

# Disciplinary perspectives on archiving qualitative data: Can we, should we, how do we?

23 June 2022 • 1:09:08

## SUMMARY KEYWORDS

archiving, data, researchers, participants, research, qualitative data, reuse, consent, people, qualitative, thinking, point, concerns, sharing, qualitative researchers, investments, datasets, required, regulatory frameworks, ethical

## SPEAKERS

Rachel Ayrton,Carolynn Low,Susie Weller,David Zeitlyn

### **Rachel Ayrton** 00:04

Today we're looking at disciplinary perspectives on archiving qualitative of data - can we, should we, how do we? And I first tried to organise this QUEST seminar around March 2020, but a few things came up and somehow it was delayed. So I'm really delighted that we're actually getting to begin this discussion today. And I became interested in these issues because of my ESRC postdoctoral project, which is with the South Sudanese diaspora in the UK looking at national identity and belonging. And as an ESRC-funded postdoc, I'm expected to archive my research data. But I was grappling at that time with producing my data management plan, and some of the complex ethics due to working with a small conflict-affected group and discussing a politically sensitive topic. And so I'm really pleased to welcome Carolyn, Susie and David today to share some of their experiences and insights into the archiving of qualitative data for reuse. Each of our speakers is going to present for about 25 minutes, which will then leave us with a good half an hour for questions and discussions. So without further ado, I'm going to introduce our first speaker, who is Carolynn Lowe. Carolynn is a postgraduate researcher at the South Coast Doctoral Training Partnership. And she's also got a vast amount of professional experience working with refugee asylum-seeking and migrant communities in the voluntary sector. She's currently undertaking data collection for her PhD, which looks at refugee entrepreneurship. And I'm sure she's going to tell us a lot more about in what follows. So thanks for being here Carolynn, and I'll pass over to you.

### **Carolynn Low** 01:46

Okay, well, thank you, firstly, for inviting me along. I'm really excited to be here today. And in my presentation, I'm going to be looking at the "how do we?" aspect of this and talk through some of the ethical and logistical challenges that I've experienced in my research so far, in terms of research design and starting data connection. I'm not going to go into any epistemological concerns of data archiving today, but suffice to say I do come from a largely interpretivist position. And so I see data as co-

constructed and myself as a researcher intimately embedded within that data. And also come from a sociological point of view. I guess the takeaway point from today, from my experience is just how time consuming and thought provoking and how much thoughts required thinking about archiving data for reuse. And I think ultimately, today, I'm really presenting a set of questions.

**Carolynn Low 02:49**

So to give you a bit of background to my research. As Rachel mentioned, I have a background in working with refugees. And my research is on refugee entrepreneurship, and looking at how that make contribute to processes of social and economic inclusion. And my research is taking place in a small city. And it's largely ethnographic in the approach. Primarily, I'm focusing on in-depth interviews. But I also have some observations and informal conversations with participants. And my participants are both entrepreneurs and professionals working to support them. And at the moment, as Rachel said, I am at the point of doing my fieldwork, so I'm about halfway through the PhD, which I'm doing part-time and it is funded by the Economic and Social Research Council. So there is a strong preference that I would archive my data for reuse at the end, but it isn't actually compulsory, as a PhD student.

**Carolynn Low 03:53**

So at the beginning, so at the research design stage, I found that data reuse was presented somewhat unproblematically, initially, just as part of the research process, but coming from a practice background, it immediately raised quite a lot of questions and concerns for me. And although these aren't new questions and concerns, they're, you know, well documented in the literature, I think they're really pertinent to my research. So I guess three really key concerns came up. Firstly, guaranteeing anonymity is extremely challenging if there's access to my full research data. My research is with a small group of participants who are a small and quite distinct population within the city that I'm focusing my research. I'm doing in-depth interviews that contain a lot of biographical information. And so I feel that the potential for exposing identity is really quite high, especially if it's linked to me as the researcher and also linking bits of data together that may be archived. And I do feel that once the place of research is identified, it's not that difficult to work out who the participants involved are, especially if someone's familiar with the context. And even more so if contextual information is archived as well. The second concern that a lot of people have raised, but is so important is how can we consent to the future use of our data when the purpose cannot be anticipated. And I think this is particularly important in the area of migration that I'm researching. It is a highly politicised hostile environment, particularly around refugees and very much currently. And so there are added sensitivities. And a key part for me of encouraging participants to participate has been assuring them of the approach I'm taking, and that I'm looking really at the contribution that refugees and migrants make to UK society. But I can't offer such assurances for any future research project using their data. I don't know the angle or the position that future researchers would take. I don't know how participants accounts would be interpreted. And while from an academic point of view this may be justified, I feel for participants, it could feel like betrayal. They may not want to consent to their data being used in a way that doesn't fit with their own values and beliefs, for example, used to feel anti-migrant sentiments. And I think this speaks to the point that Carusi and Jirotko make that the risks of research can obviously include psychological harm, discomfort or distress, and this includes harm to personal beliefs and values. And also there is an indirect risk through harming the reputation of refugees and migrants as a group. And to be fair, myself, particularly coming from a practice background, I'm not very confident that guidelines, agreements, legislation will

not change in the future or be manipulated, that could compromise current agreements, for example, to preserve the confidentiality of individuals, if it was somehow deemed to be of national interest. So although this might not be as concerning in my research, because I am looking primarily at people with status - I think it would be even more concerning if they were undocumented people I was working with - there may be mentioned of illegal practices, for example. And I think we know that the hostile environment is continually being extended into unanticipated areas requiring data sharing. And it does make me question whether research is immune from this. And therefore there could be a direct physical risk to participants.

**Carolynn Low 07:59**

And my final concern, or major concern at this point was also not recreating the bureaucratic violence of the asylum process. So the silent process, as many of you will probably know, involves repeated recounting of personal stories, intense questioning, frequently disbelief, for example, recalling previous things that people have said, and pointing out inconsistencies. And I'm concerned that the research process doesn't recreate this or trigger these experiences. And arguably, archiving data could be tantamount to dispossession of one's story, as happens in the asylum process, leaving it open to intense scrutiny by an unknown person. So this could be a secondary researcher, or in the asylum process, home office officials who may form an alternative, possibly hostile interpretation than that that was intended by the participant. And I think although research could potentially do this, I am concerned that maybe there the danger is greater in secondary analysis, because there isn't the interaction between the participant and the researcher directly, and they don't have the same bond of trust formed and understanding of context, so again, the risk of a feeling of betrayal.

**Carolynn Low 09:34**

And so at this point, my concerns were primarily ethical. How could I explain the rather opaque process of data archiving in a simple way that would enable, as much as possible, participants to provide informed consent to their data reuse? So, at the research design stage, it did take a lot of researching what this could entail and thinking about it and reflecting on it. I had numerous exchanges with my supervisors, with my funders, with UK Data Service, with the research management team at Southampton, with reviewers in my ethics application process. And I found initially, I could not get guarantees as to the level of access to the data that would be set at that point. And it seems that this is normally something that comes later on at the point of archiving and as a discussion between the researcher in the institution and crucially missing participants. For UK Data Service, although I did get assurances that it was highly unlikely my data would be open access, I don't know at this stage, whether it would be safeguarded, or controlled access. And that decision, they explained to me comes later. They did confirm however, I could deposit partial datasets. So where participants haven't consented to their data being reused at Southampton university, I was actually able to get assurances that in my particular case, the data could be archived with them with restricted access, so only accessible by myself and the research team. And this was really important, I think, because it crucially gave participants a choice, a genuine choice, to not have their data reused. And so in my consent forms, I have a clearly marked optional opt-in consent for archiving with the UK Data Service, I have the possibility that otherwise they can have their data archived at the University of Southampton with restricted access, and accompanied by a rather lengthy, probably still not sufficient explanation in the participant information sheet about what data archiving means. But overall, I would say the process to

get here was really important, but it was time consuming, it was lengthy, at times quite unclear and required negotiation. And I think this really highlights for me the point that these conversations need to happen at the start at research design stage. Now, this obviously still doesn't overcome the point as Parry and Mauthner make that controlling access is not the same as controlling presentation and processing of data. And it also doesn't ensure that you're obtaining genuinely informed consent. And that's what I'm going to discuss next. It arose during my field work.

**Carolynn Low 12:43**

So for me, I think the biggest concern barrier would be language. So most of all, well, all of my participants English was not their first language for a lot of them, they were not fluent in English. And so explaining the complexities of data reuse is really difficult when English is limited. The nuances of what this entails and the potential consequences are really lost in translation. And I find that the participant information sheet is actually pretty inaccessible anyway. And then when you add in data archiving, it just makes it more. A second problem is a lot of my participants were not, are not, very familiar with research. So to be honest, they found the idea of research in general in itself quite difficult to comprehend, particularly because it doesn't lead to any direct, tangible outcome for them. And I got the impression that talking about a data archive was really more than they could imagine and envisage. And obviously, this was compounded by language as well. Thirdly, time, participants are busy like the rest of us. And it's already been a huge challenge to recruit people to my research for various reasons, one of which is time. I find that explaining the research process takes a long time, usually at least half an hour. And I tend to do these in separate meetings. And this is you know, partly because of the language issue. But then when you add in obviously data archiving it can get very complicated and to be quite honest, off putting.

**Carolynn Low 14:28**

And so I guess linked to these points, in some way is suspicion. And I have found that quite a few of people I've tried to engage are really quite suspicious about the research process and data archiving as well. And, to be honest, my reassurances may mean very little when they come from contexts where their information has been collected, often without their consent, and used and distorted in ways that can harm them, both in their countries of origin but also through the assignment process in the UK. The idea of retaining information about them for future use in ways that cannot be anticipated, arouses more fear and suspicion. If you compare this to the Home Office, for example, keeping information on people for many years, which they then reuse at a later point, in another context to their detriment. And I think reassurances as well that data reuse will be carefully managed may not be believed, particularly as the person that will be managing this is not myself as the primary researcher, but an unknown other. So, finally, it kind of leaves me in the position of quite ambiguous concern. I have real concerns that my participants understand data reuse and the implications and are therefore providing informed consent as much as it can ever be for anyone. In several cases, my participants have said that they don't know whether they want their data stored with UK Data Service for reuse, even after we've had fairly lengthy discussions about it. Now, in a few of these cases, I do think that is because they they didn't understand. But in other cases, I think it was really because they could not anticipate what future use of their data may look like. And I've had quite a few requests that I myself decide on their behalf. I think they trusted me as a researcher, but not the research process. So in these cases, we've actually opted not to consent to data reuse, because I felt to do otherwise would not have been informed consent. But

it raises the difficult issue of decision making and the burden of this being displaced to the researcher. And also, as Carusi and Jirotko raise, there is the risk of paternalism in this that I determining on behalf of the participant what is needed to protect them. So there's the risk of being under- or over-protective.

**Carolynn Low 17:16**

So in terms of where I'm at now, thinking about going forward in my research process, I still have more questions. In terms of where participants have consented to the reuse of their data, I feel that it would be best practice to go back to them. And check once again before their data is archived, thus treating consent as an ongoing process. So ideally, I'd like to take their transcripts or field notes, or observations about them, and ask if they are happy for those to be archived, and give them an opportunity to redact any parts they wish. And I think that's important, because they'll know at that point, what they're consenting to. At the beginning, when you ask this question, it's very difficult to answer because you don't know what you're going to say during the research. And whilst this would be a great thing to do, and I would hope I will do this, it does raise the issues of participants having time to do this to read through lengthy information. And also think about what that might mean, but also the difficulty of language. But I do still feel they should have this option. The second issue is, as many people have talked about, the time required for data preparation for archive, it's going to be vital that I obviously remove direct identifiers, but lots of indirect identifiers, and probably will have to redact parts where there are personal stories about the participants or about other people. And I'll also need to remove information about myself as well, which is present in the data. And also when you think about the data set as a whole how when you combine different bits of data, whether that increases the risk of identification, and also a need to think about what data should be deposited. Should I be putting in my process notes, my personal reflections. These have great contextual information within them and including my positionality but then would very significantly increase the risk of identification. And so thirdly, this comes down to the question of data quality as a result of the above. So does depositing a partial dataset make any sense? And this would be partial in terms of chunks of data that have been removed from the data that is deposited, and also the absence of whole transcript from interviews or observations with participants who have not consented to their data reuse. And I wonder if this would be helpful for future researchers. And in this case, what is the main purpose of data depositing? And I think this compound compounds the point that Feldman and Shaw make, that from an interpretive viewpoint, the extracted data can only ever be partial, because it's intimately linked to the researcher in the context. So secondary data analysis will inevitably produce an alternative, possibly very different interpretation from my research, which is fine, but with these interpretations, be fair to participants, given the partial data?

**Carolynn Low 20:46**

So that leaves me with a few concluding thoughts. I very much agree with the point made by Feldman and Shaw that trust that's built between the researcher and the participant is critical, is a critical relationship in terms of consent. And data reuse can really challenge this. As participants trust me, the researcher and the information that I'm giving them about data reuse, being true and accurate to the actual outcome. But how can I give them reassurances when I'm not sure myself? And as both Feldman and Shaw, and Parry and Mauthner, say, trust is crucial to building rapport and the data that's generated in the research. And obviously thinking about other people, reading this data, and analysing and reusing the data does shape the information that people are willing to share in the research, and

also the questions that I'm willing to ask as a researcher, which has obvious implications for data generation. But equally, I'm concerned that participants may forget this point and share more than they would want shared with people beyond me as the researcher.

**Carolynn Low** 22:04

And so I think, really, these discussions highlight the critical importance of context, considering the specific situation of participants in terms of their abilities and positions in society, and how data reuse may impact on them. And it makes me think does there have to be an element of thinking on behalf of participants here, because they themselves may not be in a position to fully understand potential consequences and impacts of data reuse, if we can even ourselves. And obviously, I think we need to take seriously the risk of data archiving and reuse exploiting participants. And then the last couple of points are really the practical ones, just that I think that we need to think about data reuse from the start of the process, there needs to be clear options for participants to allow them to not consent to their data reuse, but still participate in the research and to consider consent as much as possible as a process. Thank you very much. I look forward to the discussion later.

**Rachel Ayrton** 23:10

Thank you so much Carolynn. That was really fascinating and very thought provoking. Really wonderful. Thank you. We're going to move straight on to our second presentation now, which is going to be given by Susie Weller. Susie is a senior research fellow at the universities of Southampton and Oxford. Over the past 20 years, she's developed an expansive research portfolio, focusing on research with children, youth and families. She's also an avid methodologist, including work on the breadth and depth approach to large-scale qualitative analysis, and using remote interview methods in qualitative work. So we're really looking forward to what you're going to share with us about data archiving and reuse today, Susie. Thank you.

**Susie Weller** 23:53

Thank you, Rachel. And thank you very much for inviting me along today. And thank you to Carolyn as well, a really interesting presentation. And I think hopefully, what I will present follows on nicely from that. I'm going to be looking a little more at the place of qualitative researchers in the process, the investments that qualitative researchers make, and how we might think about reframing qualitative data sharing plans, policies and practices.

**Susie Weller** 24:23

So within the UK, as well as the USA and other parts of Europe, making datasets available for reuse has progressively become a requirement enshrined in the policies of research funding bodies, universities and publishers. Sharing data is seen as a means of maximising the potential of datasets using public funds assiduously and assuring that research endeavours are not duplicated. But beyond this guidance for data depositors on compliance to regulatory frameworks or the practicalities of preparing material, discussions regarding the investments made by qualitative researchers, I think, are markedly absent So just before I move on to outlining the project on which I will be drawing, I just wanted to spend a minute or two reflecting on the implications of the wider neoliberal reforms in higher education because to me this is incredibly important kind of backdrop to this work. And this includes an increasing emphasis on auditing, accountability, surveillance, marketisation, massification,



managerialism and the casualization of work, which I think is particularly important in the context of this presentation, and this growing propensity for researchers to work on often multiple fixed term or short term contracts that can affect their capacities to accommodate all of these expectations around data archiving and sharing. So in this presentation, I want to highlight some of the implications of power dynamics for qualitative researchers, particularly those at earlier stages in their careers, and start thinking about, you know, enhancing the debate around what constitutes good ethical practice in qualitative data sharing.

**Susie Weller 25:58**

So I'm just going to start by providing a brief overview of the project on which I'm going to draw. I'm then going to outline three interrelated sets of personal and professional investments that I think are often overlooked. And then I want to propose a means to reframing procedures for sharing and reusing qualitative data sets that forge more ingrained habits of care that are more attentive to and value the investments that researchers make in generating rich data and how vital these are to qualitative secondary analysis, to good ethical practice and to wider knowledge production. So I'm going to draw on my experiences of working on the ESRC Timescapes initiative. I don't know how many of you have come across this, but it was the first large scale qualitative longitudinal study to be funded by the ESRC in the UK. And it was conducted across five universities, and it comprised seven empirical projects, each concerned with documenting change and continuity and key relationships and identities across the life course, or using a range of qualitative methods. And there were three overarching strands relating to archiving, secondary analysis and knowledge transfer that united these seven empirical projects, and much time was invested in creating the Timescapes longitudinal data archive, so lots of discussions around the practicalities of preparing qualitative data for deposit, as well as shaping procedures that are sensitive to qualitative research. Each of the empirical studies comprise a small team of researchers led by one or two professors, with early to mid career researchers carrying out the bulk of data generation analysis and curation. And most of those researchers were on fixed term contracts. And my role evolved from early career researcher to senior researcher and my relationship with the data also shifted over time from data generator and curator to that of secondary analyst. So I worked on one of the precursors study to Timescapes. And I collaborated with Ros Edwards on one of the core Timescape studies, Siblings and Friends. And we've followed the lives of 50 young people living in the UK for 12 years in the end. And then I received some additional funding as PI which enabled us to generate another wave of data with the same young people, and after which I co-led the day to day management of the study looking at the feasibility of conducting large scale qualitative analysis using the Timescapes datasets with Ros Edwards, Lynn Jamieson and Emma Davidson. So that gives you a little background to the research on which I draw.

**Susie Weller 28:27**

So now what to look at these three kind of interrelated sets of investments and the first one is time and hidden labour. I think a lot of this probably reinforces what Carolyn has already said really. The Siblings and Friends study initially involved preparing three waves of data from nearly 50 young people for deposit in the Timescapes Archive. So the resultant 145 transcripts, audio files, of the in-depth interviews, 50 activity sheets, contextual notes, this all required around 600 hours of preparation, including detailed removal of personal, direct and indirect identifiers digitisation, checking and editing, as well as comprehensive metadata documentation. This time, this kind of estimate of the time spent

did not include decision making around pseudonymisation, balancing data integrity with concerns about confidentiality, decisions around controls on access to or the safeguarding of specific cases or data files, the development of project guides and other resources to help promote and support qualitative secondary analysis.

**Susie Weller 29:33**

So in essence, then I believe there's a real lack of appreciation for the value of such work, the work of preparing data for archiving, and a lack of appreciation for how labour intensive it is. Carolynn has already said it, sort of, should start from the outset of the project and go right through to data deposit and beyond potentially as well. So whilst the tasks associated data curation overlap with good data management practices, well-managed, compliant qualitative datasets are not necessarily ready for sharing without further investments of time. The labour of preparing qualitative data for use by others involves a range of complex ethical and epidemiological decisions, not provisioned for and standardised data management guidance. Rather than being a relatively simple administrative tasks that it is often purported to be there, much of the work revolves around the ethics of balancing data integrity, with concerns about key issues such as anonymity and confidentiality. So, kind of, this hidden labour really often falls to more those in more junior positions and the prevalence of short-term research contracts may mean that researchers do not have the time to have reasonable first use of the data before a project ends and sharing is expected. Also, creating datasets is not held in the same esteem as the production of publications in high-impact journals for instance, or gaining prestigious research grants or in showing work has impact outside academia. So this hidden labour of preparing datasets for sharing may potentially be undertaken at the expense of other career enhancing tasks. It also doesn't help that unlike other research outputs, in most cases the intellectual property of research data rests with the institution where the researcher was employed at the time of production, rather than with the researcher themselves. So there's issues around transfer of credit if the researcher needs to credit for doing all of this work and research institution. So in short then, the task of preparing data archiving and reuse is often hidden, underestimated and undervalued and completed by researchers who may be vulnerable to precarious employment situations.

**Susie Weller 31:44**

The second set of investments involves the emotional labour, concerns the emotional labour involved. So, qualitative materials often regarded as co-constructed - the production of interaction between researchers and participants. And the relational nature qualitative knowledge production means that researchers' emotional investment, the labour and manifest in the data archive and in the data archived, and in the process of curating material for reuse. Data offer for use is shaped by researchers' experiences and the intellection investments of the wider team. Those engaged in empirical work have to share aspects of their own lives as part of the process of building trust, rapport and developing research relationships. So for instance, I spent hours with the Siblings and Friends participants, over a 12-year period visiting their homes, meeting family members and friends, engaging in lengthy discussions that also captured aspects of my own personal or professional trajectories. Regulatory frameworks fueled by a concern that data protection or mitigation against breaches describe the need to remove personal, direct or indirect identifies to ensure individuals and collectives cannot be identified. Indeed, procedures for gaining consent for reuse position researchers as responsible for ensuring participants understand to what they are agreeing, as Carolynn has just outlined. And and



indeed these kinds of procedures... Researchers are not however subject to the same consideration and are not anonymous in this process. Given them the emotional investments paid by researchers and research managers, some may be more vulnerable or exposed personally or professionally, so subject to scrutiny.

**Susie Weller 33:25**

The identities of researchers and what they reflexively reveal of themselves, how they interact with participants, the techniques and approaches and just the general messiness of qualitative work are laid bare in the artefacts of qualitative data. So even with a commitment to data sharing, willingness is often not without reservation. And researchers often report key concerns as comprising challenges to individual or group credibility, to standing or reputation, and concerns about a lack of duty of care to the way the data may be handled or represented by others. There's also the moral obligations to participants. And it's often those with a direct relationship with participants who feel a sense of moral responsibility towards their rights and wellbeing and any data relating to their lives and how it might be repurposed. In Siblings and Friends, for instance, decisions about data curation are guided by a relational and contextual feminist ethic of care. So accordingly, the inclusion of material is not only determined by formalised consent processes, but through continuous dialogue as participants' reflections regarding the wider context of participants' lives, and that through our own sense of moral responsibility, so many of the young people involved were keen advocates of data reuse, and participants' support for these endeavours places of further ethical duty on researchers and institutions to make data available. Nevertheless, over time, some participants divulge aspects of their lives that could severely compromise key relationships or cause moral harm. And I felt very protective of these aspects of their stories and in consultation with the team, decisions were made on a case by case basis to embargo segments of transcripts. Researchers then, regulatory frameworks and data management policies gloss over the challenges that relationships and practices of care bring to the process of preparing qualitative data sets for reuse.

**Susie Weller 35:19**

So the final sort of set of investments I want to draw on is the more marginalised aspect of collaboration collegiality. So Timescapes adopted this stakeholder approach to archiving, which sort to not only forge close working relationships between researchers, but also between researchers, archivists, data managers, data reusers, ensuring that procedural matters were sensitive to and appropriate for complex qualitative data. That said, collaboration is not generally a requirement or expectation many data sharing policies. The emphasis on open access by centralised repositories imposes a formalised and perhaps more distanced means of data sharing rather than through personal or professional collaborations of the researchers' choosing. And there's a general assumption that reusers will work with archive material independently of the original teams. So new users may not regard the explicit involvement of primary researchers as necessary or desirable, instead of viewing data as embodying new knowledge or alternative insights. Moreover, it may simply not be possible to work collaboratively, especially in the case of more historic datasets, or where researchers have retired, moved on to new projects or embarked on new careers. There is a small but growing movement towards more collective endeavours. And in our own work, members of the two of the original Timescapes project teams collaborated to conduct a large scale qualitative secondary analysis across six of the Timescapes data sets. And we spent a lot of time thinking about what constituted good secondary analytic practice.

We're guided by our perceived duty of care and shaped by our understandings of the temporal and emotional investments involved in qualitative work. We consulted our former timescales, colleagues ever plans to repurpose a dataset. Initially, we did this by email and then we moved on to a kind of, and this involved - although we tried to be inclusive - this generally involved liaising with one or two members of the original teams, those with whom we had strong professional relationships. Later, we took a more formalised approach using an anonymous online consultation to garner insights. So it's interesting because for former Timescapes colleagues, sustained collaboration was not uniformly decided, expected or possible to accommodate, with some constrained by time, funds, workload or just the longevity of their employment contract. And it was interesting that early career researchers, those who've kind of interviewed people and built those relationships with participants, are often best placed to enlighten secondary analysis on the minutiae of a project, but are likely to be hindered by time pressures and competing demands. And we also became concerned that we were relying on goodwill and this could result in exploitation of colleagues as well, especially those in precarious employment situations.

**Susie Weller 38:11**

So, given all of these issues then, what are the necessary conditions for the ethical and responsible and respectful sharing of qualitative data? And how can we try to ensure the investments made by qualitative researchers are not hidden or lost amidst regulatory frameworks and obligations? And so I found it particularly helpful to look at the ethics of care literature, to consider some of the ways in which data sharing policies and practices might be reframed to respect value and care, not only the investments of participants but also researchers. So although not originally applied to data sharing. Joan Tronto's focus on habits of care, or the disposition to care, for me, has been particularly valuable, and I feel this could help develop ethical practices to data sharing that are more attuned to the needs of qualitative researchers, qualitative data and participants. So Jane Tronto and Berenice Fisher outlined four aspects of care - caring about, caring for, caregiving and carer receiving, to which there are four related ethical values. The first of those is attentiveness or the disposition to identify care needs, and appreciation that that care need needs to be met. The second is willingness, that's willingness to assume responsibility for and attends to that care need. The third is competence and that competence relating to caregiving or responding pragmatically to those needs. And the fourth is responsiveness or the way that care is received.

**Susie Weller 39:40**

So within the context of data sharing, I think there are multiple levels at which care needs could be addressed. And I'm just going to outline some of those examples now. One that I've identified through my work at Timescapes is carrying forward about investments in qualitative data production. So we need To be more attentive to the intellectual, emotional temporal investments made by researchers and participants, caring for and about investments in qualitative data production creation requires a willingness by the institution's driving data sharing agendas to take responsibility for the way such work is represented and valued. This involves not only regarding qualitative endeavours in high esteem, but also looking beyond procedural data management elements of the task to acknowledge that well managed compliant qualitative datasets are not necessarily ready for sharing without further investments of time. And to achieve this requires competency carry for qualitative researchers and data by elevating in for example, funding application procedures and training opportunities the work

associated to recognise it is a complex interpretive process with epistemological ethical and epistemological logical implications.

**Susie Weller 40:52**

The second example, is thinking about thinking with care about researchers differing positions within the higher education landscape. So given that it's often researchers in more junior positions who are likely to be investing heavily in the preparation of qualitative data for archiving. We also need to be attentive to the ways in which wider trends in the higher education landscape, notably short termism impact on academics at different career stages. Obligations make data available with an acceptable timeframe often impact on researchers capacity to complete tasks regarded as vital to career progression. There are also pressures on senior researchers face the hidden labour involved in mentoring and supporting colleagues particularly on an informal basis. So taking responsibility for this requires structural changes to address the precarious employment situations of many. In terms of competence, competence, the caregiving element concerns a provision of adequate terms of employment, including sufficient time and resources to prepare qualitative data sets, as well as having the opportunity to conduct analysis publish and apply for funding. For all career stages training, data archiving, sharing and qualitative secondary analysis must transcend procedural matters to focus on the intellectual, ethical and emotional demands.

**Susie Weller 42:08**

Example is being careful about the place of researchers in the production of qualitative data. So we need to think carefully about the CO constructed nature of qualitative data, and how data management policies, protocols and practices might be framed to be more advantageous to the rights of those involved. This means caring for the nature qualitative work for participants for the position of researchers whose identities professional standing or personal lives are likely to feature as an integral part of the data and potentially subject to exposure or scrutiny. assuming responsibility requires a shift in regulatory frameworks and data management forces away from subjecting qualitative material to the same procedures, quantitative data sets in which researchers are often more invisible. And this view complicates both notions of data ownership and brings into question whether consent archiving procedures not only address participants, right that those are researchers.

**Susie Weller 43:07**

In terms of competence, then providing successful care necessitates the institution's driving the Open Access agenda, along with data repositories develop in collaboration with key stakeholders, policies, protocols and practices that reflect an understanding of data as constituted in different ways. For example, then I want to offer is a relational view of care and responsibility with research teams. I think we need to be attentive to power dynamics within research teams and the ways in which is often early career researchers complete many of the tasks associated with datasets. Recognition needs to be given to the pressures placed on more senior colleagues as well. They were often grappling with competing demands. And in thinking about responsibility, I found it useful to draw Junior McLeod's framing of responsibility as a resignation or disposition, as she argues that attention needs to shift away from neoliberal ideas of individual responsibility and, and all the inequalities of power and beauty in such nations to more relational view. So in these terms, then competency includes fostering more collegial practice a data archiving and sharing in which interdependence servitor tasks and decision

making thrive, and this can include the development of codes of practice of data preparation are shared by professional bodies. Furthermore, access to reuse must not be at the expense of more marginalised members of the research team.

**Susie Weller 44:32**

And the final example, then is thinking carefully about where and how qualitative data is archived. Whilst there are considerable benefits to making data available through centralised repositories, we need to be attentive to the implications or potentially restrictive nature of their structures. The organisation of such repositories and standardisation of procedures may not be fitting for or flexible enough to accommodate the complex, evolving, emerging, evolving and emergent nature of qualitative data. Taking responsibility for this necessitates a willingness by, for example, funding bodies to consider other means of sharing and repurposing data as both valid and valuable. And this is likely to involve thinking carefully, critically and creatively about what constitutes an archive, and what it means to co construct and curate qualitative data. And this would certainly require some degree of support for researchers to guarantee things like compliance with data protection, and care amongst both original researchers and new use of data new users of the material to ensure good ethical practice. And competence in these terms concerns enhancing researcher skills in developing research and participant led archiving practices and opportunities and encouraging alternative means, such as CO constructed community archives. In conclusion, then, design largely with quantitative data management strategies in mind the policies that govern data sharing and not attentive to the CO constructive nature of qualitative data. Moreover, regulatory frameworks often focus on procedural matters at the expense of considering the investments necessary for knowledge creation and creation in qualitative work.

**Susie Weller 46:07**

Considering these investments within the context of neoliberal reforms, illuminates the implications particularly for those in more junior positions, and the increasing number of staff on short term contracts. Integrating and adapting ideas from ethics of care literature refrains a discussion, diverting attention from purely procedural matters to encourage us to think with care about the particularities of qualitative data, and emotional, intellectual and temporary investments made by qualitative researchers. It's, I think, it's essential to recognise that the responsibility for infusing habits of care into the process not only live with individuals or research teams, but more importantly, wider structural change, as such means a vital to driving forward ethical practice in qualitative secondary analysis and wider knowledge production. And I just wanted to flag up that I have, there's a longer version of this paper available in qualitative research if you're interested in reading more about it. So again, thank you, Rachel, for inviting me to be part of

**Rachel Ayrton 47:05**

Thank you so much, Susie. That was fantastic. And a lot of food for thought and things that we can expand on. Hopefully, in the discussion, there's like loads in there that I'd love to pick up. So that's great. But we'll hold that thought, because we want to hear from our third speaker first. Professor David Zeitlyn is a professor of social anthropology at the University of Oxford. He specialises in the anthropology of religion and visual anthropology, in particular photography. And he's worked over many years in western central Africa, particularly Cameroon. And he's written extensively on data archiving, as well as being practically engaged in archiving projects. He is also interested in multimedia and how

internet technologies can shed light on and enable access to museum collections. So there are a lot of different angles to his connection with this topic. So I'll hand over to David. Thanks for joining us.

**David Zeitlyn 48:00**

Thank you very much. Let me quickly make a couple of comments on the previous two presentations. I think that qualitative researchers are often in a completely conflicted, completely contradictory set of pressures about archiving. We've heard some of the entirely legitimate reasons why archiving might be a bad thing. But I think we also need to be very conscious that not archiving can be a very bad thing. And one of the things I'm wary of, I suppose I want to say, is that a reflex to say, "archiving is dangerous, it is unethical, we shouldn't do it" means that the contents of future archives will have an enormous, an even greater bias built into them than the alternative. So I am very wary of arguments that say, "just don't do it". And at the end, I'll come on to an argument that might provide a way forward, which also has the benefit that it addresses some of the issues about how time consuming it is to anonymise and actually kind of going to Carolyn's question about if you've redacted so much, where's that data quality? For an awful lot of my work. I would say that it isn't there, that basically if it's going to be that heavily redacted, why bother? So the other kind of preamble, I suppose, about Susie's talk is that Timescapes is going to make another reentry, which is a nice bit of symmetry.

**David Zeitlyn 50:32**

So next slide, please. Uh huh, right. Okay. So yeah. Conflicting. I mean, I started by talking about conflicts. We have conflicting obligations, first and foremost to our informants, collaborators, colleagues, colleagues with a big C, colleagues with a small C, I suppose. We also have obligations through ourselves, because as it says at the bottom of the screen, the person the people most likely to reuse our data, thinking from the point of view of a doctoral student or junior postdoc, is someone just like ourselves, but even with the same name, or a closely related name, but 10, 15, 20 years older. So actually, Susie is now working on her own data. So, I mean, as she said, she is reanalysing her own stuff. One of the things I did in during lockdown, I ended up listening to some tape recordings I made - oh do I want to admit to this? - but yes, in late 1986, as part of my, no not late '86, early '86, late '85, '86, part of my doctoral research. And I went back to those recordings, and I'm currently - long story which I won't get into - but I am currently reanalysing some of my very early data. So, and then we have to start thinking about the future. Archives, as historical repositories are intensely future oriented. So we have to think about how our students, perhaps future postdocs on projects we might be lucky enough to direct, or even further ahead, how they can manage and use our material, the material we have gathered with people typically outside academia. And in some cases, in many cases, we have to think not only about those individuals, but their descendants. So and it's particularly the descendants of our informants, that I think the contradictions come to roost, as I will illustrate in a bit.

**David Zeitlyn 53:30**

So next slide, please. So first thing then, just to get rid of some of the nonsense, I think. GDPR isn't a problem. It is being cited by bureaucrats as excuses to destroy data, as excuses to inhibit the collection of data, but the people who developed GDPR were absolutely clear that it is not seen as an impediment to doing what in Europe is called scientific research, science in the widest possible sense, so absolutely interpretive research, qualitative research, and similarly no barrier to long term archiving. And I mean, the UK Data Archive are clear about this. But one of the kind of big worries is the way that citing GDPR,



an increasing amount of bureaucratic oversight is coming to bear. And this is a... I find very worrying, although of course, I have to say I am not a lawyer.

**David Zeitlyn 54:52**

So next slide, please. Consent for archiving came up already. And this is a real issue. If we can't explain what future people will make of stuff, then how can we get consent? Well, yeah, absolutely. There's a real, real challenge here. So and this has been addressed by Timescapes. And it's ideas like processual archiving, which means - and the stakeholder approach - which means that part of the importance of archivists in managing future access has to be central, in order to maintain the ongoing ethical chain. And one kind of thought that came up in my mind, actually, as a way of addressing - and it came to my mind during the previous presentations - was that one way of managing this might be to require that even long after we've retired, we are, or anything even more permanent, we are named as coauthors on any work using this, which means a different sort of bureaucratic element comes in, which would be copyright consents, because anybody who wants to publish a reanalysis of work I've been involved would have to get permission from me my literary estate, etc, etc, which could be delegated to the archivists managing the archive. But that would give a sticking point or catch point to main to ensure that the ethics have been observed. And, as I've already said, we are the main consumers of our own previous research, but well, have I got consent to reuse my own data? I don't know. And perhaps this is a real can of worms that we don't want to ask, don't want to open, but I'm going to flag it slightly at any rate.

**David Zeitlyn 58:01**

So yeah, that's Next slide, please. So I want to think of some of the unethical implications of being cautious. So one of the things that really winds me up when I see people's ethics forms, is statements such as "this data will be retained until three years after publication". Well, A, there's a question of how many publications are going to come out of one bit of research. So that sort of statement is complete wind up and more or less meaningless, because of course, you can always defer and say, well, there's more publications to come. But then hang on, haven't you already destroyed it because of the fifth publication is more than three years after the first and so people are in danger of making promises that they can't, can't keep, which I find very worrying. And real anonymisation of all the data is a, in effect, a power grab, because you're saying, if you can't get back to those people and, if you can get back to them, they haven't been anonymised, then you are the only person who can make decisions about publication. So I'm very worried about that.

**David Zeitlyn 59:54**

Next slide. Thank you. So case in point, famous book by Michel Foucault I, Pierre Rivière, having killed my mother, my sister, and my brother. And I'm absolutely fine with that sort of book, but I don't want them to have to be published as "I, homme, zero trois, zero six etc, having killed my mother, sister and brother". I think that would be fatuous. But that's what ethics committees seem to want to the future to be full of. And I think that's a bad thing, a mistake.

**David Zeitlyn 1:00:39**

Next slide, please. Okay, concrete example. Field photographs. I don't know why I took field photographs. Well, I'd sort of do, but I sort of don't. And anything that was true in 1986, is all the more



true now, when everybody's got phone phones. If you are doing qualitative research and somebody says, let's have a take a selfie, fine. But how is that listed in your consent forms? Probably not. But it's field data. So here's a photograph I took in, when was it? First of May 1986. About half those people are now dead, including the young woman in a headscarf who sadly, sadly, died relatively young. And only a few years ago, I got a female out of the blue from someone saying, "I've been given your contact by someone else and I come from the village where this photograph was taken". So we chatted away. And I said, well, you know, in the end, I said, "Well, I must have known your parents or grandparents". And so he said, he gave me the name of his family tree, basically. And I said, "Yeah, fine. I know, you know, I knew your grandparents, I knew your mother and I think I've got a photograph of them". And he said, "That's wonderful. I don't... My mother...". What he said was "My mother died, very young. I never really knew her. I don't have a photograph of her". So I was very happy to send this photograph. And in return, he kind of filled in some of the names. The ones who are still alive have signed consents, I should say, although that was all done retrospectively. And but what I want to ask you is to consider some, what I regard as, unethical responses, if he had asked for a photograph.

**David Zeitlyn 1:03:23**

Next slide. So I could have said I had to destroy it. Or I may have it but refused to send it because I don't have any permission to share it. Or I might insist that I blur her face. And I would regard all three of those responses as straightforwardly unethical.

**David Zeitlyn 1:03:52**

Next slide. So my generic solution is simply to wait. It's a version of "put things in drawers". And I don't know if it's a jokey title now, but it is based on St. Augustine. We can make me open. I can share everything but just not yet. So it's taking... Its embargos on steroids. And I'm very happy to say we can have exceptions. People might want to have their material shared, but if you want a default look to the census. Why not just wait for 100 years? Most but not all, but an awful lot of the worries about sharing go away once everyone's dead, I'm afraid. And a concrete case of this has recently been published, inadvertently, as it were. Alfred Haddon, a very early anthropologist, one of the people who invented field research actually long before Malinowski, worked in the Torres Strait in the late 19th century. And various researchers have been back working with local researchers, who are the - I don't remember how many generations - descendants of the people that hadn't worked with. And together, they have produced this book, which is an edition of Haddon's journals. And as it were, discussion of what happens next. So that's how material can be reused over the long term. And I think that, granted, the problems of being, of sensitive, of archiving sensitive material in the short term, better to wait, I suppose is my is my general thought.

**David Zeitlyn 1:06:16**

Next slide, please. I said in my title, about studying archives, so I put in one little slide about this. I mean, it's a very different tack. But when thinking about should we archive, perhaps those discussions might be helped if we actually did a bit more research on archives themselves, on what archivists do and what researchers actually do in archives. And so as a parallel to museology, the study of museums, I'm suggesting we do a bit more archivi-ology. It doesn't trip off the tongue, unfortunately. So there's, there are a few nice instances. Ciaran Trace, who's I think in University of Texas, Austin. And there's a lovely early paper by Bonnie Nardi, who these days is better known for studying World of

Warcraft with one of those great titles, what we learnt in the library. Very nice. And this also goes back to Harold Garfinkel, the good organisational reasons for bad records. I mean not everybody's into ethnomethodology, but that sort of work can help the discussions and what we do with our own qualitative records, I think. And yeah, I've got to leave room for discussion.

**David Zeitlyn** 1:08:13

So next slide. And yeah, that's me. Those of you with cameras to hand can scan the QR and get... I think it's quite cool, having ORCID's as QRs. A few references there. And yeah, let's, in the words of Mrs Merton, let's have a heated debate. Thank you very much.

**Rachel Ayrton** 1:08:40

Wonderful. Thank you so much for that, again, really, really thought provoking. And yeah, I'm really looking forward to hearing some of the responses from our participants. My apologies now. I realised at the beginning that I made a, you know, a good point of welcoming the speakers but not enough of a point of welcoming all of our participants. So thank you all for being here.