative data sharing are supported at the Qualitative Data Repository (qdr.org) at Syracuse University.

[3] Another helpful device for illustrating what may be missing from a data project is an interview methods appendix (see Bleich & Pekkanen, 2013). Such an appendix lists all interview participants (or their pseudonyms) with basic demographic information, noting for each participant whether their interviews are shared or not.

How Data Management Helps with ERR

At its core, the main obstacles to seeing qualitative data sharing and reuse as a normal aspect of scholarship are cultural (Teperek & Dunning, 2018), as also reflected in our discussion of the literature. But there are new practices that are now more likely to accompany qualitative data work and are likely to shift this kind of thinking and the low-sharing and low-reuse culture of the present.

One of the key recent developments that facilitates data sharing in general is a focus on data management during the course of a research project. Such an approach applies well to qualitative data collection and sharing—and so lays the foundation for their potential responsible reuse as well. Researchers generating empirical materials have, of course, always had to "manage" resources in some way, but historically, most figured out how to do this on their own, in an ad hoc manner, and through time-consuming trial and error. The typical dissertation—usually the first comprehensive substantive project of primary data collection for many in the social sciences—is often the worst managed project of one's career. As a result, its data are least appropriate for broader sharing and reuse despite their uniqueness and significance as a foundation of one's own intellectual and professional trajectory.

Recent attention to data management has been both spurred from "above," in the form of funder requirements for formal data management plans (DMPs) as required by the National Science Foundation (NSF) in the US, the Economic and Social Research Council (ESRC) in the UK, and assisted by the availability of online tools (such as dmptool.org) and data professionals located at university libraries and social science repositories who can guide a researcher toward a concrete plan for data management steps. Having thought about the series of mundane but useful choices regarding the organization, safety, security, and overall intelligibility of the empirical artifacts they produce in the course of data collection from the beginning, researchers have a much easier path to carrying those out systematically in the course of the actual project.

Good data management practices lay the critical base for sharing data in a rich and contextualized way that facilitates meaningful secondary reuse. Data management makes explicit, and thus intellectually *traceable* by others, what are often implicit, and known only to the original data collector and analyst, micro-steps of the work she performed. While secondary readers/users can arrive at different interpretations on the basis of a given data collection, qualitative data should not be seen as obscure or unintelligible. In fact, empirical materials that are part of a published data project go beyond data. Documentation to accompany shared projects, such as a research design, protocol for data collection, or an interview guide, lend additional layers of information about how, where, when, and by whom any one piece of evidence was gathered, all of which cumulatively enhance the transparency of a research project and the reusability of its data. When Yoon (2014, pp. 2-3) surveyed secondary users of qualitative data, many of their responses underlined the importance of documentation about the data collection process for effective reuse.

Good data management practices are good data collection practices. A data collection process that uses clear standards and conventions for organizing the work and its resulting outputs and documents key events and decisions in the course of data collection, also has a built-in quality assurance mechanism. It is both methodologically rigorous and epistemically responsible.

Researchers who keep careful notes about the immediate context of their data collection choices and experiences on the ground are generating the basis for rich metadata which allow reusers to follow the thought process that led to the availability, for example, of certain interviews but not others, or to understand the political context within which a set of collected focus group replies makes sense. On an even more basic level, when a project contains files

that are named according to a logical naming convention, and it is clear what the latest version of each is, when each one contains key information regarding the date and location of its creation as well as its author and topic, secondary users can more seamlessly enter the organizational framework of a project and thus grasp the processes through which the final results were derived.

Scholars who work in teams recognize the need for enabling such access to each other's thought processes and daily practice early on. Even close collaborators familiar both with the topic and methodology of a project need explicit signposts along the way to follow the train of thought that allows them to contribute most fruitfully to each other's discoveries and conclusions. This mapping of process can then become the template for allowing secondary reusers to track the research journey as well.

When seen from this perspective, good organization, detailed and in-the-moment documentation and making habitual steps and mental shortcuts explicit for others (i.e., good data management practices) are first and vitally helpful to the original researchers. As an added benefit, they also facilitate broader data sharing and epistemically responsible reuse.

How Collaborative Curation Supports ERR

Sharing research data in trusted repositories has the additional benefit of working collaboratively with trained curators. Data curators, like publication editors, act as a rehearsal audience standing in for the user community a repository serves. The Qualitative Data Repository (QDR) refers to its approach as "collaborative curation." Curators experienced with social science research ask the questions that others might have and guide prospective depositors in providing informative answers. Since curators are focused on the data (while for a researcher, the data, while important, are a means to a different end goal), they assist with incremental improvements to the presentation of data and, especially, documentation. In the aggregate, these details enhance the value of data and improve the potential for reuse. Curators do not alter any content or make any decisions without input from depositors, but they do give suggestions on organization, documentation to include and, critically, on human participant and copyright concerns. Ideally, the collaborative interactions begin while the research project itself is in the planning stages, so that various necessary choices can be made early on (and documented), thus making the actual data collection and deposit after it ends more efficient and ethical.

Many qualitative researchers find that reflexivity and awareness of positionality—taking oneself and one's identity into account as an integral part of their investigations—leads to work that is not only more honest but also richer and more revealing (Woolgar, 1988). Working within a collaborative curation relationship can, in an unexpected way, promote a different aspect of reflexivity with deeply illuminating results. Curation can spur additional reflection on the intellectual needs of the potential readership/audience of a research project and guide researchers with prompts through which reflexivity is documented. Thus, curation can aid the practice of reflexivity by eliciting additional documentation that helps inform reinterpretation and reanalysis.

Collaborative curation can also help to address one of the frequent criticisms of qualitative data sharing, the potential to divulge sensitive or other personal information of the human participants who took part in a research study. Curators can assist in planning to share data ethically and responsibly. In ongoing conversations, a curator can prepare and motivate a researcher to follow three important processes to this end: (1) assessing the nature of any actual sensitivities, the degree of possible harm, and the likelihood that this harm can result from sharing; (2) deciding what measures in the course of data collection and management can minimize the risks including, crucially, sharing those plans with the human participants

themselves and gaining their informed consent; and (3) keeping detailed records on the file (or, if necessary, even more granular) level about additional steps required before data publication. Thus, sharing qualitative data via a repository provides much more than technical tools (digitization of analog materials that otherwise might get lost, digital object identifiers, data cataloging and indexing, full-text search options, or video file previews), important as these are for findability and reuse. The iterative curation conversation that accompanies preparing data for sharing also includes support in selecting and formatting data, redaction and de-identification according to systematic protocols, and setting appropriate access controls.
Without proper management, and ideally sharing, data remain the proverbial iceberg under the tip of academic publications. Just like art curators, data curators inspect the final product that will be shared with the public and put the finishing touches that will allow it to be presented in its best light. They help to contextualize and frame shared data, facilitating meaningful data reuse. And they ensure their safety and preservation, and in the process add long-term, archival value. Whether secondary reuse of qualitative data takes the form of further analyses by the original researchers, substantive reanalysis by others, or methodologically motivated reanalysis, as illustrated in this special issue, curators consider all options and try to enrich and enhance the data public goods.

How Technological Advances Can Support ERR

Reliably building upon previous research depends, at a minimum, on trustworthy access to shared data. The technical architectures developed to make this a reality are increasingly sophisticated in their ability to securely preserve data, record and communicate provenance (the history of data ownership), as well as provide ready access to rich data documentation (e.g., metadata, research protocols, etc.). As described above, documentation, in particular, can help a qualitative researcher to establish a sense of how and in what ways data were produced, analyzed, and are amenable for reuse. We focus on three areas of recent technical advance that allow for qualitative data to be shared and displayed in a richer way, making epistemically responsible reuse possible: improvements in repository infrastructure, advances in sharing data analyzed with qualitative data analysis software (QDAS), and possibilities associated with linking articles and data using web annotations (Annotation for Transparent Inquiry, ATI).

At the QDR, we have developed a number of important curatorial interventions that aid researchers in preparing their data for publication (described in detail above), as well as a technical architecture that can facilitate trustworthy access and reuse. Central to our technical efforts is the open-source data repository software project Dataverse (dataverse.org). In addition to maintaining an instance of Dataverse, we have also contributed a number of new features to the software that customize it for reliable discovery and reuse of qualitative data. This includes the ability for researchers to pose a query and retrieve results based on a full-text index of all data stored in QDR. In practice, this means that any keyword or text string that is entered in a search prompt can be retrieved for a user and allows a registered user to select individual files or entire data collections for closer inspection. Since qualitative data is very frequently text-based (e.g., interview transcripts, scanned archival documents), this greatly improves the ability to discover and work with such data.

To improve the repository for visual, audio, and audiovisual data, we have also developed a feature to preview data files within a browser window, which means that instead of downloading and then locally unpacking and searching qualitative data collections, a registered user can instead inspect individual files or even play a preview of multimedia files within their browser. Full-text searching and previewing help to ensure that researchers are reliably accessing and potentially using data that meet the specific needs of their research project in ways that have previously been available for quantitative data only. Such improved

discovery, retrieval, and trustworthy access are a baseline for epistemically responsible data sharing and reuse.

Qualitative Data Analysis Software (QDAS) is a key tool used by social science researchers in many epistemic traditions. QDAS empowers researchers to systematically analyze and thematically code qualitative data collected through interviews, focus groups, archival documents, field notes, and open-ended survey questions. QDAS applications such as NVivo and ATLAS.ti are used widely across the health and social sciences (Woods, Paulus, et al., 2016). QDAS research objects (the combination of data, annotations or codes, and resulting analytic outputs) are critical to qualitative scholarship, but they are rarely shared. The sharing of QDAS outputs has been blocked, partially, by the fact that most QDAS are proprietary, and each different QDAS application depends upon a unique data model (Corti & Gregory, 2011). Incompatibilities between these proprietary data models have made it impossible to reuse or combine QDAS outputs from different software applications. For example, it has been impossible for researchers to export a QDAS project from ATLAS.ti and import the same project in NVivo. However, the developers of leading QDAS applications recently agreed on an open data exchange standard, REFI-QDA, that allows the outputs of any QDAS application to be ingested in, and reused by, another QDAS application. The REFI-QDA standard has the potential to catalyze an extraordinary increase in qualitative data sharing and the effective reuse of previous QDAS analyses (see Evers et al., 2020 for an introduction of the REFI-QDAS standard and an initial assessment). The ability to include codes and their descriptions, coded texts, and author memos allows authors to easily share rich details, and even reflexivity-related components (see Woods, Macklin, et al., 2016) related to their data with potential reusers, facilitating the responsible reuse of such data.

Innovative forms of supplementary data and materials are another area of recent technological innovation facilitating the epistemically responsible reuse of qualitative data. Annotation for Transparent Inquiry (ATI, https://qdr.syr.edu/ati), developed by QDR based on early work by Moravcsik (2010) on active citations, allows researchers to annotate individual passages in their articles and directly link them to underlying data (Karcher & Weber, 2019). ATI supports responsible reuse in at least two important ways:

1. It allows authors to directly support individual empirical claims with underlying data and use the annotation to further explicate the relationship between specific claims and evidence/data. Some articles rely on dozens, sometimes hundreds, different data sources, and the ability to precisely specify which claims or sections in a publication a source supports, aids readers in understanding it in context and in reusing it responsibly.
2. ATI annotations also provide authors with the ability to add a layer of explanation to an article. In several cases, authors have used these to shine additional light on the logistics, motivations, and details of data collection and analysis. For example, introducing a series of extended interview excerpts, Yeh (2019) explains:

> After analyzing our initial data, we were particularly interested in further probing herders' views of the determinants of grassland productivity. Author Volkmar conducted follow-up interviews in Gouli in 2014. Here are some additional responses to the question, "if you were to herd more than the number of livestock you have indicated as the maximum number you would want to graze, what if anything would be the effect in that year and in subsequent years?"

Technological advances by themselves cannot ensure epistemologically responsible reuse, but together with improvements in data management and advances in collaborative curation, they

provide unprecedented opportunities to present qualitative data in a way that allows and promotes responsible reuse. But is responsible reuse possible in practice and if so, how does it look? We turn to these questions in the final section of our paper.

Reusing Qualitative Data: Some Examples

The reuse of qualitative data has a short but fruitful history in qualitative inquiry. Here we discuss a few illustrative examples of secondary analysis of qualitative data, in order to highlight key features of successful reuse. We consider examples that map onto some of the more prominent objectives for epistemologically responsible reuse of qualitative data: reanalysis and repurposing, including the reuse of interview studies; methodological contributions; and teaching and learning qualitative research methods.

Looking at data anew, Medjedović and Witzel (2008) reused original longitudinal interview data on school-to-work transition of young adults in Germany ("Status Passages into Gainful Employment" at Bremen University's Collaborative Research Centre; for example, Heinz et al., 2004; Witzel & Kühn, 1999). The study (n=2,230) included a micro-panel (n= 198) where data were collected via semi-structured problem-focused and biographical interviews, of which 91 cases were retained over three waves of qualitative interviewing. Witzel was one of the primary investigators, while Medjedović was acknowledged for her contribution to data analysis (see Heinz et al., 2004, pp. 201-216 for details). In the reanalysis of those biographical interview data, Medjedović and Witzel sought to investigate the construction of work-process knowledge, a concept not examined in the original study. They reused the theoretical concepts of the original study to parse out the data along different, though related, dimensions. They also reused the coding scheme and retrieval functions of the original study's QDAS project. And while they uncovered the prevalence of an apprentice model and learning-by-doing processes, they retrieved evidence of some rejection of the usefulness of practical learning in favor of rules learned in vocational school. This reuse case highlights the promise of technological advances in QDAS data sharing, coupled with the reuse of the original study typology, enabled by thorough data management. This allowed the reusing researchers to engage the data responsibly, and to make new, meaningful interpretations.

Bishop (2016) recounts Bornat et al.'s (2012) reuse of oral history interview data collected to examine the development of geriatric medicine specialty in the UK. In the early 2010s, Bornat et al. re-examined and re-conceptualized the role of South Asian doctors in the advancement of the geriatric medical specialty, based in part on interviews for an original study by Margot Jeffery in the early 2000s. The reuse of the data informed the sampling frame for a new sample of South Asian doctors that would have been practicing geriatrics at the same time as the white British doctors interviewed for the original study. Bornat and colleagues had taken steps to consider the writings of Jeffery at the time, situated the study in her larger body of work, sought out other informants about Jeffery's work, biographical trajectory, and positionality. These steps illustrate that even in the case of reusing interviews, "[a]rchived data clearly have value as historical documents for studying the past. Methodologically, they can provide insights about where and how researchers were positioned in relation to theoretical, epistemological, methodological and substantive issues of the time of the research" (Mauthner et al., 1998, p. 743). Bornat and colleagues were able to "...link interviewees across time and juxtapose accounts in a way that has enriched the process of reuse and in so doing present a more complex and more richly contextualised understanding of a particular dialogue" (Bornat et al., 2012, sec. 9.6), responsibly preserving and enriching the data lifecycle of both datasets. This reuse example illustrates the opportunities of well-documented data, both in their life-cycle trajectories, and in enabling dialogue with and across data.

Reuse of qualitative data for teaching and learning methods is one of the aims of this special issue. In a blog post, Karcher (2016) describes how International Relations professor Robert Adcock of American University relies on data reuse in his classroom. In the context of teaching historical methods, Adcock chose Saunders' (2015) data project to create teaching exercises for his class. Saunders had assembled an "active citation collection" (a predecessor to ATI discussed above) that includes nearly 50 digitized primary sources, many of which from the seldom used pre-presidential papers. The collection was analyzed in a chapter on President Kennedy, in her book *Leaders at War: How Presidents Shape Military Interventions* (Saunders, 2011). Adcock believed that these richly documented archived primary sources provided an opportunity to teach about how authors make claims and about the evaluability of evidence in qualitative research. Saunders argues that JFK saw a central source of threats to US security in the internal politics of other states prior to his presidency—and not just as rhetorical devices to justify otherwise-motivated actions. Adcock divided up the footnotes of the sources among groups of students and had each group access and review the resources and evaluate claims in light of the full primary texts. This classroom exercise, about which Adcock reports that "intellectually it turned out even better than I expected," demonstrates the value of innovative technology such as active citation / ATI for data reuse. By seamlessly connecting data and text, the archival data used became not just available, but also reusable, even by comparatively inexperienced reusers such as Adcock's undergraduate students.

Our examples demonstrate the wide range of reuse of qualitative data. It occurs within and across disciplines, by researchers within the same research project and by complete strangers. The examples we discuss also highlight the importance of the three advances towards epistemically responsible curation that we discuss in this article: improved data management, collaborative curation, and technological advances such as ATI and QDAS data sharing.[4]

Conclusion

Qualitative researchers have looked at the reuse of qualitative data with some skepticism. In this article, we make the case that such reuse is not just possible but can be done appropriately. We argue that careful curation of data can enable what we call "epistemically responsible reuse"—reuse that aims to understand as much of the original data and context as possible but doesn't make claims beyond what the data can provide. Epistemically responsible reuse is facilitated by curation that follows three central tenets: it provides as much context as possible around the data and their collection; it is explicit about what is missing from the data; and it seeks to capture well the voice of the original researchers and their vision in the data, their documentation, and their mode of presentation.

We describe three broad recent trends that, in our assessment, will greatly enhance the ability to share data to this standard and thus the availability of reusable qualitative data and, ultimately, cases of reuse. The increasing sophistication of data management by qualitative researchers, both due to better training and due to funder demands, allows for more context to be recorded during research and thus be available as documentation for sharing. Consultative and collaborative curation, as practiced by QDR's staff, allows for a dialog between specialist curators and researchers, and thus helps to identify and describe key contextual knowledge, missing pieces of information, and ensure that data are displayed in a manner that reflects the

[4] For other examples, see Haaker & Morgan-Brett (2017) who document their experience creating pedagogical resources during their time at UK Data Services, with an example avaialble at https://ukdataservice.ac.uk/teaching-resources/folk-devils.aspx, as well as Hsiung's (2016) online resources on qualitative interviewing from archived data, *Lives and Legacies* available at http://www.utsc.utoronto.ca/~pchsiung/LAL/home.

researchers' perspective. Finally, technical advances in repository technology and new standards such as REFI-QDA and ATI allow for data to be more richly displayed by repositories and explored by researchers, thus enabling more responsible reuse.

The contributions to this issue are an example of the increasing practice of, and the possibilities for, reusing qualitative data. We are thankful to the editors for this creative initiative and for recruiting the contributors and could not be more excited that they make use of data curated and published by QDR.

References

Alexander, S. M., Jones, K., Bennett, N. J., Budden, A., Cox, M., Crosas, M., Game, E. T., Geary, J., Hardy, R. D., Johnson, J. T., Karcher, S., Motzer, N., Pittman, J., Randell, H., Silva, J. A., da Silva, P. P., Strasser, C., Strawhacker, C., Stuhl, A., & Weber, N. (2020). Qualitative data sharing and synthesis for sustainability science. *Nature Sustainability*, *3*(2), 81-88. https://doi.org/10.1038/s41893-019-0434-8

Bishop, L. (2005). Protecting respondents and enabling data sharing: Reply to Parry and Mauthner. *Sociology*, *39*(2), 333-336. https://doi.org/10.1177/0038038505050542

Bishop, L. (2016). Secondary analysis of qualitative data. In D. Silverman (Ed.), *Qualitative Research* (4th ed., pp. 395-411). Sage.

Bishop, L., & Kuula-Lummi, A. (2017). Revisiting qualitative data reuse: A decade on. *Sage Open*, *7*(1). https://doi.org/10.1177/2158244016685136

Bleich, E., & Pekkanen, R. (2013). How to report interview data. In L. Mosley (Ed.), *Interview research in political science* (pp. 84-106). Cornell University Press; JSTOR.

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. The MIT Press.

Bornat, J., Raghuram, P., & Henry, L. (2012). Revisiting the archives: A case study from the history of geriatric medicine. *Sociological Research Online*, *17*(2), 1-12. https://doi.org/10.5153/sro.2590

Chukwuma, A., Mbachu, C., Cohen, J., McConnell, M., & Bossert, T. (2017). *A qualitative study of perspectives on postnatal care referrals by traditional birth attendants in Ebonyi State, Nigeria* [Data set]. Qualitative Data Repository. https://doi.org/10.5064/F67H1GGS

Corti, L., & Gregory, A. (2011). CAQDAS comparability. What about CAQDAS data exchange? *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *12*(1). http://www.qualitative-research.net/index.php/fqs/article/view/1634

Corti, L., & Thompson, P. (1997). Latest news from the ESRC Qualitative Data Archival Resource Centre (QUALIDATA). *Social History*, *22*(1), 83-86. https://doi.org/10.1080/03071029708567990

Cramer, K. (2015). Transparent explanations, yes. Public transcripts and fieldnotes, no: Ethnographic research on public opinion. *Qualitative & Multi-Method Research*, *13*(1), 17-20. https://doi.org/10.5281/zenodo.893069

DuBois, J. M., Strait, M., & Walsh, H. (2018). Is it time to share qualitative research data? *Qualitative Psychology*, *5*(3), 380-393. http://dx.doi.org/10.1037/qup0000076

Evers, J., Caprioli, M. U., Nöst, S., & Wiedemann, G. (2020). What is the REFI-QDA Standard: Experimenting with the transfer of analyzed research projects between QDA software. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *21*(2), Article 2. https://doi.org/10.17169/fqs-21.2.3439

Feldman, S., & Shaw, L. (2019). The epistemological and ethical challenges of archiving and sharing qualitative data. *American Behavioral Scientist*, *63*(6), 699-721. https://doi.org/10.1177/0002764218796084

Fielding, N. G., & Fielding, J. L. (2000). Resistance and adaptation to criminal identity: Using secondary analysis to evaluate classic studies of crime and deviance: *Sociology*, *34*(4), 671-689. https://doi.org/10.1177/S0038038500000419

Guillemin, M., & Gillam, L. (2004). Ethics, reflexivity, and "ethically important moments" in research. *Qualitative Inquiry*, *10*(2), 261-280. https://doi.org/10.1177/1077800403262360

Haaker, M., & Morgan-Brett, B. (2017). Developing research-led teaching: Two cases of practical data reuse in the classroom. *SAGE Open*, *7*(2), 2158244017701800. https://doi.org/10.1177/2158244017701800

Hall, R. J., & Johnson, C. R. (1998). The epistemic duty to seek more evidence. *American Philosophical Quarterly*, *35*(2), 129-139. JSTOR.

Heinz, W. R., Kühn, T., & Witzel, A. (2004). A life-course perspective on work-related learning. In M. Fischer, N. Boreham, & B. Nyhan (Eds.), *European perspectives on learning at work: The acquisition of work process knowledge* (pp. 196-215). Office for Official Publications of the European Communities.

Hsiung, P.-C. (2016). Lives & legacies: A digital courseware for the teaching and learning of qualitative interviewing. *Qualitative Inquiry*, *22*(2), 132-139. https://doi.org/10.1177/1077800415617205

Johnson, G. F., Burns, M., & Porth, K. (2017). *A question of respect: A qualitative text analysis of Canadian parliamentary committee hearings on PCEPA* [Data set]. Qualitative Data Repository. https://doi.org/10.5064/F6Z31WJ1

Kapiszewski, D., & Karcher, S. (2020). Making research data accessible. In C. Elman, J. Gerring, & J. Mahoney (Eds.), *The production of knowledge: Enhancing progress in social science* (pp. 197-220). Cambridge University Press. https://doi.org/10.1017/9781108762519.008

Karcher, S. (2016, September 28). Teaching with qualitative data: An example. *QDR Blog*. https://qdr.syr.edu/qdr-blog/teaching-qualitative-data-example

Karcher, S., & Weber, N. (2019). Annotation for transparent inquiry: Transparent data and analysis for qualitative research. *IASSIST Quarterly*, *43*(2), 1-9. https://doi.org/10.29173/iq959

Kornblith, H. (1983). Justified belief and epistemically responsible action. *The Philosophical Review*, *92*(1), 33-48. JSTOR. https://doi.org/10.2307/2184520

Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., & Wutich, A. (2019). Qualitative data sharing: Data repositories and academic libraries as key partners in addressing challenges. *American Behavioral Scientist*, *63*(5), 643-664. https://doi.org/10.1177/0002764218784991

Mauthner, N. S., Parry, O., & Backett-Milburn, K. (1998). The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology*, *32*(4), 733-745. https://doi.org/10.1177/0038038598032004006

Medjedović, I., & Witzel, A. (2008). Secondary analysis of interviews: Using codes and theoretical concepts from the primary study. *Historical Social Research / Historische Sozialforschung*, *33*(3 (125)), 148-178. JSTOR.

Moravcsik, A. (2010). Active citation: A precondition for replicable qualitative research. *Political Science & Politics*, *43*(01), 29-35. https://doi.org/10.1017/S1049096510990781

Parry, O., & Mauthner, N. (2005). Back to basics: Who reuses qualitative data and why? *Sociology*, *39*(2), 337-342. https://doi.org/10.1177/0038038505050543

Pollock, K. (2012). Procedure versus process: Ethical paradigms and the conduct of qualitative research. *BMC Medical Ethics*, *13*(1), 25. https://doi.org/10.1186/1472-6939-13-25

Saunders, E. N. (2011). *Leaders at war: How presidents shape military interventions*. Cornell

University Press. https://muse.jhu.edu/book/57661

Saunders, E. N. (2015). *Data for: John F. Kennedy* [Active Citation Compilation]. Qualitative Data Repository. https://doi.org/doi:10.5064/F68G8HMM

Teperek, M., & Dunning, A. (2018, November 14). The main obstacles to better research data management and sharing are cultural. But change is in our hands. *Impact of Social Sciences*. https://blogs.lse.ac.uk/impactofsocialsciences/2018/11/14/the-main-obstacles-to-better-research-data-management-and-sharing-are-cultural-but-change-is-in-our-hands/

Tsai, A. C., Kohrt, B. A., Matthews, L. T., Betancourt, T. S., Lee, J. K., Papachristos, A. V., Weiser, S. D., & Dworkin, S. L. (2016). Promises and pitfalls of data sharing in qualitative research. *Social Science & Medicine, 169,* 191-198. https://doi.org/10.1016/j.socscimed.2016.08.004

Witzel, A., & Kühn, T. (1999). *Berufsbiographische Gestaltungsmodi: eine Typologie der Orientierungen und Handlungen beim Übergang in das Erwerbsleben (Sonderforschungsbereich 186 Arbeitspapier Nr. 61).* [Modes of professional biography: A typology of orientations and actions during the transition to working life (Collaborative Research Center 186 working paper No. 61)]. Universität Bremen.

Woods, M., Macklin, R., & Lewis, G. K. (2016). Researcher reflexivity: Exploring the impacts of CAQDAS use. *International Journal of Social Research Methodology*, *19*(4), 385-403. https://doi.org/10.1080/13645579.2015.1023964

Woods, M., Paulus, T., Atkins, D. P., & Macklin, R. (2016). Advancing qualitative research using Qualitative Data Analysis Software (QDAS)? Reviewing potential versus practice in published studies using ATLAS.ti and NVivo, 1994-2013. *Social Science Computer Review*, *34*(5), 597-617. https://doi.org/10.1177/0894439315596311

Woolgar, S. (Ed.). (1988). *Knowledge and reflexivity: New frontiers in the sociology of knowledge.* Sage.

Yardley, S. J., Watts, K. M., Pearson, J., & Richardson, J. C. (2014). Ethical issues in the reuse of qualitative data perspectives from literature, practice, and participants. *Qualitative Health Research*, *24*(1), 102-113.

Yeh, E. T. (2019). *Data for: Pastoralist decision-making on the Tibetan plateau* [Data set]. QDR Main Collection. https://doi.org/10.5064/F6WACFVR

Yoon, A. (2014). "Making a square fit into a circle": Researchers' experiences reusing qualitative data. *Proceedings of the American Society for Information Science and Technology*, *51*(1), 1-4. https://doi.org/10.1002/meet.2014.14505101140

Author Note

Sebastian Karcher is Research Assistant Professor of Political Science at Syracuse University and Associate Director of the Qualitative Data Repository (QDR). In his research, he investigates the intersection between social science, information science, and technology. Please direct correspondence to skarcher@syr.edu.

Dessislava (Dessi) Kirilova currently serves as the senior curation specialist at the Qualitative Data Repository, having joined QDR at its origin in 2012. In that time, she has contributed to shaping the repository's policies, acquisitions outreach and provided ongoing curatorial support for depositors (including especially in dealing with human participant and copyright constraints).

Christiane Pagé is Associate Director for the Center for Qualitative and Multi-Method Inquiry, Syracuse University, and a data specialist for the Qualitative Data Repository (www.qdr.org). She has published in the areas of QDAS projects data sharing, international

communication and information policy, leadership styles of NGOs, and public-private partnerships.

Nic Weber is Assistant Professor in the Information School at the University of Washington. He is also the technical director of the Qualitative Data Repository. He has a Ph.D. in Information Science from the University of Illinois.

Acknowledgements: The contributions to this issue are an example of the increasing practice of, and the possibilities for, reusing qualitative data. We are thankful to the editors for this creative initiative and for recruiting the contributors, and could not be more excited that they make use of data curated and published by QDR. We also thank the editors and two anonymous reviewers for helpful comments that significantly improved this manuscript. This article is based upon work supported by the National Science Foundation under Grant No. 1823950.

Article Citation

Karcher, S., Kirilova, D., Pagé, C., & Weber, N. (2021). How data curation enables epistemically responsible reuse of qualitative data. *The Qualitative Report, 26*(6), 1996-2010. https://doi.org/10.46743/2160-3715/2021.5012