HHS Public Access

Author manuscript

Am Behav Sci. Author manuscript; available in PMC 2019 November 05.

Published in final edited form as:

Am Behav Sci. 2019 May ; 63(5): 643–664.

Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges

Sara Mannheimer,
Montana State University

Amy Pienta,
Inter-university Consortium for Political and Social Research

Dessislava Kirilova,
Qualitative Data Repository

Colin Elman,
Qualitative Data Repository

Amber Wutich
Arizona State University

Abstract

Data sharing is increasingly perceived to be beneficial to knowledge production, and is therefore increasingly required by federal funding agencies, private funders, and journals. As qualitative researchers are faced with new expectations to share their data, data repositories and academic libraries are working to address the specific challenges of qualitative research data. This paper describes how data repositories and academic libraries can partner with researchers to support three challenges associated with qualitative data sharing: (1) obtaining informed consent from participants for data sharing and scholarly reuse; (2) ensuring that qualitative data are legally and ethically shared; and (3) sharing data that cannot be deidentified. This paper also describes three continuing challenges of qualitative data sharing that data repositories and academic libraries cannot specifically address—research using qualitative big data, copyright concerns, and risk of decontextualization. While data repositories and academic libraries can't provide easy solutions to these three continuing challenges, they can partner with researchers and connect them with other relevant specialists to examine these challenges. Ultimately, this paper suggests that data repositories and academic libraries can help researchers address some of the challenges associated with ethical and lawful qualitative data sharing.

1. Introduction and background

With the growth of data-intensive research and "big science" (Hey & Trefethen, 2003), data are being increasingly aggregated and mined from new sources. "Big data" is still an ill-defined term, but generally refers to large-scale datasets from networked technologies (Metcalf & Crawford, 2016). Big data—from sources such as credit card transactions,

Corresponding Author: Sara Mannheimer, Montana State University, Bozeman, MT 59717, USA. sara.mannheimer@montana.edu.

website clickstream tracking, mobile device location tracking, fitness tracking apps, Internet of Things sensors, social media, and blogs—reflect human behavior and interactions. Consequently, big data and big data analytics have altered the landscape of industry research. As the Economist suggested in 2017, "data are to this century what oil was to the last one: a driver of growth and change" (Economist 2017). In academi a, the idea of data as a valuable commodity has taken hold in the form of data sharing. Data sharing in academia can accelerate the pace of research, encourage new research questions and design, help to avoid duplication of research, provide resources for student research, and reduce the burden on research subjects (Borgman, 2015; Lyon, 2016). Data sharing can also promote research transparency and reduce misconduct, increase researcher visibility and research partnerships, and maximize the payoff of public investments in research and education (Fry et al., 2009; Piwowar & Vision, 2013; Perrino et al., 2013). Consequently, data sharing is on the upswing across a range of scholarly communities, with further encouragement from policies instituted by federal funding agencies (NSF, 2011; NIH, 2003), private funders (Wellcome Trust, 2017; Gates Foundation, 2015) and scholarly journals (Dryad, 2011; PLOS, 2014; Taichman et al., 2017). In the social sciences, most secondary analysis has been conducted with quantitative data such as survey data. However, qualitative data are increasingly seen as having value for reuse and secondary analysis, especially as a way to produce new insights while requiring less burden on respondents (Heaton, 2004; Bishop & Kuula-Luumi, 2017). This perceived value, in conjunction with data sharing policies, has led to an increasing number of qualitative data collections being shared.

However, sharing qualitative data poses especially difficult epistemological and ethical challenges. Regarding epistemological challenges, some qualitative researchers have voiced concern that the legitimacy of the data will be compromised if they are removed from their original context. There is also concern that data will lose value without the knowledge and expertise of the researchers who designed and implemented the original research project, and prepared and analyzed the original data (Walters, 2009).[1] Regarding ethical challenges, qualitative researchers often view data as being co-created by the researcher and the research participant, which would suggest that releasing the data for secondary use is not a decision that can be made by the researcher alone (Moore, 2007). Other scholars have cited practical ethical challenges surrounding informed consent, confidentiality, and anonymity when sharing qualitative data for secondary use (Neale, 2013; Bishop, 2009; Ruggiano & Perry, 2017). As Broom Cheshire, & Emmison write, "the idea that data can be neutralized and deposited into an archive, ready to be 'picked up' by others, sits uncomfortably for many" (2009, p. 1164). On the other hand, some scholars suggest that qualitative data sharing and secondary use can be facilitated with increased planning, research rigor, transparency, and ethical interrogation (Thorme, 1994; Elman, Kapiszewski, & Vinuela, 2010). This paper supports the idea that ethical qualitative data sharing is desirable and often possible, and suggests that data repositories and academic libraries can partner with qualitative researchers to promote ethical and lawful sharing of qualitative research data, when possible.

[1]For further discussion, see "Decontextualization," below.

Secondary use of qualitative research data

Publicly available qualitative research data can be valuable resources for secondary analysis, especially when curated, documented, and preserved by a data repository. Evidence from the Inter-university Consortium for Political and Social Research (ICPSR), the largest data repository in the social sciences, suggests that there is increasing demand for qualitative data by secondary data users. For example, over the last five years, there has been a steady increase in the number of searches done on the ICPSR website that included the terms "qualitative" or "mixed method." For searches containing the terms "mixed method" or "qualitative", there has been a large 253.5% increase in the number of searches performed (142 searches in 2010 compared with 360 searches in 2014) on the ICPSR website.[2] Additionally, qualitative data that are fully deidentified and made publicly available receive considerable use at domain repositories. For example, the National Archive of Criminal Justice Data (NACJD) at ICPSR disseminates 5 studies with qualitative public-use data where datasets have been downloaded between 42 and 185 times in the past three years. Below, we provide two illustrative examples of secondary use of qualitative data. The first example, the Human Relations Area Files, are an ethnographic archives that provides qualitative data for secondary use. The second example, Parenthood in Early Twentieth-Century America Project, is a single project that has seen substantial reuse.

As these examples show, making qualitative data accessible beyond the immediate researcher and their project is an established practice, although it is still not as widespread as quantitative data sharing. Increasing the rates of data sharing by social scientists would provide a number of benefits to individual scholars and to the research enterprise as a whole. These benefits include: increasing transparency and the reliability of the evidentiary based used in publications; allowing for access to information about research contexts that other scholars might not have directly (not simply due to resource constraints, but also because events about which data were collected are in the past); and facilitating the teaching of research methods.

Data repositories

Technically, there are several options for a scholar who wants to share their research data. Until about fifteen years ago, it was not uncommon to signal one's willingness to share the data used for a publication by including a note in the text that they are available "upon request." While this approach might have seemed progressive at the time, in reality it only facilitates data sharing on an ad hoc basis, with unpredictable outcomes. Long term access to data "upon request" is far from guaranteed. The original data collectors may be hard to locate after several years. The data themselves may have been changed since the publication without a clear versioning record, or the data formats may have become obsolete, or data may be lost altogether. Moreover, without extensive documentation, the original data collectors may have difficulty remembering—let alone explaining to a secondary user unfamiliar with the original project—details of organizational or analytical choices made. In short, while admirably telegraphing one's support for transparency, a researcher who limits the sharing to "upon request" leaves too many factors vulnerable to chance and time.

[2]Between 2010 and 2014, there has been a 25.1% increase in the total number of ICPSR site searches done annually.

Another approach is to share data as downloadable files on a website, either personal or journal-sponsored.[3] This approach was common through the mid-2000s, but is no longer considered a best practice for data sharing. Most of the downsides of ad hoc sharing described above remain present in this scenario as well. Even with the more solid institutional infrastructure of journal supplements, the chance of broken links is high (Klein et al., 2014). Additionally, there is no systematic option for searching for any such materials even when they are available on the internet. A potential secondary user may or may not come upon them by chance, which drastically limits many of the benefits of data sharing.

With the help of ongoing technological and infrastructural improvements since the early days of the internet, the best current option for sharing scholarly data is to make use of professional repositories.[4] There are several different kinds of repositories, but the main advantage among all of them over the other possible venues for sharing lies in the long-term preservation they all offer, as well as the guaranteed attention to metadata (data about the data), which further enables discovery, versioning and citation. Some repositories are fully self-service, for example figshare[5] and Zenodo,[6] neither of which specialize in data only, but allow the uploading of any form of scholarly output. Other repositories employ professional staff that have deep disciplinary expertise and offer levels of curation for individual deposits, as well as guidance for data preparation within specific contexts. For example, social science qualitative data are found at the ICPSR[7] at the University of Michigan, the Qualitative Data Repository (QDR)[8] at Syracuse University, and the University of North Carolina (UNC) Dataverse at the Odum Institute for Research in Social Science at UNC Chapel Hill.[9] In addition to professional curation, a key option offered by such repositories is access restriction for data which might not otherwise be ethically or legally possible to share. But unlike in the ad hoc scenario described above, the conditions under which a legitimate researcher can seek access are pre-specified and published as part of the terms of use for a given collection.[10]

A special subset of repositories includes institutional repositories (IRs) affiliated with a university. They vary greatly depending on the data policy choices the university has made, but are typically based at the library level; employ professional staff which might or might not be dedicated to the IR operations only; are meant to house any form of scholarly output, which only sometimes includes data; and limit their services to the faculty and students on campus. Depending on both financial resources of the institution and the professional priorities of their leadership, they might or might not offer some of the more advanced

[3]This also includes the special case of institutionally backed databases or archives such as: http://www.lanic.utexas.edu/la/cb/cuba/castro.html (Fidel Castro speeches in English), http://www.cs.cornell.edu/home/llee/data/convote.html (US Congressional speeches), http://www.intereuro.eu/public/data (published documents and interviews from interest groups involved in European Union policy-making), which exhibit the same technical limitations described above and are often grant-funded and so frozen in time despite the institutional support.

[4]In a 2015 article, Swauger and Vision found that the three top reasons their respondents gave for selecting one type of repository over another were specialization, ease of use, and trust. While some of these and the remaining motivations could be conflated in the minds of researchers, nonetheless, the optimal combination of the three is probably represented by professional domain repositories as well.

[5]https://figshare.com/about

[6]http://about.zenodo.org

[7]https://www.icpsr.umich.edu/icpsrweb

[8]https://qdr.syr.edu

[9]http://odum.unc.edu

[10]For more information on restricted access, see “Restricted Access,” below.

options of other repositories, such as differential access conditions, substantive curation, or permanent identifiers.

Additionally, about seventy data repositories of various kinds (including national, domain-specific, and IRs) have currently qualified for a certification widely known as "Data Seal of Approval."[11] The self-assessment process needed to gain this sign of data management quality documents that the data archived by a particular organization can be found, understood and used in the future. In other words, the dependability and sustainability of data access remain the umbrella challenges in this sphere, and digital repositories which spend resources on both human curation and consistent maintenance of technical infrastructure are in the best position to provide the necessary assurances.

Qualitative researchers are increasingly faced with data sharing expectations from federal funding agencies, private funders, and journals. In response, data repositories and academic libraries are working to meet qualitative data sharing needs. In the next section, we outline three challenges surrounding sharing qualitative data that can be addressed through partnerships with data repositories and academic libraries. In section 3, we identify three continuing challenges surrounding qualitative data sharing. While data repositories and academic libraries don't provide solutions to these challenges, they can act as advisors and sounding boards for examining these continuing challenges, and they can connect researchers with other relevant specialists to discuss potential solutions.

2. Three challenges that can be addressed by data repositories and academic libraries

Challenge 1. Obtaining informed consent from participants for data sharing and scholarly reuse

Response: Data repositories and academic libraries can educate IRBs and researchers about planning for appropriate informed consent processes—The laws that require—and ethical imperatives that influence—the protection of the human subjects whom scholars involve in their research, represent one of the central challenges to wider sharing of qualitative data. Scholars rarely consider the ethical issues discussed in detail in their mandatory IRB application in conjunction with the possibility of sharing the data at the end of a project.[12] To the degree some do, the most likely outcomes are default assertions that collected data cannot be shared due to IRB concerns. Most IRBs, risk-averse and institutionally protective by design, remain satisfied when scholars withhold or even promise to destroy their research data. While the interplay between IRBs and funder-required data management plans (DMPs) could generate a virtuous cycle supportive of sharing data, to date the opposite has been the case. As a result, the status quo in which most

[11] As of 2017, this designation has been replaced by the *CoreTrustSeal* Trustworthy Data Repository certification which combines the earlier efforts of Data Seal of Approval (DSA) and World Data System (WDS). https://datasealofapproval.org/en/news-and-events/news/2017/9/11/coretrustseal-certification-launched

[12] For an interesting but not atypical case where an attempted deposit without real human participant concerns could not be processed merely due to such lack of advance planning, see here: https://qdr.syr.edu/qdr-blog/participant-protection-informed-consent-and-data-sharing

social science data (and possibly an even larger share of collected qualitative data) are not shared is still firmly in place.

Some exceptions to this general description exist. Some forward-looking IRBs have begun to consider the interaction between the imperatives of data sharing, research transparency, preserving confidentiality and ensuring informed consent. Cornell University's IRB's recent revisions to the consent script language it offers to its social and behavioral researchers are one example. In a dedicated Data Sharing section,[13] the suggested wording is premised on the understanding that data will be made available in an appropriate form. Importantly, the wording directly invokes two critical tools for managing the risks of sharing sensitive qualitative data: deidentification and differential data management (specifically when recordings might be made of participants). Yet such bright spots continue to be the exception.

An ongoing empirical analysis (Elman et al., 2017) of IRB guidance documents by the fifty United States universities that received the highest total amounts of National Science Foundation Social, Behavioral & Economic Sciences awards during 2016 (i.e., whose researchers are most obviously under the interacting imperatives listed above) suggests that while most of these IRBs might not promote data sharing, few of them issue explicit blanket prohibitions. Thus, the ultimate solution for this dilemma lies less in changing any formal rules than in educating actors from across the scholarly domains and coordinating their efforts on specific projects. In a related initiative, QDR is currently organizing a series of workshops that bring together IRB staff from research universities, journal editors, public and private funders, and representatives of social science associations to discuss how ethical human subjects data sharing can occur throughout the research lifecycle. The key planned outputs of the initiative are template texts for informed consent that scholars can use, which spell out the details for data sharing in a variety of contexts (with or without access restrictions, after de-identification, under a timed embargo, etc.). This is an educational and bridge-building role that other data repositories and academic libraries are well-positioned to fulfill.

Challenge 2. Ensuring that qualitative data are legally and ethically shared

Response: Data repositories and academic libraries can provide guidance and technical infrastructure—Many data repositories and academic libraries provide services that can facilitate legal and ethical qualitative data sharing, including guidance on data management planning, data deidentification, metadata and description, and terms of use.

Planning for data sharing: Data repositories and academic libraries can help researchers write a data management and sharing plan—a formal document that outlines how research data will be managed during data collection, generation, and analysis, both during a research project and once the project has concluded. While data management and sharing plans are required by some funding agencies (NSF 2011; NIH 2003), the value of a data management

[13] https://www.irb.cornell.edu/forms/#dThree

and sharing plan extends beyond simply fulfilling a requirement. This document functions as a roadmap for ethical and efficient research, including information about data access and sharing; potential secondary users; procedures for selecting data for archiving; data retention periods; procedures in place or envisioned for long-term archiving and preservation of the data; and informed consent and privacy considerations. Working with an academic library or data repository from the planning stages of their projects encourages researchers to examine and document how research data will be managed during each phase of a research project, under the guidance of a data professional. If a researcher refers to their data management and sharing plan while they collect and generate data, then relevant data management steps can be implemented as they arise, rather than retroactively. Planning for data management and data sharing also encourages organized workflows, promotes efficiency, facilitates analysis and writing, and facilitates ethical data sharing at the end of a project (Qualitative Data Repository, 2017). Thus individual researchers can pursue their professional goals and contribute to scholarship more broadly, while satisfying transparency expectations.

Data repositories and academic libraries increasingly provide data management planning guidance to researchers and in this way facilitate the achievement of these dual benefits. California Digital Library's DMPTool[14] and Digital Curation Centre's DMPonline[15] both provide online tools to facilitate data management planning, particularly in response to funder requirements. In addition to online resources, data repositories and academic libraries often provide one-on-one consultation services for researchers writing data management and data sharing plans.

Planning for curation: Even with such advance planning, curating data for sharing can be time consuming and often requires a specialized set of skills. Whether a researcher is planning for a new data collection or deciding how to share data that have been sitting for years in a file cabinet, there will be effort and time needed to prepare the data for sharing. The time and cost of curation increase if a well-documented plan did not exist or was not implemented during the data collection stage. Understanding the resources required to share data is important for planning whether the curation work will be conducted by the research team, a professional curator, or through iterative interactions between the two. A clear understanding of required data curation resources is also important when preparing a grant budget. Allocating funds to cover curation costs ensures that resources are available for the work.

Curation of qualitative data files involves documentation and organization to support future use; curation sometimes also includes deidentification guidance. Levels of required resources to address curation work are summarized in Table 1. Both number and length of files within a study increase required effort. More files require more effort to produce metadata or documentation such as understandable file names and/or a file list to help users identify and select relevant files. Length of files affects the amount of time required to review the data and remove identifying information if necessary. However, if data will be shared under restricted access conditions, then identifiable information can remain entirely

[14] https://dmptool.org
[15] https://dmponline.dcc.ac.uk

or partly in the file. Therefore, planning and preparing data for restricted access might take considerably less time than sharing the data publicly after all the necessary processing. Paper records, outdated data formats, and certain proprietary data formats add complexity and require much more effort and cost to share data. Finally, better-organized files and files that include structured elements are easier for users to work with and reduce the curation work required to make files available for secondary use. A professional data curator can improve the organization and structure of the data, thus providing better context and usability for potential secondary users.

Data Deidentification: Ideally, prior to submitting qualitative data to an archive, data contributors would remove any information that directly or indirectly identifies study participants. A best practice is that an anonymization plan is created prior to data collection and anonymizing the data occurs as qualitative files are created for analysis (ICPSR, 2012). The following are examples of modifications that can be made to qualitative data to ensure respondent confidentiality (Marz and Dunn, 2000): (1) replacing actual names with generalized text (e.g. “Mrs. Briggs” to “teacher”); (2) replacing dates, especially those referring to specific events, such birthdates; and (3) removing unique and/or publicized items. A number of tools and services exist to support the systematic deidentification of qualitative data, including within well-known software packages such as Atlas.ti and Nvivo, and the advice of a data professional well-versed in them is likely to shorten the time a researcher needs to implement this step.

Metadata and description for qualitative data: This is even more true when it comes to creating metadata on the project and file levels of a collection being prepared for sharing. Most individual scholars do not need to know the ins and outs of the structured information that is used to describe in a machine-readable (and partially human-readable) way their digital collections. What they do need is to provide detailed narrative documentation that will allow the staff of a library or data repository to create such metadata, enabling discovery and proper long-term preservation of the data. Several relevant metadata standards are applicable to qualitative data, encoding the descriptive, administrative, and structural levels of metadata. Specific to the social sciences, the Data Documentation Initiative (DDI), though created for quantitative data, is applicable, at the study-level, to describe qualitative and mixed-method studies.

Special issues that may arise with metadata of qualitative data include complex study designs and relationships between files, the need to preserve the hierarchical structure of codes, and the attachment of comments or memos to specific segments of text or to codes. Repositories that work heavily with qualitative data (for example, the UK Data Archive which has been a leader in qualitative metadata preparation) are currently working to the develop a new schema capable of incorporating object and sub-object level metadata in addition to DDI study-level metadata to address this challenge.

Thus, in order for qualitative data to be findable by and intelligible to secondary users, it is extremely important that the data are well-documented. Any information that could provide context and clarity should be provided to the data repository including: research methods and practices; copy of informed consent form with IRB approval number; details on setting

of interviews; details on selection of interview subjects; instructions given to interviewers; copies of data collection instruments; steps taken to remove direct identifiers in the data; problems that arose during the selection and/or interview process and how they were handled; and interview roster (see ICPSR, 2012 for more information). An experienced data professional consulting a depositing scholar would know which specific items to suggest be included with a given qualitative project, easing to a large degree the "decontextualization challenge" (also discussed below).

Terms of use: Data contributors generally work with data repositories to determine how data should be disseminated and under what conditions. Depending upon the repository, the legal framework guiding data sharing may allow the data contributor to select a license to document permitted uses of the data. Creative Commons is a nonprofit organization that has developed several such licenses that are appropriate for research data. For example, researchers may select a Creative Commons Zero (CC0) license to release their data to the public domain, or researchers may select a Creative Commons Attribution (CC BY) license to make data freely available for redistribution and unconstrained use, with the requirement of author attribution. On the other end of the spectrum, custom deposit agreements and data dissemination agreements designed by data repositories are often used to structure the flow of rights and responsibilities to the repositories to manage, curate, and disseminate data, but at the same time allowing for limitations and restrictions on data use and redistribution to be specified in the agreement.

Secondary data users downloading data from a repository must follow the terms in a Creative Commons or other license regarding attribution and placing additional restrictions on the data. In the case of a repository that has crafted a unique deposit agreement, users follow repository terms of use for the data that specifically prevent attempts to identify research participants, restricts the data to research use, and/or prevents redistribution (see ICPSR study number 20460[16] for an example). For restricted-access data with disclosure risks, repositories typically require that secondary users sign legal agreements that the restricted-use data will be securely stored and accessible only to authorized people. These agreements also outline the consequences of non-compliance.

Challenge 3. Sharing data that cannot be deidentified

Response: Data repositories can provide restricted access—As mentioned above, some data repositories can provide restricted access for sensitive data that cannot be deidentified. This option is useful for data that cannot be modified to protect confidentiality without significantly compromising the research potential of the data. The specific implementation of restricted access can differ depending on the entity, but it includes some combination of timed embargoes, online or offline enclaves, and secure downloads to authorized recipients only. Online enclaves offer remote access to restricted data and both online and offline enclaves typically feature third party vetting of all output before any information leaves the enclave.

[16] http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/20460/terms

Restricted access techniques can be applied either to individual files or whole projects, and are augmented by depositor and specialized end-user data use agreements as mentioned above. Such agreements are signed by the requesting investigator and the requestor's institutional representative. A typical agreement might also require the investigator requesting access to the data to obtain IRB approval for their research. Where non-deidentified, proprietary or otherwise sensitive data are involved—as is the case in much qualitative research involving human participants—such specialized management is crucial and can only be achieved through institutional sharing of the data via professional repositories.

3. Continuing challenges

In addition to the challenges that can be addressed through partnerships with data repositories and academic libraries, we suggest three key challenges that remain to be fully solved to enable data sharing and secondary use. First, as qualitative data sources increasingly include big data sources such as social networking sites and blogs, there are issues around privacy and ethics that are still unresolved. Second, textual and visual qualitative data are often constrained by copyright, raising concerns about how qualitative data can be shared while respecting proprietary rights. Third, there is a risk of decontextualizing a study through the data sharing process. While data repositories and academic libraries don't provide simple solutions to these challenges, they can partner with researchers and connect them with other relevant specialists to examine these continuing challenges and discuss potential solutions.

Continuing challenge 1. Qualitative data from big data sources

An additional challenge surrounding qualitative data reuse is the availability of "big data." While most big data is used to conduct quantitative analysis, big data from social media such as social networking sites and blogs can be used for qualitative analysis, and sharing these data sources for secondary use present as-yet-unsolved ethical challenge for qualitative researchers. Items posted on social media and blogs are unique types of qualitative data that don't neatly fit into the traditional definition of human subject data. Such data, often mined from the web without explicit consent from research subjects, have additional considerations that are not addressed by traditional ethical frameworks such as the Common Rule, and may not be subject to IRB oversight (Metcalf & Crawford, 2016; Shilton & Sayles, 2016). The ethical considerations for social media data generally relate to sensitivity of topics, vulnerability of populations, informed consent, expectation of privacy, and social media platform terms of service (Mannheimer & Hull, 2017). While posts to social networking sites and blogs can be analyzed using conventional social science research methods like ethnographic observation and close reading, they can also be mined and analyzed on a large scale using computational methods (Bruns, 2013). When conducting such large-scale analysis, obtaining informed consent from each social media user becomes impractical, if not impossible. Additionally, while social media content is often posted publicly to the web, social media users may not intend for their posts to be seen beyond their immediate community (Marwick & boyd, 2014), and they are likely not aware that their posts can be collected and used for research purposes. Most social networking platforms require that

users agree to terms of service that include consent to data mining, analysis, and research. However, even if the consent language is read and understood by social media users,[17] a blanket consent statement does not allow users to be informed about each research project that uses their data. Lastly, users are obliged to agree to terms of service in order to use social media platforms and other online services; in a society that increasingly relies on social media as a social commons for personal and professional connections, it is not reasonable to expect users to opt out of social media altogether in order to preserve their privacy (Tufekci, 2010). The issues described above have been demonstrated by several high-profile examples of social media data use in recent years, including the "emotional contagion" study in which researchers tweaked Facebook timelines in an attempt to influence users' emotional wellbeing (Kramer, Guillory, & Hancock, 2014; Meyer, 2014), an incident in which a researcher scraped data from OKCupid and shared them without any attempt at deidentification (Kirkegaard & Bjerrekær, 2016; Zimmer, 2016), and the scandal that erupted after the firm Cambridge Analytica obtained personality quiz data from tens of millions of Facebook users, and then used that data to serve targeted advertisements to Facebook users, potentially influencing voter opinions during the 2018 United States presidential election (Rosenberg, Confessore, & Cadwalladr, 2018).

Some ethical frameworks have been developed to guide researchers working with and sharing qualitative social media data (van Wynsberghe, 2013; Mannheimer, Young, & Rossmann, 2016; Weller & Kinder-Kurlanda, 2016, Mannheimer & Hull, 2017). These frameworks generally provide structures for researchers to consider issues surrounding informed consent and privacy in context[18]—including the norms of each specific social media platform and disciplinary norms in the researchers' fields. Most frameworks also encourage researchers to conduct a risk-benefit analysis, weighing the benefits of the research against the potential privacy risks to users.

More research is needed to better understand the ethical implications of social media research, and the research community needs to establish new rules of ethics that apply to research using "passively collected" data such as social media content. As privacy advocates and data professionals, librarians and data repository personnel can work with researchers to examine the ethics of qualitative social media research and sharing social media data.

Continuing challenge 2. Copyright

Scholars may be constrained from sharing data if they belong to someone else. This is most patently the case where proprietary data are provided under a user agreement that specifically limits further distribution. For example, replication in disciplines like economics face significant obstacles because of the widespread use of proprietary quantitative data which are not easily accessible by third parties.[19] For qualitative data, similar issues arise when scholars use databases of text and images, with terms of use that restrict what they are allowed to do with the material they download. Likewise, visitors to archives are often

[17]It is uncommon for users to read and fully understand terms of service. See Acquisti & Grossklags, 2005 and Steinfeld, 2016.
[18]See Nissenbaum, 2009.
[19]The editors of the *American Economic Review*, for example, report that 34 out of 83 papers received a data exemption for proprietary data (Goldberg, 2016, p. 703, Table 7).

required to agree to significant constraints on what they can do in the archive (e.g., whether they can photograph documents) and afterwards (e.g., whether they can share materials further).

Even where researchers do not explicitly opt in to restrictions by agreement, ownership rights may raise legal impediments on what they are permitted to do. Copyright is a particular intellectual proprietary right which is especially applicable to qualitative data. In the United States, statute establishes that copyright "subsists…in original works of authorship fixed in any tangible medium of expression."[20] The categories include literary, dramatic, pictorial, graphic and sculptural works. Copyright holders have exclusive rights to distribute and use the works. Per this form of intellectual property protection, when someone else holds the copyright in some of a scholar's data and she was not legally assigned that right, her ability to grant others access to those data may be limited.

While scholars must of course only make data available in ways that do not violate the law, there will often be solutions that allow the sharing of copyrighted sources. In the best circumstances, rights holders may be willing to grant permission to further share their copyrighted work for pedagogical or research purposes. Even absent such permission, however, researchers may be able to rely on the "fair use" exception.[21] As Hirtle, Hudson and Kenyon (2009, p. 89) note, fair use:

> "…ensures that the balance between the interests of copyright owners and users can be maintained and that copyright law does not stifle the very creativity it is intended to foster. On a very practical level, it provides important protections to libraries, archives, and nonprofit educational institutions. When those organizations have a reasonable belief that their use of a copyrighted work is a fair use, many of the most stringent remedies in copyright law cannot be applied."

Some types of data sharing by researchers may be more likely to fall under "fair use." For example, it is arguable that when copyrighted materials (and associated documentation describing them) are deposited for sharing in a data repository, they are being put to a new purpose. Almost universally, researchers (both those who share copyrighted sources they have used in their work and those who use copyrighted sources shared by others) will use them for scholarly (i.e., academic), educational, and/or non-commercial (i.e., non-profit-making) purposes. Moreover, if limited portions of an original are deposited because they support particular claims in a published work, then those selections may qualify under both the amount and substantiality, and the market and value, factors of fair use.

To be sure, there are some usages that would be a more challenging fit for "fair use." The wholesale reproduction of a commercial text database, for example, would raise serious

[20] 17 U.S. Code § 102-Subject matter of copyright: In general. Retrieved from https://www.law.cornell.edu/uscode/text/17/102

[21] Fair use is a limitation on the otherwise exclusive right held by a copyright owner to reproduce an original work that allows others to use a portion of that work without permission. Section 107 provides that "the fair use of a copyrighted work… for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright." The statute provides that whether or not the case falls under fair use depends on four factors: (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.

concerns. Where not as much material is employed, it is used in ways that are different from the original, and the selections do not undercut the value of the source material, the case is much easier to make. For example, a new approach to providing data for qualitative research, Annotation for Transparent Inquiry (ATI), uses "open annotation" to enrich online articles. ATI builds on "active citation," an earlier approach to achieving transparency in qualitative research pioneered by Moravcsik (e.g. 2010, 2014a, 2014b, 2014c). Scholars who use ATI produce a "data supplement" to their publication that includes digital annotations (with information about how data were generated and analyzed) as well as the underlying data sources themselves (when possible). Even where the sources are not wholly sharable, ATI encourages the inclusion of an excerpt of the text in the body of the annotation.

Librarians and repository personnel can assist scholars with finding a reasonable compromise between complying with copyright and sharing data in some form. In many cases, a repository itself might consult with a copyright librarian or lawyer in finding a creative way to allow access to the underlying data without infringing upon rights.[22]

Continuing challenge 3. Decontextualization

Another challenge to qualitative data sharing is that of decontextualization. "Context" in qualitative analysis generally refers to information beyond the text or interview that is meaningful to the analysis, ranging from rich socio-cultural histories to the micro-characteristics of the interviewer (van der Berg, 2008; Bishop, 2006, 2007). Decontextualization occurs during all primary data collection and coding, of course, but may become particularly problematic in secondary analyses if key information is not accessible to the analyst (Bernard et al., 2016; Hammersley, 2010; Bishop, 2009; Fielding, 2004). Recontextualization, or the reconstruction of data contexts, is a primary challenge in all qualitative data analysis, and may pose significant difficulties in re-analyses of qualitative datasets (Hammersley, 2010; Moore, 2007, 2006; Temple et al., 2006; Blommaert, 1997).

Social inquiry is a multifaceted enterprise (Elman, Kapiszewski, & Lupia, 2018), and the challenge of recontextualization manifests differently in different analytic traditions. In oral history, for example, there is an assumption that data archiving and sharing will be as complete as possible, and respondent consenting and consultation procedures have developed to minimize decontextualization (Parry & Mauthner, 2004). In some linguistic traditions, analysis may be confined to text generated in focal interactions, and the need for contextual information is minimal (Schegloff, 1997; van der Berg, 2008). Yet, in some analytic traditions, data deidentification may remove information that is essential for meaningful analysis and interpretation (Parry & Mauthner, 2004). In ethnography, for example, researchers are expected to gain significant contextual knowledge through long-term engagement with research communities and participants (Hammersley, 1997). In such cases, qualitative researchers may feel that key contextual information may not be fully documentable, much less transferrable (Broom et al., 2009; Mauthner et al., 1998). Ultimately, the feasibility of recontextualization in a secondary analysis depends on the research methods and aims, and the kinds of data being used (van der Berg, 2008; Bishop,

[22]For a real-life example of how the copyright issue was handled in one recent project, see Cassese, 2018, especially "Structure of the deposit" section in the Data Narrative documentation file.

2007; Moore, 2007). Our view is that data sharing and re-analysis should be instantiated in ways that fit the context of particular research traditions (Lupia & Elman, 2014; Elman & Kapiszewski, 2014).

Data repositories and academic libraries can assist researchers in understanding and minimizing the problems of decontextualization in several ways. First, librarians and data repository personnel can educate qualitative researchers about different approaches to dealing with the challenge of decontextualization, such as the well-developed methods used by oral historians. Second, librarians and data repository personnel can inform qualitative researchers of best practices in archiving the contextual information required to support secondary analyses of qualitative data (e.g., Bishop 2006, van der Berg 2008: 190–191)[23]. Third, librarians and data repository personnel can help researchers determine if a specific qualitative dataset is appropriate for archival and secondary analysis, given concerns around decontextualization and recontextualization.

Data repositories and academic libraries can also offer new and different uses of original data that are informed by the primary use. Secondary uses of qualitative data may instil some level of objectivity and reinterpretation that add further value and impact to the original research. This is important as there is growing recognition in many domains that participants in research studies provide information to researchers in exchange for the offer that that their information will be protected but that it also will be used in maximal ways to advance scientific knowledge and accelerate discovery. So, while it is true that the ubiquity of data sharing has some potential to change the very nature of some kinds of qualitative data collection efforts as researchers must pre-meditate disseminating their data and methods, there is also an ethical response to maximize data use, responsibly. Given the wide range of approaches taken by archives and repositories to embargo and/or restrict use of the data (described above), it is often possible to both ensure the integrity of various in-depth field approaches to data collection, and to share data for secondary use.

4. Conclusion

Qualitative data are valuable for a number of uses. This paper suggests three key challenges to sharing qualitative data that can be addressed by data repositories and academic libraries. To address challenge 1, obtaining informed consent from participants for future uses beyond the original research team, data repositories and academic libraries can provide guidance for working with IRBs to ensure that, to the extent possible, informed consent language includes explicit provisions for data sharing and secondary analysis. To address challenge 2, ensuring that qualitative data is ethically shared, data repositories and academic libraries can assist in creation of data management and data sharing plans, assist with deidentifying data, and assist with creation of metadata. And to address challenge 3, data that cannot be anonymized, data repositories can provide layers of restricted access. This paper also suggests three continuing challenges to sharing qualitative data that data repositories and academic libraries can discuss with researchers: qualitative big data, copyright, and risk of decontextualization. While data repositories and academic libraries cannot provide easy

[23]See also “Metadata and description for qualitative data,” above.

solutions for these challenges, they can partner with researchers to examine the complexities of these continuing challenges, and can connect researchers with other relevant specialists to discuss potential solutions.

When designing research and preparing grant budgets, researchers should consider including data repositories and academic libraries as partners, including budgeting for curation costs. Data repositories that provide high-quality curation services often charge for their services, and researchers should be prepared to budget accordingly. Data repositories and academic libraries are key partners for preparing to manage data from the outset of their project and share data effectively upon the project's completion. Ultimately, this paper proposes that qualitative data can be shared ethically and lawfully, and positions data repositories and academic libraries as key partners for qualitative researchers addressing challenges surrounding data sharing.

References

Acquisti A, & Grossklags J. (2005). Privacy and rationality in individual decision making. IEEE Security & Privacy, 3(1), 26–33. 10.1109/MSP.2005.22

Bill & Melinda Gates Foundation. (2015). Bill & Melinda Gates Foundation open access policy. Retrieved from http://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy

Bishop L. (2006). A proposal for archiving context for secondary analysis. Methodological Innovations Online, 1(2), 10–20. 10.4256/mio.2006.0008

Bishop L. (2007). A reflexive account of reusing qualitative data: Beyond primary/secondary dualism. Sociological Research Online, 12(3), 1–14. 10.5153/sro.1553

Bishop L. (2009). Ethical sharing and reuse of qualitative data. Australian Journal of Social Issues, 44(3), 255–272. 10.1002/j.1839-4655.2009.tb00145.x

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

Blommaert J. (1997). Whose background? Comments on a discourse-analytic reconstruction of the Warsaw Uprising. Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA), 7(1), 69–81.

Borgman CL (2015). Big data, little data, no data: Scholarship in the networked world. Cambridge, MA: MIT Press.

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. 10.1177/0038038509345704

Bruns A. (2013). Faster than the speed of print: Reconciling 'big data' social media analysis and academic scholarship. First Monday, 18(10). 10.5210/fm.v18i10.4879

Cassese E. (2018). Monstrosity and dehumanization in the 2016 U.S. presidential contest. Qualitative Data Repository. 10.5064/F6TB14TP

Dryad Digital Repository. (2011). Joint data archiving policy. Retrieved from http://datadryad.org/pages/jdap

The Economist. (2017, 5 6). Fuel of the future: Data is giving rise to a new economy. The Economist 423(9039), 22 Retrieved from https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy

Elman C. & Kapiszewski D. (2014). Data access and research transparency in the qualitative tradition. PS: Political Science & Politics 47(1), 43–47. 10.1017/S1049096513001777

Elman C, Hoelter LF, Kapiszewski D. & Kirilova D. (2018). IRB Guidelines and Data Sharing in the Social Sciences: Tensions and Strategies to Address Them Presentation C5, November 5, 2017.

Social, Behavioral, and Educational Research Conference of PRIM&R; San Antonio, TX 10.6084/m9.figshare.5969104.v1

Elman C, Kapiszewski D, & Lupia A. (2018). Transparent Social Inquiry: Implications for Politics Science. Annual Review of Political Science, forthcoming.

Elman C, Kapiszewski D, & Vinuela L. (2010). Qualitative data archiving: Rewards and challenges. PS: Political Science & Politics, 43(1), 23–27. 10.1017/S104909651099077X

Ember CR (2007). Using the HRAF collection of ethnography in conjunction with the standard cross-cultural sample and the ethnographic atlas. Cross-Cultural Research, 41 (4), 396–427. 10.1177/1069397107306593

Fielding N. (2004). Getting the most from archived qualitative data: epistemological, practical and professional obstacles. International journal of social research methodology, 7(1), 97–104. 10.1080/1364557030001640699

Fry J, Lockyer S, Oppenheim C, Houghton J. & Rasmussen B. (2009). Identifying benefits arising from the curation and open sharing of research data. UK Higher Education and Research Institutes. Retrieved from http://ie-repository.jisc.ac.uk/279/

Goldberg PK(2016). Report of the Editor: American Economic Review. American Economic Review, 106(5), 700–712. 10.1257/aer.106.5.700

Hammersley M. (2010). Can we re-use qualitative data via secondary analysis? Notes on some terminological and substantive issues. Sociological Research Online, 15(1), 1–7. 10.5153/sro.2076

Hammersley Martyn (1997). Qualitative data archiving: some reflections on its prospects and problems. Sociology, 31(1), 131–142. 10.1177/0038038597031001010

Heaton J. (2004). Reworking Qualitative Data. Thousand Oaks, CA: Sage.

Hey T, & Trefethen A. (2003). The data deluge: an e-Science perspective. In Berman F, Fox G, & Hey T. (eds.), Grid computing, 809–824. 10.1002/0470867167.ch36

Hirtle Peter B., Hudson Emily, and Kenyon Andrew T.. (2009) Copyright and cultural institutions: guidelines for digitization for U.S. libraries, archives, and museums. Ithaca, NY: Cornell University Library Press, Forthcoming; University of Melbourne Legal Studies Research Paper No. 434. Retrieved from https://ssrn.com/abstract=1495365

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 (5th ed.). Ann Arbor, MI. Retreived from http://www.icpsr.umich.edu/files/deposit/dataprep.pdf

Klein M, Van de Sompel H, Sanderson R, Shankar H, Balakireva L, Zhou K, & Tobin R. (2014). Scholarly context not found: one in five articles suffers from reference rot. PLOS One, 9(12), e115253. 10.1371/journal.pone.0115253

Kirkegaard EO, & Bjerrekær JD (2016). The OKCupid dataset: A very large public dataset of dating site users. Open Differential Psychology. 10.26775/0DP.2016.11.03

Kramer AD, Guillory JE, & Hancock JT (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788–8790. 10.1073/pnas.1320040111

LaRossa R. (2009). Parenthood in Early Twentieth-Century America Project (PETCAP), 1900–1944 (ICPSR 6876). 10.3886/ICPSR06876.v2

Lupia A, & Elman C. (2014). Openness in political science: Data access and research transparency. PS: Political Science & Politics, 47(1), 19–42. 10.1017/S1049096513001613

Lyon L. (2016). Transparency: the emerging third dimension of open science and open data. Liber quarterly, 25(4). 10.18352/lq.10113

Mannheimer S, Young SWH, & Rossmann D. (2016). On the ethics of social network research in libraries. Journal of Information, Communication, and Ethics in Society 14(2), 139–151. 10.1108/TICES-05-2015-0013

Mannheimer S, & Hull EA (2017). Sharing selves: Developing an ethical framework for curating social media data. International Journal of Digital Curation 12(2).

Marwick AE, & boyd d. (2014). Networked privacy: How teenagers negotiate context in social media. New Media & Society, 16(7), 1051–1067. 10.1177/1461444814543995

Mauthner NS, 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. 10.1177/0038038598032004006

Metcalf J, & Crawford K. (2016). Where are human subjects in big data research? The emerging ethics divide. Big Data & Society, 3(1). https://doi.org/2053951716650211

Meyer R. (6 28, 2014). Everything we know about Facebook's secret mood manipulation experiment. The Atlantic. https://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/

Moore N. (2006). The contexts of context: Broadening perspectives in the (re) use of qualitative data. Methodological Innovations Online, 1(2), 21–32. 10.4256/mio.2006.0009

Moore N. (2007). (Re) using qualitative data? Sociological Research Online, 12(3), 1–13. 10.5153/sro.1496

Moravcsik A. (2010). Active Citation: A Precondition for Replicable Qualitative Research. PS: Political Science & Politics 43(1):29–35. 10.1017/S1049096510990781.

Moravcsik A. (2014a). Transparency: The Revolution in Qualitative Political Science. PS Political Science & Politics 47(1):48–53. 10.1017/S1049096513001789

Moravcsik A. (2014b). One Norm, Two Standards: Realizing Transparency in Qualitative Political Science. The Political Methodologist. Retrieved from https://thepoliticalmethodologist.com/2015/01/01/one-norm-two-standards-realizing-transparency-in-qualitative-political-science

Moravcsik A. (2014c). Trust, but Verify: The Transparency Revolution and Qualitative International Relations. Security Studies 23(4): 663–688. 10.1080/09636412.2014.970846

Murdock GP (1961). Outline of cultural materials. New Haven, CT: Human Relations Area Files, Inc.

National Institutes of Health. (2003). NIH data sharing policy and implementation guidance. Retrieved from http://grants.nih.gov/grants/policy/datasharing/datasharingguidance.htm

National Science Foundation. (2011). Dissemination and sharing of research results. Retrieved from http://www.nsf.gov/bfa/dias/policy/dmp.jsp

Neale B. (2013). Adding time into the mix: stakeholder ethics in qualitative longitudinal research. Methodological Innovations Online, 8(2), 6–20. 10.4256/mio.2013.010

Nissenbaum H. (2009). Privacy in context: Technology, policy, and the integrity of social life. Palo Alto, CA: Stanford University Press.

Parry O, & Mauthner NS (2004). Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data. Sociology, 38(1), 139–152. 10.1177/0038038504039366

Perrino T, Howe G, Sperling A, Beardslee W, Sandler I, Shern D, Pantin H, Kaupert S. Cano N, Cruden G, Bandiera F, & Brown CH (2013). Advancing science through collaborative data sharing and synthesis. Perspectives on Psychological Science, 8(4), 433–444. 10.1177/1745691613491579 [PubMed: 24244216]

Piwowar HA, & Vision TJ (2013). Data reuse and the open data citation advantage. PeerJ, 1, e175 10.7717/peerj.175 [PubMed: 24109559]

PLOS. (2014). Data availability. Retrieved from http://journals.plos.org/plosone/s/data-availability

Qualitative Data Repository (2017). Planning Data Management. Retrieved from https://qdr.syr.edu/guidance/managing/planning-data-management

Rosenberg M, Confessore N, & Cadwalladr C. (3 17, 2018). How Trump consultants exploited the Facebook data of millions. The New York Times. https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

Ruggiano N, & Perry TE (2017). Conducting secondary analysis of qualitative data: Should we, can we, and how? Qualitative Social Work. 10.1177/1473325017700701

Schegloff EA (1997). Whose text? Whose context?. Discourse & society, 8(2), 165–187. 10.1177/0957926597008002002

Shilton K, & Sayles S. (2016). "We aren't all going to be on the same page about ethics": ethical practices and challenges in research on digital and social media. 2016 49th Hawaii International Conference on System Sciences (HICSS), 1909–1918. 10.1109/HICSS.2016.242

Steinfeld N. (2016). “I agree to the terms and conditions”: (how) do users read privacy policies online? An eye-tracking experiment. Computers in human behavior, 55, 992–1000. https://doi.org/10.1016Xj.chb.2015.09.038

Swauger S, & Vision TJ (2015). What Factors Influence Where Researchers Deposit their Data? A Survey of Researchers Submitting to Data Repositories. International Journal of Digital Curation 10(1), 68–81. 10.2218/ijdc.v10i1.289

Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, Hong S, Haileamlak A, Gollogly L, Godlee F, Frizelle FA, Florenzano F, Drazen JM, Bauchner H, Baethge C, Backus J. (2017). Data Sharing Statements for Clinical Trials—A Requirement of the International Committee of Medical Journal Editors. New England Journal of Medicine, 376, 2277–2279. 10.1056/NETMe1705439 [PubMed: 28581902]

Tufekci Z. (2010). Facebook, Network Externalities, Regulation. Technosociology: our tools, ourselves. Retrieved from http://technosociology.org/?p=137

Van den Berg H. (2008). Reanalyzing qualitative interviews from different angles: The risk of decontextualization and other problems of sharing qualitative data. Historical Social Research/Historische Sozialforschung, 6(9), 179–192. 10.17169/fqs-6.1.499

van Wynsberghe A, Been H, & Keulen M. (2013). To use or not to use: guidelines for researchers using data from online social networking sites. UK: RRI, Rict Responsible Innovation Retrieved from https://research.utwente.nl/en/publications/to-use-or-not-to-use-guidelines-for-researchers-using-data-from-o

Walters P. (2009). Qualitative archiving: Engaging with epistemological misgivings. Australian Journal of Social Issues, 44(3), 309–320. 10.1002/j.1839-4655.2009.tb00148.x

Wellcome Trust. (2017). Policy on data, software and materials management and sharing. Retrieved from https://wellcome.ac.uk/funding/managing-grant/policy-data-software-materials-management-and-sharing

Weller K, & Kinder-Kurlanda KE (2016). A manifesto for data sharing in social media research. Proceedings of the 8th ACM Conference on Web Science: 166–172. 10.1145/2908131.2908172

Wutich A. and Brewis A.. (2014). Food, water, and scarcity: toward a broader anthropology of resource insecurity. Current Anthropology, 55(4). 10.1086/677311

Zimmer M. (2016). OkCupid study reveals the perils of big-data science. Wired Magazine. https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science

Human Relations Area Files

The Human Relations Area Files (HRAF) are the oldest ethnographic archives in the United States (Murdoch, 1961). Founded in 1935, HRAF contains ethnographic data collected from over 300 world cultures. Rather than archiving raw field notes, each entry contains a longitudinal record of field reports and ethnographic writings that contextualize and interpret rich participant-observation data. All entries are then coded using the Outline of Material Cultures, a coding scheme that covers a wide range of cultural topics (Murdock, 1961). Over the years, this has facilitated hypothesis-testing quantitative analyses on varied topics including warfare, ethnomedicine, and climate change (Ember, 2007). Yet, the data are also suitable for qualitative analyses, such as an exploratory analysis of household responses to extreme water scarcity (Wutich & Brewis, 2014). Difficulties of working with HRAF data are well-documented, and include missing data, observer bias, and decontextualization (Heaton, 2004). Nevertheless, HRAF remains a unique and valuable resource for secondary analyses of cross-cultural ethnographies.

Parenthood in Early Twentieth-Century America Project (PETCAP)

Another example of qualitative data reuse is the Parenthood in Early Twentieth-Century America Project (PETCAP), a large qualitative study funded by the National Science Foundation (LaRossa, 2009). In 1996, Ralph LaRosa of Georgia State University deposited PETCAP at ICPSR. The study provided information on parenting, especially fathers' roles, in the early part of the twentieth century in the United States. The collection comprised transcriptions of original handwritten and published materials relating to infant and child care dating from the turn of the century into World War II and includes: (1) popular magazine articles, (2) letters to educator and author Angelo Patri (1876–1965) and his replies, and (3) letters to the United States Children's Bureau, along with the Bureau's replies. This large data collection consists of 1,428 text files. The data collection was first released by ICPSR in April 14, 1997. Over the past 20+ years, the files (data and/or documentation) have been downloaded 1,118 times.

Table 1.

Expected Resources Required to Share Qualitative Research Data

	Least Effort	Moderate Effort	High Effort
# of Files	<50	50–100	>100
Length of Files (average per file)	Short (<20 pages)	Medium (20–60 pages)	Long (>60 pages)
Anticipated Release Level When Shared	Data Enclave	Restricted Access Data with Data Use Agreement	Standard Public Download
Original Format	Plain Text, Rich Text	Portable Document Format (PDF), Microsoft Word	Proprietary (e.g., Microsoft Excel, Microsoft Access, Atlas.ti, NVivo) or obsolete format (e.g., WordPerfect)
Electronic vs. Paper	Text-searchable electronic file	Non-searchable scanned image or Electronic file-paper mix	Paper only
Study File Organization	Highly Organized	Moderately Organized	Poorly or Not Organized
Internal Structure of Files	Structured	Semi-Structured	Unstructured