Scoping Review: 43 actions to overcome eight barriers to qualitative data sharing

Evangeline Gowie, Conceptualisation, Methodology, Analysis, Investigation, Writing

Anna Tsakalaki, Conceptualisation, Methodology, Writing, Supervision

Ruth Davies, Analysis

Chloe Beesley, Analysis

Etienne Roesch, Conceptualisation, Supervision

Contact: e.gowie@reading.ac.uk

Method

Search terms	(“qualitative” OR “field Notes” OR “interview”) AND (“open” OR “shar*” OR “repositor*” OR “archiv*” OR “reus*”)
Databases	Pubmed, Web of Science, Psychinfo and the Cochrane database of systematic reviews, OpenAIRE and BASE
Additional search methods	We searched the reference list of all included items.
Search dates	13th February 2023 and 4th June 2023
Inclusion	Suggests an action that an individual researcher can take to overcome a barrier to sharing interview transcripts.
Exclusion	Raw data set, no English language version available

Contact: e.gowie@reading.ac.uk

Results

114 items

48 actions

27 barriers

8 Themes

Themes

Confidentiality

Consent

Misappropriation of data

Context

Copyright

Difficulty obtaining ethical approval

Researcher distress

Time and money

Contact: e.gowie@reading.ac.uk

Making interview transcripts open: Preliminary results from a scoping review

Evangeline Gowie, Anna Tsakalaki, Etienne Roesch

bitly

Introduction

Making research data open is recognised as improving both the quality of individual research outputs, and the effectiveness of the perpetuated 'self-correcting' goals of research overall. Many funders now mandate open data as a condition of their grants. This can be difficult for qualitative researchers, whose data poses different obstacles to the quantitative data that these policies seem to be aimed at.

In preparing our own interview transcripts for deposit in an archive, we struggled to identify guidance that addressed the practical issues we were facing at a detailed enough level to be helpful to us. We began to sort through the extensive literature on the subject, and reasoned that it may be beneficial for other researchers to have access to the results of this exercise.

We therefore began a formal scoping review, focusing specifically on the practical steps that researchers can take towards opening their transcripts. We used the search terms ("qualitative" OR "field Notes" OR "interview") AND ("open" OR "shar*" OR "repositor*" OR "archiv*" OR "reus*"), and searched six online databases covering both published and grey literature, plus the reference list of all included items. The search was conducted between 13th February 2023 and 4th June 2023. The full review will be made available at a later date.

From each included article, we extracted specific actions that a researcher could take to overcome an identified barrier. This pre-print represents the information extracted by only one coder (the author). At the time of release, two more coders are double extracting information from items identified during the literature search. The content or organisation of the output may therefore differ to that of the final publication.

We acknowledge that there are many important philosophical objections to archiving transcripts that are not covered in this review. We do not intend to dismiss these, but we write for the researcher who a) is mandated to make their data available or b) has examined these objections and has determined that the benefit of opening data exceeds the concerns raised in the instance of their own dataset.

We also note that there may be further practical barriers for which no suggestions were made, and I therefore did not extract.

Finally, we intend for the scoping review to make the decision process easier for researchers, but not to make any decisions for them. Many suggestions are contradictory, or only relevant to specific types of data. It is up to the reader to critically assess which are potentially appropriate for their project.

How to use this resource

Barriers and corresponding suggestions are summarised into the table for each theme, and further explained in the text below. Click on the barrier to go to the corresponding text explanation. In the suggested reading column, we provide links to literature that we think will help interested readers to understand each individual suggestion.

Not all suggestions have recommended reading. This is because some suggestions were only mentioned very briefly, and we do not feel that any of the sources we found during the review will contribute to understanding more than our short summary.

If you are interested in the full references, we have made our NVivo output available on the Open Science Framework.

Contact: e.gowie@reading.ac.uk

Theme 1: Confidentiality

Barrier	Suggestion		Suggested Reading	Content
1A: Identifying identifiers	1A(i): De-identification analysis		Campbell, 2023	Develops and validates a three phase analysis protocol for de-identifying dyadic qualitative data
			Campbell, 2024	A webinar on the above analysis protocol
			Dodds et al, 2020, supplementary material	An example of a de-identification analysis protocol.
			Saunders et al, 2012	A reflective account of conducting a de-identification analysis based on six key areas (e.g. family relationships, religious background)
1B: Population insider knowledge	1B(i): Consult insiders during the identification process via…	Member checking		
		Verbal discussions with participants	Campbell et al, 2023, paper and supplementary material	The supplementary material contains a script for discussing de-identification with participants, and the paper discusses the authors' experience of using it
		Stakeholder consultations	Campbell, 2023	Explains the process of, and learnings from, consulting relevant professionals during de-identification protocol development
			Campbell et al, 2024	A webinar on the above process.
1C: Identifier remediation methods	1C(i): Blurring		Dodds et al, 2020, supplementary material	Example project protocol of what to blur and how to blur it
			Dunning and Camp, 2015, anonymisation protocol	Example project protocol of what to blur and how to blur it
			Saunders et al, 2012	A reflective account of de-identification, discussing blurring examples and the decision process behind them
			Campbell, 2024	A webinar including reflection on the thought process behind how to blur various categories of data
	1C(ii): Removal		Dunning and Camp, 2015, anonymisation protocol	Example project protocol of what to remove and how to present this in text
			Campbell, 2024	A webinar including personal decisions to protect identities over the value of the data set for future researchers
	1C(iii): Fictionalisation		Saunders et al, 2012, pg 672	Short discussion on the decision to fictionalise some information
1D: Human error	1D(i): Minimise the amount of identifiers collected			
	1D(ii): Third party checks		Campbell, 2023, pg 9-10	The section "Phase 3: Assessing the Validity of the De-Identification Analyses" covers the decision process behind selecting an appropriate third party
1E: Data cannot be de-identified	1E(i):Access controls			
	1E(ii): Request consent for specific, non-removable data		Anonymising qualitative data, UK data service	Briefly discusses a collection of identifiable transcripts, with a description of the edits that still
	1E(iii): Embargo			
	1E(iv): Share data in disembodied forms		Karcher et al, 2021, 40:48 onwards	Addresses what sharing disembodies forms of data may look like in response to a question at the end of this webinar.
	1(i): Anonymise both the participant and the interviewer		Yardley et al, 2013, pg 110	Includes an extract from the interview in which a participants made this suggestion, including a rough structure of how this would work.

Barrier 1A: I don't know how to identify 'identifiable information'.

Suggestion 1A(i): Approach "de-identification as a type of qualitative analysis in its own right[1]"

By dividing transcripts into 'data points'[1] then systematically examining each point to locate identifiers. We extracted four 'research questions' to consider for each data point:

- "Who else knows that information"[1]?
- "What is known about [potential readers'] intentions, motivations, and behaviors?[1]"
- "What other records contain this information—and how could the information be cross-linked across records?[1]" . Example records include court records[1,2], media coverage, social media or linked outputs from the research project.
- What do datapoints reveal when viewed in combination with one another?[3]

One could use MS Word comment boxes or qualitative data analysis software (QDAS) to conduct this analysis. This allows "cleaning while you cook"[4]; whereby the researcher can tag items for consideration in the de-identification analysis throughout data analysis.

Barrier 1B: Insider knowledge

Suggestion 1B(i): Consult insiders during the de-identification process

Insider expertise can be leveraged during the de-identification process through:

- Member checking. After researchers have de-identified a transcript, the clean version could be sent to the participant along with a description of further protective measures that will be undertaken (e.g specific access controls), allowing them to highlight any overlooked risks.
- Checking if participants want certain sections removed at the end of an interview/focus group. This less burdensome alternative to member checking involves asking immediately after a study whether there are any extracts that participant(s) would like remediated.
- Stakeholder consultations. A third option is to consult members of the target population or subject matter experts as a proxy for the research participants themselves.

Barrier 1C: What do I do with identifiable information?

Suggestion 1C(i): Blur it

Blurring is removing the precision, so as to prevent subject identification, but retaining important information for future readers . Authors usually view blurring as an ideal, recommending moving on to the more invasive methods below only when blurring is not judged sufficient.

Suggestion 1C(ii): Remove it

To maintain usability as far as possible, authors suggest retaining a very high level summary of redacted text.

Suggestion 1C(iii): Fictionalise it

Researchers could fictionalise participants names, with the option for participants to choose their own pseudonym[2,5]. Other details may also be fictionalised to make identifying a participant from the remaining information more difficult.

Barrier 1D: There is a risk that we miss an identifier before sharing

Suggestion 1D(i): Minimise the identifiers your transcripts contain to begin with

"Concrete actions can include greeting the participant by name before turning on the recorder, instructing the participant to avoid mentioning potentially identifying information unless pertinent to the question at hand, and to not ask about potentially sensitive information unless this is clearly motivated by the research question."[6]

Suggestion 1D(ii): Validate by asking experts to re-identify the participants from the clean data

A researcher could ask an expert in the subject area[42] and/or qualitative research[46] to review their anonymisation.

Suggestion 1D(iii): Repository staff may check

We recommend that researchers check the policy of their

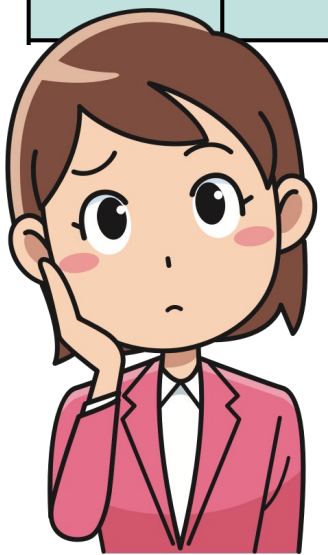
bitly

Contact: e.gowie@reading.ac.uk

How do I know what is an identifier within different departments?

Theme 1: Confidentiality

Barrier	Suggestion		Suggested Reading	Content
1B: Population insider knowledge	1B(i): Consult insiders during the identification process via…	Member checking		
		Verbal discussions with participants	Campbell et al, 2023, paper and supplementary material	The supplementary material contains a script for discussing de-identification with participants, and the paper discusses the authors' experience of using it
		Stakeholder consultations	Campbell, 2023	Explains the process of, and learnings from, consulting relevant professionals during de-identification protocol development
			Campbell et al, 2024	A webinar on the above process.



Alice's Plan

Member checking (opt-in)
Verbal Discussion (adapted Campbell's script)
Consulting Professors

And:
For departments without social contacts, over cautious
Removed 'identifiable speech patterns'

Barrier 1B: Insider knowledge

Suggestion 1B(i): Consult insiders during the de-identification process

Insider expertise can be leveraged during the de-identification process through:

- Member checking. After researchers have de-identified a transcript, the clean version could be sent to the participant along with a description of further protective measures that will be undertaken (e.g specific access controls), allowing them to highlight any overlooked risks.
- Checking if participants want certain sections removed at the end of an interview/focus group. This less burdensome alternative to member checking involves asking immediately after a study whether there are any extracts that participant(s) would like remediated.
- Stakeholder consultations. A third option is to consult members of the target population or subject matter experts as a proxy for the research participants themselves.

Contact: e.gowie@reading.ac.uk

Thank you

Contact: e.gowie@reading.ac.uk

Search: Evangeline Gowie on the Open Science Framework (OSF) for full Nvivo project