Sharing qualitative interview data in dialogue with research participants

AUTHORS SECTION

Kvale, Live Håndlykken University of Oslo, Norway | kvalel@ub.uio.no

Pharo, Nils OsloMet- Oslo Metropolitan University, Norway | nilsp@oslomet.no

Darch, Peter University of Illinois at Urbana-Champaign, United States | ptdarch@illinois.edu

ABSTRACT

Research data sharing is embedded in policies, guidelines and requirements commonly promoted by research funding organizations that demand data to be "as open as possible, as closed as necessary" and FAIR. This paper discusses the challenges of balancing privacy protection with data sharing in a PhD project involving long-tail, small-sized qualitative human subjects' data. Based on experiences and feedback from project participants, we argue that privacy protection is about respecting the participants and their self-image. This can be achieved through dialogue and involvement of the participants building on the principles of shared stewardship. Further, we suggest that de-identification and plain language consent materials are better at protecting privacy than anonymisation, which in a digital data environment is difficult to achieve and not necessarily a sensible approach for qualitative data, where the gold is in the details. The literature indicates that it matters to participants whether data are reused for research or other purposes, and that they trust the institutions. This supports our claim that research data services must find better solutions for restricted sharing when necessary.

KEYWORDS

Data Sharing, Research Ethics, Privacy, Qualitative Research, Research into Practice

INTRODUCTION

Open research data and sharing of research data is a key pillar of Open Science, promoting transparency and trust in scientific institutions and the advance of scientific research (UNESCO, 2021). Institutional and international policies for open science increasingly require datasets collected for the purpose of research to be made available "as open as possible, as closed as necessary" (European Commission, 2016), with the aspiration of making them FAIR (Findable, Accessible, Interoperable, and Reusable) (Wilkinson et al., 2016).

Researchers conducting human subjects' research must abide by research ethics guidelines, national and international data legislation (Ursin & Bentzen, 2021), and by university and funding agency policies (The Norwegian National Research Ethics Committees, 2014). For instance, the Belmont Principles that govern human subjects' research draw attention to 1) respecting human subjects, by use of informed consent and protecting privacy; 2) protecting research subjects from the risk of harm; and 3) addressing issues of justice (U.S. Department of Health, Education, and Welfare, 1979). However, each of these three imperatives is made more complicated by opening human subjects' data: 1) human subjects may struggle to imagine how data about them could be used in new contexts, undermining their ability to give truly informed consent for reuse (Wilbanks, 2014); 2) their data could be combined with other available datasets to generate new insights about them that could be used for harmful purposes (Ohm, 2014); 3) marginalized groups of people may be more liable to harms from misuse of data about them (Carroll et al., 2020).

86th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – 31, 2023 |London, United Kingdom.

Balancing the realization of the benefits of Open Science with adherence to ethical and legal standards is a "problem of many hands," that require action by many research data curation stakeholders and at all stages of the research lifecycle. Decisions taken early on in a research project have significant implications for subsequent possibilities for sharing and for privacy protection. For instance, researchers need to design and carry out data collection processes with sharing in mind, choosing metadata schema and creating documentation that will enhance reusability of the dataset, while carefully minimizing the data they collect to enhance privacy protections (Tibor, 2021, p. 54). However, little training currently exists for social science researchers about how to approach these topics effectively. Data management plans, increasingly required by funding agency, have the potential to help researchers in designing and carrying-out data collection, but only if requirements for these plans are well-formulated and effectively enforced. A third key group of stakeholders is research support service providers, for instance in university libraries and disciplinary data repositories, who can provide consultational and informational services, but who require training and resources to provide adequate support to researchers.

This paper builds on the first author's PhD project, a study of 24 expert stakeholders involved in research data curation in Norway (L. Kvale, 2022), to explore the challenges in making social scientific human subjects' data as open as possible while respecting study participants' privacy. The study data were collected through a modified Delphi study, comprising interviews (n=48) and surveys (n=24). The first author aimed for these data to be made as open as possible in line with open science expectations from domestic science funding agencies (The Research Council of Norway, 2020), while working within domestic and international ethical and legal constraints, such as the European Union's General Data Protection Regulation (GDPR) and Norway's Personal Data Act and research ethical guidelines (GDPR, 2016; The Norwegian Personal Data Act, 2018; The Norwegian National Research Ethics Committees, 2014).

It is this tension we explore through addressing the following question**:**

Can sharing of non-sensitive de-identified human subjects' data be practiced while respecting the research participants' right to privacy?

Drawing from analysis of feedback from participants, we provide experienced-based recommendations for researchers and data management professionals to handle the difficulties involved in balancing opening data with respecting privacy. There is no one size fits all for translating policies for open science into research practice. The challenges researchers encounter when dealing with human subjects data, have until recently not been addressed in the literature on research data management. De-identification describes the removal of identifiers that are either linkable or directly identifiable. De-identified interview transcripts could have value for reuse for instance, in teaching research data curation to information management professionals, to facilitate future research in the area of scholarly communication, and historically as a documentation of strategies and decisions in the digitalisation of research, or simply for machine learning purposes.

Our study is relevant for research support service providers struggling to balance expectations and requirements of sharing with protection of privacy and research ethical norms, providing a case of how this can be solved. It is also relevant to researchers, in particular researchers conducting research in Europe or collaborating with researchers in a European context, as an example of an approach to how open sharing of interview transcripts and personal privacy is perceived amongst expert stakeholders of data sharing. We recognize that while general norms for privacy and research ethics have a global scope, they are perceived and implemented in quite different ways. Subsequently knowledge of these differences is crucial for international research collaborations.

PREVIOUS RESEARCH

Research participants' attitudes towards open data sharing has not yet been explored extensively. A study based on interviews with qualitative researchers, Institutional Review Board members and data curators in the US revealed that researchers are the least knowledgeable and are often unfamiliar with the concept of sharing qualitative data in a repository (Mozersky, Walsh, et al., 2020). Several countries have traditions of sharing of qualitative data via restricted access repositories with different standards for de-identification. However following the argument of open science as a way to foster public trust in research (DuBois et al., 2018; UNESCO, 2021) it would benefit science to share more qualitative data openly with the consent of the research participants.

Current studies of data sharing and reuse explore restrictions on sharing and highlight how the type of reuse matters to research participants (Hardy et al., 2016; Mozersky, Parsons, et al., 2020; Shah et al., 2021). Mozersky, Parson and colleagues (2020) find that the majority of the research participants were positive to sharing of qualitative data given that the sharing and reuse is limited to research purposes and that their identity is concealed. Shah and colleagues find similar attitudes towards sharing and reuse of health data; the research context matters to the participants (Shah et al., 2021). Hardy and colleagues (2016), on the other hand, argue that sharing and reuse of the data collected would be unethical as the conditions for access to the participants were that certain issues would not be investigated even if the data could have provided the possibility. This is echoed by researchers expressing concerns that it would be unethical to share qualitative data collected in a relation of trust without an element of informed consent (DuBois et al., 2018). While researchers refer to the relation of trust between researchers and participants, the participants generally trust institutions rather than the individual researchers (Guillemin et al., 2018).

Within the scholarship of human computer interaction, privacy by design is used to describe how technology can be designed in compliance with privacy (Hoepman, 2021). This way of thinking privacy embedded in the methods and data collection is also useful when designing and conducting research projects. It does however require awareness of different perceptions of privacy. Gürses group privacy by design in three categories according to how privacy is understood: 1) hard privacy approaching privacy as confidentiality, 2) soft privacy addressing privacy as control, and 3) contextual privacy addressing privacy as a practise where negotiations of social boundaries are taking place (Gurses, 2014). Later privacy by compliance has been added, referring to how GDPR has led organizations to aim at compliance with the corresponding jurisdiction rather than providing individuals with control over their data (Fiebig et al., 2022). The conflict between GDPR and data sharing initiatives within health research has been adressed by Ursin and Bentzen (2021) who call for harmonization of data protections standards to allow global sharing of research data.

In qualitative research, member checking is sometimes used as a technique to validate results by either returning interview transcripts or analysis of the data to the participant for comments, corrections and verification (Creswell & Miller, 2000; Forbat & Henderson, 2005). Karhulathi (2022) propose using member-checking for sharing qualitative data, using co-curation as a term to describe how member-checking allows participants to edit trancripts as part of preparing them for open sharing. While there are studies on how participants approach transcripts when asked to review these for errors (Birt et al., 2016; Carlson, 2014; Forbat & Henderson, 2005), the authors are not familiar with literature addressing participants' experience of member checking for publication. Also Huma and Joyce highlight the lack of expertise amongst researchers and ethical comities on how to proceed with sharing of qualitative data while safeguarding research participants (Huma & Joyce, 2022). A synonym for co-curation is shared stewardship, a concept drawing from studies of indigenous data sovereigny, that embodies the data subjects rights and interest in their material (First Archivist Circle, 2007; Leopold, 2019). Within research data management, indigenous communites have developed the CARE principles to address their interest and rights over data governance and reuse (Carroll et al., 2020), within which shared stewardship conceptually embeds the practice of shared control and governance.

METHOD

This paper describes the process of sharing data from a Delphi study, which is characterised by using experts on a particular topic as participants and collecting data in multiple rounds focusing on solving an issue or developing a policy (Ziglio, 1996). A Delphi study is designed to find agreements and common understandings amongst involved stakeholders. In a rapidly developing domain such as that of data sharing, the Delphi method offers a way of systematically collecting solution-oriented opinions. A Delphi study typically contain three sequential phases: 1) the exploration phase, 2) the evaluation phase, and 3) the concluding phase (Ziglio, 1996). In each phase data are collected and analyzed and the intermediate results are used in the development of the next phase of data collection.

Participants were drawn from four stakeholder groups involved in research data sharing: policymakers, infrastructure providers, research support staff, and researchers. 24 participants took part in the study, which makes it similar in size to other Delphi studies in library and information studies (Lund, 2020).

The data collection took place in three phases (Figure 1), where the first and last phase involved conducting interviews (n=48.) First phase interviews were about one hour long, while last phase interviews lasted approximately 30 minutes. At the recruitment stage, participants were informed of plans to make interviews openly-available in a public repository. The interviews were transcribed by the first author, yielding 313 pages of transcripts. During transcription, the first author de-identified participants by removing directly identifiable information such as names and workplace. Nevertheless, most of the participants remained identifiable as they in many cases represent the only domain expert with their background and experience. Transcripts were sent back to the participants for verification, and participants were also asked to mark sequences of the interview they did not want openly shared (Birt et al., 2016). These sequences were included in data analysis but removed from the published transcripts, extra care have been taken to conceal the participants identity where quotes from marked sequences are included in this and other publications from the study.

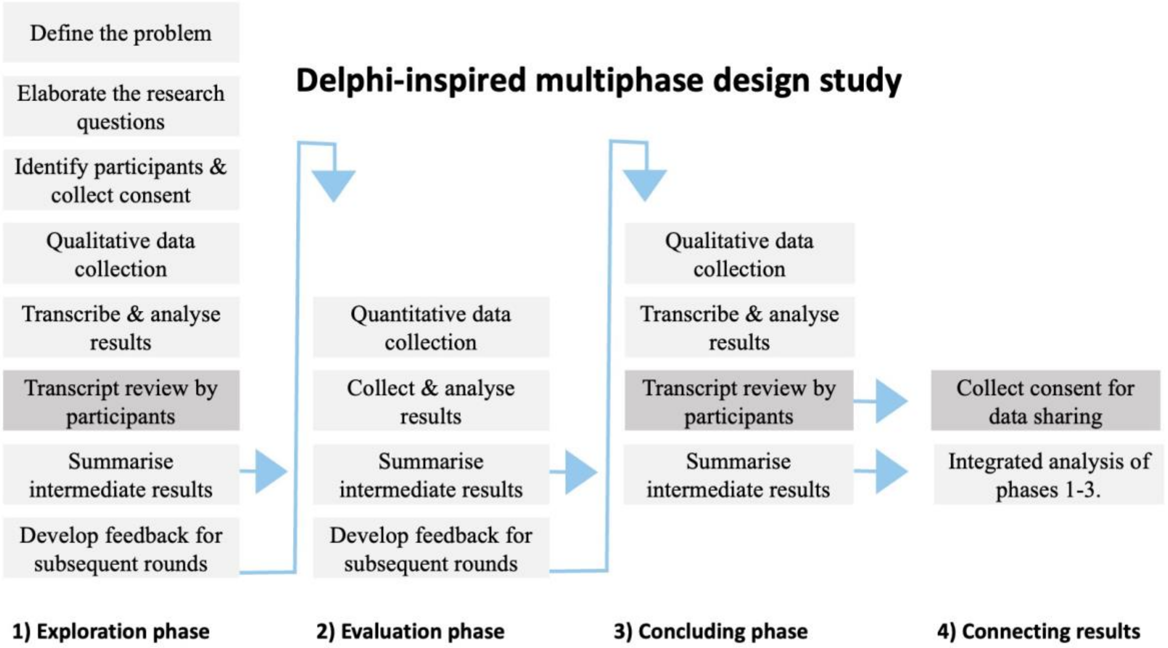

Figure 1 The research design (L. Kvale & Pharo, 2021)

Following interviews, participants were given the opportunity to indicate their consent for sharing their data openly. Feedback on a draft version of the consent form given to participants was solicited from the Norwegian National Privacy Advisors' Services (NSD/SIKT) which had no objections or comments on the form. In this consent form, the participants were also asked how they felt about participating in the study (L. Kvale, 2021). 21 participants granted permission to share all or parts of their data, while three participants did not respond to the consent form so their data were not shared.

The aim of sharing de-identified data and the review of transcripts were discussed with the participants multiple times throughout the study, and multiple participants expressed interest in, and opinions on, these topics. Initially, the plan was to anonymize the data. As the first round of interviews were transcribed, it became evident that the data could not be anonymized, even with directly identifiable information removed. Information about stakeholders' roles, selection criteria, and participants' expressed opinions and experience made it likely that some participants could be identified readily. When reading the transcripts, the first author could recognize participants based on their language and metaphors or by descriptions of their work with research data.

True anonymity is difficult to achieve, and may not always be the most appropriate approach for privacy protection; instead, an approach that focuses on what needs to be protected and how context shapes human subjects' privacy expectations is preferred (Barocas & Nissenbaum, 2014). The first author followed this approach by informing participants about the aim to share data in an open repository, and allowing them to review their data and select what to share openly.

Code	Description
Experience as participant	Referring to own experience as research subject.
Reflections	Sharing of thoughts or reflections on the method used in the project.
My data sharing	Thoughts or experiences with the data sharing in the project.

Table 1. Examples of methods related qualitative codes used in the analysis of the interviews

This paper builds on the discussions of methods in the first author's PhD thesis (L. Kvale, 2022) . The findings are derived from different parts of the study including the interviews, material retracted from sharing in the interview transcripts, the first author's notes taken after each interview, and comments provided by the participants in the consent form for data sharing. This material has been structured and analysed qualitatively by exploratory thematic coding (Saldaña, 2016) during the first three phases of the study (Figure 1.), followed by themathic grouping (S. Kvale, 2007) and meaning synthetization as part of the integrated analysis (Creswell & Plano Clark, 2018) after data collection was completed (Figure 1, phase 4).

In the paper, quotes from each participant have been assigned one of the following pseudonyms, suggested as gender-neutral by Google: Andrea, Leigh, Ashley, Frankie, Jackie, Parker, Sam, Linden, Bobby, Chris, Dylan. In some cases, pseudonyms have been associated with professional roles; in other cases, they have not where the risk of reidentification was judged to be too high.

FINDINGS

The findings raise awareness about how different research participants experience issues related to privacy in diverse ways, and how researchers need to address these differences as part of privacy protection and ethical considerations regarding data sharing. We have categorised our findings into two parts: 1) how participants control their self-presentation through reviewing the de-identified data, and 2) how the participants experience data sharing.

Control over self-presentation by reviewing de-identified data

When reviewing their transcript, each participant could select parts of the interview that were not to be published in the open dataset. Analysis of what participants requested removed from the transcripts before publishing show two tendencies, concealments of a) relational conflicts, or b) personal identity.

Concealments of relational conflicts or criticism of organisations

Participants requesting removal of text containing descriptions of conflicts in collaborations or criticisms of their own organisations. For instance:

"We collaborate with [other research support services at the university] and there are different cultures, different attitudes towards different things, and when there is a need for distribution of responsibility it is typically different perspectives on who should do what and who should be responsible for what. So, this can occasionally be quite demanding." (Leigh, research support)

The quote from Leigh refer to the challenges of collaborating with other research support services in the university, a challenged echoed by several other interviewees. However, this participant did not want others to know they had discussed these challenges due to concerns it could negatively affect an existing collaboration.

Two other interviewees, Ashley and Frankie, requested removal of text that could be perceived as damaging to their own, fragile organizations, as illustrated here:

"Right now, we are just trying to promote our self and our services, we are in a situation where we need to make sure we as organisation are still relevant". (Ashley)

"I would ask you to be quiet about these things for a while as I do not yet know to what extend this is public knowledge" (Frankie)

Both quotations describe organisations in limbo, discussing dilemmas in making their services relevant to users and describing changes that are not yet formalised. Frankie's quotation could reflect negatively on them for speaking out of turn by revealing confidential details about their organisation, while Ashley expressed concerned about publicly underestimating the value of their organisation in a competitive environment.

These three quotes describe different types of tensions in the organisations where the participants work. Removing these details from the published data obscured part of the tension within organisations regarding directions and responsibility for data management services and tools. In all three quotes the participants opened up to the interviewer and shared their thoughts in confidence, based on trust between the interviewer and interviewee. This closeness places an ethical responsibility on the interviewer, which is fulfilled by sharing these quotes with as little context as possible to respect the participants' control over their identity.

Concealment of personal identity

The second tendency was several participants' requests to hide their identity further by removal of specific information regarding their work. This tendency included terms such as, "university library," "data curation," and "language researcher." The possibility of identification was something the participants showed understanding of, as the researcher Andrea points out:

"If people know that you came to this institute for an interview with people who created a DMP, people who know that I am here on a fellowship, they will find out that these are my answers. But others will not know it." (Andrea, researcher)

Jackie, a high-profile researcher, removed multiple details about their research practice that could make them easily identifiable. Still, enough details remain for someone with close knowledge of the domain to potentially identify Jackie.

Parker, another researcher, revealed a reluctancy to comply with privacy requirements at an earlier point in their research career (L. Kvale & Darch, 2022). Revealing Parkers identity could potentially cause harm to their reputation as researcher. Parker requested many of the details which could identify their identity to be removed from the transcripts. A combination of details reported in published articles, combined with the information retained in the transcript, could, however, still make Parker identifiable. The text Parker removed, along with dialogue with the first author, indicated that Parker did not wish to be identified in the open dataset. In this case, the first author decided not to publish the interview transcripts at all, despite Parker's consent to publication, to protect Parker's dignity and integrity.

In the published transcripts, several other study participants are likely to be identifiable by someone with knowledge of their research. Further, knowledge of selection criteria for study participants, as described in existing publications(L. H. Kvale, 2022), could also contribute towards identification of participants.

Participants' experiences of sharing de-identified data

During the final round of interviews, several participants shared their opinions about their experiences of reviewing their first round interview transcripts and sharing their data. These opinions were noted in the first author's interview notes. This section first addresses how participants described the experience of reading their transcripts before considering participants' and perceptions relating to sharing the study data. Participants diverged in how they experienced the process of making their data open. Some expressed that reviewing their own words made them feel vulnerable, while others appreciated the possibility to review transcripts. 13 participants did not remove any text but consented to publishing the transcript in full.

Personal emotional experiences

Having one's conversation laid out in text gives a sensation of exposure. One participant, Linden, mentioned how they felt "so stupid" when reading their incomplete sentences. In her notes, the first author describes being surprised about Linden's reaction:

"I was surprised that Linden, of high status, position and with extensive research background experienced [reading their transcript] so harshly. I assumed that my participants are all aware of being intelligent, still reading themself on paper is experienced by Linden so brutally. Linden described a real emotional breakdown at 'sounding so incoherent.' " (notes from interview with Linden, policymaker)

This sentiment was echoed in several second-round interviews:

"After the interview, they talk about the discomfort of reading themself orally on paper, with lines of reasoning with no clear start or ending" (notes from interview with Sam, researcher)

Other participants highlighted the burdensome workload of reviewing their transcripts, including the time taken correcting grammatical errors and understanding sentence fragments. One of the researchers suggested that more elaboration of the text in the transcripts would have been useful to increase readability of the interview transcripts:

"One recommendation for later projects is to modify transcripts so that they consist of complete sentences; this would not reduce the truth content in the reporting… and would probably increase the understanding and ensure more unambiguous interpretation." (Sam, researcher) ([author], 2022 p 90)

In this case the transcripts were verbatim, including filler words, false starts, incomplete phrases and off-topic comments. Editing the transcripts so that each sentence had a start and an end would increase readability for the participants when going through the text.

Chris points to how the review of transcripts is overwhelming for a research participant:

"As participants we can all say: now this is getting too demanding. That is a right the participant has, and in that way all informants have absolute security. But on the other side take qualitative methods and informants, if every quote has to be approved by the informant. Then the informant is interfering with the analysis in a way that would interfere with the research quality. So, something will have to rely on trust to the researchers understanding of the issues explored, and I don't think it has to be that difficult" (Chris, Researcher)

Chris was themself considering dropping out of the study as they perceived the workload in participating high, at the same time as they did not consider themself an expert. Both one of the policymakers, Billie, and researcher Criss address how privacy protection and mutual trust is important in regard to requirement of participants. This trust also entails the protection from unwanted negative consequences, when participants express uncertainty or feel they don't really have reflected enough on a topic. In a broad yet specialised topic such as data sharing, inducing the voice of multiple stakeholders were important to include the opinions of researchers "who do not think of data management plans 24/7" (Parker).

Context related emotions

The research support staff Bobby and Dylan described feeling hesitant about publishing their transcripts, Bobby points to how research as context makes them think differently about their statements. Awareness of how their statements are part of a dataset to be coded, analysed and discussed in publications makes Bobby more sceptical to sharing the transcripts in full. They again find this experience useful in their work with data management services. For some, this was their first hands-on encounter with qualitative research, and expressed that their experiences as research participant would making them more cautious in advicing anonymisation as strategy for researchers to share qualitative interview data in the future. The self-image the research participants felt embedded in interviews as research data, also increased awareness on the importance of consent and dialogue with participants around sharing for ethical reasons. It also provided awareness on the relation of mutual trust between researcher and participant, by first-hand experience. This was shared by research support staff both in the context of the second interview and elaborated on in the consent form for data sharing:

"And this was quite an interesting experience for my own sake, when I went through the interview and thought – do I stand for this? Do I want this removed?" (quote from the interview with Bobby, research support staff)

"Dylan also shared thoughts of why they might not want to share the transcripts. The fact that they make several claims regarding issues they might not know that much about, things that come up in the conversations but which they haven't actually reflected on." (Notes from interview with Dylan, research support staff)

"It is interesting to be in the information position and relate to one's own statements. Even if the information I have provided is in no way sensitive or provoking, I feel some hesitation in accepting publication of the interview material, even in de-identified form. This is interesting and will be a useful experience to carry along with in the work with support services for those working with qualitative data". (comment in the consent form from Bobby, research support staff) (L. Kvale, 2022, p. 91).

Bobby also addresses how they experience contextual differences from expressing something in their private sphere, awareness of the data being qualitatively analysed and deiminated in publications makes them hesitant about publishing the transcripts in full. For Bobby the control of what information to share in which context is a central part of their privacy perception:

“When we discussed anonymizing and sharing the research data, as soon as you are an embedded part of the analyzed material, in a way you get categories to yourself, then I feel much more skepticism toward getting this published. Even if I could have said many of the same things in my private Facebook account or in presentation, but something about being part of a study makes it different.” (Bobby, research support)

Elliott addresses how the experience of being a research participant provided a new perspective on data sharing:

“Interesting theme and interesting question, which initiated reflections and ideas. Considering my work, participating in the study felt relevant and useful.” (Elliott, research support)

This aspect of participating in a research study was shared also Bobby and Dylan who like Elliott work in research support services advising researchers on data sharing as part of their job. Sitting in the research participant chair made them more aware of how sharing is perceived from the participant’s side of the table. As research support services, research participants are not part of their sphere, still they are likely to advice researchers on how to best balance privacy and data sharing in research.

DISCUSSION

To live up to the ideals of Open Science and at the same time securing research participants’ privacy and integrity is challenging when data are collected using qualitative interviews (de Koning et al., 2019; Mozersky, Parsons, et al., 2020). We find that anonymization of interviews within a small community of domain experts would require the removal of significant amounts of useful contextual information, in addition to modifications of expressions that make individuals identifiable. For qualitative data this approach could easily ruin the data quality and hence the value of the shared data.

Removing context and details in qualitative data is likely to lower the quality of the data shared for future qualitative analysis. We have studied stakeholders involved in data sharing curation and demonstrated how removing details about their practices has revealed the complexity of creating anonymous qualitative data. Anonymizing data is often presented to qualitative researchers as an option for making it possible to archive and share qualitative data openly (Huma & Joyce, 2022; Mozersky, Parsons, et al., 2020), this was not perceived as a realistic possibility in this material.

Anonymous data can be shared without conflicting with the GRDP requirements. Still research ethical requirements for research on human subjects would apply (The Norwegian National Research Ethics Committees, 2014). Scholars have previously argued that sharing of human subjects data from health research is not compliant with GDPR (Ursin & Bentzen, 2021). Analysis of the interviews from the current project made it clear that identification of participants would be possible even if directly identifiable information such as names and affiliation were de-identified.

The debate regarding anonymity shows that it is hard to achieve, and not always the appropriate approach to privacy (Barocas & Nissenbaum, 2014). Gürses identifies multiple different approaches to privacy (Gurses, 2014), highlighting how privacy are different things to different people, and that GDPR has introduced a compliance approach to privacy which does not aim at providing individuals with control of their data (Fiebig et al., 2022). According to Nissenbaum and Barocas, it is important to discuss what privacy is protecting, and the context in which human subjects data are managed, rather than focusing on finding solutions to ensure anonymity. Our study shows it is possible to re-identify research participants in qualitative interviews, either by close knowledge of the domain investigated or by access to additional data. As qualitative researchers we continuously strive to balance details with confidence. Open sharing in a repository while guaranteeing participants anonymity in the interview transcripts is often not possible. Therefore, the first author pursued a different strategy of consent through dialogue and review of the data with each participant building on the principles of privacy in context (Nissenbaum, 2010).

To balance sharing and privacy without focusing on anonymity, the first author selected a strategy of informing participants and allowing them to review and select what to share. This, we believe, is a fruitful way in reaching the goal of making research data “as open as possible, as closed as necessary”. Member-checking and the relation between researchers and participants can be used as an asset to achive data sharing (Karhulahti, 2022) and enable shared stewardship. It does, however, raise some new challenges. The findings show that reviewing transcripts is not as straight forward as it can appear, further the removal of elements in the text can affect adepts’ attempts at reanalysing the material, and,finally, the usage of context and quotes in publications can come in conflict with aims at protecting the participants identity.

Soundness, transparency and accessibility are important Open Science principles, but, as we have shown, they are not easy to support in qualitative research (Huma & Joyce, 2022). Huma and Joyce argue that the contextual dependency of qualitative data “warrants a rethink of whether and how such data can ever be reused in secondary analyses” (Huma & Joyce, 2022). We will, on the other hand, argue that transparency with respect to how conclusions have been made is important and that secondary use of data might be of a different nature. Omitting details from the interview transcripts weaken the data quality and affects potential re-analysis of the material. Transparency may be strengthened by indicating where text is removed. We recommend that the researchers review the request for omitting text carefully and expand on what is omitted where necessary.

The time and context of qualitative interview data may also be relevant, so that issues that are considered sensitive may change over time. Qualitative interview data may thus be released at a later date. On the other hand, data may also become more sensitive with time so that researchers’ knowledge of the context and ethical training is essential for evaluating what is the correct approach (de Koning et al., 2019). Un-sharing published data is difficult, caution from the researchers side is therefore essential to maintain participants trust and willingness to participate in research.

How research participants experience sharing their data is important.The vulnerability participants felt when reading their own transcripts is relevant for personal privacy and integrity. Making participants “feel stupid” is a negative consequence, which, according to general research ethical norms should be avoided. In this study participants were resourceful adults and experts on research data sharing, we therefore considered the study to be conducted in an ethical and responsible way. In contexts including vulnerable participants we would not recommend the reviewing of complete transcripts as a strategy to share data openly, neither would we recommend it for transcripts where potential sensitive issues are investigated.

Some participants in our study emphasize the workload of reading and deciding on what to share in the transcripts. The literature on member-checking also address the presumption that participants have time and capacity to read through the transcribed text (Carlson, 2014). To reduce the workload and exposure transcribing with complete sentences is one option. An alternative is to write extensive summaries of each interview for participants to verify and share as open data, while the complete transcripts remain closed.

Qualitative interview data have the qualities of observational data and are unique and irreproducible and thus valuable data to keep (Borgman, 2015). Their potential future value include historical research and longitudinal studies documenting changes in societies. The general hesitance multiple participants express towards data sharing illustrate that the ideal “as open as possible and as closed as necessary” in the context of qualitative interview data is not easily fulfilled, but require options for restricted access sharing. Because how and what data are used matters to research participants (Hardy et al., 2016; Shah et al., 2021), we call for broader exploration of qualitive data sharing with predefined purposes. Involving research participants in a dialogue on how to share data require extra work from both researchers and participants, but may be of great value to the research community as well as society, in general. Within archival studies, shared stewardship is used as a concept to describe involvement and dialogue with the population from who the material originates(First Archivist Circle, 2007). Applying this concept to research data management can contribute to the acknowledgement of the participants as stakeholder. Shared stewardship would call for dialogue with and involvement of participants in research, and would help researchers arrive at the right approach to privacy in each project. We therefore recommend further studies on shared stewardship and sharing of human subjects data which includes the view of research participants.

For sharing qualitative human subjects data, it is necessary for research data services to provide guidance for researchers on how to balance sharing of data with respect for the participants. While clear and explicit consent for data sharing makes open sharing of human subjects’ research data possible (The Research Council of Norway, 2021), each case will require individual assessments of what is ethical.

CONCLUSION

Identification and self-image are two different aspects of research participants’ privacy. Identification often receives more attention but protecting the participants’ self-image is equally or possibly more important to research participants. Our findings show how anonymisation can be impossible to achieve in interview transcripts, without losing value. Anonymising participants is particularly difficult when conducting qualitative research within a small community (Haugen & Skilbrei, 2021). Rather, protection of participants’ privacy can be achieved through protection of their self-image. This corresponds with the fundamental ethical guidelines for research to respect the people who participate in research, and further that “researchers shall seek to ensure that their activities produce good consequences and that any adverse consequences are within the limits of acceptability” (The Norwegian National Research Ethics Committees, 2014).

The exposure participants felt when reading their own transcripts is an important addition to the debate of personal privacy from a research data management perspective, leading to the question of whether there can be too much sharing and too much transparency? We argue that transparency and dialogue with the participants is important both for privacy protection and research quality. However, dialogue involves listening to the participants' wishes and adapting to their requests. Shared stewardship is a concept from archive sciences embodying the participants' rights and interest in their own material, and ensuring shared stewardship is relevant to research data sharing. The research data sharing community could by learning from indigenous communities by using shared stewardship to empower the data subjects.

De-identification and plain language consent forms are better strategies for protecting privacy than anonymisation, which is increasingly difficult to achieve. Protection of self-image and "good consequence" (The Norwegian National Research Ethics Committees, 2014) according to general research ethical guidelines, are important perspectives in research data sharing. However, open data sharing exposes human subjects to further risks as their data is available to all members of society, for whom ethical norms of the research community may not apply. Therefore, data sharing services should address the sharing of human subjects' data with care, and set their own ideals of openness aside when necessary. By sharing experiences gained in the sharing of qualitative human subjects' research data, we wish to illustrate how there is not a one-size-fits-all solution for privacy protection, but rather a need for a respectful and reflective dialogue.

ACKNOWLEDGMENTS

We wish to thank the participants of the study for their time, commitment, and valuable contributions.

REFERENCES

Barocas, S., & Nissenbaum, H. (2014). Big data´s end run around anonymity and consent. In J. Lane, V. Stodden, & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good: Frameworks For Engagement* (1 edition, pp. 44–75). Cambridge University Press.

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? *Qualitative Health Research*, *26*(13), 1802–1811. https://doi.org/10.1177/1049732316654870

Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press.

Carlson, J. (2014). Avoiding Traps in Member Checking. *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2010.1332

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, *19*, 43. https://doi.org/10.5334/dsj-2020-043

Creswell, J. W., & Miller, D. L. (2000). Determining Validity in Qualitative Inquiry. *Theory Into Practice*, *39*(3), 124–130. https://doi.org/10.1207/s15430421tip3903_2

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3rd ed.). SAGE Publishing.

de Koning, M., Meyer, B., Moors, A., & Pels, P. (2019). Guidelines for anthropological research: Data management, ethics, and integrity. *Ethnography*, *20*(2), 170–174. https://doi.org/10.1177/1466138119843312

DuBois, J. M., Strait, M., & Walsh, H. (2018). Is It Time to Share Qualitative Research Data? *Qualitative Psychology (Washington, D.C.)*, *5*(3), 380–393. https://doi.org/10.1037/qup0000076

European Commission. (2016). *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. European Commission.

Regulation (EU) 2016/679, Pub. L. No. 32016R0679, 119 (2016). http://data.europa.eu/eli/reg/2016/679/oj/eng

Fiebig, T., Gürses, S., & Lindorfer, M. (2022). Position Paper: Escaping Academic Cloudification to Preserve Academic Freedom. *Privacy Studies Journal*, *1*, 51–68. https://doi.org/10.7146/psj.vi.132713

First Archivist Circle. (2007). *Protocols for Native American Archival Materials*. https://www2.nau.edu/libnap-p/protocols.html

Forbat, L., & Henderson, J. (2005). Theoretical and Practical Reflections on Sharing Transcripts With Participants. *Qualitative Health Research*, *15*(8), 1114–1128. https://doi.org/10.1177/1049732305279065

Guillemin, M., Barnard, E., Allen, A., Stewart, P., Walker, H., Rosenthal, D., & Gillam, L. (2018). Do Research Participants Trust Researchers or Their Institution? *Journal of Empirical Research on Human Research Ethics*, *13*(3), 285–294. https://doi.org/10.1177/1556264618763253

Gurses, S. (2014). Can you engineer privacy? On the potentials and challenges of applying privacy research in engineering practice. -*Communications of the ACM*, *57*(8), 20–23.

Hardy, L. J., Hughes, A., Hulen, E., & Schwartz, A. L. (2016). Implementing qualitative data management plans to ensure ethical standards in multi-partner centers. *Journal of Empirical Research on Human Research Ethics*, *11*(2), 191–198. Scopus. https://doi.org/10.1177/1556264616636233

Haugen, H. Ø., & Skilbrei, M.-L. (2021). *Håndbok i forskningsetikk og databehandling* (1. utgave.). Fagbokforlaget. https://www.nb.no/search?q=oaiid:"oai:nb.bibsys.no:999920160576102202"&mediatype=bøker

Hoepman, J.-H. (2021). *Privacy Is Hard and Seven Other Myths: Achieving Privacy through Careful Design*. The MIT Press.

Huma, B., & Joyce, J. B. (2022). ‘One size doesn’t fit all’: Lessons from interaction analysis on tailoring Open Science practices to qualitative research. *British Journal of Social Psychology*, *15*(1). https://doi.org/10.1111/bjso.12568

Karhulahti, V. (2022). Reasons for qualitative psychologists to share human data. *British Journal of Social Psychology*, bjso.12573. https://doi.org/10.1111/bjso.12573

Kvale, L. (2021). *Data from a three-phase Delphi study used to investigate Knowledge Infrastructure for Research Data in Norway, KIRDN_Data v.2* [Data set]. OsloMet - Oslo Metropolitan University; Zotero. http://doi.org/10.5281/zenodo.5582714

Kvale, L. H. (2022). *Perspectives, roles and knowledge transfer amongst stakeholders of research data sharing* (Vol. 2022 nr.15). Department of Archivistic, Library and Information Science, Faculty of Social Sciences, OsloMet - Oslo Metropolitan University.

Kvale, L. H., & Darch, P. (2022). Privacy protection throughout the research data life cycle. *IR Information Research*, *27*(3). https://doi.org/10.47989/irpaper938

Kvale, L. H., & Pharo, N. (2021). Understanding the data management plan as a boundary object through a multi-stakeholder perspective. *International Journal of Digital Curation*, *16*(1). https://doi.org/10.2218/ijdc.v15i1.729

Kvale, S. (2007). *Doing interviews*. SAGE.

Leopold, R. (2019). What Is Shared Stewardship? New Guidelines for Ethical Archiving. *Folklife - Smithsonian Center for Folklife and Cultural Heritage*. https://folklife.si.edu/news-and-events/shared-stewardship-new-guidelines-for-ethical-archiving

Lund, B. D. (2020). Review of the Delphi method in library and information science research. *Journal of Documentation*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/JD-09-2019-0178

Mozersky, J., Parsons, M., Walsh, H., Baldwin, K., McIntosh, T., & DuBois, J. M. (2020). Research Participant Views regarding Qualitative Data Sharing. *Ethics & Human Research*, *42*(2), 13–27. https://doi.org/10.1002/eahr.500044

Mozersky, J., Walsh, H., Parsons, M., McIntosh, T., Baldwin, K., & DuBois, J. M. (2020). Are we ready to share qualitative research data? Knowledge and preparedness among qualitative researchers, IRB members, and data repository curators. *IASSIST Quarterly*, *43*(4), 1–23. https://doi.org/10.29173/iq952

Nissenbaum. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University press.

Ohm, P. (2014). Changing the Rules: General Principles fro Data Use and Analysis. In J. Lane, V. Stodden, & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good: Frameworks For Engagement* (1 edition, pp. 96–111). Cambridge University Press.

Lov om behandling av personopplysninger, LOV-2018-06-15-38 (2018). https://lovdata.no/dokument/LTI/lov/2018-06-15-38

Saldaña, J. (2016). *The Coding Manual for Qualitative Researchers* (3rd ed.). SAGE Publishing.

Shah, N., Viberg Johansson, J., Haraldsdóttir, E., Bentzen, H. B., Coy, S., Mascalzoni, D., Jónsdóttir, G. A., & Kaye, J. (2021). Governing health data across changing contexts: A focus group study of citizen's views in England, Iceland, and Sweden. *International Journal of Medical Informatics*, *156*, 104623. https://doi.org/10.1016/j.ijmedinf.2021.104623

The Norwegian National Research Ethics Committees. (2014). *General guidelines*. Forskningsetikk. https://www.forskningsetikk.no/en/guidelines/general-guidelines/

The Research Council of Norway. (2020). *Forskningsrådets policy for åpen forskning [Pollicy for open science]* [Policy]. Norges forskningsråd. https://www.forskningsradet.no/siteassets/tall-og-statistikk-seksjonen/apen-forskning/nfr-policy-apen-forskning-norsk-ny.pdf

Tibor, K. (2021). *Research Data Management and Data Literacies*. Elsevier Science & Technology.

UNESCO. (2021). *UNESCO Recommendation on Open Science*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000379949

Ursin, G., & Bentzen, H. B. (2021). Open science and sharing personal data widely – legally impossible for Europeans? *Acta Oncologica*, *60*(12), 1555–1556. https://doi.org/10.1080/0284186X.2021.1995894

U.S. Department of Health, Education, and Welfare. (1979). *The Belmont Report* (p. 10).

Wilbanks, J. (2014). Portable Approaches to Informed Consent and Open Data. In J. Lane, V. Stodden, & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good: Frameworks For Engagement* (1 edition, pp. 234–252). Cambridge University Press.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Ziglio, E. (1996). The Delphi Method and its Contribution to Decision-Making. In M. Adler & E. Ziglio (Eds.), *Gazing into the Oracle—The Delphi Method and its Application to Social Policy and Public Health* (pp. 3–33). Jessica Kingsley.