Published in final edited form as:

J Empir Res Hum Res Ethics. 2018 February ; 13(1): 61–73. doi:10.1177/1556264617744121.

Examining Data Repository Guidelines for Qualitative Data Sharing

Alison L. Antes,

Department of Medicine, Washington University School of Medicine

Heidi Walsh,

Department of Medicine, Washington University School of Medicine

Michelle Strait,

Department of Medicine, Washington University in St. Louis

Cynthia R. Hudson-Vitale, and

Data & GIS Services, Washington University in St. Louis Libraries

James M. DuBois

Department of Medicine and (secondary) Department of Psychology and Brain Sciences,
Washington University in St. Louis

Abstract

Qualitative data provide rich information on research questions in diverse fields. Recent calls for increased transparency and openness in research emphasize data sharing. However, qualitative data sharing has yet to become the norm internationally, and is particularly uncommon in the United States. Guidance for archiving and secondary use of qualitative data is required for progress in this regard. In this study, we review the benefits and concerns associated with qualitative data sharing, and then describe the results of a content analysis of guidelines from international repositories that archive qualitative data. A minority of repositories provide qualitative data sharing guidelines. Of the guidelines available, there is substantial variation in whether specific topics are addressed. Some topics, such as removing direct identifiers, are consistently addressed, while others, such as providing an anonymization log, are not. We discuss the implications of our study for education, best practices, and future research.

Keywords

qualitative research; qualitative data; data sharing; data archiving; secondary data use; data repositories; research ethics

Qualitative methods—for example, in-depth interviews, focus groups, and ethnographic observation—are commonly used in fields as diverse as social work, nursing, anthropology, psychology, bioethics, and clinical research (Gibson, Timlin, Curran, & Wattis, 2004; Merriam, 2002; O'Brien, Harris, Beckman, Reed, & Cook, 2014; Saldaña, 2013). The

qualitative data generated in these fields provide rich information to explore complex questions not adequately addressed via quantitative methods. Qualitative data are essential for understanding the experiences and perspectives of individuals and communities (Corbin & Strauss, 2014; Hardy, Hughes, Hulen, & Schwartz, 2016).

Qualitative researchers vary greatly in terms of approaches to gathering and analyzing qualitative data and in the research goals pursued through qualitative methods (Roller & Lavrakas, 2015). However, the shared feature of qualitative data is that they are non-numeric (though qualitative researchers may produce numeric data such as frequencies and percentages based on coded data). The primary type of qualitative data is ordinary language data; for example, audio recordings of verbal communications (e.g., in a focus group or interview), type-written text (e.g., transcribed audio recording or responses to open-ended questions in a survey), or original text from archival material (e.g., public records, field notes, or court transcripts) (Jackson, 2015; Schwandt, 2014).

The purpose of this study was to explore existing guidance on sharing qualitative data. Before examining available guidance for sharing qualitative data, we summarize advantages of sharing qualitative data, as well as potential obstacles.

Benefits Associated with Sharing Qualitative Data

Sharing qualitative data presents at least six benefits to various stakeholders in research. First, sharing data supports transparency in research—a growing concern among the scientific community in light of concerns about data reproducibility (Begley & Ellis, 2012; B. A. Nosek et al., 2015; Nuzzo, 2015). Generally, concerns with reproducibility have focused on quantitative studies and disciplines. Given the variety of statements that participants might make during an interview, the variety of methods of coding (Saldaña, 2013), and the creativity involved in theory building (Corbin & Strauss, 2014), the strict notion of “replication” or “reproducibility” of findings discussed in the context of quantitative research does not apply to qualitative research. However, sharing qualitative data might allow others to ensure that researchers have adequate justification and evidence to support their claims (DuBois, Strait, & Walsh, 2017). Researchers have significant discretion and flexibility in performing research, and this feature is even more prominent in qualitative research (DuBois et al., 2017; Simmons, Nelson, & Simonsohn, 2011). DuBois et al (2017) relate the following experience of publishing qualitative data in a medical journal with standard word limits:

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)

Qualitative researchers serve as filters of information—they decide which themes are present in interviews and serve as gatekeepers of which statements are published as representative of a theme. Accordingly, even if one espouses the view that reproduction in the strict sense does not apply to qualitative data analysis, the transparency enabled by data sharing still supports efforts to ensure the trustworthiness of data analysis. This supports the aim of fostering trust in research and could potentially enhance research quality.

Second, sharing data supports compliance with the requirements of some of the largest funding agencies and journals. In the US and the UK, the largest funding agencies require researchers to describe their plan for sharing data (Economic and Social Research Council (ESRC), 2015; National Institutes of Health, 2015; National Science Foundation, 2014). These agencies are the largest funders of qualitative research, yet in the US, qualitative data sharing has yet to become the norm (Corti, 2000; DuBois, Strait, & Walsh, 2016). In the UK, advances have been made in qualitative data sharing but are still ongoing (Corti & Van den Eynden, 2015). In addition to expectations of funding agencies, journals are increasingly calling for data sharing to foster transparency and scientific integrity (Nature, 2016; Science Magazine, 2016; Taichman et al., 2016). Many journals have signed on to the Transparency and Openness Promotion (TOP) guidelines from the Center for Open Science initiative (Center for Open Science, n.d.; B. Nosek et al., 2015).

A third advantage of qualitative data sharing includes that it permits new research with existing data. This provides a way of fostering more knowledge generation with the same resources, thereby maximizing the impact of limited grant funding and the time of participants (Kuula, 2011).

Fourth, researchers who deposit data also benefit by enjoying a higher rate of citation (Piwowar, Day, & Fridsma, 2007). This makes sense because it is routine to cite publications associated with a dataset when publishing new analyses of the data.

Fifth, approaches have recently been developed to support meta-analyses of qualitative datasets (Musheke et al., 2013; Walsh & Downe, 2005). Meta-analyses can be valuable because evidence gathered from multiple studies is typically stronger than evidence gathered from a single study—particularly when sample sizes are modest.

Finally, students often lack data to analyze when learning qualitative data analysis approaches; data sharing provides them with high quality data in areas of interest to them for purposes of training (King, 2006).

Concerns Associated with Qualitative Data Sharing

Despite the many advantages of sharing qualitative research data, the practice appears to be relatively uncommon. In 2016, the Inter-university Consortium for Political and Social Research (ICPSR), the largest repository for social science data, had fewer than 50 (out of more than 10,000) datasets consisting of qualitative research data, and most of these have omitted actual transcriptions of interviews or other original data, presenting only their interview guides, codebooks, and data counts (Interuniversity Consortium for Political and

Social Research (ICPSR), 2016a). This may be due to a series of ethical and logistical concerns with sharing qualitative research data.

Some have observed that preparing data to share is time consuming (Saunders, Kitzinger, & Kitzinger, 2015). It takes time to anonymize data, digitize data (if necessary), transfer data to file formats that best preserve it and that are accepted by a repository, and ensure adequate data documentation (i.e., “metadata”) (Cliggett, 2013; Corti, 2000; Saunders et al., 2015). Particularly when study sites are published, data anonymization requires more than removal of direct identifiers such as names, but also indirect identifiers—details such as being a psychiatric nurse from Honduras working in a small hospital in DeKalb, IL may serve to identify a participant even though they meet HIPAA safe harbor standards for de-identification. However, when such details, which may provide important contextual information, are removed, it is also necessary—and timeconsuming—to create a log of changes made to the data (whether deletions or substitutions) to preserve the integrity of the data. Furthermore, data curation requires skills researchers may not have and institutional resources may be limited (Jahnke & Asher, 2014).

Some researchers have expressed concern that qualitative research is contextspecific and relationship-dependent: How could someone who did not participate in data collection competently code the data (Broom, Cheshire, & Emmison, 2009; Yardley, Watts, Pearson, & Richardson, 2014)? There is some merit to this concern: secondary analysis is always different from primary analysis—different in what it can accomplish, and typically different in the questions it investigates (Kuula, 2011). Yet, the novel purposes of secondary data analysis may still offer value to the world of qualitative research, particularly if data are shared following best practices such as sharing not only transcripts, but also data collection instruments, codebooks, notes on difficulties encountered, a log of any changes made to the data, and publications resulting from the project.

Qualitative data contains rich details that can be sensitive; thus participant privacy, confidentiality and consent become concerns (Parry & Mauthner, 2004). The potential misuse of data in future research is a particular concern, especially when data come from vulnerable communities (Guishard, 2017; McCurdy & Ross, 2017). For example, McCurdy and Ross (2017) discuss international research on criminal activities, which could expose re-identified participants to risk of blackmail, arrest, or death. Even identifying study sites could put participants at significant risk by leading law enforcement to locales frequented by participants. Thus, not all data should be shared, and when data are shared researchers must attend to these concerns in their consent forms and through the terms for secondary use of data (Hardy et al., 2016).

These various concerns raise a series of questions about the guidance offered by data repositories. Do repositories provide guidance or tools to make the data anonymization and preparation processes as efficient as possible? Do repositories offer guidance on whether potentially identifying data points should be retained, dropped, or fictionalized? Are repositories prepared to accept identifiable data and share it only in a protected manner within a restricted data use agreement? Is participant consent or institutional review board

(IRB) permission required to deposit data? What conditions must secondary users agreed to in order to access data?

Analyzing Available Guidance for Qualitative Data Sharing

Our aim in this study was to assess the existing body of guidance available from data repositories for addressing ethical, legal, and logistical matters related to depositing and secondary use of shared qualitative data. Researchers need guidance ranging from what items should be considered for deposit to approaches for anonymizing data and expectations for participant informed consent and even permissible file formats (Corti, Van den Eynden, Bishop, & Woollard, 2014; Smioski, 2011).

Through a content analysis of topics addressed by data repository guidelines, we map out the existing guidance provided for archiving and re-using qualitative data. In particular, we were interested in the topics addressed in the guidelines and whether these are shared across different repositories that archive qualitative data.

Method

We conducted a content analysis of English-language guidelines provided by social science and multidisciplinary data repositories. We coded separately two kinds of data repository guidelines: Those for the original researchers who wish to deposit data, and those for researchers interested in secondary use of qualitative data that are archived in the repository.

Content analysis is an appropriate methodology for understanding and reducing information from archival or existing data sources such as repository guidelines. Because no taxonomy of qualitative data sharing guidelines exists, we used an inductive structural coding approach, which involved identifying concepts (codes) that may apply to large segments of text, and which enable comparison of frequencies across cases (Saldaña, 2013).

Our project proceeded in five stages: (1) identification of guidelines; (2) compilation and management of guideline documents to enable uniform mark up during structural coding; (3) development of structural codes (i.e., lists of topics addressed) inductively; (4) determining the number of guidelines that address specific topics; and (5) verification of the accuracy of guideline coding through member checking, or when not possible, double coding. Below we describe each of these five stages in further detail.

1. Identification of Guidelines

Our aim was to review English-language guidelines from repositories that accept social science data or multidisciplinary data. To identify the topics addressed by data repository guidelines, we searched all international repositories listed with the Open Access Directory (2016) (OAD). OAD is an index of open access science materials maintained by the open access community.

2. Compilation and Management of Source Materials

Guidelines varied in their electronic format: Html, Word, and pdf. Using Adobe Acrobat Pro DC, we created one electronic PDF file for each set of guidelines (merging multiple files when necessary to enable coding of the entire set of guidelines from a given repository). This enabled us to document codes by highlighting text within one file for each set of guidelines.

3. Development of Codebooks

Our aim was to identify the topics addressed by data repository guidelines and to document the number of data repositories whose guidelines address specific topics. We developed two codebooks: one focused on guidelines for *depositing data* and the other on *secondary use* of deposited data. Initially, one member of our team (H.W.) reviewed the data depositing guidelines, while another team member (M.S.) reviewed the secondary use guidelines. They inductively generated structural codes to capture the contents of the guidelines (Saldaña, 2016). Inductive coding means that codes reflect the content found in the source documents, rather than categories pre-established by the researchers (e.g., based on existing taxonomies or common concepts found in the literature). They began by reviewing the two longest sets of guidelines from the Interuniversity Consortium for Political and Social Research and the UK Data Service, which were collectively 203 pages in length. As they read the guidelines, they extracted the requirements, recommendations, and topics addressed in them, and represented them with a short phrase (structural code) and a description in the codebook. After developing an initial codebook, a second rater (H.W. or M.S.) coded the same 203 pages of guidelines. Discrepant ratings were discussed during several team meetings with the principal investigator (J.M.D.) Initially topics were coded as requirements, options, prohibitions, not mentioned. This proved impossible to use reliably because guidelines varied in terms of nuance, with some tailoring recommendations to special factors such as the sensitivity of data, whether consent for deposit was obtained, or whether the secondary users were students or faculty. We thus resorted to a simpler system of coding topics as “addressed” or “not mentioned.” Consistent with the aims of structural coding, which focuses more on content-based or conceptualbased codes as opposed to specific words, we combined several codes to yield a simpler coding structure. For example, we combined “de-identify data” with “remove direct identifiers” and combined “data documentation” and “metadata.”

4. Determination of the Number of Guidelines that Address Specific Topics

After finalizing the codebook, we coded all repository guidelines. If a given guideline document addressed the topic, it was coded as “addressed,” otherwise the code was marked as “not mentioned.” Coders annotated the electronic PDF guidelines to indicate where they found a topic addressed. Additionally, to ensure the comprehensiveness of the first-round codebooks, the coders reviewed the remaining guidelines for additional content not represented in the codes. This process did not reveal any missing codes.

5. Verification of Findings

After our team coded the depositing and secondary data use guidelines, we sought to verify the coding by asking a representative at each of the repositories to review the accuracy of our coding. We emailed representatives to request that they confirm that we had accurately depicted the contents of their guidelines. To do this, we emailed them the section of the table for their guidelines and the PDF guideline documents with annotations.

Representatives at four of the repositories agreed to review our coding of their guidelines. The representatives engaged in both phone and email conversations with H.W. to discuss their guidelines and our coding. If they disagreed with our coding, we asked them to indicate their rationale and where that topic was addressed in the guideline document.

Our interactions with the repository representatives revealed that, in practice, many guidelines are unwritten and applied through the exchange between the depositor or secondary user and the repository curator. This highlights the critical role of repository curators in decisions about data depositing and secondary use, and the importance of the interaction between those submitting or using archived data and the repository. This led us to add an asterisk in our results tables to indicate that the repositories address the topic during the data curation process, but not explicitly in their guidelines.

Although it was ideal for a representative at each repository to confirm our codes (or suggest supplemental information), we anticipated that not all repositories would have adequate staff time to devote to our request (and some might not respond at all). Thus, when we emailed our request for verification of our codes to each repository, we stated that if they did not wish to review the codes themselves, we would have a second coder from our team independently apply the codes as a check on the first round of codes. Representatives from 3 repositories specifically requested that we apply the codes in a second round of coding. Representatives at 5 repositories did not respond to our follow up inquiries. Therefore, H.W. coded secondary use guidelines and M.S. coded data depositing guidelines for those 8 repositories. Next H.W. and M.S. met to discuss any discrepancies and reached consensus on each discrepancy; this process was relatively straightforward, as coders were expected to indicate specific, annotated passages within guidelines to support their application of a code.

Additionally, a code representing whether a fee was required by repositories was added after a representative at one of the repositories mentioned a required fee for depositing data. However, not all guidelines indicated whether a fee for depositing or for use was required. Thus, we ascertained this information by emailing the repository representatives and searching websites. It is possible that some repositories required a fee for depositing, secondary use, or both, but did not respond to our query or did not indicate whether or not a fee applied.

Results

In what follows, we first present our findings from the coding of guidelines for depositing data, and then present our findings from the coding of guidelines for the secondary use of

data that are archived with repositories. Complete, detailed results are presented in tables; in the text below, we highlight the most noteworthy findings presented in tables.

Identification of Guidelines

As shown in Figure 1, as of June 2015, the Data Repositories Index from the OAD listed 105 data repositories throughout the world. (Twenty-nine listings were excluded as duplicates.)

Of the 105 repositories, 25 were indexed as social science or multidisciplinary. A member of our research team visited the websites for each of these 25 social science and multidisciplinary repositories to obtain guidelines. Through representatives of two of these repositories, we identified 12 additional social science data repositories. Five of these 37 repositories operate exclusively in a non-English language. We identified and tried to contact the remaining 32 English-language (or multilingual) social science data repositories. Of these, we found that 9 repositories do not accept ordinary language qualitative research data; 11 had no qualitative data sharing guidelines; and 12 had written guidelines for qualitative data sharing. When guidelines were not openly published online, we asked representatives to share their repository guidelines with us.

Our sample of data sharing guidelines included guidelines from 12 global repositories. Australia Data Archive (2016); Czech Social Science Data Archive MEDARD Catalog (2016); Dataverse US Murray Research Center at Harvard (2016) Databrary at New York University and Penn State (2016); The Dryad Repository at North Carolina State University (2016), Finnish Social Science Data Archive (2016); Swiss Centre of Expertise in the Social Sciences (2016); Interuniversity Consortium for Political and Social Research at the Institute for Social Research at the University of Michigan (2016b); Norwegian Social Science Data Services (2016); Qualitative Data Repository at Syracuse (2016); Slovenian Social Science Data Archives (2016); UK Data Archive (2016). Of the 12 repositories, 2— The Czech Social Science Data Archive/MEDARD Catalog and the Qualitative Data Repository at Syracuse University—were exclusively for qualitative data.

Source Materials

The 12 sets of guidelines varied greatly in their length and detail, ranging from one 2-page document to a series of 11 documents totaling 122 pages. We had a total 549 pages, which provided the source material for our content analysis.

Data Depositing Guidelines

Table 1 presents the topics that appear in guidelines for depositing data and provides a definition of each. The final codebook for guidelines for depositing data included 18 topics in three categories: participant protections, documentation to be submitted, and fees. For example, removing or replacing indirect identifiers as part of anonymization was a code associated with participant protections.

Table 2 presents the number of guidelines that address each of the 18 topics. All but one repository offers depositors the option to embargo sharing their data for a specified amount of time. All but one repository also addressed the need for metadata describing the files and

materials deposited. All but two of the guidelines addressed access levels for data, which allowed depositors to specify open access or restrictions to access of the data. For example, access to the data may require permission from the depositors, or data access could be restricted to investigators at institutions holding memberships with the archive and who provide a research protocol that has been approved by an institutional review board (IRB) or research ethics committee (REC).

The most commonly addressed topics focused on protecting participants (e.g., deidentification or permission to archive) or ensuring the usability and value of the deposited data (e.g., metadata, bibliography, research methods, instruments, codebook, and DDI compliance).

None of the data depositing topics were mentioned by all repository guidelines. In fact, only half of the guidelines (6/12) addressed several topics relevant to successfully navigating ethical and regulatory issues such as: whether or how to deal with indirect identifiers; providing an anonymization log that clarifies the extent and nature of changes made to data; and providing proof of IRB/REC permission or waiver to deposit data.

Secondary Data Use Guidelines

Table 3 presents the topics address by guidelines for the secondary use of archived data. The final codebook included 14 topics in three categories: type of access, before access, and after access. For example, before accessing data, researchers might be required to explain why the data were being requested; after accessing the data, there might be time restrictions on its use.

Table 4 presents the number of guidelines that address any of the 14 topics related to secondary use of data. All of the guidelines for secondary data use mentioned that the secondary user must cite the data source in publications. In addition, all but one repository offers open access data; however, the majority (10/12) of the repositories also permit restricted access terms to be applied to secondary use of data. All but one also stated that secondary users must affirm that they would maintain data anonymity (i.e., not attempt to re-identify participants). At least 9 out of 12 repositories provided additional specific requirements of secondary users, including: requiring an application to access data with an explanation of how data will be used; prohibiting redistribution of data to other potential users; prohibiting use of data for commercial purposes; and fulfilling confidentiality and data protection requirements for a restricted data use agreement.

Discussion

We identified data repositories that accept qualitative data and have English language guidelines for depositing data or secondary use of data and analyzed the content of the guidelines. Several overarching lessons emerged from our investigation. First, many repositories did not have guidelines addressing qualitative data sharing. Only 38% of social science data repositories had written guidelines for depositing and secondary use of qualitative research data. Thus, researchers who wish to share qualitative data may be without adequate guidance to do so. To share qualitative data, researchers may engage in

guesswork about how to address practical and ethical concerns; this may undermine the usefulness of their data, or the protection of human participants. Alternatively, lack of guidance may curtail sharing of qualitative data all together.

Second, there are gaps in the topics addressed by repositories: Only one topic was addressed by 100% of the guidelines. Only half of the repositories requested an anonymization log, which is instrumental to understanding changes made to data when identifiers have been edited or removed. Additionally, while 10 of the 12 repositories mentioned removing direct identifiers only half addressed removing or replacing indirect identifiers. Moreover, one repository expressly prohibited or advised against removing or replacing indirect identifiers, citing that it could influence interpretation of the data, while others explicitly suggested it. Generally, guidance provided to qualitative researchers regarding de-identification of data should adhere to existing guidelines and laws for protecting and sharing data (e.g., the Safe Harbor Standard of the HIPAA Privacy Rule) (Malin, Benitez, & Masys, 2011). However, these legal considerations are specific to individual nations or specific types of data, which presents a unique concern when sharing data in the international context.

Some variability in how guidelines approach archiving data may be appropriate. While ensuring legal compliance, guidelines may also need to offer researchers and repositories flexibility in their approaches to sharing qualitative data, particularly given the nature of different types of data. For instance, field notes raise concerns about the privacy of the researchers, data may range from less sensitive to highly sensitive, and participant preferences regarding data sharing or even being identified may differ. This point, however, only reinforces the need for clear guidance, particularly for addressing sensitive data or data with special legal protections. This guidance is essential for investigators in the earliest stages of planning their research, as it influences decisions about the informed consent process, informed consent, and IRB/REC review.

A third lesson gleaned from our investigation is that guidelines only go so far. We learned through our interactions with repository curators that they review requests to deposit data and apply the guidelines on a case-by-case basis. When data are sensitive or confidentiality is a concern, sometimes they impose additional requirements beyond those noted in the guidelines. We anticipate that context-specific application of the guidelines is appropriate, but it underscores the importance of the decision-making and judgments of curators and investigators in this process. Again, some consensus on key considerations for qualitative data sharing would facilitate this process.

Overall, we conclude that some variation in the application of guidelines is appropriate, but that the current guidance for qualitative data sharing requires more examination and clarification. Many questions remain salient. Some guidelines required or recommended approval (either participant permission, or a research ethics committee waiver) to archive data—that is, to store it—and some also required or recommended separate approval to share it—that is for secondary users to access it. When is each of these approvals appropriate? Are research ethics committees prepared to make such decisions? Should such questions be answered simply as matters of compliance with relevant laws and regulations, or also as ethical questions, which might merit community engagement? The majority of

repositories noted that such permission may be necessary to archive data. However, just over half of the repositories mentioned requesting proof of permission (or a waiver) to share the data with others. Acquiring such permissions is easier to do during the initial informed consent process, so data producers need to be aware of this recommendation before they produce informed consent protocols (Australian National Data Service, 2016; UK Data Archive, 2016). Similar questions exist regarding the best strategies for addressing the confidentiality of data, which could range from deidentifying and publicly posting non-sensitive data to requiring a restricted use data agreement and IRB approval prior to permitting use of sensitive data.

In addition to the issues that relate primarily to the protection of participants, many repositories provide guidance on documenting data and research materials. Such practical guidance is necessary at the outset of research planning and influences whether data can be appropriately interpreted upon secondary use. There appears to be some variability in guidance in this regard. Not all guidelines require compliance with DDI standards, which provide a consistent approach for labeling and documenting data (Hoyle et al., 2012).

Before exploring what our findings suggest in terms of best practices, future research, and education, a few limitations of our approach should be noted. First, we only analyzed English-language guidelines that were provided to us by repository representative or were discoverable on repository websites. Second, our analysis of the guidelines was limited to what was written explicitly in the guidelines; some guidance may be provided to researchers by data curators only once the depositing process begins. To mitigate this concern, we sought verification from repository curators regarding our interpretation of their guidelines. We held validation conversations with four of the repositories. This is a strength in that it allowed us to verify our interpretation of some of the guidelines and to identify new topics that they address internally as they curate data; but it is also a weakness in that we did not hold validation conversations with the remaining eight repositories and thus know little about their internal processes. Indeed, our interactions with curators led us to add some codes, such as a requirement of a fee; therefore, we know some of the guidelines do not capture all details a researcher may need to know. Additionally, we coded guidelines simply as addressing or not addressing a topic, but we provided no evaluation of the quality of the guidance. In part, this is due to the fact that guidance needs to be somewhat context-specific, tailored to national standards and the kinds of data curated. Nevertheless, the project does not speak to the quality of guidance.

Best Practices

Planning upfront for archiving and future uses of data can save investigators significant trouble and time (Corti et al., 2014). Study protocols and data sharing plans should take into consideration institutional policies and repository guidelines that could hinder or facilitate data archiving and sharing. We recommend contacting institutional and repository representatives from the outset to work with them as projects commence. Notably, investigators should make appropriate preparations for data sharing in planning their human subject protection plan and their informed consent procedure (Corti et al., 2014; UK Data Archive, 2011).

Institutional best practices include acquiring requisite expertise in issues surrounding sharing and secondary use of qualitative data to provide investigators with adequate guidance. Institutions may support qualitative data sharing by publishing online their own written guidance and policies to expedite investigators' discovery of guidance.

Repositories should work with data producers to identify options for protecting participants while maintaining the usefulness of data and following legal and institutional requirements. Qualitative data sharing could be quite intimidating without clear guidance. In the United States, some data (e.g., genetic, education records, and protected health information) enjoy special legal protections (Solove & Schwartz, 2015). Depending on the level of de-identification or anonymization, the law or institutions may require data use agreements. It is further unclear what options if any, researchers have for archiving and sharing when templates for consent forms state that data will not be shared beyond the research team. Repositories sometimes have solutions to these challenges, but they are not always apparently in their written guidelines.

Research Agenda

A number of issues arise as we seek to move this line of work forward. First and foremost, some of the research ethics questions raised by this project require empirical investigation. We need to identify standards and sources of guidance on such issues as how to determine the sensitivity level of data, when data should be anonymized and how, and what level of access is appropriate for different types of data. A Delphi or other consensus process with qualitative researchers, research ethicists and lawyers, data curators, and research participants might allow us to come to agreement on basic standards for guiding sharing of qualitative data (Fink, Kosecoff, Chassin, & Brook, 1984).

It would also be beneficial to conduct research on participant preferences regarding the use of their qualitative research data. At least some evidence exist that participants in qualitative research may actually wish to be associated with their responses (DuBois et al., 2016; Kuula, 2011; Yardley et al., 2014). Nevertheless, preferences may differ significantly depending on a variety of variables such as culture, the sensitivity of data, and whether data are shared publicly or through restricted access data use agreements; few data speak to these specific variables.

Obtaining data on researchers' preferences may be as important as obtaining data on participant perspectives. We suspect there may be field differences regarding whether it is acceptable to change variables (e.g., a participant's age or profession) in order to anonymize data or even whether one's primary data are suitable for sharing; for instance, anthropologists may see their field notes as their own intellectual property or highly private reflections and thus may not be keen on sharing them.

Educational Implications

In the process of developing this project, we identified more than 70 unique, in print qualitative research methods textbooks on a popular online bookstore's website. None of the tables of contents dedicated a chapter to sharing qualitative research data. Ideally investigators would learn early in their training issues in data sharing so that they are more

prepared to address these issues in their research. Additionally, investigators and students could be prepared to use archived data in a manner that allows the research community to draw knowledge from resources already expended. Along these same lines, investigators should be well-trained to recognize and articulate limitations in secondary data use.

Educational implications for ethics committee members include preparing members to review qualitative protocols that call for data sharing. Additionally, curators will need to be knowledgeable about human participant regulations in various nations, and they may need to understand common institutional policies and legal requirements on particular types of data.

Another pertinent question is at what point should funding agencies begin to require data sharing of qualitative data? With the increasing call for data sharing by funding agencies and journals, such a call may be just around the corner. Can, and should, grant reviewers play a role in advancing qualitative data sharing, for example, by expecting plans for sharing not only quantitative but also qualitative data in grant proposals? Growth in expectations for sharing qualitative data would require grant reviewers to be prepared for this activity and educated appropriately to review plans for sharing qualitative data.

Conclusion

Sharing qualitative data presents several advantages, but several concerns must be addressed to do it well. We analyzed the existing qualitative data sharing guidance from international data repositories. Reviewing their guidelines for depositing and secondary use of qualitative data, we learned that there are gaps in the topics addressed by repositories and significant variability in the guidance offered. Although we expect that flexibility is necessary to meet the needs of different qualitative research projects and researchers, we outline several questions and future topics for investigation. Notably, to move the research enterprise forward with qualitative data sharing both practical and ethical considerations must be addressed. Moreover, the viewpoints of multiple stakeholders in the global research enterprise—research participants, researchers, research ethics committee members, repository curators, institutional officials, funding agency officials, the public—all require consideration.

Acknowledgments

This research was supported in part by a grant from the National Center for Advancing Translational Science UL1 TR000448. The effort of A.L.A. was supported in part by the National Human Genome Research Institute, K01HG008990. The authors would like to thank our repository contacts for their contributions and feedback.

References

- Australian Data Archive (ADA). Australian Data Archive. 2016. Retrieved from <https://www.ada.edu.au/ada/about>
- Australian National Data Service. Ethics, consent and data sharing. 2016. Retrieved from <http://ands.org.au/guides/ethics-consent-and-data-sharing>
- Begley CG, Ellis LM. Raise standards for preclinical cancer research. *Nature*. 2012; 531:531–533.
- Broom A, Cheshire L, Emmison M. Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology*. 2009; 43(6):1163–1180.

- Center for Open Science. The Transparency and Openness Promotion Guidelines. n.d. Retrieved from <https://cos.io/top/>
- Cliggett L. Qualitative data archiving in the digital age: Strategies for data preservation and sharing. *The Qualitative Report*. 2013; 18(1):1–11.
- Corbin, J., Strauss, A. Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications; 2014.
- Corti L. Progress and Problems of Preserving and Providing Access to Qualitative Data for Social Research —The International Picture of an Emerging Culture. *Forum: Qualitative Social Research*. 2000; 1(3)
- Corti L, Van den Eynden V. Learning to manage and share data: jump-starting the research methods curriculum. *International Journal of Social Research Methodology*. 2015; 18(5):545–559. DOI: 10.1080/13645579.2015.1062627
- Corti, L., Van den Eynden, V., Bishop, L., Woollard, M. Managing and Sharing Research Data: A Guide to Good Practice. Vol. 2015. London, UK: SAGE Publications Inc; 2014.
- Czech Social Science Data Archive (CSDA). Czech social science data archive (CSDA) Institute of Sociology. 2016. Retrieved from <http://archiv.soc.cas.cz/o-ceskemsocialnevednim-datovem-archivu>
- Databrary at New York University and Penn State. Databrary: About the Project. 2016. Retrieved from <http://databrary.org/about.html>
- DuBois JM, Strait M, Walsh H. Is it time to share qualitative research data? *Qualitative Psychology*. 2016; 3(2)
- DuBois JM, Strait M, Walsh H. Is It Time to Share Qualitative Research Data? *Qualitative Psychology*. 2017; doi: 10.1037/qup0000076
- Economic and Social Research Council (ESRC). ESRC Research Data Policy. 2015. Retrieved from <http://www.esrc.ac.uk/files/about-us/policies-and-standards/esrcresearch-data-policy/>
- Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: Characteristics and guidelines for use. *American Journal of Public Health*. 1984; 74(9):979–983. [PubMed: 6380323]
- Finnish Social Science Data Archive (FSD). About Us : Finnish Social Science Data Archive. 2016. Retrieved from <http://www.fsd.uta.fi/en/organisation/index.html>
- Gibson G, Timlin A, Curran S, Wattis J. The scope for qualitative methods in research and clinical trials in dementia. *Age Ageing*. 2004; 33(4):422–426. DOI: 10.1093/ageing/afh136 [PubMed: 15226116]
- Guishard, MA. Now's not the time! Qualitative data repositories on tricky ground. Comment on DuBois et al. 2017. *Qualitative Psychology*, Advanced online. 2017. doi:<http://dx.doi.org/10.1037/qup0000085>
- Hardy LJ, Hughes A, Hulen E, Schwartz AL. Implementing Qualitative Data Management Plans to Ensure Ethical Standards in Multi-Partner Centers. *J Empir Res Hum Res Ethics*. 2016; 11(2): 191–198. DOI: 10.1177/1556264616636233 [PubMed: 27074911]
- Hoyle, L., Corti, L., Gregory, A., Martinez, A., Wackerow, J., Alvar, E., ... Zuell, C. A Qualitative Data Model for DDI. 2012. Retrieved from <http://www.ddialliance.org/system/files/AQualitativeDataModelForDDI.pdf>
- Institute for Quantitative Social Science (IQSS), Harvard University Library, & Harvard University Information Technology Organization. The Dataverse Project: About the Project. 2016. Retrieved from <http://dataverse.org/>
- Interuniversity Consortium for Political and Social Research (ICPSR). 2016a. from University of Michigan <https://www.icpsr.umich.edu/icpsrweb/landing.jsp>
- Interuniversity Consortium for Political and Social Research (ICPSR). ICPSR About. 2016b. Retrieved from <https://www.icpsr.umich.edu/icpsrweb/content/about/>
- Jackson MR. Resistance to qual/quant parity: Why the “paradigm” discussion can’t be avoided. *Qualitative Psychology*. 2015; 2(2):181–198. DOI: 10.1037/qup0000031
- Jahnke, LM., Asher, A. The problem of data: Data management and curation practices among university researchers. Council on Library and Information Resources; 2014. Retrieved from <http://www.clir.org/pubs/reports/pub154/problem-of-data>

- King G. Publication, Publication. PS: Political Science and Politics. 2006; 39(01):119–125.
- Kuula A. Methodological and ethical dilemmas of archiving qualitative data. *Iassist Quarterly*. 2011; (34):12–17.
- Malin B, Benitez K, Masys D. Never too old for anonymity: a statistical standard for demographic data sharing via the HIPAA Privacy Rule. *J Am Med Inform Assoc*. 2011; 18(1):3–10. DOI: 10.1136/jamia.2010.004622 [PubMed: 21169618]
- McCurdy SA, Ross MW. Qualitative Data Are Not Just Quantitative Data With Text but Data With Context: On the Dangers of Sharing Some Qualitative Data: Comment on DuBois et al. 2017. *Qualitative Psychology*. 2017; doi: 10.1037/qup0000088
- Merriam, SB. *Qualitative Research in Practice: Examples for Discussion and Analysis*. 1. Jossey-Bass; 2002.
- Musheke M, Ntalasha H, Gari S, McKenzie O, Bond V, Martin-Hilber A, Merten S. A systematic review of qualitative findings on factors enabling and deterring uptake of HIV testing in Sub-Saharan Africa. *Bmc Public Health*. 2013; 13doi: 10.1186/1471-2458-13-220
- National Institutes of Health. NIH Sharing Policies and Related Guidance on NIH-Funded Research Resources. 2015. Feb 26. Retrieved from <https://grants.nih.gov/grants/sharing.htm>
- National Science Foundation. Dissemination and Sharing of Research Results. 2014. Dec 26. Retrieved from <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>
- Nature. Nature Author Policy: Availability of data, material and methods. 2016. Retrieved from <http://www.nature.com/authors/policies/availability.html>
- Norwegian Social Science Data Services (NSD). Feb 16, 2017. 2016Retrieved from <http://www.nsd.uib.no/nsd/english/index.html>
- Nosek B, Alter G, Banks G, Borsboom D, Bowman S, Breckler S, ... Christensen G. Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science (New York, NY)*. 2015; 348(6242):1422.
- Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Scientific standards. Promoting an open research culture. *Science*. 2015; 348(6242):1422–1425. DOI: 10.1126/science.aab2374 [PubMed: 26113702]
- Nuzzo R. Fooling ourselves. *Nature*. 2015; (526):182–185.
- O'Brien B, Harris I, Beckman T, Reed D, Cook D. Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine*. 2014; 89(9):1245–1251. [PubMed: 24979285]
- Open Access Directory. Data Repositories Index. 2016. Jun. 2015Retrieved from http://oad.simmons.edu/oadwiki/Main_Page
- Parry O, Mauthner N. Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data. *Sociology*. 2004; 38(1):139–152.
- Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. *PLOS One*. 2007; 2(3):1–5. DOI: 10.1371/journal.pone.0000308
- Qualitative Data Repository (QDR) at Syracuse. About the Qualitative Data Repository. 2016. Retrieved from <https://qdr.syr.edu/>
- Roller, MR., Lavrakas, PJ. *Applied qualitative research design: A total quality framework approach*. New York: Guilford Press; 2015.
- Saldaña, J. *The Coding Manual for Qualitative Researchers*. 2. Thousand Oaks, CA: Sage Publications Ltd; 2013.
- Saldaña, J. An Introduction to Codes and Coding. In: Seaman, J., editor. *The Coding Manual for Qualitative Researchers*. 3. Los Angeles, CA: Sage Publications Ltd; 2016. p. 1-42.
- Saunders B, Kitzinger J, Kitzinger C. Anonymising interview data: challenges and compromise in practice. *Qual Res*. 2015; 15(5):616–632. DOI: 10.1177/1468794114550439 [PubMed: 26457066]
- Schwandt, T. *The Sage Dictionary of Qualitative Inquiry*. 4. Los Angeles: Sage; 2014.
- Science Magazine. Science: Editorial Policies. 2016. Retrieved from <http://www.sciencemag.org/authors/science-editorial-policies>

- Simmons J, Nelson L, Simonsohn U. False-positive psychology: Undiscolored flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*. 2011; 22(11):1359–1366. [PubMed: 22006061]
- Slovenian Social Science Data Archives (ADP). 2016. Retrieved from <http://www.adp.fdv.uni-lj.si/eng/>
- Smioski A. Archiving qualitative data: Infrastructure, acquisition, documentation, distribution. Experiences from WISDOM, the Austrian Data Archive. *Forum Qualitative Social Research*. 2011; 12(3)
- Solove, DJ., Schwartz, P. Information privacy law. 5. New York: Wolters Kluwer; 2015.
- Swiss Centre of Expertise in the Social Sciences. FORS About. 2016. Retrieved from <http://forscenter.ch/en/about-us-2/>
- Taichman DB, Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen JM, ... Wu S. Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors. *Annals of Internal Medicine*. 2016; doi: 10.7326/m15-2928
- The Dryad Repository at North Carolina State University. DataDryad About. 2016. Retrieved from <http://datadryad.org/>
- UK Data Archive. Managing and Sharing Data: Best Practices for Researchers. Vol. 3. Wivenhoe Park, Colchester, Essex: University of Essex; 2011.
- UK Data Archive. Create and Manage Data: Consent and Ethics. 2016, 2017. Retrieved from <http://www.data-archive.ac.uk/create-manage/consent-ethics/consent>
- UK Data Archive (UKDA). 2016. Retrieved from <http://www.data-archive.ac.uk/>
- Walsh D, Downe S. Meta-synthesis method for qualitative research: a literature review. *Journal of Advanced Nursing*. 2005; 50(2):204–211. [PubMed: 15788085]
- Yardley SJ, Watts KM, Pearson J, Richardson JC. Ethical issues in the reuse of qualitative data: perspectives from literature, practice, and participants. *Qual Health Res*. 2014; 24(1):102–113. DOI: 10.1177/1049732313518373 [PubMed: 24374332]

Biographies

Alison L. Antes is Assistant Director of the Center for Clinical and Research Ethics at Washington University School of Medicine where she co-directs a course on the responsible conduct of research. Her research emphasizes ethical decision-making in research, leadership and professionalism in research, and instruction on the responsible conduct of research. She contributed to analysis and interpretation of the data and wrote the first draft of the manuscript.

Heidi Walsh is project manager in the Professional & Social Issues lab in the department of General Medical Sciences at Washington University School of Medicine. Her research and professional interests include clinical and research ethics and improving health literacy. She contributed to the data collection and data analysis, and she edited the manuscript.

Michelle Strait is a graduate of the Master of Social Work and Public Health program at Washington University in St. Louis. Her research and professional interests include mental health and mental illness prevention. She contributed to the data collection and data analysis, and she edited the manuscript.

Cynthia R. Hudson-Vitale is the Data Services Coordinator in Data & GIS Services at Washington University in St. Louis Libraries. In this position, Cynthia leads research data services and curation efforts for the libraries. Her interests include data sharing, data management, and increasing the transparency of research studies. She contributed to the editing of the manuscript.

James M. DuBois is Director of the Center for Clinical and Research Ethics and of the Professionalism and Integrity in Research Program (PI Program) at Washington University School of Medicine. He conducts social science research on mental health research ethics, instruction on the responsible conduct of research, and factors that foster or impair research integrity. He helped to conceptualize the study, contributed to development of the codebook and the interpretation of data, and edited the manuscript.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
Repository Inclusion Search Strategy

Table 1

Codes for Topics Addressed in Data Depositing Guidelines

Topic	Definition
Participant protections	
Anonymization: Remove or replace direct identifiers	Direct identifiers—variables that explicitly identify individuals (e.g., a person's name)—are removed and annotation indicates the deletion or are replaced with pseudonyms or terms that describe individuals generally, such as mother or Friend1 and Friend2.
Anonymization: Remove or replace indirect identifiers	Indirect identifiers—variables that when combined with other attributes (e.g., race, profession) might identify individuals—are removed or replaced.
Anonymization log	Documentation of the steps taken to remove direct identifiers to achieve anonymization; such documentation provides context to secondary users.
Set access controls: Specify if access is restricted (and how)	Certain data require special handling and access restrictions due to confidentiality concerns. Generally, other than open access, repositories allow depositors to specify restricted or special access.
Set access controls: Dissemination is delayed (embargo)	For data that may pose a confidentiality risk when released (or when journals require it), depositors may request an embargo period, whereby no access to the data is permitted until after a specified date.
Adhere to requirements of ethics committee (e.g., IRB, REC)	Adherence to the requirements of committees, such as Institutional Review Boards (IRBs) or Human Research Ethics Committees (RECs), formally designated to review, approve, and monitor research involving humans.
Proof or waiver of REC permissions <i>for archiving</i>	Formal documentation indicating that all individuals, such as interview participants, have agreed to the terms set for <i>archiving</i> or that an ethics committee deemed the study exempt from such permission.
Proof or waiver of REC permissions <i>for sharing/ regulating access</i>	Formal documentation indicating that all individuals, such as interview participants, have agreed to the terms for <i>sharing</i> or that an ethics committee deemed the study exempt from such permission.
Documentation to be submitted	
Research methods and practices	Study title and summary, protocols detailing the research plan, including research methods, practices, and study methodology (e.g., sampling, hypotheses, aims), and information about the informed consent process, if applicable.
Data collection instruments	Data collection instruments (e.g., surveys or interview questions).
Data collection process/approach and problems	Information about issues in collecting or managing the data that impact the data and interpretation of the data (e.g., information about data quality assurance, data cleaning, missing data, changes made to the dataset, or problems that arose, such as attrition or issues in the interview process, and how they were addressed).
Codebook	Instructions for when and how to apply predetermined codes in the analysis qualitative data such as transcribed interview data.
Data documentation/metadata	Description of all files and content deposited which the repository uses to index the data and helps secondary users discover the data.
Bibliography	References relevant to the data collection or that resulted from the data.
Liability waiver	Certification that the repository is not liable for the content of the deposited materials or for the loss of data or data documentation.
DDL-compliance/clearly labeling data files	Data Documentation Initiative (DDI) is an international standard for describing surveys, data, and study information to allow for consistency in labeling and storing data to facilitate its preservation and reuse. (Gregory, 2011)
Fees	
Fee required for depositing some or all data files	Fee may be required to submit data; may cover storage costs or curator services.

Table 2

Topics Addressed in Data Depositing Guidelines

Topic	1	2	3	4	5	6	7	8	9	10	11	12	%
Participant protections													
Anonymization: Remove or replace direct identifiers	-	-	-	-	-	-	-	-	*	-	-	-	83
Anonymization: Remove or replace indirect identifiers	-	-	-	-	-	-	-	**	*	-	-	-	50
Anonymization log	-	-	-	-	-	-	-	-	-	-	-	-	50
Set access controls: Specify if access is restricted (and how)	-	-	-	-	-	-	-	-	-	-	-	-	83
Set access controls: Dissemination is delayed (embargo) [†]	-	-	-	-	-	-	-	-	-	-	-	-	92
Adhere to requirements of ethics committee (e.g., IRB, REC)	-	-	-	-	-	-	-	-	-	-	-	-	58
Proof or waiver of REC permissions for archiving	-	-	-	-	-	-	-	-	-	-	-	-	83
Proof or waiver of REC permissions for sharing/regulating access	-	-	-	-	-	-	-	-	-	-	-	-	50
Documents to be submitted													
Research methods and practices	-	-	-	-	-	-	-	-	-	-	-	-	75
Data collection instruments	-	-	-	-	-	-	-	-	-	-	-	-	75
Data collection process/approach and problems	-	-	-	-	-	-	-	-	-	-	-	-	67
Codebook	-	-	-	-	-	-	-	-	-	-	-	-	67
Data documentation / metadata	-	-	-	-	-	-	-	-	-	-	-	-	92
Bibliography	-	-	-	-	-	-	-	-	-	-	-	-	83
Liability waiver	-	-	-	-	-	-	-	-	-	-	-	-	75
DDI-compliance/clearly labeling data files	-	-	-	-	-	-	-	-	-	-	-	-	75
Fees													
Fee required for depositing some or all data files [‡]	-	-	-	-	-	-	-	-	-	-	-	-	25

Key: 1= Australian Data Archive; 2 Slovenian Social Science Data Archives; 3= Czech Social Science Data Archive MEDARD Catalog; 4= Databrary at New York University and Penn State; 5= Dataverse US Murray Research Center at Harvard; 6= The Dryad Repository at North Carolina State University; 7= Swiss Centre of Expertise in the Social Sciences; 8= Finnish Social Science Data Archive; 9= Interuniversity Consortium for Political and Social Research at the Institute for Social Research at the University of Michigan; 10= Norwegian Social Science Data Services; 11= Qualitative Data Repository at Syracuse; 12= UK Data Archive

= Guidelines address topic; - = No mention of topic in guidelines

* Repository requires vaguer descriptors, but discourages pseudonyms and fictionalizations

** Repository discourages removing or replacing indirect identifiers

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

⁷ Delaying is an available option for depositing researchers, but not required for deposit

^{7/8} Repository Contact indicated that the topic is addressed internally rather than in the written guidelines.

⁸ Some repository contacts mentioned fees for depositors, but these were not written in the guidelines.

Table 3

Codes for Topics Addressed in Secondary Data Use Guidelines

Topic	Definition
Published Requirements Before Access	
Complete application to access data	Forms or applications documenting that the user agrees to the “after access terms” are required to access data.
Clarification given about potential fees	Fee required for access to the data.
Data producer permission for some or all datasets	Access to the data may require permission from the original research team that deposited the data.
Fulfill special conditions to access restricted use data	Access to data is restricted (e.g., use by principal investigators is permitted but access is not granted for students) or embargos on datasets may delay access.
Justify purpose for data use	Must explain aims and purpose of the proposed secondary use.
Adherence to the requirements of research ethics committee	The secondary data may be required to adhere to the requirements of their ethics committee (e.g. IRB, HREC, IEC, ERB, REB, etc).
After Access Topics	
Time limit on data use	Access to data is only available for a limited period.
Commercial use prohibited	Use of data is allowed only for non-profit purposes.
Continued data anonymity required	Secondary users must not attempt to re-identify participants.
Redistribution of downloaded data prohibited	Secondary users will not share the data with, or provide access to, unauthorized individuals.
Cite data source in publications	Give credit to the repository and original research team when publishing results from secondary use of the data by referencing the bibliographic citation in any of the secondary users’ publications.
Secondary user accountable for new discoveries	New discoveries or claims stemming from secondary use of the data are the responsibility of the secondary researcher.
Send citations or notify repository about publications written on the basis of the downloaded data	References to publications from secondary use of the data must be sent to the repository for posting with original dataset.
Share new dataset created from secondary use	New datasets created through secondary use of the data may be added to the original data deposit to supplement the materials.

Table 4

Topics Addressed in Secondary Data Use Guidelines

Topic	1	2	3	4	5	6	7	8	9	10	11	12	%
Published Requirements – Before Access													
Complete application to access data	-	-	-	-	-	-	*	-	-	-	-	-	67
Clarification given about potential fees	-	-	-	-	-	-	-	-	-	-	-	-	67
Data producer permission for some or all datasets	-	-	-	-	-	-	-	-	-	-	**	-	83
Fulfill special conditions to access restricted use data [†]	-	-	-	-	-	-	-	-	-	-	-	-	83
Justify purpose for data use	-	-	-	-	-	-	-	-	-	-	-	-	75
Adherence to the requirements of research ethics committee	-	-	-	-	-	-	-	-	-	-	-	-	25
After Access Topics													
Time limit on data use	-	-	-	-	-	-	*	-	*	-	-	*	17
Commercial use prohibited	-	-	-	-	-	-	-	-	-	-	-	*	75
Continued data anonymity required	-	-	-	-	-	-	-	-	-	-	-	-	92
Redistribution of downloaded data prohibited	-	-	-	-	-	-	-	-	-	-	-	-	83
Cite data source in publications	-	-	-	-	-	-	-	-	-	-	-	-	100
Secondary user accountable for new discoveries	-	-	-	-	-	-	-	-	-	-	-	*	58
Send citations or notify repository about publications written on the basis of the downloaded data	-	-	-	-	-	-	-	-	*	-	-	-	58
Share new datasets created from secondary use	-	-	-	-	-	-	-	-	-	-	-	-	25

Key: 1= Australian Data Archive (ADA); 2= Slovenian Social Science Data Archives; 3= Czech Social Science Data Archive MEDARD Catalog ; 4= Databrary at New York University and Penn State; 5= Dataverse US Murray Research Center at Harvard; 6= The Dryad Repository at North Carolina State University; 7= Swiss Centre of Expertise in the Social Sciences; 8= Finnish Social Science Data Archive; 9= Interuniversity Consortium for Political and Social Research at the University of Michigan; 10= Norwegian Social Science Data Services; 11= Qualitative Data Repository at Syracuse; 12= UK Data Archive

= Guidelines address topic; - = No mention of topic in guidelines

* Repository Contact indicated that the topic is addressed internally rather than in the written guidelines

** Depositors grant Interuniversity Consortium for Political and Social Research permission to handle/disseminate the data

[†] Examples include affiliation with an institution that is a member of the repository or lead researcher status