



HHS Public Access

Author manuscript

Qual Psychol. Author manuscript; available in PMC 2019 January 18.

Published in final edited form as:

Qual Psychol. 2018 November ; 5(3): 380–393. doi:10.1037/qup0000076.

Is It Time to Share Qualitative Research Data?

James M. DuBois,

Department of Medicine, Washington University in St. Louis

Michelle Strait, and

Department of Medicine, Washington University in St. Louis; Heidi Walsh, Department of Medicine, Washington University in St. Louis.

Heidi Walsh

Department of Medicine, Washington University in St. Louis.

Abstract

Policies by the National Institutes of Health and the National Science Foundation, as well as scandals surrounding failures to reproduce the findings of key studies in psychology, have generated increased calls for sharing research data. Most of these discussions have focused on quantitative, rather than qualitative, research data. This paper examines scientific, ethical, and policy issues surrounding sharing qualitative research data. We consider advantages of sharing data, including enabling verification of findings, promoting new research in an economical manner, supporting research education, and fostering public trust in science. We then examine standard procedures for archiving and sharing data, such as anonymizing data and establishing data use agreements. Finally, we engage a series of concerns with sharing qualitative research data such as the importance of relationships in interpreting data, the risk of re-identifying participants, issues surrounding consent and data ownership, and the burden of data documentation and depositing on researchers. For each concern, we identify options that enable data sharing or describe conditions under which select data might be withheld from a data repository. We conclude by suggesting that the default assumption should be that qualitative data will be shared unless concerns exist that cannot be addressed through standard data depositing practices such as anonymizing data or through data use agreements.

Keywords

qualitative research; research ethics; data sharing; confidentiality

This paper explores the idea of sharing qualitative research data. We begin by sharing why we believe it is time to dedicate special attention to the topic. We then consider what are qualitative research data, before examining arguments in favor of and against sharing qualitative research data.

A Reflexive Introduction

In qualitative research it is standard to disclose characteristics of the research team and to engage in reflexivity (O'Brien, Harris, Beckman, Reed, & Cook, 2014; Tong, Sainsbury, & Craig, 2007). In this spirit, we begin by relating three experiences of the first author that have inspired and shaped this paper.

Experience One

Our team conducted a series of seven 90-minute focus groups with a total of 70 people on attitudes toward an organ donation protocol. Discussion was rich. When we published the paper in a leading medical journal, many of our illustrative quotes were cut or shortened. In the end, we published 14 quotes from the 7 focus groups totaling 577 words: That translates to one quote for every 45 minutes of conversation with a highly engaged group; or, one quote from 20% of participants. Eighty percent of participants never had any of their words shared in print. (DuBois, Waterman, Iltis, & Anderson, 2009)

As researchers, we served as filters of content. We made decisions about which themes were most prominent, and we decided which voices were heard in excerpt. At the least, what we shared was incomplete—it necessarily excluded the nuanced personal experiences of *most* of our participants. We wished that we could publish our transcripts in their entirety, because the discussions were interesting, complex, and somewhat surprising.

This experience inspired the creation of a new journal co-edited by the lead author, *Narrative Inquiry in Bioethics*, which publishes collections of first-person narratives from stakeholders on a shared theme such as being a living organ donor. Such an approach to understanding complex issues grants much more control to storytellers than does the traditional qualitative interview because all of the words a storyteller chooses to share end up in print, and readers can decide upon the central themes. Of course, this ordinarily comes at a price: Storytellers are rarely anonymous, and not everyone feels comfortable going public with their stories.

Experience Two

The lead author was asked to peer review a qualitative research article. The study involved a qualitative content analysis of publicly available blogs by women writing on reproductive issues. The article used quotes from the blogs, but provided no references to the sources. Some of the claims made about the women's experiences seemed to lack nuance and balance. So the reviewer did a Google search using excerpts from quotes and quickly found some of the original sources. The reviewer found that important themes went unaddressed, perhaps to the point of distorting the main conclusions of the paper. The reviewer recommended the urls be published for the blogs to enable readers to verify the interpretation of data. After all, the study did not involve a second coder with inter-rater reliability, or any form of member checking. And there is precedent for publishing urls for blogs that are analyzed for research purposes (Fox, 2015).

This led to a lively exchange among various parties: Does sharing one's data sources violate privacy—if not in a legal sense, then in some more personal sense—when the data are public? Does it matter if the data are sensitive? Would consent from the blog authors be necessary, ethically ideal, or unreasonably burdensome?

Discussion of this experience led the authors of this paper to consider whether sharing qualitative data would improve the analysis of qualitative data by increasing transparency: Authors would know that others could re-analyze their data, which might increase care to generate codes and themes that are warranted and complete. Additionally, those who study the same topic could explore new themes in the data, which might lead to richer explorations of experiences shared by research participants (or by those publishing stories—such as bloggers).

Experience Three

The lead author served on the American Psychological Association's (APA) Committee on Human Research (CHR). CHR contributed to the Board of Scientific Affairs' examination of data sharing, which culminated in a recent APA publication, "Data Sharing: Principles and Considerations for Policy Development" (American Psychological Association, 2015). Members of the board clearly saw advantages to data sharing. The document notes that data sharing "... promotes scientific progress," "... enables replication of analyses for verifying empirical findings," and "... promotes aggregation for the purposes of knowledge synthesis" (American Psychological Association, 2015).

However, the document also expresses concerns with sharing qualitative research data in particular:

"Appropriate access levels, methods, formats, and timing of data sharing vary with the type of data collected. For example, procedures for sharing data that have the potential for identifying individual human participants, violating confidentiality, or identifying sites of illegal or stigmatized behavior need to be carefully designed and monitored. *Of particular concern is the sharing of qualitative data...*" [emphasis added] (American Psychological Association, 2015).

While serving on CHR, the lead author heard stories from individuals who conducted qualitative research with clients who were suicidal or with individuals in international settings who engaged in sexual activities that were not only illegal but sometimes also punished by death. This experience of engaging other researchers in crafting the APA statement on data sharing tempered—but did not entirely quell—the enthusiasm for data sharing that developed across the first two experiences.

In this paper, we explore what are qualitative data and what makes them unique; present how data repositories generally operate; put forth arguments for sharing qualitative data; consider arguments against such sharing; and conclude by suggesting that there should be a presumption that qualitative data should be shared unless reasonably anticipated harms cannot be avoided or data cannot be meaningfully interpreted.

Contextualizing the Problem

Data are shared all the time. Large online data repositories exist where researchers can upload data files for use by other researchers (Inter-university Consortium for Political and Social Research (ICPSR), 2012; UK Data Archive, n.d.-a). Some repositories contain thousands of data sets. The largest funding agencies in the US and UK require data sharing plans for research grants (Economic and Social Research Council (ESRC), 2015; National Institutes of Health, 2015; National Science Foundation, 2014) So why does the idea of sharing qualitative data in particular merit special consideration (American Psychological Association, 2015)? Before answering this question, it is necessary to consider what constitutes qualitative data.

What Are Qualitative Data?

Qualitative researchers use diverse methods to *gather* data: Focus groups, in-depth interviews with individuals, literature searches for archival material, and longitudinal ethnographic observation of communities or individuals are among the most common (Merriam, 2002).

Qualitative researchers *analyze* data in radically different ways. Saldaña (2013) identifies 24 kinds of “first cycle” coding including: 3 grammatical methods—attribute coding, magnitude coding, and simultaneous coding; 5 elemental methods—structural coding, descriptive coding, in vivo coding, process coding, initial coding; 4 “affective methods”—emotion coding, values coding, versus coding, evaluation coding; 4 literary and language methods—dramaturgical coding, motif coding, narrative coding, verbal exchange coding; 3 exploratory methods—holistic coding, provisional coding, hypothesis coding; and 5 procedural methods—protocol coding, outline of cultural materials coding, domain and taxonomic coding, causation coding, and theming the data. All of these forms of coding occur prior to “secondary coding” methods that may focus on patterns, relationships, theory, and longitudinal factors.

Finally, different qualitative researchers set fundamentally different *goals*, often reflecting different understandings of science. Some reject the idea of contributing to generalizable knowledge, and instead focus on understanding the particular in unique contexts (Van Manen, 1990); some aim to build theory by analyzing subjective experience and exploring relationships among large number of variable (Corbin, Strauss, & Strauss, 2008); others wish to identify themes, concepts, and research questions in relatively uncharted territories prior to studying them using quantitative methods (Creswell, 2014); others seek to gather qualitative data following a quantitative study in order to explore more deeply the meaning of their findings (Aarons, Fettes, Sommerfeld, & Palinkas, 2012); still others view qualitative research as a primary means of activism, gathering data to effect social change (Steinberg & Cannella, 2012).

In the context of such fundamental differences in approaches to data gathering, data analysis, and research goals, it is no wonder that some have rejected the idea of defining qualitative research (Madill, 2015). In fact, defining it is not only difficult, but also fraught with practical risks:

...It is unhelpful to decide on features uniquely defining qualitative research. The potential negative implications of such a practice are that it excludes, marginalizes, and/or devalues methods falling outside or near the edge of presented definitions, the fields in which these methods tend to be used, and the research questions they are best placed to address. (Madill, 2015, p. 215)

However, if we want to explore the unique challenges of sharing qualitative data, it is necessary to find a way of delineating the scope of qualitative research and qualitative data. For our purposes, qualitative research projects also share in common something crucial: The data they analyze are non-numeric. Qualitative researchers might generate some numeric results (e.g., counts of codes, inter-rater reliability coefficients, and even statistics that test for differences between groups), but the fundamental data analyzed are non-numeric (Greener, 2011). We might further suggest that subjects rather than investigators determine the initial conceptual categories of responses to research questions. For example, in quantitative survey research, investigators provide 5 – 7 Likert-type response options; but in qualitative research, the subjects produce responses to questions in an open-ended fashion or generate narratives or archival material spontaneously, and investigators must then decide how best to “reduce” data to meaningful results.

In what follows, we focus on ordinary language data. While ordinary language data are not the only kind of qualitative data, they are the primary kind (Jackson, 2015). More importantly, they are the primary kind of data that can be productively shared—whether as audio files (e.g., from recorded focus groups), transcribed text (e.g., from recorded interviews), or as primary text (e.g., from blogs or archives).

Even as we assume that it is possible to identify shared features across qualitative approaches and to define qualitative research as research involving non-numeric data, we also assume that in discussing seriously the idea of sharing qualitative data, unique ethical and scientific concerns may arise from differences in the methods used to gather data (e.g., observations may influence the interpretation of interview data), the relationships built while gathering data (e.g., participants might trust a researcher to interpret data in a non-stigmatizing manner but not trust others), and research questions (e.g., participants may have consented to have data analyzed for one purpose but not another). So while it is important to consider common denominator factors in qualitative research, it is also important to consider specifics of research projects.

The Advantages of Sharing Qualitative Data

Sharing qualitative data offers several benefits to science. We consider these benefits before engaging a series of concerns with sharing qualitative data in order to clarify what motivates the entire project.

Transparency Provides the Opportunity to Verify Warrant

The world of social psychology was recently rocked by the failure to reproduce key findings of seminal studies. The Open Science Collaboration found that only 39% of 98 quantitative

studies in social and cognitive psychology could be replicated (Nosek, 2015). Moreover, in successful replications, the effect sizes were generally smaller than in the original studies.¹

There are many reasons for failures to replicate quantitative studies such as deviation in methods, being underpowered, publication bias in favor of positive findings, and reporting bias (which leads to under reporting of negative findings) (Nosek, 2015); in rare cases, it may be due to data fabrication (Fanelli, 2009; Titus, Wells, & Rhoades, 2008). Accordingly, there are many partial solutions to the problem.

The Open Science Collaboration has developed a series of Transparency and Openness Guidelines (TOP) meant to improve “public trust in science, and science itself” (B. A. Nosek et al., 2015, p. 1425). The TOP standards include posting data, codes, and materials to trusted open access repositories, and preregistering studies and analysis plans to avoid publication of significant post-hoc analyses as though they supported the originally theorized hypotheses. This is meant to reduce opportunities for p-hacking or data dredging (Ioannidis, 2005).

In one sense, there is no reason to drag qualitative research into the mire of reproducibility. Most studies that use saturation as a criterion for sampling will not have samples representative of a larger population—which is rarely the purpose of qualitative studies (Vogt, Vogt, Gardner, & Haeffele, 2014). Even with the same dataset (e.g., transcriptions) there is no reason to think that any two researchers would come to the same conclusions when conducting qualitative research on the same research questions—recall that Saldaña identifies 24 ways of doing initial coding. Qualitative studies that report high levels of inter-rater reliability typically achieve such levels of rater agreement *after* coming to a consensus on a codebook and training raters in the operationalization of codes. Our own studies provide examples of this (DuBois, 2013; DuBois & Ciesla, 2002). Even Kuula (2011), a strong proponent of sharing qualitative data, notes that “re-use of qualitative data is never a replication of qualitative research” (p. 14). Moreover, there is little evidence that qualitative research data are ever fabricated or falsified. We searched for such instances using multiple databases and found no guilty findings pertaining specifically to data fabrication or falsification, though at least one case involved multiple instances of questionable research practices (Baud, Legêne, & Pels, 2013).

Nevertheless, reminiscent of Experience 1 shared above, we think qualitative research by its very nature involves a dynamic that may interfere with research *quality*—and this is a notion that applies equally to quantitative and qualitative research.

The dynamic of concern is well described by Simmons, Nelson, and Simonsohn (2011):

The culprit is a construct we refer to as *researcher degrees of freedom*. In the course of collecting and analyzing data, researchers have many decisions to make: Should more data be collected? Should some observations be excluded? Which conditions

¹This is not so different from the field of biomedical research: several years ago, in their attempts to replicate findings from 53 landmark cancer studies, scientists at Amgen succeeded replicating findings in only 11% of cases (Begley & Ellis, 2012). But as Mom always said, “Two wrongs don’t make a right.”

should be combined and which ones compared? Which control variables should be considered? (P. 1359)

Although TOP standards were developed with quantitative studies in mind, in one fashion or another, they can be applied to qualitative studies. Qualitative studies could in principle post results, data (e.g., transcripts), codebooks, and original materials such as interview guides to repositories; and the aims and research questions guiding a study could be preregistered.

In addition to the notion of “trustworthiness,” which has gained some traction in qualitative research (Lincoln & Guba, 1985), warrant refers to having justification or evidence to support one’s assertions (reference e.g. to a philosophical dictionary). Applied to qualitative research, this means that each code in a codebook should be supported by passages of text. This is not to say that others would necessarily generate the same codebook or identify the same primary themes; but rather that they would find warrant for these in the text. Naturally, disputes about claims of warrant and comprehensiveness in the context of qualitative data analysis and reporting cannot be resolved using statistical methods (or those alone); rather, much like the philosophical notion of warrant, they must be resolved through description and argumentation, and not everyone will arrive at the same conclusion. Enabling others to test our warrant and comprehensiveness by examining our data may improve the quality of research by enabling correction and increasing attention to detail.

In the field of moral psychology, data sharing among post-doctoral trainees enable such exercises to occur. For example, (Gibbs, Basinger, & Fuller, 1992) reanalyzed the original data obtained from the open-ended Moral Judgment Test developed by Lawrence Kohlberg and found that four of Kohlberg’s six stages of moral development had strong evidence; however, two stages reflected a very different dynamic, and were at best supported on philosophical grounds.

Sharing Enables New Research with Existing Data

As noted above, there are many different ways of coding the same data, for example, by focusing on emotions, relationships, deductive concepts found in the literature on a subject, or causal factors (Saldaña, 2013). Sharing data enables other investigators to conduct novel research with the same dataset, thus maximizing the scientific findings from a project. (Corti & Thompson, 2007; Wiles, 2013) One example is the Timescapes Study, the first of its kind undertaken in the United Kingdom. Spanning five years, it explored how personal and family relationships developed and changed (Timescapes, 2012). A separate aim of the Timescapes study was to begin developing “resources” aimed at facilitating the future reuse of qualitative longitudinal data (Fielding, 2004). The project resulted in major publications and a publically available dataset for other researchers to utilize and examine

Research on Existing Data is Economically Responsible

Gathering data is often time-consuming and expensive. Costs are involved in recruiting participants, conducting interviews or focus groups, and transcribing recordings. In contrast, re-analysis of data avoids all of these financial and time investments. Permitting re-analysis of data—either to verify warrant or to explore new research questions—is simply cost-effective. It’s a form of good stewardship of what others provided to the researcher—

whether this be the time donated by participants or the funding provided by private foundations or tax-payer supported government agencies.(N. Mauthner, 2012; N. S. Mauthner & Parry, 2009; Yardley, Watts, Pearson, & Richardson, 2014)

Trainees Benefit from Honing Skills on Existing Data

It is impossible to learn how to code qualitative data without having access to data. Sharing data enables students to learn coding skills. Sharing data may also enable masters and doctoral candidates to do sophisticated, novel analyses of existing robust datasets, datasets that they may have been unable to gather due to a lack of connections to a community or lack of funding.

Transparency and Good Stewardship Foster Public Trust

As noted above, transparency and openness are essential to permitting peer and public oversight of research, which in turn may foster public trust in research. Perhaps the public also knows that even if no one actually opens a dataset to check the warrant of claims, simply believing one is being watched improves the integrity of behavior (Mazar, Amir, & Ariely, 2008).

Government agencies are also the largest funders of qualitative research. With annual battles over government budgets, it seems more important than ever to convince legislators and the public that their investments generate reasonable returns. The return on investment in a qualitative project will almost certainly be higher if we can create a culture of secondary data analysis and of training researchers using existing data sets. At a minimum, the number of publications per funded projects can be expected to increase.

How Do Repositories Protect Qualitative Research Data?

Before engaging concerns with qualitative data sharing, it is useful to clarify standard procedures for sharing qualitative data. Qualitative data are not commonly shared in the US; however, repositories do exist that have guidelines for qualitative data sharing, and in some countries, hundreds of data sets are archived (UK Data Archive, n.d.-b). Exploring common practices enables us to see what sorts of measures are taken to address obvious concerns. Accordingly, when evaluating arguments against qualitative data sharing, we can consider whether any of them are not readily addressed by standard repository policies and practices.

What Kinds of Materials are Archived?

Archiving qualitative data is not as simple as it sounds. In addition to archiving transformed data—the data produced by researchers and partially presented in research publications—it is common to archive any materials that are necessary to make sense of the data. Table 1 provides a summary of such materials.

In general, protections for deposited materials exist at three levels: Repository practices; depositing practices by the original researcher; and data re-use agreements by secondary data users. Table 2 summarizes protections offered at each of these three levels. In what follows, we focus primarily upon the protection of raw data.

Repository Practices

Data repositories protect data in at least three ways. First, they follow best practices for data security. Data are stored on secure servers behind firewalls using encryption technology (National Institutes of Health, 2015). Second, they establish policies that must be followed by researchers in preparing data for archiving and sharing; and third, they establish policies that secondary data users must follow in order to access archived data. The next two sections elucidate such policies.

Depositing Practices by Original Researcher

As a general rule, in order to obtain REC approval, data either need to be de-identified prior to archiving, or participant consent must be obtained to archive and share the data. De-identification can be accomplished in a variety of manners, and repositories frequently provide guidance (Finnish Social Science Data Archive (FSD), 2015; Inter-university Consortium for Political and Social Research (ICPSR), 2012; Swiss Foundation for Research in Social Sciences (FORS); UK Data Archive, n.d). Such strategies may involve omitting characteristics such as last names; recoding values (e.g., making specific values more general such as changing “teacher” to “professional” or “AIDS” can be changed to “serious, long-term illness”); designate sensitive sections that cannot be quoted or mentioned; fictionalize non-essential information (e.g., change profession, age, or number of children); and when an individual’s information cannot be adequately de-identified, omitting the case from the dataset (Saunders, Kitzinger, & Kitzinger, 2015).

Data Use Agreements by Secondary Researchers

Secondary users of archived data sets typically need to sign a data use agreement. The terms of such agreements usually stipulate that they must not attempt to re-identify participants; must not present in publications any identifiable information (if it exists in data set)—adhering to the same requirements of confidentiality original research had; must have REC approval prior (particularly of data are identifiable or re-identifiable); must establish adequate data protection procedures (e.g., password protection); and must cite data source (Inter-university Consortium for Political and Social Research (ICPSR), 2015; UK Data Archive (UKDA), 1967).

Engaging Concerns with Data Sharing

Concerns with qualitative data sharing generally fall into four broad categories: (a) scientific quality; (b) permission; (c) confidentiality; and (d) burden on researcher. In what follows, we present concerns in each category and offer preliminary responses.

Scientific Quality

Concerns about the quality of qualitative data analysis typically focus on the subjectivity of some approaches to qualitative research, the importance of context in analyzing qualitative data, or the risks of anonymizing data.

Qualitative Research is Subjective.—Roller and Lavrakas (2015) refer to the “researcher as the data-gathering instrument” (p. 5). Others speak of the [expand with at

least 2 other very brief quotes and citations] Given the subjectivity of the qualitative research enterprise, it may seem strange to post data to repositories as though they are objective data points that could be analyzed objectively.

Response.—First, much of the subjectivity of qualitative research has to do with what questions are asked, how they are asked, and what information is provided. But once data exist, it is possible for others to code the data in novel ways. In fact, the subjectivity and potential bias that researchers bring to a qualitative research project increases, rather than decreases, the potential value of having others examine the data. This insight underlies common practices for ensuring quality in data coding—such as using multiple coders, discussing interpretations with groups, or member checking (Barbour, 2001).

Kuula (2011) also observes a certain irony in this objection to sharing qualitative data, which is frequently associated with claims that qualitative research rejects positivist and post-positivist epistemologies (Mauther & Parry, 2009):

The perception behind the idea that the original researcher is the only one capable of analyzing the data correctly means that the original methodology is the orthodox way to understand research data. What this implies is that the original researcher has an exclusive right to define the characteristics and nature of the empirical world under investigation. That is an odd presupposition for a research paradigm that often accuses quantitative research of naïve realist epistemology. There are few empirical methods in social sciences that can be defined as neutral or unbiased. Even the ethnographic gaze is always partial, not all-embracing. P. 14

Context and Relationships Matter.—One of the guiding assumptions behind qualitative research is that context and relationships matter (Roller & Lavrakas, 2015). Qualitative research often involves triangulation—examining one kind of data in relationship to other data sources, including research observations (Barbour, 2001). Researchers often keep journals or field notes to help them control for bias (Kawulich, 2005) and to keep the context of statements in mind. Thus, it may be questionable whether a secondary data user—someone who was not part of the process of gathering data, can approach the data without risking misinterpreting it or at least interpreting it in an incomplete manner (Broom, Cheshire, & Emmison, 2009; Slavnic, 2013; Yardley et al., 2014).

Response—First, it would be silly not concede that data analysis is enriched by the ability to triangulate with observations and field notes. Reflexivity is severely limited when one has no role in gathering the data that are analyzed. Accordingly, it seems appropriate for secondary data users to disclose the limitations of secondary data analysis. Second, when depositing qualitative data, researchers will best serve future uses by sharing additional materials: Interview guides, field notes, codebooks, sources of triangulation, and references to publications resulting from the original study. Third, it is worth noting that the importance of context cuts both ways: When researchers share key quotes or significant statements that are excerpted from hours of conversation, they risk removing such statements from their context. Providing access to full transcripts gives others a much better chance of understanding significant statements in their fullness.

Some Forms of Anonymization Are Misleading.—One common approach to anonymizing data is to use pseudonyms rather than proper names for participants and locations, or to change details (e.g., gender, age, or ethnicity) to make identification difficult. While this approach may succeed in anonymizing data, it also introduces inaccuracies into the data and risks confusing the context of statements; for example, gender may be highly relevant to the individual's experience (Hammersley & Traianou, 2012).

Response.—First, it should be conceded that preparing data to be shared may involve some compromises that need to be identified as limitations. Sharing data may offer some significant advantages over not sharing data; but it is naïve to think that secondary users will have access to the same level of information that original investigators have.

Different anonymization schemes can be effective in de-identifying participants to a reasonable degree. In some cases, redacting by deleting information may be appropriate. At other times, specific information can be made more vague (e.g. “Washington University School of Medicine in St. Louis” becomes a “an academic medical center in a Midwestern city”). Anonymization schemes should be selected carefully and perhaps in dialogue with participants and RECs. Anonymization is less difficult when an anonymization scheme is created ahead of data collection (UK Data Archive, n.d.-a).

Permission and Ownership

Questions about authorization to share qualitative research data may arise with regard to the permission of research participants or the ownership of research data.

Permission from Participants as an Element of Informed Consent.—Researchers may object to sharing data if they have not asked their participants for permission. Sharing data might be viewed as a threat to privacy or a breach of trust within the interviewer-interviewee relationship.

Response.—As a general rule, it does make sense to request participant permission to share qualitative data. This can be done within the course of the informed consent discussion. In a series of focus groups with qualitative researchers, Yardley et al. (2014) discuss the widespread use of general consent to re-use of data, and stated that feedback from their center's researchers is that “the majority of participants have been willing to consent to this” (p. 111).

When such a request was not made during the consent process, data sharing may still be possible. First, if data are adequately de-identified, RECs may not require permission. Participants are unlikely to feel that trust is violated if they cannot recognize themselves in the data that are reported in re-use studies. Second, if data cannot be de-identified to that standard, it is always possible (though time-consuming and rarely 100% successful) to re-contact participants to request permission. In a study of participant attitudes toward data archiving and sharing, 169 participants from 4 qualitative research studies were contacted; 98% permitted their data to be archived (Kuula, 2011). The author speculates that when people consent to be in a study, it's because they want to support research on the topic; for most, archiving data is viewed as an extension of this original decision. This is analogous to

what our team found studying organ donation: Among those who joined a donor registry, the decision to donate an organ after death to help another person was central; they were relatively uninterested in variations in organ donor protocols (DuBois et al., 2009).

Permission Based on Data Ownership.—Broom et al. (2009) state that interviews “are the intellectual property of both the interviewer and the interviewee, creating complex issues around allowing wider access to the data” (p. 1166). APA guidance further states that:

Sharing of data must conform to applicable statutes and regulations and to prior agreements with other parties, including participants’ consent and—in the case of community-based participatory research—agreements concerning community ownership or control of the data. (American Psychological Association, 2015)

Response.—While property laws vary across nations, as a general rule, in the US, when a study is funded by a grant, then the institution owns the data. Some consent forms clarify this matter. However, as APA notes, researchers do sometimes enter into special agreements with communities that may create complex ownership or control arrangements, and special permissions may be needed to archive and share data.

It is important to recall institutional RECs and intellectual property offices may have guidelines on matters of data ownership and the permissions needed prior to sharing data.

Confidentiality and Harms

Concerns about the confidentiality of qualitative data may arise from the perspective of privacy protection for its own sake or from the risk of disclosing sensitive information that could harm participants in various ways.

Identification as Privacy Invasion.—How identifiable are qualitative data? Let’s say a data set identifies a nurse with a son who is twelve with Down’s syndrome. If the researcher publishes the location of the study (e.g., St. Vincent’s hospital in New York), it is possible that people who work at the hospital or people who generally know the woman could identify her by inference, even though not one of the HIPAA safe harbor identifiers was included in the data. Again, in a small institution simply mentioning nation of origin (Guam) and profession (phlebotomist) is enough to identify someone, even though neither piece of information is a HIPAA safe harbor identifier. De-identifying qualitative data will require more discretion.

Response.—As noted above, when depositing data in an established repository, data protections are put in place involving the repository, the original researcher and any future secondary data users. Admittedly, REC consent forms will often nevertheless disclose a risk that confidentiality could be breached despite best efforts. But arguably, a serious data breach is more likely to occur with the data held by the original investigators—which may be highly identifiable, including participant names and audio files (voice prints)—than with data have been prepared for archiving.

Many participants consider confidentiality to be important (Yardley et al., 2014). However, Silverman's book, "Doing Qualitative Research," shares a challenging perspective:

One thing that I encountered doing oral history interviewing for my PhD was that many of my interviewees did not want to be anonymized. They regarded their interview as public testimony and stated that they were looking forward to seeing their names in print in my book. When I put it to the interviewees that I would have to change their names and hide their identity they became quite upset, and one of them said that she would not have let me interview her if her identity was to be concealed.

Which made me think: who are we trying to protect? Once it was anonymized, there was nothing to stop me from tampering with the testimony knowing that there could be no comeback from the respondent. (Patrick Brindle quoted in Silverman, p. 167).

As noted above (Kuula 2011), a study of participants in four different qualitative studies found that 98% were willing to share their data. It may be that individuals who are deeply concerned about privacy simply decline to participate in research studies.

Participants may also devise ways to protect their privacy themselves in the course of interviews. For example, Graham, Grewel and Lewis (2006), who conducted in-depth interviews with 50 participants in social science studies, found that:

interviewees had ways of withholding information if they so wished even though they had not said explicitly "I do not want to answer or discuss this topic". Participants told how they had given misinformation and how they sometimes had held back or gave an outline of a reply but no details. In addition, they explained how behaving in certain ways, for instance, showing discomfort, affected the interaction and pushed the interviewer to move on so they did not need to reveal personal information concerning the issue at hand. (cited in Kuula 2011, p. 15)

Risk of Harms Beyond Privacy Invasion.—Sometimes, concerns about confidentiality represent concerns about more than privacy insofar as linking some information to individuals can lead to significant legal, social, or even physical harms. Qualitative researchers frequently study sensitive issues such as domestic abuse, substance use, and sexual practices that can transmit HIV. Disclosure of information from such individuals could lead to social stigmatization, loss of employment opportunities, legal problems, and damage to personal relationships (DuBois, 2008; Sieber, 1992). In international contexts, disclosures of homosexual activity can be punished by death (Rupar, 2014). Because qualitative research data often contain very detailed, open disclosures, participants make themselves uniquely vulnerable to harm if the confidentiality of data is not adequately protected.

Response.—There may be times when data should not be shared because the risks to participants are simply too great. NIH's current data sharing policy allows investigators to refuse to share data on grounds that privacy cannot be adequately protected. They offer the following example:

Example 1. The proposed research will involve a small sample (less than 20 subjects) recruited from clinical facilities in the New York City area with Williams syndrome. This rare craniofacial disorder is associated with distinguishing facial features, as well as mental retardation. Even with the removal of all identifiers, we believe that it would be difficult if not impossible to protect the identities of subjects given the physical characteristics of subjects, the type of clinical data (including imaging) that we will be collecting, and the relatively restricted area from which we are recruiting subjects. Therefore, we are not planning to share the data. (National Institutes of Health, 2003)

However, sensitive data can sometimes be de-identified to a reasonable degree and then further protected through other means. The NIH's second example illustrates this.

Example 2. The proposed research will include data from approximately 500 subjects being screened for three bacterial sexually transmitted diseases (STDs) at an inner city STD clinic. The final dataset will include self-reported demographic and behavioral data from interviews with the subjects and laboratory data from urine specimens provided. Because the STDs being studied are reportable diseases, we will be collecting identifying information. Even though the final dataset will be stripped of identifiers prior to release for sharing, we believe that there remains the possibility of deductive disclosure of subjects with unusual characteristics. Thus, we will make the data and associated documentation available to users only under a data-sharing agreement that provides for: (1) a commitment to using the data only for research purposes and not to identify any individual participant; (2) a commitment to securing the data using appropriate computer technology; and (3) a commitment to destroying or returning the data after analyses are completed. (National Institutes of Health, 2003)

Researcher-Focused Concerns

Data archiving and sharing comes at some expense to researchers in terms of opportunities, time, and perhaps budget.

It's My Work.—Traditionally, scholars have tremendous latitude over what is done with their intellectual products. They do not seek institutional permissions to publish articles and books. In focus groups on data sharing, qualitative researchers have expressed the sentiment that they worked hard to get their data and the data are theirs (Broom et al., 2009). Moreover, while collaboration and collegiality are norms of science, limited grant funding and limited space in high impact journals contribute to competition among researchers, which may limit the desire to share unique resources.

Response.—Ordinarily, researchers do not own their research data when they either receive funding from a government or conduct their work as employers (Clinical Tools Inc. & US Office of Research Integrity, 2006). Research grants are typically awarded to institutions, not individuals, and accordingly, institutions own data. This is clear in US law when government agencies fund research. Nevertheless, researchers deserve credit for

generating data used in published studies. Secondary data users should cite the source of the data analyzed (B. Nosek et al., 2015).

De-Identifying and Archiving Data is Time-Consuming.—De-identifying and documenting data can be time-consuming and complicated involving decisions about what variables to leave, remove, or change in order to balance anonymization with data integrity (Saunders et al., 2015). While search and replace functions can assist in de-identifying data, they can also lead to problems as described in the Finnish guidelines for anonymizing qualitative data:

Computers provide ways to carry out fast anonymisation operations. Yet one should be very careful when using find and replace techniques, and preferably replace one item at a time. Names may form part of other words: for example, when ‘Tom’ is replaced by ‘Jack’ using the ‘Replace all’ command, ‘atomic’ is also changed to ‘aJackic’. Therefore, it is best to apply changes one by one. Prior to the anonymisation, one must also check whether the same person is referred to by using different names (e.g. Tom as Thomas or Tommie).

Response.—Many research tasks are time-consuming. The key is to include such tasks in timelines and to create budgets to support them. Once de-identifying and archiving data become routine tasks, researchers will include them in their milestones and budgets. Already, the NIH treats expenses related to sharing data as allowable costs (National Institutes of Health, 2003).

Moving the Debate Forward

We have argued that there are several notable benefits to sharing qualitative research data. We have examined concerns with data sharing, and argued that most concerns are addressed by current policies and practices of data repositories, and in rare cases, data should not be shared due to risks to participants.

In some ways it is surprising that the American Psychological Association would especially flag concerns with sharing qualitative research data. In contrast to the average qualitative interview, medical records typically contain far more identifiers as a matter of routine, such as name, address, birth date, and social security number. More importantly, they typically contain the most sensitive information about an individual; for example, if someone suffers from bipolar, abuses alcohol, is gay, and has contracted HIV, that’s all in the medical record. Yet APA does not single out concern with sharing data from medical record research—presumably because we have found ways of de-identifying data and restricting access to data that satisfy reasonable concerns for privacy and confidentiality. That is to say, sometimes, ethical concerns can be allayed by experience and technological solutions. At least some data repositories appear to have the experience and technology to enable the responsible sharing of qualitative data. In our view, it is time to establish a new presumption among qualitative researchers: Data will be shared unless concerns exist that cannot be addressed by any of the options for de-identifying data and establishing data use agreements.

Acknowledgments

This research was supported in part by a grant from the National Center for Advancing Translational Science UL1 TR000448.

References

- Aarons GA, Fettes DL, Sommerfeld DH, & Palinkas LA (2012). Mixed methods for implementation research: application to evidence-based practice implementation and staff turnover in community-based organizations providing child welfare services. *Child Maltreat*, 17(1), 67–79. doi: 10.1177/1077559511426908 [PubMed: 22146861]
- American Psychological Association. (2015). Data sharing: Principles and considerations for policy development. Retrieved from Washington DC: <https://www.apa.org/science/leadership/bsa/data-sharing-report.pdf>
- Barbour RS (2001). Checklists for improving rigour in qualitative research: A case of the tail wagging the dog? *British Medical Journal*, 322(7294), 1115–1117. [PubMed: 11337448]
- Baud M, Legêne S, & Pels P (2013). Circumventing reality: Report on the anthropological work of Professor Emeritus M.M.G. Bax. Retrieved from Amsterdam: http://www.vu.nl/en/Images/20131112/Rapport/Commissie/Baud/Engelse/versie/definitief_tcm12-365093.pdf
- Begley CG, & Ellis LM (2012). Raise standards for preclinical cancer research. *Nature*, 531, 531–533.
- Broom A, Cheshire L, & Emmison M (2009). Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology*, 43(6), 1163–1180.
- Clinical Tools Inc., & US Office of Research Integrity (Producer). (2006, December 7, 2015). Guidelines for responsible data management in scientific research. Retrieved from <https://ori.hhs.gov/education/products/clinicaltools/data.pdf>
- Corbin JM, Strauss AL, & Strauss AL (2008). *Basics of qualitative research : Techniques and procedures for developing grounded theory* (3rd ed.). Los Angeles, Calif: Sage Publications, Inc.
- Corti L, & Thompson P (2007). *Secondary analysis of archived data Qualitative Research Practice*. London: SAGE.
- Creswell JW (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks: SAGE Publications.
- DuBois JM (2008). *Ethics in mental health research: Principles, guidance, and cases*. Oxford; New York: Oxford University Press.
- DuBois JM (2013). Understanding the severity of wrongdoing in healthcare delivery and research: Lessons learned from a historiometric study of 100 cases. *American Journal of Bioethics, Primary Research*, 4(3), 39–48.
- DuBois JM, & Ciesla JE (2002). Ethics education in US medical schools: A study syllabi. *Academic Medicine*, 77(5), 69–74. doi:10.1097/00001888-200205000-00019
- DuBois JM, Waterman A, Iltis A, & Anderson J (2009). Is rapid organ recovery a good idea? An exploratory study of the public's knowledge and attitudes. *American Journal of Transplantation*, 9, 2392–2399. [PubMed: 19681823]
- Economic and Social Research Council (ESRC). (2015, 2015). ESRC Research Data Policy. Retrieved from <http://www.esrc.ac.uk/files/about-us/policies-and-standards/esrc-research-data-policy/>
- Fanelli D (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLOS One*, 4(5), e5738. doi:10.1371/journal.pone.0005738 [PubMed: 19478950]
- Fielding N (2004). Getting the most from archived qualitative data: Epistemological, practical and professional obstacles. *International Journal of Social Research Methodology*, 7, 97–104.
- Finnish Social Science Data Archive (FSD). (2015, August 10, 2015). Anonymisation and Personal Data. Retrieved from <http://www.fsd.uta.fi/aineistonhallinta/en/anonymisation-and-identifiers.html>
- Fox RC (2015). *Doctors without borders: Humanitarian quests, impossible dreams of medecins sans frontiers*: Johns Hopkins University Press.

- Gibbs JC, Basinger KS, & Fuller D (1992). *Moral maturity: Measuring the development of sociomoral reflection*. Hillsdale, N.J.: L. Erlbaum Associates.
- Greener I (2011). *Designing social research: A guide for the bewildered*. Thousand Oaks, CA: Sage.
- Hammersley M, & Traianou A (2012). *Ethics in qualitative research: Controversies and context*. Thousand Oaks, CA: Sage.
- Inter-university Consortium for Political and Social Research (ICPSR). (2012). *Guide to social science data preparation and archiving: Best practice throughout the data life cycle (978-0-89138-800-5)*. Retrieved from Ann Arbor, MI: <http://www.icpsr.umich.edu/files/deposit/dataprep.pdf>
- Inter-university Consortium for Political and Social Research (ICPSR). (2015). *Terms of use*. Ann Arbor, MI: University of Michigan.
- Ioannidis JP (2005). Why most published research findings are false. *PLoS Med*, 2(8), e124. doi: 10.1371/journal.pmed.0020124 [PubMed: 16060722]
- Jackson MR (2015). Resistance to qual/quant parity: Why the “paradigm” discussion can’t be avoided. *Qualitative Psychology*, 2(2), 181–198. doi:10.1037/qup0000031
- Kawulich BB (2005). Participant observations as a data collection method. *Forum: Qualitative Social Research*, 6(2). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/466/996>
- Kuula A (2011). Methodological and ethical dilemmas of archiving qualitative data. *Iassist Quarterly*, 12–17.
- Lincoln YS, & Guba EG (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.
- Madill A (2015). Qualitative research is not a paradigm: Commentary on Jackson (2015) and Landrum and Garza (2015). *Qualitative Psychology*, 2(2), 214–220. doi:10.1037/qup0000032
- Mauther NS, & Parry O (2009). Qualitative data preservation and sharing in the social sciences: On whose philosophical terms? *Australian Journal of Social Issues*, 44(3), 291–307.
- Mauthner N (2012). Accounting for our part of the entangled webs we weave: Ethical and moral issues in digital data sharing. *Ethics in Qualitative Research* (2nd ed., pp. 157–175): SAGE.
- Mauthner NS, & Parry O (2009). Qualitative data preservation and sharing in the social sciences: On whose philosophical terms? *Australian Journal of Social Issues*, 44(3), 291–307.
- Mazar N, Amir O, & Ariely D (2008). More ways to cheat: Expanding the scope of dishonesty. *Journal of Marketing Research*, 45(6), 650–653. Retrieved from <Go to ISI>://WOS:000261527000004
- Merriam SB (2002). *Qualitative research in practice: Examples for discussion and analysis*, 1st Ed. Jossey-Bass.
- National Institutes of Health. (2015). NIH security best practices for controlled-access data subject to the NIH Genomic Data Sharing (GDS) policy (pp. 8).
- National Institutes of Health. (2003, March 5, 2003). NIH data sharing policy and implementation guidance. Retrieved from http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.html
- National Institutes of Health. (2015, February 26, 2015). NIH sharing policies and related guidance on NIH-funded research resources. Retrieved from <https://grants.nih.gov/grants/sharing.htm>
- National Science Foundation. (2014, December 26, 2014). Dissemination and sharing of research results. Retrieved from <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>
- Nosek B (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251).
- Nosek B, Alter G, Banks G, Borsboom D, Bowman S, SBrecklet S, & al, e. (2015). Scientific standards: Promoting an open research culture. *Science*, 348 (6242), 1422–1425. doi:10.1126/science.aab2374 [PubMed: 26113702]
- O’Brien BC, Harris IB, Beckman TJ, Reed DA, & Cook DA (2014). Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med*, 89(9), 1245–1251. doi:10.1097/ACM.0000000000000388 [PubMed: 24979285]
- Roller MR, & Lavrakas PJ (2015). *Applied qualitative research design: A total quality framework approach*. New York: Guilford Press.
- Rupar T (2014, February 24). Here are the 10 countries where homosexuality may be punished by death. *Washington Post*. Retrieved from <https://www.washingtonpost.com/news/worldviews/wp/2014/02/24/here-are-the-10-countries-where-homosexuality-may-be-punished-by-death/>

- Saldaña J (2013). *The coding manual for qualitative researchers* (2 ed.). Thousand Oaks, CA: Sage Publications Ltd.
- Saunders B, Kitzinger J, & Kitzinger C (2015). Anonymising interview data: Challenges and compromise in practice. *Qual Res*, 15(5), 616–632. doi:10.1177/1468794114550439 [PubMed: 26457066]
- Sieber JE (1992). *Planning ethically responsible research: A guide for students and internal review boards*. Newbury Park, Calif: Sage Publications.
- Simmons JP, Nelson LD, & Simonsohn U (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. doi:10.1177/0956797611417632 [PubMed: 22006061]
- Slavnic Z (2013). Towards qualitative data preservation and re-use—Policy trends and academic controversies in UK and Sweden. *Forum: Qualitative Social Research*, 14(2). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/1872/3520>
- Steinberg SR, & Cannella GS (Eds.). (2012). *Critical qualitative research reader*. New York: Peter Lang.
- Swiss Foundation for Research in Social Sciences (FORS). Qualitative data archiving at FORS – Policy and procedures. Retrieved from http://forscenter.ch/wp-content/uploads/2013/11/FORS_policy_and_procedures_on_qualitative_data1.pdf
- Timescapes. (2012, 2012). Timescapes: An ESRC qualitative longitudinal initiative. Retrieved from <http://www.timescapes.leeds.ac.uk/>
- Titus SL, Wells JA, & Rhoades LJ (2008). Repairing research integrity. *Nature*, 453(7198), 980–982. doi:10.1038/453980a [PubMed: 18563131]
- Tong A, Sainsbury P, & Craig J (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19(6), 349–357. [PubMed: 17872937]
- UK Data Archive. (n.d). Anonymisation. Retrieved from <http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation>
- UK Data Archive. (n.d.-a). Preparing data for deposit. Retrieved from <https://www.ukdataservice.ac.uk/manage-data>
- UK Data Archive. (n.d.-b). Reusing qualitative data. Retrieved from <https://www.ukdataservice.ac.uk/use-data/guides/methods-software/qualitative-reuse>
- UK Data Archive (UKDA). (1967). Retrieved from <http://www.data-archive.ac.uk/>. from University of Essex <http://www.data-archive.ac.uk/>
- Van Manen M (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. Albany, NY: SUNY.
- Vogt PW, Vogt ER, Gardner DC, & Haeffele LM (2014). *Selecting the right analyses for your data. Quantitative, qualitative, and mixed methods*. New York: Guilford Press.
- Wiles R (2013). *Where next for research ethics? What are qualitative research ethics?* Great Britain: Bloomsbury Academic.
- Yardley SJ, Watts KM, Pearson J, & Richardson JC (2014). Ethical issues in the reuse of qualitative data: Perspectives from literature, practice, and participants. *Qualitative Health Research*, 24(1), 102–113. doi:10.1177/1049732313518373 [PubMed: 24374332]

Table 1**Materials that Are Commonly Archived***

Interview Guide: A list of questions posed to study participants.

Raw Data: E.g., video, audio files, archival text, or transcriptions of interviews.

Codebook: A list of codes with definitions.

Research Ethics Committee/IRB approval: Documentation of study approval and archiving approval by an oversight board.

Transformed Data: E.g., raw data tagged with codes; key quotes illustrating themes; code counts; inter-rater reliability data. Such data are often produced using computer assisted qualitative data analysis software (CAQDAS).

Field Notes: E.g., in ethnographic studies, researchers may keep a diary of their thoughts on data as they are collected, their frustrations or biases, or new questions to explore.

Research Methods Protocol: A description of the methods used to collect and analyze data.

References: References to publications reporting on the qualitative data, or, if copyright laws (or open access agreements) permit, copies of the publications themselves.

Note. Not all studies will deposit all of these materials. Some materials such as field notes may not exist or may be deemed too personal to post. Other materials might be available only upon request by contacting the original researcher.

Table 2

Standard Protections of Qualitative Data in Repositories

1. By Repositories	2. By Depositing Researcher	3. By Secondary Data Users
Server is secured by firewalls and encryption Depositing researchers must follow policies aimed at protecting participant information and scientific integrity Secondary data users must follow policies aimed at protecting participant information and scientific integrity	Data must be de-identified—removing or changing as little as possible and as much as necessary For very sensitive data, entire cases may be removed Or If data are identifiable, REC approval and participant consent must be documented For very sensitive de-identified data, this may also be required	Must not attempt to re-identify data that have been de-identified Must securely store data May not publish identifiable information May not redistribute data May not use for non-research or commercial purposes Must cite original data source

Note. Protections are cumulative across all 3 columns. Not all repositories have the same policies—these are sample elements of repository policies and practices.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript