QR01\_transcript\_deidentified

**SUMMARY KEYWORDS**

data, people, qualitative data, interview, qualitative, participants, redact, qual, digital preservation, informed consent, study, set, research, transcripts, qualitative researcher, terms, curator, published, qualitative research

**SPEAKERS**

Sara Mannheimer, QR01

**Sara Mannheimer** 00:00

If that's okay with you...

**QR01** 00:02

Yeah, totally.

**Sara Mannheimer** 00:04

And I'll also be taking some notes as we go through this. So basically, my research is about, I'm seeing connections between qualitative data reuse and big social data research. And I am doing these interviews because I want to learn more about data curation practices and research practices, qualitative and big social research practices. So I can understand better like if there are areas where they can connect. And so we'll move down to page two. So tell me about the type of qualitative research you do. And then what type of data you generally produce.

**QR01** 01:29

Yeah, so typically, my research is multi-method. But I have been veering more and more into just straight up qualitative work, in, in in recent times. So the last qualitative project I did that was fully qualitative was called [project name redacted], and that data set is [shared in a repository]. Basically, we had identified... or the [national organization] puts out a survey every five years, and there was about like a 15 point drop in workplace satisfaction amongst [workers in a specific field] in that five year span. But obviously, the survey can only tell us that that happened, it doesn't tell us the why. So my co authors and I set out to interview people in service of answering the "Why are we so dissatisfied?" We had a bunch of a prioris of course, it's things that we see in our own workplace. But we were able to talk to about 21 people. We did in depth interviews, and it was really interesting. We analyzed them; we took a phenomenology approach. And I usually do inductive qualitative research, so I usually don't go in with a codebook. I usually make one and my co authors make one as we go. So that was a, that was straight up qual. And then you mentioned [a second research project]. So [that project] is multi method. So we did focus groups, a survey and then interviews. We used the outcomes of the focus groups to inform the survey. And then we use the survey to recruit our interview participants. So we basically said at the end of the survey, "Hey, if you want to, if you're cool with being interviewed for about an hour about these topics, hit, say yes. And you'll be redirected to like the other informed consent." We had about [number] people say yes, which is wild.

**Sara Mannheimer** 03:19

Wow.

**QR01** 03:20

And we interviewed [number] of them. So those are the last two like projects that have major qualitative components.

**Sara Mannheimer** 03:28

Perfect. Great. So and then I also asked you to identify a specific example, did you select one of those that you just told me about as your example?

**QR01** 03:38

For qualitative data? Yeah, I thought it'd be interesting to talk about [the first project mentioned above].

**Sara Mannheimer** 04:18

So let's look at... I have a couple like sub questions just to get into more details on your example. Was this part of a grant funded project that required like specific treatment like a data management plan?

**QR01** 04:43

No, it was not. But [the second project mentioned above], my multi method study was.

**Sara Mannheimer** 04:47

Okay.

**QR01** 04:48

But it was a funded study that didn't have a DMP. So that's kind of funny too.

**Sara Mannheimer** 04:53

So there wasn't...

**QR01** 04:54

It was funded by [funder], so.

**Sara Mannheimer** 05:01

So you didn't have like an official data management plan when you started out doing this research?

**QR01** 05:07

No, I. And it's funny because [I have expertise in data management]. They were like, "[Name] is here, we don't have to, [they'll] just let us know when we're messing up." And we're very close group. So I also like renamed files without telling people like there's some of that latitude that I had that doesn't exist in other research groups.

**Sara Mannheimer** 05:53

Yeah, right. Okay. And then you published the data in [a repository], you said, right?

**QR01** 06:02

Yeah. And that was chosen...

**Sara Mannheimer** 06:06

Go ahead.

**QR01** 06:07

I was gonna say we chose [that repository] because, one, we're all big believers in openness and transparency. But also because we can lock the files and have them be available by request only because people did like, you know, talk smack about their bosses. And it's, there's rife with repercussions there. So we negotiated with [the repository], because something that they care about is also being able to give access if like we die or retire. And so we basically negotiated with them a right of first refusal for anyone who requests to see the actual interview files. And if we don't respond within like two weeks, then they can use their judgment. And we've sort of outlined some of the things that we would want them to look out for. It was a very collaborative negotiation for the terms of use and stuff around the data. So that's, that's why we picked [that repository].

**Sara Mannheimer** 07:00

Nice, that is really nice. And so do you have plans for storing retaining and deleting data in the future? Either the stuff that you've published in [a repository] or our other data that you have locally?

**QR01** 07:16

Yeah, so we told our participants that we would keep the transcripts. And that's what we've done. And we've deleted the audio. We did local recordings, as well as Zoom recordings. And so we had to, at the end of our study, go into Zoom and delete it. Or when we had a successful transcription, we deleted everything off Zoom. And so however long that's on their servers, it's on there. By now it's for sure deleted because it's a year later. But I am keeping a copy personally on my server of the transcripts of our notes of the participants like names and aliases. And that's under some security protocols that I've put in place with I use [cloud system] for my personal file management. So that's how we've approached retaining the data for ourselves. And we actually didn't keep unredacted transcripts either. We just what we did is when we got the transcripts, when we created the transcripts, we sent them to the participants and said, put brackets around anything you want us to redact anything that you think would be like an indirect identifier. And we would take it out of the transcripts. And especially because during the consent process there, the participants had three options. They could choose to participate in the study and have their data go into [the repository]. They could choose to participate and not have it go in [the repository] or they could choose not to participate, of course. And everybody chose to have it go into [the repository]. So we were really keen on having them be a part of the redaction process because of that. And so those are the only transcripts we've actually retained. We don't have the raw transcripts or the raw audio anymore.

**Sara Mannheimer** 09:03

Cool. And do you just you have access to the data or do other members of your research team as well?

**QR01** 09:12

Yeah. So I was tasked with the again, because I'm the data management person I was tasked with with stewarding it, but everybody has access.

**Sara Mannheimer** 09:20

Okay.

**QR01** 09:23

And the way that... Yeah, the way that it works is basically they can view it, but it's not downloadable. And we've sort of all agreed to that. Again, we're a very close research group. And we were really keen to limit human error as much as possible. So we've sort of like negotiated that out as a team. It's also unlikely we'll revisit this data unless like [the next version of the survey from the national organization] comes out and there's another dramatic drop and we might want to do a longitudinal study. So that's pretty much the only reason why we're even retaining the data in the first place. And of course, for verification in case people, for some reason, based on our [access conditions] thinks that our data doesn't exist or you know, whatever.

**Sara Mannheimer** 10:07

Yeah. And do you have a certain time that you'll store it for? Is it sort of indefinite? Or do you have? Yeah. Okay, cool. All right.

**QR01** 10:16

So [name redacted] succeeds me in taking over the server, should anything happen. And I don't have a will or anything. But that's actually a really interesting point, what happens to the research data after someone dies or retires? Or I guess dies in this case? I'll write that down here, I want to think about that.

**Sara Mannheimer** 10:33

Yeah. So from here, I have these six issues that I identified with a lit review that I want to just go through one by one. So the first one is context. So I have a little quote at the top there. "Qualitative research is a process that may include deep and prolonged contact and connection with research subjects are attempting to understand subjects within their own context. So qualitative data is highly context dependent. And context is the source of data meaning and understanding and ignoring it or not recognizing when one's own context driven perspective will result in incomplete or missed meaning and a misunderstanding of human phenomena." So that's sort of where we're coming from in terms of context. Can you tell me about a time during your example with the...

**QR01** 11:24

[First project discussed above]?

**Sara Mannheimer** 11:27

Yeah, when you considered the issue of understanding, maintaining or communicating your data's context?

**QR01** 11:34

Yeah, so that's a really good question. So you mean the context of like the interviewee that's what I got from the passage, not necessarily our own context for like being a part of the data collection or in the community that we're studying?

**Sara Mannheimer** 11:49

I guess I'm thinking about like, when, basically like when other people are getting your data from [the repository]. What context do they need to know? And did you think about adding additional information to the data [in the repository], trying to communicate all of this contextual information about who the people where the people were coming from? What other factors were contributing to what they said during the interview?

**QR01** 12:17

Yeah, so that's a really good question. In [the repository], we do have a lot of our supporting documentation, we have our interview guides, we have our workflow of how we actually like literally scheduled things we have our informed consent, we have, I think our participant demographics, like at a high level, that are behind lock and key. So that's something you'd have to request. But it's like the alias, the type of institution they're at and the type of work they were doing. And then at the end of these interviews, we asked people to self identify, like their age, their ethnicity, a few other demographic questions that we included in the codebook. And we did that, because we asked them, "Do you think any of your demographic, your demographics impacted the answers you gave to our interview?" Because it's, of course, a big part in [the field in which the study participants work], that it's really, you know, heavily white, heavily cis, heavily male dominated and [subfield] in particular suffers from these issues, that, you know, these issues become compounded. So, we have a lot of documentation there. But in terms of what people need to know, I think that's there, because we also put a methodology statement in there of how we chose to collect the data, how we chose to form the interview, how we chose to analyze it, we also explicitly say we're trying to answer the why, to what the [national organization] survey has, you know, explained, that is the what, if that makes sense. So yeah, I actually feel like it's there. The challenge, of course, is like getting all those materials ready to share, which is also something that I did. [My university] facilitated the deposit to [the repository]. So that also was like a big part of my work on the project, excuse me, making sure that we have sort of that supporting information there. And the other the selfish reason was to provide context, but also because we want people to extend this study to different types of [organizations and fields]. Like in particular, [the subfield addressed in the study] has this problem in particular, of attrition and of poor working environments. But it is certainly not the only one. Of course. So and our data set, our interviewees skewed like white and middle aged, and as most of [the broader field addressed in the study] do. And so part of the reason why we wanted to put the context out there is so that people have enough information to pick up our work and continue it with underserved, you know, communities in our, in [the broader field addressed in the study], so that's sort of why the challenges I guess were like, you know, prepping that stuff, making sure that if I say, they're like, you know, I don't know, a Black [job title] that it's not, that's a fake example, to be clear, like that wouldn't be directly identifiable. So there is like, the indirect identifiers, were really something we paid attention to a lot in preparing that supporting documentation.

**Sara Mannheimer** 15:27

Okay.

**QR01** 15:28

Does that answer your question?

**Sara Mannheimer** 15:30

I think so. Yeah, this is good. And we can talk more about sort of the de-identifying process. I have a privacy question later on. Okay. So during your example, what quality issues or concerns arose, so thinking quality I'm thinking like missing data, or bias, or the quality of your method? And then yeah, and so then how can I ask these follow up questions if I need to.

**QR01** 16:02

Yeah, so it's a good question we wrote in the paper, I don't know if it's actually with the data in [the repository]. The data in [the repository] is linked directly to our paper, like we cite the data set in our paper, and in [the repository], there's an actual link back to our published work on this, which is open access. And in the paper, we describe basically, our limitations, and one of the limitations is that our data set is such a narrow demographic. Um, and that was something we definitely said like would probably have an impact and and could possibly change is definitely a bias in the data set, that I can say that, um, I think our methods fit for the study. So I don't actually think there's any quality issues with that. We spent a long time talking about methods. And we're really invested in inductive research. And so we were trying to find good inductive methods for what we wanted to do and think phenomenology is pretty close is to what exactly we want to do. Um, that said, trying to think about, we didn't really have any missing data. We'd like the interviews, that recordings went off without a hitch. The transcriptions were good, we did them by hand as a group. So it was really hands on with everything. Mmm, hmm. Yeah, no.

**Sara Mannheimer** 17:29

That's good. Sometimes I feel like quality can be more of a concern if you're using other people's data and trying to...

**QR01** 17:35

Absolutely.

**Sara Mannheimer** 17:36

...assess quality. So it's okay, if the answer is there wasn't much considering.

**QR01** 17:41

I think the bias is really—the bias in our data set is really the the defining factor in our data set, but also with us, like we as a group tend to tend to value more highly the opinions of non-managers, I want to say: we have managers in our data set, and they're lovely people. But part, the other part of the impetus for the study is we're really sick of just seeing the [national organization] reports about [high level managers], saying "the future of [the field]," talking about labor without actually like, doing labor, or caring about employees. So that is also like I guess, an unsaid, more unconscious bias. And the reason of who we picked, who we talked to, because it was people that we people who worked at [national organization] member institutions that also narrowed things down. We tried to like cherry pick people based on some defining factors, like demographics was one of them, we didn't want to have a mostly white data set. It ended up that way. Which sucks. But so yeah, I think like the biases are mainly... and I would say I would generalize that to all of qual.

**Sara Mannheimer** 18:56

Yes. And did you like, did you think about writing a statement or some way like creating, trying to communicate that to future users? Or is that in your paper?

**QR01** 19:10

Yeah, I guess the limitation section of our paper definitely talks about the bias in the data set. But in terms of our own biases, I'm not sure. I don't even know if my collaborators would agree that those biases were present.

**Sara Mannheimer** 19:24

Yeah.

**QR01** 19:24

Not about the biases about our data set, but about our internal, like, anti-managers bias. I don't know if I don't know if they would see it as impacting our our, work. It's definitely on the forefront. They would call it maybe a value, like these are our values going in, but not necessarily a bias. Yeah, so yeah, I don't know. I kind of want to ask them now.

**Sara Mannheimer** 19:56

Yeah, it's an interesting idea to say like, here's who we are as researchers, and here are that potential, like, here's sort of the context and the, like, biases that we're bringing in.

**QR01** 20:10

Well, I think about that, especially because we had published a competency profile for [workers in the subfield addressed in the study], as a group, the same group of co authors and I. And we did it by talking basically to [workers] and explicitly not talking to people in more heavily or like admin or supervisory roles. And a [high level manager] at a conference was like, "Why don't you talk to [high level managers]?" And we were like, "Cause do you guys do anything like with [the subfield addressed in the study]?" And so that's why I bring it up. It's like very much, it's been a part of the conversation around all of our scholarship that we've done together, just like the [group] of us.

**Sara Mannheimer** 20:52

So that's great. This is all great stuff. Alright, let's move to data comparability. So during your example, did you compare or combine multiple qualitative data sets? Or did you consider the comparability or interoperability of your data set with others?

**QR01** 21:13

Yes, in that we wanted the format to be very similar. So all of our data sets have the same format. If you see something redacted, it all appears the same like from, I guess, a machine actionable standpoint, they're all very interoperable. In terms of like combining with other interview studies, it's not actually something we ever considered doing as a part of this work. But it would be really interesting to think about, especially because in light of our, our conversation just now about our biases, I wonder how that would impact how we would be able to cross compare.

**Sara Mannheimer** 21:48

Yeah.

**QR01** 21:49

The other thing is, we did this study, because we didn't see similar studies going on in the field, we got a lot of feedback that like, basically, that [subfield addressed in the study] isn't special in being unsatisfied in the workplace. And it's not worse there than it is in other places. So this was a criticism of our work that I think would actually benefit from some cross comparisons and studies. Because I don't actually think that's true. There are obvious labor issues that everybody deals with. But as a research group, we kind of do posit that the workplace satisfaction for [workers in this subfield] is just a lot worse than the rest of the [broader field]. And so we didn't consider that. But now that I'm thinking about, it would actually be a great future direction to actually test that criticism a little bit by conducting interviews with a wider array of [workers]. I guess the other thing is, with interoperability, we did publish our interview guide. And I think that actually goes a long way in facilitating interoperability because people will be able to see the direct questions we asked and will be able to see whether or not the potential answers would be able to mesh with, for instance, other interview or, or particular survey data. Um, so yeah, I think that's a good point. We definitely I always just, I want to optimize everything I do for interoperability. And so I often like over document in service of that. So I would say like, we didn't do it for our study, but it's definitely something we want to encourage.

**Sara Mannheimer** 23:25

Yeah, it's an issue, of course, for qualitative data, because there's just so it's so complex, so you can't like just go do some search for an interview question and be like, oh, someone else's exact interview question, then I can now compare it to all of these other answers.

**QR01** 23:44

Yeah, I think really, the topics will be the most important part, especially because like, you know, with in depth interviews, the guide is really just what it says it's a guide. It's not like a one to one question answer. So that also can sometimes be a problem with interoperability and qualitative spaces. Like you can look at our guide. Some people spent, like we had interviews that went on for two hours, because it was just cathartic for people, uh, our guide was pretty good at keeping people in one hour unless they wanted to keep going. So we often, like have a lot of topics that aren't present in our documentation, you know, and I think that actually hurts interoperability but is, like you say, pretty common in qualitative, especially because a guide is a guide, not one to one question answer.

**Sara Mannheimer** 24:30

Yep. Yeah. Wonderful. Thanks. Okay. Let's move to informed consent. Tell me about a time if any, during the process of your example, when you consider the idea of consent, particularly consent for future use, so you've already discussed this a little bit. But, um.

**QR01** 24:47

Yeah, it [potential future use of data] was an explicit option in our consent. And we, we in our consent form went down the specific things that we were going to do like most consent forms do. We say you're going to talk to us for an hour, we're going to record it on Zoom, we're going to do this in that for anonymization. We said in the informed consent that we're going to send you a copy of the transcript that you will be able to redact. So that was very upfront with the participants. We said, "We are really hoping that you will allow us to put this data in [this repository], here's how it would look by going into [the repository], it wouldn't just be openly available, people would have to request it from us." So we outlined sort of the risk mitigation that we were doing, I guess, by depositing in [the repository]. And then like I said, we gave people those three options to to participate and deposit, participate and not deposit or not participate. And everyone chose to deposit which was really encouraging. So we really cared about collaborating, because, again, with these interviews, they were our colleagues, like the people we see at conferences, they're people we hit up for like documentation and other stuff. And even beyond that we care deeply about giving our participants a voice in the research, particularly in like how their own data is going to be used. So having them involved in the redaction process was really important to us, and offering them an out if they were super uncomfortable with the open part of it was also really important to us. And so that's how we set it up. We were kind of worried that the consent form was... it was a little over two pages, or no, sorry, a little over one page. And usually our consent forms are one page even. And so people were a little worried that it would impact, you know, participation, but I'm really happy that it didn't. And that's actually the informed consent is publicly available with our [repository] deposit.

**Sara Mannheimer** 26:44

Nice. And you walk....

**QR01** 26:50

Yeah, yes, we did walk them through it.

**Sara Mannheimer** 26:53

Alright, and did you consult with anyone or consider other research projects or refer to literature, etc. When you were coming up with these strategies?

**QR01** 27:16

For specifically the informed consent?

**Sara Mannheimer** 27:18

Yeah.

**QR01** 27:20

Oh, no. So and we use IRB has informed consent generator now. It's really funny to me, they're wonderful, we collaborate on a lot of a lot of services. And, and they have like a little button that you click if you're planning on sharing it in a repository. But it wasn't those three options, it was just like this data may be shared in a repository like one sentence and the informed consent. So I was like, that doesn't really cut it for us. [The university] where one of my co-authors was had their own like, strict template they wanted us to use. So we sort of just blended the informed consent generator output from [my university], with the template from [my co-author's university]. And we sort of added in those options that we saw as important. So we kind of mesh together templates, and then eventually just put in those three like checkboxes, on our own.

**Sara Mannheimer** 28:13

You just sort of came up with those about like, having people redact things they didn't feel comfortable with and that wasn't something you'd seen another research group to?

**QR01** 28:24

No. But I did hit up my friend who is a who's she's has a PhD in [discipline] and used to [work with me at my university]. And I said, "Would it totally invalidate our study if we let our participants redact their own transcripts?" And she's like, "No." So I just took her word for it.

**Sara Mannheimer** 28:43

Yeah, you used your network.

**QR01** 28:44

Yeah, I thought it was...

**Sara Mannheimer** 28:47

Nice.

**QR01** 28:48

I used one node in my network. But I was like, I yeah, I, I was I was a little bit worried that it would other qualitative, like, serious, I consider myself I'm self-taught qualitative research. When I did my project [as part of a fellowship program], it was a qualitative project. When like, the [group] of my co-authors and I had met, and I had to, like teach myself qualitative. I lived off of different trials of NVivo. Like I had to create email addresses to get new trials, it was like a whole mess. So as a as someone who taught themselves, I just wanted to double check with a seasoned researcher that other serious qual people wouldn't see that we did that and discount our whole process. So that was mainly what I was worried about is sort of, would it invalidate the rigor of the study? Because we did put so much effort into choosing the specific methodology that we used. And yeah, so that was mainly the concern, I guess.

**Sara Mannheimer** 29:48

Right. Okay, privacy and confidentiality. I meant to tell you at the beginning that we have seven questions, so now we're on number five, just to give you a little check. So tell me about a time if any during your process what of the of the example you're you've been discussing this whole time when you considered issues of privacy and confidentiality.

**QR01** 30:15

Yeah, a lot because again, these are people like talking smack about their bosses and we already live in in such a precarious [subfield addressed in the study] in particular particular is so precarious. The jobs are usually like two to three years. It's rough. So we heavily considered this from the beginning. Especially because like A) we wanted people to talk to us and B) we didn't want to get our our colleagues in trouble, you don't want to be a narc. So we were really, really interested, invested in having a private conversation. Yeah. So what we did, you know, the publishing side of it, we published it by request, we negotiated the terms with [the repository] such that we had that right of first refusal. If you go to download or request access, you have to tell us what you're going to use with it, you have to explicitly say you will not seek retribution. Like we added that line into [the repository's] terms, because we were really concerned, particularly with that issue at like, somebody's boss had talked to us... we did a panel at a conference and somebody's boss was like, "Oh, well, we don't have a problem here at my institution do we?" Like they just say the most awkward things. So we, again, we were just really nervous about that. So the publishing side, it was restricted access. All the documentation is open. Now [the repository we used] is nice, because you can have a mix of public and private. So all the supporting docs are there. What we did further, we assigned everybody a pseudonym [details of pseudonym creation redacted]. We do have a record of who those names correspond with. They're on my server under those security protocols that I outlined earlier. That information is also I don't actually remember if it's on [the repository], I should have looked, I could I can look and get back to you on it. I'm I'm 60% sure that it is in [the repository] under lock and key. Um, but I'll have to check. And then for the other privacy strategies. I'm trying to think. I think that mainly it like, it was really the redaction process and then making sure that aliases can't be linked with people very readily. Also, when we were, when we incorporated quotes, even if those quotes were... had some redacted pieces, we really picked and chose because some things are obvious, or at least obvious to us. Like this one person described a particularly terrible outburst at their [workplace] that had gotten some press attention. And while we had taken out a lot of the indirect identifiers, and this person didn't specifically, you know, put brackets around that quote, we chose not to use it even though it was really impactful because we were nervous that people would recognize it from those press briefings like [number] years ago or whatever. We have a lot of memories in [our field]. So yeah, that's something we also considered I guess, with respect to privacy, we really carefully picked and picked and chose our quotes that we made public but we do have a we do have a set of quotes we, [the repository] asked us for basically as a part of our codebook... the tag, the qualitative tags that we used, how we describe them and then one quote that corresponded to each so we do have like a set of quotes published but again, we really carefully picked and chose them.

**Sara Mannheimer** 34:41

Okay, so it sounds like [the repository] gave you some guidance as well on some of this stuff. Did they give you guidance about deidentifying the data too?

**QR01** 34:53

Not from this project, but I can talk about [the second project discussed at the beginning of the interview] if you want, because they did... We went through heavily back and forth with them on [that project], which is meant to be more public. So for data that's, by default restricted, they sort of were hands off. But with the [second project's] data, we wanted to make our interviews fully publicly available. Because like A) that's what the participants consented to, and B) it's the community that we're studying. [The study population is focused on open practices] and they expect things to be openly available. And I agree with them. So but they, [the repository] did not allow for making those interviews publicly available. Because they said they don't have the legal infrastructure right now to actually allow us to make that qualitative data fully openly available. And because we wanted to put no restrictions on it, they also like went through line by line of each transcript, and pointed out things that could be potentially re-identifying. And the [second project's] data set is not that controversial. If people talk about [uncontroversial topic], it's not going to affect their daily lives at all. So that but they went through it with a fine tooth comb specifically because we didn't really want to put that many restrictions. But for the [first project that has been the main project discussed in this interview], because we had all these restrictions, they were... they did not do that.

**Sara Mannheimer** 36:16

Okay.

**QR01** 36:17

But the level of curation on the [second project] was impressive, like a lot. They did a lot of stuff for that data.

**Sara Mannheimer** 36:50

Yeah. Okay, on to our last question about intellectual property. Can you tell me about a time during your [first project that has been the main project discussed in this interview], when you considered intellectual property concerns, especially if you publish the data? So in this case, you did.

**QR01** 37:11

Yeah, we did. So it's interesting because [the repository] has different terms for the open data than the closed data. So the data by request is under a different set of licensing terms, than the materials that are open. My [group of] co-authors and I again, we're very close, we're fully happy to share copyright amongst us. And we said, like, you know, this is all of our IP, and we stand by that. When we published it, the open data, I believe, is [published under a specific license]. If I remember correctly. The the closed data is subject to the to [the repository's] specific terms of access, plus whatever we've added on to it. But yeah, so the the actual IP remains with the [group] of us, the actual content with the [group] of us.

**Sara Mannheimer** 38:09

What about intellectual property... I guess, because it was so important for you to de-identify everything and keep the people anonymous, like did you think about intellectual property from the participant point of view? You know, like when you had them sign their consent form? Was there something about like you're also waiving your intellectual property to the what you've created here? I've never seen that but...

**QR01** 38:37

Me either. Because I think actually, by default, they don't have, like, I don't, I can't claim copyright to the transcript of this interview. I don't actually think that's how research data is allowed to be copyrighted. So for, like, if we were talking about social media data it would be totally different. But for qualitative or survey, I don't think you can actually go back and I don't think you the participants are allowed to under current U.S. law to claim copyright over this. And it was never a consideration for us. And a, that's a good point for equity, actually,

**Sara Mannheimer** 39:14

Yeah. And fair use, you know, if people are using this for sort of research purposes, or educational purposes did you think about is do you think Fair Use plays in here?

**QR01** 39:25

No. Also, because we've made a section of our quotes available, that like we already have a public subset of the data. So I don't think in this particular case, fair use would apply. Yeah, um, yeah. That's my take, I guess.

**Sara Mannheimer** 39:46

Okay. That's great

**QR01** 39:47

For projects that have like fully who don't make that available who have fully closed data. I think that's actually a pretty that's a gray area legally. I don't see it as a gray area ethically like I would never try and claim Fair Use over interviews that could potentially harm someone if an educator helps information get out. Because [I teach], I think deeply about that stuff and like the role of the educator and setting classroom norms and socialization. And, you know, I just couldn't imagine trying to claim Fair Use if that were somebody's deeply identifiable, like talking smack about bosses research data. Yeah, but people are wild. So probably someone is going to. In which case, I'd tell them don't sue me, or [the repository] would probably actually shoulder a lot of that. If it's about the specific files, because they are under their specific terms of access. They have like a standard download agreement that outlines some stuff.

**Sara Mannheimer** 40:58

Yeah, okay.

**QR01** 41:00

And we did of course, grant them rights, [the repository]. We did grant them rights to store it and transmit metadata about it. But still, like we are still the copyright holders, we never signed that over.

**Sara Mannheimer** 41:12

Okay. Great. Okay. Are there any other issues or challenges that arose during your example that we haven't talked about yet?

**QR01** 41:25

Well, I guess in respect to like prepping things, or with the whole process?

**Sara Mannheimer** 41:32

Yeah, um...

**QR01** 41:33

Yeah.

**Sara Mannheimer** 41:40

Do you know, I don't know, like if you had any challenges that you had to discuss as a research group about publishing the data, or if there were any ethical or sort of, or epistemological things that you considered, as you were going...

**QR01** 41:53

Yeah.

**Sara Mannheimer** 41:53

...that we haven't discussed yet.

**QR01** 41:55

Well, I guess the most in depth, we went with that everyone was pretty much on board with publishing. That's not an argument I had to make to this research group. The hardest part was coming up with terms that we all agreed with, and that [the repository] agreed with for our restricted files. So the unrestricted files are [published with an open license]. So that's pretty easy. And we all had a conversation about what that means. And we all agreed to it. And that was a really easy conversation. But then, in terms of the restricted data, [the repository] allowed us to enter in our own terms at the first deposit. And so we spent probably like a good two hours talking about the potential restrictions and terms that we would want to put on people using this data, the situations where we would say yes, and the situations where we would say no, and a lot of that made it in all of it made it into our final agreement. A lot of the stuff that we wanted was actually in [the repository] standard download agreement. So a lot of the stuff that we were advocating for, they were like, "Yeah, it's already..." we don't have to say it twice. It's in our standard download agreement. Some things were added on. And I guess we were sort of nervous at first, when [the repository] was like we need terms that we can act on if we can't get in touch with you. And this is what made us really, really nervous. And so this is why we negotiated for a right of first refusal, which they were happy to give us by the way, they were really collaborative in that respect. So that was really nice. Yeah, I guess I could... something I could give a different example because actually, with [the second project], the collaboration with [the repository] stretched on for a really long time. There actually a lot more issues because we wanted to make all the data completely openly available. And they had no legal or technical mechanisms to let that be so so we actually had to get like, and we got an agreement with them that says, as soon as they are legally allowed, because they run through [a parent organization] and have to go through [the parent organization's] general counsel and their hands are sort of tied in that regard. But they were like, "As soon as we are legally allowed to have public qualitative data, like without restriction, then we will go back to your [second project] and make the data publicly available." Because right now the stuff is you have to log into [the repository] to be able to download it and we specifically didn't want that. Because we... this data set was meant to be completely open. So we actually had like a 30 email thread with them going back and forth about this. [A higher level administrator of the repository] got involved and we came to this agreement that basically like when that when that happens, you we will we will go back and make things open. And we actually talked about that [in a public forum]. Because we knew that our participants were gonna be like, "Why isn't this open? You guys said it was gonna be open." And so we had a little paragraph that's like, here's what it means. Here's what the standard download agreement means. Just know that as soon as they're able to, to make a completely open, they will make you're completely open. And this was, yeah, a big part of the hindrance in using [the repository] for qualitative, non-private data. So yeah, really, really good for the private side, but less they have less mechanisms for the public side.

**Sara Mannheimer** 45:18

Did you consider putting data from [the second project] somewhere else like in just like a GitHub or something to make it more open?

**QR01** 45:26

Yeah, we did consider that. Um, we wanted to stick with [the repository]. One, because it's just like, it's [a known repository that holds] qualitative data, there's a discovery aspect to it. And two, we had already sort of started the deposit process, and they had already gone through and done like quite a lot of curation for us. So I'd kind of feel like a jerk if they had put in all that labor that I took what they did, and went to a different repo. And it's but that said, like, it's behind the data wall. And that's at least better than some other walls that exist.

**Sara Mannheimer** 46:02

So true.

**QR01** 46:04

I'm... that is not my best case scenario, but it's not my worst case scenario. So I was pretty happy with where it where it landed.

**Sara Mannheimer** 46:12

Yeah. Yeah.

**QR01** 46:16

I have a lot of opinions [because of my expertise in data curation] about some of these questions, which we can talk about any other time. But I will say, the Qualitative Data Repository (QDR) has made helping people be compliant with grants easier. So like, again, I'm sorry, this is a little bit of an aside, but maybe it would be helpful thinking about the curation side. We so [our library data management services] review data management plans for people submitting grants. We have like five PhD students from [a qualitative discipline] get their dissertation fieldwork grants get sent back because they said, "We're keeping this data under lock and key forever." And the NSF panel said "That's unacceptable," even if like... they had pretty sensitive data too I was actually shocked that they had said it for one study, which dealt with people talking about politics in Turkey. And they were like it has to, the metadata has to be available at least. So because we're a member, we're actually able to refer like those exact students to the QDR's mechanisms for safely sharing qual data. And that has helped people become compliant with a lot of new NSF in particular mandates. QDR has helped I think, in general bolster people wanting to share qualitative data. For so long at [my university], I heard like, we just can't share qual. And I'm like, that's wrong on a lot of levels. Obviously, I'm not gonna force you if you're uncomfortable if you're like, if all your participants are very uncomfortable, like there's only so much we can do. And maybe we can talk about ways that we can make it private for 50 years, and then public. Like there's other things that we can do. But the conversation didn't really move to be honest, until the until QDR was listed up. So I hate to talk about curation, like only in respect to this one repository, because obviously, we curate things for different repositories like there's Qualidata and ICPSR. There's qualitative data all over. But the fact that you had this request button in the QDR really changed the game, I think in terms of bolstering people feeling comfortable about sharing.

**Sara Mannheimer** 48:34

Yeah, I mean, that's kind of why I wanted I asked you to be a qualitative researcher to have to wear that hat during our interview is because it is harder to find qualitative researchers who have shared their data. Because, you know, people I've reached out to just randomly have said, "Oh, no, I, I would never. So I've never even thought about it." So yeah...

**QR01** 48:58

That actually a problem for sure.

[Several minutes spent discussing who else Sara should interview]

**Sara Mannheimer** 53:31

Well, I think that's everything. Thank you so much.

**QR01** 53:37

Alright, my pleasure. If you want to follow up, feel free anytime.

**Sara Mannheimer** 53:42

Okay. All right. Thank you and have a wonderful day. I really appreciate it.