

The Ethical and Practical Challenges of Archiving Refugee Accounts: Reflections from Two Research Projects in the UK

Bulletin de Méthodologie Sociologique

2021, Vol. 150 51–69

© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0759106321995707

journals.sagepub.com/home/bms



Emma Stewart

MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, UK

Marnie Shaffer

Franklin University, Columbus, OH, USA

Résumé

Les défis éthiques et pratiques de l'archivage des témoignages de réfugié.es : Réflexions sur deux projets de recherche au Royaume-Uni. Depuis la 'crise des réfugié.es' de 2015, les chercheur.es s'intéressent de plus en plus aux demandeur.es d'asile et aux réfugié.es dans le monde entier. Or travailler dans le champ des *refugee studies* n'est pas sans poser des problèmes de méthodologie et de données. Par exemple, le manque de données statistiques détaillées sur le nombre de réfugié.es a conduit les chercheur.es à privilégier la collecte de témoignages personnels et qualitatifs auprès des populations de réfugié.es. Si cela a produit un volume important de récits détaillés, ceux-ci sont aussi bien spécifiques à une région géographique ou une période donnée. De plus, la collecte des données qualitatives est coûteuse, elle demande beaucoup de temps et d'efforts. C'est pourquoi, au-delà du caractère institutionnel et obligatoire, dans certains cas, de verser le matériel qualitatif dans des dépôts en libre accès, on accorde de plus en plus de valeur à l'archivage des témoignages de réfugié.es. Cependant, l'archivage des transcriptions des entretiens avec les réfugié.es pose de réels problèmes si l'on veut effectivement améliorer les connaissances. Dans cet article, nous examinons le processus d'archivage des témoignages de réfugié.es afin de mettre en évidence les défis pratiques et éthiques que pose le dépôt de documents sensibles. Plus précisément, nous nous appuyons sur le

Corresponding Author:

Emma Stewart, MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Berkeley Square, 99 Berkeley Street, Glasgow, G3 7HR, UK

Email: Emma.Stewart@glasgow.ac.uk

processus d'archivage qui a été exigé à l'issue de deux projets financés par le Research Council au Royaume-Uni. Ce processus a impliqué la préparation et le dépôt des transcriptions d'entretiens de plus de 100 réfugiés.es. Au nombre des principaux défis rencontrés, on évoquera la nécessité de préserver la confidentialité des personnes interrogées, le processus d'anonymisation et la détermination du niveau d'accès aux témoignages pour les futur.es utilisateurs/trice.s. Par la suite, il a fallu également répondre aux demandes d'accès aux données, autoriser la diffusion sélective de ces dernières et définir les conditions de diffusion. Nous abordons enfin plus généralement la question des tensions qui existent entre la procédure à suivre et la micro-éthique, autrement dit, comment les décisions fondées sur des règles diffèrent du jugement qu'on porte sur les cas. Ce faisant, nous mettons en évidence les processus clés et ce qui nous est apparu comme les meilleures pratiques à adopter dans le futur archivage des témoignages de réfugiés.es.

Abstract

Since the 2015 'refugee crisis', the lens of researchers has been increasingly focused upon asylum seekers and refugees around the world. Nevertheless, working in the field of refugee studies poses several methodological and data challenges. For example, there is a relative paucity of detailed statistical data on refugee stocks, which has led to researchers favouring the collection of personal, qualitative stories from refugee populations. Although this produces a substantial volume of rich narratives, these can be geographically and temporally specific. The collection of qualitative data is also expensive, time consuming, and labour intensive. Therefore, alongside the increasing institutional and mandatory demands to deposit qualitative material in open access repositories, there is growing recognition of the value of archiving refugee accounts. There are also significant challenges in archiving refugee interview transcripts to enhance broader knowledge. In this paper, we discuss the process of archiving refugee accounts to highlight the practical and ethical challenges of depositing sensitive material. Specifically, we draw upon the archival process that was required upon completion of two Research Council funded projects in the UK. This involved the preparation and depositing of interview transcripts from over 100 refugees. Key challenges that arose included the need to uphold interviewees' confidentiality, the process of anonymisation, and determining the level of access to grant future users. Subsequent issues have involved responding to data requests, permitting selective release of data, and stipulating conditions for release. We then reflect more widely upon the tensions we encountered between procedural and micro-ethics, namely the difference between decisions based upon rules rather than judgement. In doing so, we consider key processes and highlight best practice to be adopted in the future archival of refugee stories.

Mots clés

archivage, éthique, recherche qualitative, témoignages de réfugiés, transcriptions d'entretiens

Keywords

archiving, refugee accounts, interview transcripts, ethics, qualitative research

Introduction

In recent years, there has been a marked increase of secular and academic interest in refugee migration driven by the so-called ‘refugee crisis’ in 2015. The field of forced migration research, established in the 1980s, is multi-disciplinary in nature and global in focus. The methodological, ethical, and data challenges facing academic and independent researchers in this field have been well documented (Stewart, 2004; Pittaway et al., 2010; Block et al., 2013; Dona, 2007). Nevertheless, despite the extensive consideration of the practical and ethical complexities of research from design through to dissemination, there is little commentary on what subsequently happens to data. As one example, there is a lack of practical guidance available to researchers on the process of curating, depositing and archiving refugee accounts, which this paper specifically aims to address. This is important since the mandatory requirement to archive data leads to researchers and funding institutions facing everyday challenges caused by the practicalities of archiving, and the inevitable tensions that arise between procedural and micro-ethics. By its nature, qualitative research is an inductive, reflexive process that does not always produce anticipated outcomes. Accordingly, difficulties can arise since the outputs of such qualitative research may not neatly adhere to the procedures required for archival. In order to protect the ethical integrity of qualitative research, this calls for an application of micro rather than procedural ethics (Pollock, 2012). Micro ethics is based on judgement rather than rules and relies upon the discernment and integrity of researchers. Our aim is to reflect upon our personal experiences of depositing refugee accounts with the purpose of illuminating how this reflexive approach can inform national archiving guidelines for vulnerable populations.

In the UK,¹ there is a growing trend towards archiving and depositing qualitative interview data generated by social science research projects (Bishop, 2017). This has been borne out of UK data-sharing policy, reforms within UK universities, improved infrastructure to facilitate data sharing, and driven forwards by the institutional demands of funding bodies such as the Economic and Social Research Council (ESRC). In 1994, the ESRC established the Qualitative Data Archival Resource Centre (Qualidata). This was set up to facilitate and document the archiving of qualitative material whilst also highlighting its existence and potential to the wider research community.² Since 2012, Qualidata no longer exists but is now part of the UK Data Archive. This repository acts as gatekeeper for the UK’s largest digital collection of social sciences and population research data, both qualitative and quantitative³. In line with these developments, a key responsibility of all ESRC grant holders is to deposit data created in an appropriate digital repository (Economic and Social Research Council, 2018).

In this article, we consider how to translate the institutional demands to deposit research data in the context of forced migration research. The paper is based upon two ESRC-funded grant projects that involved the archival of interview transcripts. First, a six-month project entitled ‘Becoming British Citizens? Experiences and Opinions of Refugees Living in Scotland’ was undertaken in 2010 (Project 1). The aim of this project was to explore the experiences and opinions of refugees living in Scotland towards the UK citizenship process and becoming British citizens (Stewart and Mulvey, 2011). This research project produced 30 in-depth interview transcripts for deposit. Second, a

two-year project entitled ‘Moving on? Dispersal Policy, Onward Migration and Integration of Refugees in the UK’ took place from 2012–2014 (Project 2). This project mapped the geography of onward migration amongst refugees dispersed across the UK as asylum seekers to explore the main factors that influence refugees’ decisions to stay in a town or city or move on, and how this links to integration outcomes (Stewart and Shaffer, 2015). Some 83 interview transcripts were produced from across 4 different UK cities: Glasgow, Cardiff, Manchester, and London. In both research projects, we adopted a biographical approach (Halfacree and Boyle, 1993), which allowed refugees to talk through their life-history chronologically, in an open-ended way.

We use these two projects as exemplars to illustrate the requirements that are placed upon social researchers and the practicalities of meeting these demands. We argue that researchers face many dilemmas in balancing the practical and ethical demands of archiving due to the lack of guidance on how to deal with sensitive data from vulnerable groups such as refugees. This discussion is useful not only from a practical viewpoint in aiding researchers to meet archiving requirements, but the issues we confronted are likely to be faced by researchers working with other vulnerable groups or sensitive material. Our goal is to discuss the practicalities of undertaking the archiving process and how this can be managed through a data management policy, specifically with reference to refugee accounts. Through reflexive discussion of the research life cycle, we examine several issues including the anonymisation process, the importance of confidentiality, and considerations for the reuse of data through the lens of procedural and everyday micro ethics. In doing so, we highlight some concrete challenges facing researchers in this field and identify recommendations for best practice through ethical reflexivity. The paper contributes to existing knowledge on archiving within social science research and can usefully inform funding institutions and future grant applications.

Qualitative Research and Ethics

Within the social sciences, qualitative data depositing and reuse has taken longer to emerge and develop when compared to quantitative data archiving. Fundamentally, the nature of social science research and the prevailing research culture vis-à-vis qualitative data has presented challenges to archiving interview material. While humanities research, and particularly oral history projects, consider the main purpose of collecting and depositing qualitative data as securing a historical record for current and future access, social scientists have tended to regard qualitative data as a resource to generate new hypotheses, findings, and theories (Kuula, 2010; Parry and Mauthner, 2004). Accordingly, the contrast between regarding qualitative data as a communal resource *versus* the private property of the researcher has contested the prevailing stance being promoted, by bodies such as the ESRC, towards archiving and reusing qualitative data.

The methodological obstacles connected to archiving and reusing qualitative data within social science research have been well debated (Parry and Mauthner, 2004; Bishop, 2014; Roth and von Unger, 2018; Tsai et al., 2016; Yardley et al., 2014). An important set of practical and epistemological issues remain unresolved: ethical considerations, confidentiality, data protection, respondent anonymity, informed consent,

the reuse of data, misinterpretation of data, the threat to intellectual property rights, and methods of gatekeeping for access to data. Fundamentally, the context specificity and co-constitutive processes of qualitative data production raises important questions about reusing data. Some query whether qualitative material can be separated from the relations of its production, which is necessary when evidence is archived for future use by others (Feldman and Shaw, 2019). As a result, questions persist over whether institutional demands to deposit qualitative data are appropriate for the co-constitutive nature of qualitative data, and particularly knowledge that is collected from vulnerable populations.

Ethical Considerations in Forced Migration Research

The field of forced migration research has faced several methodological challenges since its inception. Epistemologically, the policy-oriented focus of refugee research has led to extensive reflection and debate (Black, 2001). There has been discussion over the value of small-scale case studies and calls for data to be comparative across disciplines and geographically in order to contribute to theory-building (see Landau and Jacobsen, 2004). More practically, the methodological and ethical challenges of conducting research with refugee populations have been examined (Bloch, 1999; Akesson et al., 2018). First and foremost, the vulnerability of refugee populations means that any research conducted is ethically fraught (Stewart, 2005; Mackenzie et al., 2007; Pittaway et al., 2010; Surmiak, 2018), and by extension creates obstacles to the archival process due to the ethical considerations required for working with refugees in conflict and crisis situations.

As refugee researchers, we constantly grapple with ethical dilemmas when collecting data, which leads to subsequent questions and difficulties during the archival process. One key dilemma is whether it is ethical to ask refugees to recount painful and distressing memories and how to handle this material once disclosed. At the data collection phase, we designed an interview schedule that was semi-structured and flexible, with topic prompts utilised as required. Combined with the biographical approach, this meant that individuals had the ability to include or exclude information as they wished. We felt this was an important way to build trust during the interview and to give power to the refugee in terms of what sensitive information they wished to disclose. Furthermore, interviews can 'provide a 'therapeutic' function, as in the process of telling, the refugee has an opportunity to try to make sense out of senseless experiences of uprooting, large scale violence, individual torture, and traumatized pasts' (Harrell-Bond and Voutira, 2007: 291). Several of the refugees we interviewed deviated from interview questions to share their narratives in a way that was healing, cathartic, and safe for them. For example, many of the Syrian refugees we interviewed were recent arrivals to the UK. They were very much processing the enormity of the losses they suffered in their homeland and spent much of the interview talking through the stages of denial, shock, and grief they experienced before leaving Syria. They also used the interview as an opportunity to share their struggles and disappointments with using their education and experience to find jobs and start over in the UK. As interviewers, we found that the experience of simply talking to someone who is a refugee, empowering them to

determine how their narrative is told, was often a release that could elicit powerful stories of their journey into the present. The disclosure of sensitive material during refugee encounters – whether it is the pain of the past or the challenges of the present or uncertainty of the future – means that the duty of confidentiality must be strictly honoured in the dissemination and archival of refugee stories. We consider the maintenance of trust between researchers and participants as crucial to the success of qualitative research with refugees and this must be maintained during the archival process (Hynes, 2003; Yardley et al., 2014).

Within refugee studies, there is a rising demand for ethical research to move beyond harm minimisation towards projects that result in reciprocal benefits for refugee participants and/or communities (Mackenzie et al., 2007). When research is conducted with sensitivity and guided by ethics, it becomes a process with benefits for both participants and researchers. In addition to the personal gratification that comes with listening to someone share intimate details about his or her life, one of the agencies that supported us (by providing access to refugees and space to conduct interviews) also used some of our data in their grant applications, hoping to raise money to continue providing services to the refugees in their city. Although depositing interview transcripts may not immediately seem to benefit refugee populations, research has found that interviewees can regard access to research data for future use as a way to engage in the advancement of science (Kuula, 2010). Interviews with refugees can be long and emotionally stressful but having invested their time and energy, it would seem to be beneficial to utilise the data for multiple projects. The archiving of refugee stories can give voice to refugee populations (Harrell-Bond and Voutira, 2007; Benezer and Zetter, 2014). As one example, the Refugee Council Archive at the University of East London was established to document the refugee and migrant experience in a central depository that is available to a broad audience. This type of data is often employed to myth-bust and challenge negative attitudes towards refugees and asylum seekers in society.

Finally, additional ethical issues must be considered when curating and archiving refugee accounts. One key subject relates to power and the relationship between the researcher/curator and interviewee. The risks of asymmetries in power between researchers and research participants have long been noted (Block et al., 2013) and this can also be applied to the curator/depositor. For example, the relationship during an in-depth interview can often be very ‘open, confessional, truth-telling, intimate and sometimes emotional’ (Kuula, 2010: 14). The desire of refugees to tell their ‘stories’ can often overcome their consideration of the potential danger to themselves and their communities (Pittaway et al., 2010). Having collected such data, we feel responsible to protect research participants and one way to do so is to prevent the depositing of data or to redact information; however, there are many uncertainties over what counts as sensitive data and from whose point of view. This highlights vital matters concerning the ownership of data, the powerlessness of research participants after the project concludes and determining what refugee voice(s) are heard. Finally, there is a need to guarantee the premises of confidentiality made to research participants. In our experience of archiving, we have been faced with questions about whether we can trust users to act responsibly or

whether more regulation is needed. In our role as curators/depositors, we have also pondered the potentially new ‘gatekeeper’ role this engenders.

Archiving Refugee Accounts

UK Data Service Guidelines for Depositing Qualitative Data

In terms of procedural ethics, the UK Data Service stipulate guidelines for data archiving and deposit. The unique nature of qualitative research means, however, that exhaustive guidelines are not provided. Below is a summary of the current advice that is applicable to interview material, taken from the online UK Data Service (no date) guidelines.

Preparing Data Files. Allow sufficient time during and towards the end of a project for these preparations. Build in quality control checks for your data capture and cleaning processes.

- Use consistent and meaningful file names, variable names, codes and abbreviations.
- Ensure variable and value labels are complete and consistent, removing any temporary/administrative or dummy variables.
- Check that the level of detail included in the data is suitable for the agreed access.
- Apply an appropriate level of anonymisation.

Sensitive and Confidential Data. These data can be shared ethically and legally by paying attention to three important aspects:

- Discuss data archiving and sharing with research participants to gain their consent for data sharing.
- Anonymise data where needed.
- Consider controlling access to data.

Below we discuss how the archiving process was completed and managed in the context of these guidelines. During the research project, we constructed the data dictionary. At the data collection stage, each transcript was assigned an identifier. This identifies the regional location of the interview, gender, and case number. Next, we categorised key socio-demographic variables to be included in the data dictionary. This data was gathered from interviewees at the beginning and end of the interview (interview schedule), and supplemented by information from the transcript. The data dictionary for Project 2 is listed below as an example. As part of the analysis, we produced an interviewee matrix to summarise the variables. The data dictionary forms the basis of the documentation that is uploaded to the UK Data Service along with the transcripts. Additionally, a Word document was included that contained the interview schedule, the participant information sheet, and the consent form.

Data Dictionary for Project 2

Description of Transcript Identifier

W, S, L, NW	Location of interview: Wales, Scotland, London, North West
F, M	Female, Male
Number	Case number

Label	Description
Interview ID	Interviewee pseudonym
Age	Age range
Gender	Gender (Male/Female)
Employed	Are you currently employed? (Yes/No)
Occupation	Current occupation (Job title)
Student status	Student status (ESOL, Student, University Student)
Marital status	Marital/relationship status (Single, Married, Separated, Divorced, Widowed)
Children	Do you have any children? (Yes/No)
Country of origin	Country of origin (Country specified)
Religion	Religion (Christian, Muslim, Other)
Year of arrival	Year of arrival to UK (Year specified)
Mover/Stayer	Refugee has moved away from (Mover) or remained (Stayer) in the original dispersal city
Interpreter used?	Use of interpreter during interview for project (Yes/No)
Place of interview	Location of interview (Wales, Scotland, North West, London)
Date of interview	Date of interview (Date specified)

After determining the data dictionary and what data would be redacted to ensure anonymity and confidentiality (which will be discussed more below), the process of preparing the actual transcripts for archiving began. This involved re-reading all of the transcripts from both projects. This included 30 interviews for Project 1, which was carried out solely by the original interviewer. For Project 2, this meant re-reading 83 interview transcripts. This task was shared between the Research Associate (interviewer) and the project Principal Investigator (PI). Each interview lasted around 1–2 hours and so produced between 10–20 pages of text. While reading through this material, the data dictionary codes were applied manually. Decisions were also taken regarding data to be redacted. For each project, an appropriate level of anonymisation has to be determined with techniques such as using pseudonyms, applying coding schemes, or removing text to address this. In our case, decisions were based upon discussions between the researchers prior to the archival process but also any other topics that may have been initially overlooked. This led to subsequent deliberations to determine if further information was to be redacted and agreement on the identifiers to be used to ensure consistency. The approach to anonymisation was both protective and balanced and resulted in a systematic convention for handling sensitive material (Surmiak, 2018). For example, names were replaced during transcription with pseudonyms, with search and replace techniques for digital text utilised once the terms to be removed had been identified. The transcripts were also proofread individually to ensure that subtle identifiers were picked up. Some interviews contained so much sensitive data that multiple reviews were necessary to ensure potential identifiers were redacted. Accordingly, this was a time-consuming component of the research projects. On average, we would estimate that the resources required are similar to those for transcription, namely 3–4 hours for each hour of an interview. One unforeseen benefit of going through this process was the re-immersion in the interview material, which

was beneficial for the PI who had not conducted the original interviews in Project 2. This was extremely valuable for the project analysis and write-up.

Discussion and Reflections

Having engaged in the process of depositing refugee interviews on two separate occasions, we have confronted several practical and ethical issues. The first important consideration is the time required to undertake this process. The priority for researchers is to complete analysis and meet deadlines in relation to dissemination, which take precedence for any PI with funding limits. Nevertheless, there is still the requirement to deposit data, which may have been neglected or even forgotten by researchers. For example, the Research Associate from Project 2 was brought back on to the project for archiving, several months after her official role ended. We found this to be a labour-intensive process that requires judgement and analysis of the situation. Indeed, we found the curation process took weeks and months as opposed to days or weeks. We would therefore recommend that researchers explicitly include this element within the research timetable and retain research personnel to work on depositing. This means it is necessary for funding institutions to recognise the time and financial commitment required by the stipulation to deposit and to fund this accordingly. For both projects, we must be candid and state that the curation and depositing process took place after the original research timetable. Learning from our experiences, we suggest that researchers should regard archiving as an integral part of the transcription and analysis rather than an add-on after the project has finished. During both projects we implemented the optimal strategy for transcription whereby each audiotaped interview is transcribed by a single professional transcriber and proofread by the interviewer. On reflection, the issue of archiving should feature more prominently during this process, although we do recognise the potential danger for censorship to occur. Furthermore, when the transcript is being coded in qualitative data analysis software and checked for accuracy, the archived version of the transcript should be prepared at this time by implementing a systematic convention for handling sensitive material. While this may add time to the analysis and coding process, this would avoid the complete re-reading of interview transcripts for deposit in studies with large samples.

Consent

There is a need to maintain the authenticity and integrity of the data while also ensuring that the original consent provided by individuals is honoured. Securing informed consent for future data use from vulnerable individuals is challenging (Feldman and Shaw, 2019). When ethical approval was sought for both research projects, the respective Ethics Committees recommended that a statement be included in the consent form to allow for the storage and future use of interview data. In Project 2, the statement on the consent form read 'I give permission for the researcher to hold the data given for this and any future study'. We feel that there are difficulties in obtaining genuinely informed consent from refugee populations. The standard interpretations of informed consent are based upon the assumption that participants are autonomous, understand

the implications of giving consent and are in relatively equal positions of power with researchers, which does not apply to crisis or conflict situations (Mackenzie et al., 2007). Previous research suggests that consent may have to be gained on multiple occasions (Block et al., 2013) and we feel that consent for data archiving may not be fully understood by some groups. Upon reflection, we endorse a more explicit reference to data archiving in consent forms. The individual may have only consented to the interview for a specific project and not fully appreciated how their data could be used by a new research team in the future. Again, therefore, we wish to advocate for more explicit terminology being utilised within consent forms and the need for researchers to fully explain this to participants. One caveat, however, is that researchers must carefully consider how changes to consent forms may potentially impact upon willingness to participate, particularly with vulnerable populations such as refugees. In order to address this potential conflict between procedural and everyday micro ethics, one solution may be to provide two options for participants: either consent for the current project or consent for both the project and future use (which would include archiving).

There are further steps that can be taken to ensure an iterative approach to gaining consent. Although participants sign the information sheet at the beginning of the interview, it is likely that they are less clear about how data will be used and stored. One practical way to address this is to return the interview transcript to participants to reiterate that the data will be used for publications and archiving, thereby ensuring fully informed consent and the opportunity to opt-out of archiving, if so desired. Indeed, previous research with refugees found that not returning transcripts or publications can be viewed as an extreme breach of trust and exploitation of privilege (Mackenzie et al., 2007). Accordingly, we recommend that completed transcripts should be returned to participants so that they are aware of what material is being deposited. For example, Project 1 transcripts were returned to participants to provide the opportunity for material to be redacted or withheld from publication. In some instances, individuals requested that personal details be withheld but otherwise there were very few requests. Given the lack of response from participants in Project 1, and due to the significantly larger sample in Project 2, transcripts were only returned to individuals who requested the material, which was less than five. We feel this low number can be explained in part by the methods of recruitment. In each of the cities we selected for interviews, refugee organisations were instrumental in our gaining access to interviewees. We established rapport and built trust over time with the organisations that helped us and the refugees who agreed to participate in the research. We also relied on snowball sampling, which again ensured feelings of trust from people who were willing to share their experiences with us. Moreover, our contact details were on the project information sheet provided to participants and those individuals would need to get in touch with us if they wanted transcripts. We suspect these extra steps deterred individuals from requesting copies, or perhaps language translation issues or losing the contact sheet reduced the number of requests. While returning material to participants can take significant time and resources (Pittaway et al., 2010), experience has taught us that this is a way of empowering refugees to make decisions about what material is archived and to ensure they are appropriately informed about what personal data is

being held. It is worth the extra effort. In the future, this could be another opportunity for refugees to determine if they would like to have their interviews archived at all and to give their informed consent.

Data Redaction to Preserve Anonymity and Confidentiality

When preparing transcripts for deposit, one key challenge is the issue of participants' confidentiality and their right to anonymity. Given the nature of an interview encounter, the archivable data file is likely to contain personal information. While preparing the transcripts for deposit, the dedicated repository provided useful guidance. The anonymity of interviewees is protected by the UK Data Protection Act (1998),⁴ with 'sensitive data' subject to stricter rules regarding processing. Archival access can therefore be provided through anonymisation with the removal of identifying details such as names and addresses. While this may satisfy anonymity requirements, it can compromise the integrity and quality of data, or even change its meaning (Parry and Mauthner, 2004). In our case, we feel that the interviews holistically tell a story, and part of that story is the time, place, and context under which those details were collected. Thus, the removal of such data during archiving significantly alters the meaning behind the stories and minimises the power of those accounts. The archiving requirement is also challenging for researchers working with refugees since not only are the stories very personal but can also contain extremely sensitive and potentially damaging or life-threatening information. In our research, we have faced the epistemological dilemma of meeting the mandatory requirement of depositing interview transcripts while having concerns about how this can be conducted ethically when working with material generated by vulnerable groups.

In order to provide access to data, major identifying details were removed from the transcripts such as specific city names, school names, addresses, streets, neighbourhoods, church or mosque names, organisations, support services, places of employment, place of birth, and so on. Our goal was to remove any pieces of information that could potentially identify an individual. Although this may suffice in some circumstances, several challenges occurred across our two projects. For example, Project 1 was a small sample (30 interviews) and geographically limited. Since all of the interviews took place in Scotland, with the majority in Glasgow city, it was possible that nationality would identify individuals due to the small population size of some nationality groups. As a result, 'country of origin' was changed to 'region of origin' (Central, East, West, North and Southern Africa, Middle East, Europe, Central and South Asia) in the transcripts. Additionally, Project 1 involved collaborative research with a refugee organisation who discussed a cautionary tale of a research participant being accidentally identified due to the combination of geographic and personal data. This was valuable knowledge that helped to guide the archiving process for Project 1 as well as the subsequent Project 2.

Indeed, in both projects we found that it was the triangulation of data that potentially threatened the confidentiality and anonymity of participants. This meant that single details were not problematic, but the cross-referencing of data had to be addressed, for example nationality and occupation. These concerns also led to the redaction of data

concerning details of pregnancy outcomes and religion. When possible, appropriately vague labels were provided, but in some instances, this was not possible. As a result, the key descriptors inserted included ‘sensitive details removed’, ‘personal details removed’ or ‘locational details removed’. The most sensitive interviews from Project 2 have pages of redacted data due to the unique nature of the details revealed in the interviews. These labels were also used to remove parts of interviews that the interviewee specifically asked to be kept confidential. To ensure openness and transparency, the removal of information was explained in the data notes submitted with the deposited files.

Taking the decision to redact certain information highlights essential topics that researchers and depositors must contemplate. The depositor must consider the reality of deductive disclosure and attend to potentially identifiable data, such as nationality and location, and determine how to maintain confidentiality and anonymity. This means the depositor must have contextual knowledge and be fully aware of the refugee situation in the study locations. Given the demands on the part of the depositor to have knowledge regarding the data collection and production as well as the refugee situation in the area, size of community, and identifiable nature of the data, we think it is preferable for the original researcher(s) and interviewer(s) to conduct the depositing rather than outsourcing this process. The most practical and obvious impact related to the redaction of details is how the data is used in the future, highlighting the epistemic distinction between inductive and deductive research that is often ignored by archiving requirements (Feldman and Shaw, 2019). In our projects, there may be certain constraints imposed on the reuse of data due to the redactions. For example, the redaction of country of origin in Project 1 means that re-analysis by country of birth would be impossible. While redaction was largely restricted to confidential and identifiable details, in a few minor cases significant sections of sensitive material were redacted, which may impact upon future usability. Cognisant of the sensitivity of refugee accounts, we do recognise the potential value of sharing knowledge and understanding through archiving and data re-use. In some cases, however, we would argue that the sensitivity of the data means that ideally it should not be deposited. In Project 2, we were faced with this ethical dilemma and were forced to adopt the restricted route (discussed more below). Upon reflection, there may be other potential ways to address such concerns.

Handling Sensitive Data

Some sensitive data was redacted from the transcripts before archiving. This included information relating to experiences of domestic abuse, persecution based upon gender, and gender reassignment surgery. In one case, a couple’s previous asylum claim had been denied and they were deported back to their home country, only to be detained upon their return and suffer tremendously as a consequence. They later arrived in the UK and were granted asylum. The couple shared intimate details of their harrowing experience during their interview, shifting the focus away from the goals of the interview and toward the cathartic experience of talking through the obstacles they faced as refugees in the UK. In another case, a male refugee’s political activities before fleeing his home country led him

to reveal details of the resistance he waged and the loss he suffered at home. There was a male refugee who discussed his feelings of guilt and the complications he faced as a married man (whose wife was not with him) involved in an extramarital relationship with a woman in the UK. In another example, a female refugee focused on the fear she felt when she ran away from a wealthy family who had brought her to the UK to work. Other refugees had lived in several UK villages and towns or had unique jobs or experiences that would identify them with minimal effort. Collectively, details from these interviews could be pieced together to reveal the individuals whose identities we promised to protect.

It was common for refugees to discuss their personal struggles with mental health, based on their experiences at home, as migrants and asylum seekers, and finally as refugees. One person, who was a hermaphrodite, spent much of the interview sharing experiences of being shunned by family and feelings of depression, loss, and desperation. A female refugee battled depression and anxiety as a result of refusing an arranged marriage and thereby cutting ties with her family. In both cases, these individuals asked the interviewer to be extremely careful with identifiable data for fear their lives could be in danger if anyone determined their identity. Another woman shared her pain of living as a closeted lesbian until she found catharsis through a support group. In a final extraordinary example, we paused an interview when a male refugee and his interpreter began sobbing as the man talked about his life in his homeland and everything he lost when circumstances forced him to flee his country. Each interview was intimate, and we believe allowing these individuals to speak freely ultimately helped them in their healing process. We eventually returned to the interview topic and finished our data collection. In these most vulnerable moments, there was a humanity that the researchers wanted to honour by omitting extraneous details from the transcripts.

Admittedly, we felt protective of those intimate stories that were shared with us and grudgingly prepared these accounts for archiving. We made the decision to remove sensitive and personal information in such cases for two key reasons. First, similar to above, the triangulation of data can potentially compromise anonymity. Second, since the studies were based upon onward migration flows within the UK and reasons for taking British citizenship, it was considered that such detailed, personal information was not a central element of the research. Indeed, when refugees engage in in-depth interviews it can often be a cathartic as well as painful experience (Harrell-Bond and Voutira, 2007) through the divulgence of traumatic experiences. In both projects, the interviews were semi-structured and allowed for participants to freely share information. Accordingly, this meant that refugees often shared very detailed life events and stories beyond what was required for analysis related to the original research questions. We felt that it would be intrusive to retain some of this very personal data that was shared during the intimate encounter of the one-on-one interview setting.

In our sample of over 80 interviews (Project 2), we identified 5–6 cases that were highly sensitive. We think that depositors should be given the freedom to withhold such cases from the deposited dataset. This would be beneficial for several reasons. First, there is a high level of redaction in sensitive cases that may render them useless for reanalysis. Second, protecting the sensitive data would allow for more free access to the remainder of the sample. Finally, in the long term, this would be less time consuming for the repository and the original researchers. If data cannot be withheld, another solution

would be to embargo data for a set time period, for example 5–10 years. Furthermore, it would be helpful if there were a specific caseworker in the archive repository who dealt with highly sensitive data cases. This would mean developing specific expertise and the subsequent ability to provide tailored advice, and deal with queries quickly and efficiently.

Data Reuse and Determining Future Access

One of the main purposes of archiving qualitative material is to facilitate the reuse of data by future users. While sensitive, personal details within the transcripts were removed to assuage privacy concerns, the usefulness of what remains of those stories depends on the goals and objectives of subsequent researchers who wish to review the dataset for their own projects. For example, one student used Project 2's interview transcripts with Eritrean refugees to write her dissertation. In her follow-up correspondence with us, she indicated the interviews suited her project better than expected. Her project explored the impact of security politics on the refugee experience, with a particular focus on methods. Others who made initial enquiries to access the transcripts did not follow through after we provided them with details regarding the redactions and limitations of the data. For us, the tension between adhering to the data depositing requirement while protecting the integrity of the data and honouring the privacy of our research participants was palpable throughout the archiving process. There were many discussions about how the power of storytelling is lost when the nuances of one's experience or specifics of one's identity are removed. But we feel this loss is the result of more than what is redacted from the transcripts. There is the moment in time when the interview took place, the intonation of the speaker, the setting of the interview, the mood of the participants, the conversations that were not recorded as part of the interview, and so on. In other words, all the details that are not captured in words on paper – in the official transcript – but exist in the memories of those who were there are important pieces of the data collection experience. This collective power of the story, of the experience of telling and hearing the story, is lost in the archiving process. We therefore question the efficacy of the transcripts for reuse in the absence of its full context, as well as our apprehension for how the remaining details might be used and interpreted by someone else.

While being minded of our concerns related to data reuse, we had to determine the level of access to grant future users. Since all data from ESRC funded research must be deposited and we could not omit the most sensitive transcripts, our preferred option, we had to reluctantly negotiate a way forward. During the curation process, the dedicated repository provided guidance on different modes of access, conditions for release, and the provision of usage reports to the depositor (Mannheimer et al., 2019; UK Data Service 2018). Due to the mandatory requirement to deposit our data, in Project 2 we felt that our only option was to utilise the restricted access route offered by the repository. This decision was taken during the preparation of the transcripts mainly due to the highly sensitive nature of certain cases and the potential harm that could result for identifiable subjects. This restricted access arrangement means that before data is released to future users, permission has to be granted by the original researcher(s). In practical terms, this means that any request for data must be considered and approved by

the repository. This is an extra administrative duty beyond the timeline of the research project, but we felt it was important to retain this input into future data access.

Since depositing the data for Project 2, we have received a total of 11 queries. In response, we have explained the sensitive nature of the data and stipulated the conditions for release (e.g. dissertation supervisors request the data for undergraduate students, data is stored on password protected devices, data is only used for the stated project, and data should be destroyed after use). We have received requests from researchers, policy makers, and students but the majority of enquires have been from undergraduate students. We have therefore suggested that a sub-sample of the data may be more appropriate since 83 transcripts may be potentially too onerous for an undergraduate project. After highlighting these issues, and as far as we are aware, around half of the queries have translated to an actual request to the UK Data Service. In terms of publications from the data requests, we have received only one completed undergraduate dissertation. Having responded to these requests, we feel convinced of the importance of the depositor's contribution in determining the level of access and data release. A dialogue with the original researchers has helped to facilitate a selective release of the data, which may be more difficult for those without knowledge of the original project. We also feel that only those with serious interest in the data have pursued access.

Selecting the restricted access route, however, does raise questions about what happens in instances when the original depositor cannot be contacted. The UK Data Service have recently implemented a programme of renegotiation of access conditions for some data collections (Moody, 2015).⁵ For example, when the depositor cannot be contacted, such as if they are retired or have passed away, often the data are considered the intellectual property of the institution where the depositor worked at the time. Rights to grant permission can also be sought from the funding body, such as the ESRC, or the depositor's Estate. Otherwise, some earlier licence agreements granted permission rights to the Director of the Archive and this clause can be used if all other routes have failed. Nevertheless, before releasing data a disclosure review would be undertaken to check the data meet current statistical disclosure standards and to release the data under safeguarded access controls (UK Data Service, 2015). Our experiences of the restricted route highlight the clear limitations of depositor-determined data. On reflection, we have identified potentially important ways forward. One simple change would be to add more explicit stipulations about data access to the public site, such as the need for dissertation supervisors to make the request and take responsibility for data, the fundamental necessity of data security and storage as well as the option of selective release. Given the multiple requests from undergraduate students, we would suggest one solution may be to provide a subset of the data that is specifically identified as suitable for this group. This would be an excellent way to promote secondary data analysis to undergraduate students and produce accessible educational resources that would be invaluable to the future generation of social researchers.

Conclusions

Despite the marked interest in refugee issues, the reality is that refugee researchers still face several practical and methodological challenges. These range from gaining access

to individuals by negotiating with gatekeepers through to issues of language and interpretation. These difficulties will also be undoubtedly heightened during the current global pandemic. As a result, accessing refugees' accounts from a central repository is one practical way of overcoming such issues, especially in cases where researchers lack resources such as undergraduate students. The cost effectiveness of data archiving and sharing by avoiding duplicating research is another obvious practical benefit, although this shift within academia towards an emphasis on value for money faces considerable resistance. In the case of refugees, we agree that the issue of cost should not be the primary focus. Instead, the provision and benefits of archiving refugee data should be regarded as one way to avoid causing harm to participants by reducing unnecessary data collection. Additionally, the establishment of archived refugee accounts is a valuable tool that can be used to combat research fatigue amongst refugee communities as well as to myth bust, challenge prejudice, and contest negative stories about refugees. Nevertheless, a key challenge remains for refugee researchers: namely to determine how the epistemic assumptions that underlie qualitative data collection can somehow be honoured throughout the curation and archival process.

Overall, in our research with refugees and the subsequent process of archiving, the procedural ethics have caused us to reflect more widely upon the everyday micro ethics of conducting research with refugees. We were reminded of the power of the researcher not only during fieldwork interactions but also as curators in the preparation of manuscripts for deposit. While we play an important role in protecting vulnerable individuals by removing sensitive material, we have significant autonomy and power to do so. This means that not only as interviewers but as depositors, we have the power to silence the refugee voice. As discussed above, we tried to negate this issue by asking participants themselves what data could be used for analysis and storage. Nevertheless, there is clearly a power imbalance at work in the process of depositing data, particularly from vulnerable groups. The researcher also has an important responsibility in acting as the gatekeeper to refugee accounts, a role which may be extended for years through the restricted access arrangements put in place with the repository. The process of depositing refugee accounts is far from simple and will continue to raise ethical questions for qualitative researchers.

Acknowledgements

We are indebted to each refugee who willingly shared their experiences and personal stories with us throughout both research projects.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the UK Economic and Social Research Council [grant numbers ES/H038345/1, ES/I010831/1] and in-kind support of the Scottish Refugee Council.

Notes

1. Corti and Thompson (2006) usefully summarise the historical development of archiving interview data as well the emergence of national archives in the UK.
2. Given the resource-intensive nature of archiving, this task is undertaken by the investigators rather than Qualidata, who now act primarily as gatekeepers.
3. The UK Data Archive is a lead partner of the UK Data Service.
4. As of 23 May 2018, the Data Protection Act 1998 was replaced by the Data Protection Act 2018.
5. Other archives have clearer procedures for different levels of access, for example see the Australian Data Archive (no date).

References

- Akesson B, Hoffman DA, El Joueidi S and Badawi D (2018) So the world will know our story: Ethical reflections on research with families displaced by war. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (online journal) 19(3): Article 5. Available at: <http://www.qualitative-research.net/index.php/fqs/article/view/3087>.
- Australian Data Archive (no date) ADA License Agreement Form, Available at: https://docs.ada.edu.au/index.php/Download_License_Agreement_Form.
- Benezer G and Zetter R (2014) Searching for directions: Conceptual and methodological challenges in researching refugee journeys. *Journal of Refugee Studies* 28(3): 297-318.
- Bishop L (2014) Re-using qualitative data: A little evidence, on-going issues and modest reflections. *Studia Socjologiczne* 3(214): 167-176.
- Bishop L (2017) Revisiting qualitative data reuse: A decade on. *SAGE Open* 7(1): <https://doi.org/10.1177/2158244016685136>.
- Black R (2001) Fifty years of refugee studies: From theory to policy. *International Migration Review* 35(1): 57-78.
- Block K, Warr D, Gibbs L and Riggs E (2013) Addressing ethical and methodological challenges in research with refugee-background young people: Reflections from the field. *Journal of Refugee Studies* 26(1): 69-87.
- Bloch A (1999) Carrying out a survey of refugees: Some methodological considerations and guidelines. *Journal of Refugee Studies* 12(4): 367-383.
- Corti L and Thompson P (2006) Secondary analysis of archive data. In: Seale C, Gobo G, Gubrium F and Silverman D (eds) *Qualitative Research Practice*. London: Sage, 297-313.
- Data Protection Act (1998) Available at: http://www.opsi.gov.uk/acts/acts1998/ukpga_19980029_en_1
- Dona G (2007) The microphysics of participation in refugee research. *Journal of Refugee Studies*, 20(2): 210-229.
- Economic and Social Research Council (2018) ESRC Research Data Policy. Available at: <https://esrc.ukri.org/files/about-us/policies-and-standards/esrc-research-data-policy/>
- Feldman S and Shaw L (2019) The epistemological and ethical challenges of archiving and sharing qualitative data. *American Behavioral Scientist* 63(6): 699-721.
- Halfacree K and Boyle PJ (1993) The challenge facing migration research: The case for a biographical approach. *Progress in Human Geography* 17(3): 333-348.
- Harrell-Bond B and Voutira E (2007) In search of 'invisible' actors: Barriers to access in refugee research. *Journal of Refugee Studies* 20(2): 281-298.

- Hynes T (2003) The issue of 'trust' or 'mistrust' in research with refugees: Choices, caveats and considerations for researchers. *New Issues in Refugee Research*, UNHCR Working Paper 98, UNHCR, Geneva.
- Kuula A (2010) Methodological and ethical dilemmas of archiving qualitative data. *IASSIST Quarterly* 34(3-4): doi.org/10.29173/iq455
- Landau L and Jacobsen K (2004) The value of transparency, replicability and representativeness: A response to Graeme Rodgers, 'hanging out' with forced migrants. *Forced Migration Working Paper Series* no. 11, University of the Witwatersrand.
- Mackenzie C, McDowell C and Pittaway E (2007) Beyond 'do no harm': The challenge of constructing ethical relationships in refugee research. *Journal of Refugee Studies* 20(2): 299-319.
- Mannheimer S, Pienta A, Kirilova D, Elman C and Wutich A (2019) Qualitative data sharing: Data repositories and academic libraries as key partners in addressing challenges. *American Behavioral Scientist* 63(5): 643-664.
- Moody V (2015) Helping depositors share data: Negotiating away difficult licence conditions. *Data Impact Blog*, UK Data Service, Available at: <http://blog.ukdataservice.ac.uk/helping-depositors-share-data-negotiating-away-difficult-licence-conditions/>
- Pollock K (2012) Procedure versus process: Ethical paradigms and the conduct of qualitative research. *BMC Medical Ethics* 13(25): doi.org/10.1186/1472-6939-13-25
- Parry O and Mauthner NS (2004) Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data. *Sociology* 38(1): 139-152.
- Pittaway E, Bartolomei L and Hugman R (2010) 'Stop stealing our stories': The ethics of research with vulnerable groups. *Journal of Human Rights Practice* 2(3): 229-251.
- Roth W and von Unger H (2018) Current perspectives on research ethics in qualitative research. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (online journal) 19(3): Article 33. Available at: <http://www.qualitative-research.net/index.php/fqs/article/view/3155>
- Stewart E (2004) Deficiencies in UK asylum data: Practical and theoretical challenges. *Journal of Refugee Studies* 17(1): 29-49
- Stewart E (2005) Exploring the vulnerability of asylum seekers in the UK. *Population, Space and Place* 11(6): 499-512.
- Stewart E and Mulvey G (2011) *Becoming British Citizens? Experiences and Opinions of Refugees Living in Scotland*, Scottish Refugee Council, Glasgow. Available at: <https://www.scottishrefugeecouncil.org.uk/wp-content/uploads/2019/10/Becoming-British-citizens-Experiences-and-opinions-of-refugees-living-in-Scotland.pdf>
- Stewart E and Shaffer M (2015) *Moving On? Dispersal Policy, Onward Migration and Integration of Refugees in the UK*. University of Strathclyde, Glasgow. Available at: <http://eprints.gla.ac.uk/209384/>
- Surmiak A (2018) Confidentiality in qualitative research involving vulnerable participants: Researchers' perspectives. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (online journal) 19(3): Article 12. Available at: <http://www.qualitative-research.net/index.php/fqs/article/view/3099>
- Tsai AC, et al. (2016) Promises and pitfalls of data sharing in qualitative research. *Social Science & Medicine* 169(November): 191-98.
- UK Data Service (no date) *Preparing Data for Deposit: Guidance on Preparing and Managing Data*. Available at: <https://www.ukdataservice.ac.uk/deposit-data/preparing-data/data-files>

- UK Data Service (2015) Personal Communication with UK Data Archive Colleagues, Email Communication.
- UK Data Service (2018) *Data Access Policy*. Available at: <https://www.ukdataservice.ac.uk/media/604526/cd239-dataaccesspolicy.pdf>
- Yardley SJ, Watts KM, Pearson J and Richardson JC (2014) Ethical issues in the reuse of qualitative data perspectives from literature, practice, and participants. *Qualitative Health Research* 24(1): 102-113.