QR05\_transcript\_deidentified

**SUMMARY KEYWORDS**

data, qualitative data, survey, qualtrics, people, campus, thinking, reuse, qualitative research, talk, institutional, project, publish, research projects, interviews, qualitative researchers, students

**SPEAKERS**

QR05, Sara Mannheimer

**Sara Mannheimer** 00:00

That should work. Wonderful. So, I gave you an overview of the research a little bit when I sent you the email. But basically, the idea is that there are key similarities between big social data like social media, blogs, online forums, stuff like that, and archived qualitative data like transcripts or field notebooks. The idea being that these are both qualitative types of data that are posted for a reason besides the reason that you're using them for. So they're sort of being reused, reused in both cases. And so I also identified similar issues between them. So I'm hoping that by talking to big social researchers talking to qualitative researchers who have published their data or reused their data, that I'll be able to identify some more practices that data curators can use to support responsible research and both of these areas. Yeah, and so the interview will take about 60 to 75 minutes, depending on how long we talk. And I have six key questions (I do have an intro question and then one little wrap up question) that are structured around six issues that I've identified, which are context, data quality, data comparability, uh, informed consent, privacy and intellectual property. And then I asked you also to come up with a recent, an example of a recent time when you prepare your own data for your own qualitative data for publication or sharing or reuse existing qualitative data. Do you have a time in mind? We can....

**QR05** 00:07

I do, yes. I'm actually currently doing that. So.

**Sara Mannheimer** 01:50

Oh, wonderful. Okay. Yeah, we don't have to stick to it perfectly. But it helps to structure the answers if we have at least one example. Okay, so tell me about your, your example.

**QR05** 02:03

Um, so this is a project that's just wrapping up now that I've been doing with [collaborator]. We have been both surveying as well as interviewing qualitative researchers in different disciplines across campus to better understand, in general, what their research needs are and what kind of support they need [on specific issues].

**Sara Mannheimer** 02:33

Great.

**QR05** 02:33

Oh, yeah. So we have like...

**Sara Mannheimer** 02:33

That's a good example.

**QR05** 02:34

Yeah, this is a better example. Um, so we have these 30 some odd interviews, that with faculty and doctoral students that we aim to deposit in a repository, and so we had to, you know, de-identify the data, um, you know, make sure our that which was, you know, the interviews, we had to, you know, spiff up our documentation, so it would be accessible or readable or understandable by others who didn't actually collect the data. And really, I think the only the de-identification was the hardest. And I'm not, you know, I'm not sure if I didn't have a group of students who I managed that it would have ever gotten done. Just because I had I had some money. I had some students, you know?

**Sara Mannheimer** 03:37

Yeah, it is difficult to de-identify when you're like talking about specific research projects and specific institutions and places.

**QR05** 04:01

Yeah, so I mean, I was able to use this as a learning experience for the students too in just looking at even though I don't think any of the data I was collecting, were ultra sensitive. I was talking to researchers that were involved in ultra sensitive qualitative projects.

**Sara Mannheimer** 04:17

Right, right.

**QR05** 04:19

As well as um, I think, um, you know, some of them were just very identifiable. So we talked about different levels of de-identification and when and kind of different. We kind of came up with our own protocol. I mean, nobody is we looked all over, there's really no protocol for de-identification.

**Sara Mannheimer** 04:46

Mm hmm.

**QR05** 04:47

Yeah, and how to do it. So we ended up after we did some training and talking, having two people deidentify, and then one person read all of them to kind of see if they, how they felt about the de-, if they could identify people from the deidentified transcript. And we did things like web searches on terms to see how identifiable some things were. Yeah, I mean, what surprised me the most was that the repository wanted the consent forms with the name, or, you know, asked for that. And I was like, "No, I don't think I'll give you that."

**Sara Mannheimer** 06:10

Yeah, that is weird. Yeah, I understand the form itself even see what people consented to. But yeah, strange to ask or all of them.

**QR05** 06:19

Yeah. So there were all the issues with that data, both in terms of, um, you know, privacy of the person themselves, they, you know, talked about subjects, they talked about other people's, so they were secondary, personal information in the interviews, because we were asking people about how they, how they learned how to do qualitative methods. And so we found out from doctoral students about advisors, and you know, about courses and about, you know, what was, you know, what they learned and what they wish they'd learned. So, yeah.

**Sara Mannheimer** 06:57

So, let's see, when you were putting together these, de-identification protocols, how did you develop them? Like, did you consult with anyone or look at other research projects that had done similar work? Or was it just sort of like talking through it?

**QR05** 07:16

We didn't find any from other research projects, we looked at repository guidelines from, you know, QDR and ICPSR, and DANS, and some UK Data Archive, so we looked at different places that we knew had some qualitative research and what they said, and they were, they were pretty general, like, you know, personal names, institution name, so you get this feeling that, you know, anything that's named, but you know, some people would say, "Well, I'm doing a project in social media, I'm using Twitter and da da da da." Well, there are hundreds of people doing that. And then somebody was doing this very specific thing on [a religion]. And and it just seemed like from everything that person described that you could have found their project pretty easily. So, um, so we kind of none of the guides talk about other information that's, that's personally identifiable, you know.

**Sara Mannheimer** 08:32

So what was what did you end up doing with those? Did you decide to restrict any of the transcript?

**QR05** 08:39

No, we we have a lot of allusions or we have like the subject is [a country, a religion] or something. We puffed it up, or sociology of religion. Yeah.

**Sara Mannheimer** 08:50

So you in gen-, you made things more general. And did you feel that you ultimately had to make any compromises to your participant privacy in order to publish? Or did you feel comfortable with the results?

**QR05** 09:07

I think I feel pretty comfortable with the results. Um, I think there is more of a detriment to reuse than the subjects.

**Sara Mannheimer** 09:16

Right. Yeah, actually, let's, let's talk about context. They're like, what are, what were your... Did you do anything to try to sort of bolster back up the context after you had removed some of these specifics that might be helpful for understanding the data?

**QR05** 09:37

No, we just tried to let people know what kinds of data we had taken out whether it was a place name or was detailed information about, you know, the research project, we tried to clue people in what wasn't there or if it was, you know, "Dr. such and such, my advisor," we made it just "my advisor." Right?

**Sara Mannheimer** 10:07

Did you have any other concerns about context, besides the info that you would take out? Like what other strategies did you use to ensure that re-users would really understand the context in which you...

**QR05** 10:22

You know, in the documentation we didn't, we didn't name [the university where we were conducting interviews], for example. But we tried to talk about the larger institutional context a bit and a bit of the context about the project as a whole and link it down to link it to we also did a more large scale survey. So we tried to link it to that documentation.

**Sara Mannheimer** 10:50

Okay. Um, let's see. So let's see. So you let people know what type of data had been taken out. And then you talked about brought things in general institutional context, in general without naming the university are your names as researchers on the project, though?

**QR05** 11:23

Yeah. So it's probably easy, you know, we say the, "the [size and research level of the] university" and everyone listed is from [that one university], so yeah, uh huh.

**Sara Mannheimer** 11:36

And what about metadata? Was there like, did you create, was all this information... Did you write it in the paper in the Methods section? Or was there like a ReadMe that you plan to include?

**QR05** 11:48

It's a ReadMe that we plan to include. We have published some things out of the data. They, the repository only asked for some sample things that we publish. Not everything.

**Sara Mannheimer** 12:03

Okay. So you have links to the articles that are related?

**QR05** 12:08

Yeah. Um, which are also, I think, I think all of them may be in the institutional repository. Another clue as to everything.

**Sara Mannheimer** 12:21

Yeah, I know, it's like, I guess. I want to hear more about your process of thinking through, like, how much to keep in and how much to take out and sort of how you weighed those decisions.

**QR05** 12:37

I guess we didn't think that it was that important if it was revealed the university that we were at, because you know, [we're looking at a fraction of the] qualitative researchers on campus, and for our survey, [about a third of the people to whom the survey was distributed responded], so I think we thought that the, the way that we were anonymizing was anonymized enough so individuals couldn't be identified. That understanding the site was less, um, was wasn't as important. If people guessed what the site was, it was it would be, I think it would be hard to reconstruct given that we took stuff out sometimes a lot of stuff out about the nature of the research project.

**Sara Mannheimer** 13:41

I'm jumping back to like the data itself, or the project. Did you was it like a grant funded project that required a data management plan, or...?

**QR05** 13:51

It was an [federally] funded project.

**Sara Mannheimer** 13:55

And they required that you publish the data is was that part of why you did it?

**QR05** 13:58

No, we just decided to do that. [Grant details redacted].

**Sara Mannheimer** 14:14

Oh, cool.

**QR05** 14:16

So the one that I was involved in was looking at qualitative researchers on campus. And so [a collaborator] and I directed this lab and students kind of rotated, rotated through it. Also, um, so [the funding agency] didn't approve the particular research project, more just the educational experience as a way to kind of acknowledge that research was an important part [students' education in this discipline]. And, you know, this, this gave students a real authentic experience of doing [real research].

**Sara Mannheimer** 14:58

Right. Cool project. And what was the repository that you're thinking of publishing in?

**QR05** 15:07

[Name of repository].

**Sara Mannheimer** 15:08

Yeah, okay. Great. All right, let's move to data quality. So during your example, what quality issues arose? And not really like, was your data good quality, but like how you communicate any potential issues to future users like to show that you're trustworthy? Like, was there missing data? Was there bias of any kind that could have come in or what was your method?

**QR05** 15:38

I don't know. I, our documentation doesn't really include any kind of reflexive discussion. We do talk about how we de-identified things where I think that's the big data quality, you know, problem that someone going through a lot of the data leaks, it's a bit you know, some of it's about data reuse, so, but it would be hard, I think, to go through and say, Interviewee A is a sociologist Interviewee B is something else, because the topics are sometimes cross disciplinary. Um, and, um, yeah, I think in a lot of those studies like to say, well, earth scientists do this, or, you know, psychologists do that. Um, I think that that approach would be difficult, kind of demographic details about our subjects would be difficult.

**Sara Mannheimer** 16:51

Yeah. And so, where are you thinking that because of that, your results like aren't as broadly like, aren't... Sorry, your conclusions have to be more limited, because you're, you can't show these demographic details?

**QR05** 17:12

I mean, I think our conclusions can kind of look cross disciplinary at it. But if someone wanted to reuse the data, um, you know, we didn't share all the demographic information we knew.

**Sara Mannheimer** 17:28

Right. Okay.

**QR05** 17:29

Particularly the survey data, because we got some institutional survey data that we could only use. So that only part of the survey data were sharing is our survey data, not what we got from the institutional data dump.

**Sara Mannheimer** 17:48

And so did you talk about like thinking from, like, data reuse, or curation perspective did you did how did you? Or how do you plan to communicate that to future users or is that not included?

**QR05** 18:02

No, in that's in the survey part, not the qualitative part. But in the survey part in the ReadMe, we talk about that we're sharing our survey, and here's our survey instrument, um, you know, in our publications, we had access to, um, institutional administrative data that...

**Sara Mannheimer** 18:22

Got it.

**QR05** 18:24

...we cannot share. Okay, except in the aggregate in our papers, for example, we got race and ethnicity data.

**Sara Mannheimer** 18:35

Okay, great. Let's move to data comparability. So during this example, this project, did you compare or combine multiple data, multiple qualitative data sets? Or did you consider like comparability and interoperability for the data?

**QR05** 18:57

We did not combine qualitative, our qualitative data with anything. We we did the interviews first, when we did the survey, we tried to parallel certain things with the qualitative data to see in the aggregate, you know, people had it, some of the same patterns were emerging.

**Sara Mannheimer** 19:22

Okay. And so, thinking about people using your data in the future. Are you thinking that others would be able to do that to to sort of extend your study by asking similar questions to your survey or by continuing the research? I guess I'm just thinking about like, part of the reason that qualitative data publishing is encouraged is trying to do more longitudinal studies from older, some archived data?

**QR05** 20:06

Yeah, I mean, we're also depositing the survey data. So people could, you know, take, take questions from either one, take questions from both and use them, you know, in a, in a new qualitative study or, you know, in a in a survey study. We did... I don't think we borrowed any survey questions, but we certainly looked at surveys by [other researchers in the field] that they had done in the past, just get a sense of what were they asking.

**Sara Mannheimer** 20:50

Interesting. So did were those surveys, did you decide those weren't similar enough to your question that you could use some, like, ask them for their data or use their data?

**QR05** 21:01

Yeah. I mean, particularly, because part of the... part of the grant was for [students to participate in research]. And particularly around qualitative research, there's a real gap in support, you know, for for exam-, and it's kind of infamous that, and I think [the collaborator] that I was working with really wants to make a case for the [the institution to] offer more support. You know, for example, there used to be one person who knew something about NVivo [on campus]. And when [they] left, no one on campus is the NVivo expert. They, so there's, there's this big tool gap about how to use tools, and why use one and why not the other and there's just a big gap, it's very hard to get into qualitative methods courses on campus, there, there are a lot of issues about supporting qualitative research on our campus that isn't as supportive as quantitative. So I think that, um, so we weren't interested... I mean, you know, I'm always interested in what [data] reusers are doing. But we really wanted to find out what was happening on this campus, like to see if an argument could be made for [more support for] tools or, you know, highlight[ing] resources in different ways.

**Sara Mannheimer** 22:54

Hmm hmm. Okay, let's move to informed consent. Can you tell me about a time, if any, during the process of your example, your project when you consider the idea of consent, particularly consent for future use of the data?

**QR05** 23:15

Yeah, I mean, we had a discussion about that, did we want to, um, did we want to ask our participants to sign a consent form that, you know, told them, you know, that we were committing to put this in a repository? And we decided, yes, we wanted to do that, um, you know, I, this is the first time I had done it, actually. And it's, um, you know, I didn't know if it would suppress participation. It didn't seem to suppress participation, but we did have a big discussion about it. And it costs money to deposit in [the data repository we chose]. And I had some other funding to use. But it's ironic that if I was just doing it out of my research account, it costs almost as much as an institutional membership. So [my collaborator was able to advocate for] an institutional account. So we've kind of used the project to... that's how we wanted to use the project to show how your research can, you know, advocate for things on campus.

**Sara Mannheimer** 24:46

Cool, that's great. Did you when you were...what, so what other issues did you discuss when you were thinking about whether you wanted to have this specific consent to future use? And did you have anybody, like, ask about it?

**QR05** 25:10

We just have a discussion of why do we want to do this. And I guess we, we thought it was the right thing to do. Um, we thought that, you know, I guess, because we didn't think our data would be particularly sensitive, although we didn't know how hard it would be to de-identify that we thought we would use it as a learning, as you know, is a learning experience. This was a learning experience for students. And we talked about how it might suppress the number of people who'd want to be interviewed. So we talked about a lot of different issues around around that and decided to do the consent that would allow us to share data in the long run, and then get it through the IRB. Of course, yeah, they had no... I was surprised. They didn't say anything.

**Sara Mannheimer** 26:15

Oh, okay. And did you use specific language... like I know the data repository... I think they have a template for consenting to future use. Did you use that or any other resources?

**QR05** 26:28

We used resources in our university. At that point, you were thinking about where we would share the data.

**Sara Mannheimer** 26:36

And did they have any...

**QR05** 26:37

Like in our institutional repository, or [at a different data repository]. We just were kind of like...

**Sara Mannheimer** 26:44

Oh, yeah, why didn't you deposit in [the different data repository]?

**QR05** 26:47

Um, I think it would have gotten lost in [the other repository].

**Sara Mannheimer** 27:23

Have you been working with them? Or?

**QR05** 27:31

Yeah, we've been doing a lot of the work ourselves. Both because we didn't want to pay them to do a lot. And because, um, you know, we have students, it's this, it, you know, this is kind of this, it's education for the students' [future careers]. So, um, partly, you know, because we were fortunate to have all this student help. So I think we've just been working through it, and just going back to them and showing them things, and they give us comments, and they want this, you know, so if we're in a back and forth process, Mm hmm.

**Sara Mannheimer** 28:16

Okay, can you talk about a time if any, during your research and your like publication experience when you considered intellectual property concerns? Like participant intellectual property, or your organization, whether [your institution] owns the data, or the idea of fair use even?

**QR05** 28:52

Um, yeah. We didn't really consider intellectual property, um, pretty much everybody at [my institution] does with their data, what they want to do with their data. Um, there's also particularly, you know, I think this is also the culture of qualitative data. I was in a discussion I don't know, probably five years ago with somebody and um, well, it was he was very concerned about data science data. And I'm like, "Well, what about other kinds of data on campus, like qualitative data," and he's like, "Oh, that can just go in the institutional repository, or it can go somewhere else." There's maybe it's because it's qualitative data. There's a very institutional laissez faire attitude about it, like I'm... which is really, because you're also talking to Big Data people, which is really I think, you know, I think big data is in the eye of the beholder. I think, you know, qualitative data can be very, you know, big in terms of how you want to mine how you want to analyze it, you know, if you're doing it by hand, certainly qualitative data is... Also video data -- I mean, you talk about big data, those are big files. So, um, yeah, I don't, because I think [my institution] tends to look the other way when it isn't patentable, or people, you know, we weren't... nobody we talked to I mean... Even the people who were doing qualitative research with like Twitter or something, and may have been using algorithms to get Twitter data legally or not, they didn't talk about the algorithm itself. We were talking more of the data as qualitative data and how they were analyzing it as qualitative data and maybe their data management problem. But we didn't talk about the algorithms of capture.

**Sara Mannheimer** 31:23

Mm hmm.

**QR05** 31:24

Um, so there wasn't there wasn't. There wasn't any monetary pressure in any of our yeah, I think that yeah, the issues were more of respecting participants, and kind of the more on privacy issues and the data that were more, I think the ethical issues, not divulging, you know, corporate secrets or anything.

**Sara Mannheimer** 31:55

Right. Yeah. Okay, let me think. I want I want to talk more.

**QR05** 32:08

I will tell you, I will tell you an off the record story. [Redacted].

**Sara Mannheimer** 36:20

Well, oh, okay. Well, what else? Like, are there, were there any other issues or challenges that came up that are coming up as you're thinking about publishing and reuse?

**QR05** 36:35

Um, not not on this project. On another project that I mean, I am very interested in this idea of privacy and respectful reuse and ethical reuse. And I'm very interested in reuse in [of videos]. And I've been trying to think through a paper about respectful reuse, because of course, in [some educational videos I have looked at], there are children, there are teachers, and we know that teachers bear the brunt of a lot of societal frustration. So I'm thinking about, you know, how can we make this stuff reusable on and how do we come to some, you know, consensus about respecting participants, you know, children who may now be adults? Yeah, teachers may have been having a bad day, you know?

**Sara Mannheimer** 37:45

Yeah. And what kind of conclusions are you coming to or what have you...

**QR05** 37:56

I don't know. I mean, I'm still trying to think through it. I guess I do believe in open data. But I think that there are there are a lot of considerations about understanding the data and placing the data in context that I think are very important. When you're, when you're looking at any kind of sensitive data.

**Sara Mannheimer** 38:27

Yeah. Yeah. And like with video, well, with any digital data that's available online, it's like, people could come across it so much more easily to, and then it's just going to be there forever. And it's like, how do you opt out? And yeah, yeah.

**QR05** 38:49

So I think, you know, I think I think the ques-, you know, as you framed your study, I think a lot of the questions that you're asking are really key moving forward with qualitative data. And in thinking about, you know, the context, quality, and privacy issues, and the kind of ethical issues around it are really are really key. You know, I think it's one thing for me to share my data, but some of the qualitative researchers we talk to, you know, I'm not sure, you know, their data is shareable, they would have to take so much out.

**Sara Mannheimer** 39:32

Right? Yeah. Yeah. Yeah, how you can protect privacy while making that useful as well. Yeah, yeah.

**QR05** 42:07

I don't think any of our interviewees did, actually... I mean, it's really, it's funny. I mean, one of the papers, or another paper I'm trying to write, is to talk about data management, data sharing and data reuse in terms of qualitative research. And what's funny is that none of the interviewees reuse qualitative data -- or if they did, it was from a friend -- but all of them have opinions, like, strong opinions. And it's, you know.

**Sara Mannheimer** 42:42

I can't wait to read the paper.

**QR05** 42:44

I mean, you know, of course they have, but it's just so funny. It's like, you know, one of those things in life, it's like, but you've never done it. So why do you think this is...? There's just this mythos around it?

**Sara Mannheimer** 42:57

Yeah. Right.

**QR05** 42:59

Well, a very negative mythos. Surrounding why it can't be done and what are the problems.

**Sara Mannheimer** 43:05

Yeah. Well, and I feel like until we have more examples of people reusing qualitative data, we only have these theoretical problems rather than, like actual examples of how the problems, you know. Yeah.

**QR05** 43:20

And there's so few published examples of reusing qualitative data too.

**Sara Mannheimer** 43:27

Yeah.

**QR05** 43:31

That it's kind of hard to figure out, what is the state of the art?

**Sara Mannheimer** 43:37

For sure. All right, thank you so much for taking the time to talk to me. I really appreciate it.