[00:00:00] Kasha Ely: Hello and welcome to Data Night with the Odum Institute. I'm Kasha Ely and with me today is our guest. Dr. Sebastian Karcher, the associate director of the Qualitative Data Repository at Syracuse University. Hi Sebastian.

Sebastian Karcher: Hey, thanks for having me Kasha.

Kasha Ely: Thanks for joining us. So, to get started, can you tell me a little bit about the Qualitative Data Repository, or QDR, and what your role is?

Sebastian Karcher: Yeah, absolutely so QDR is a data repository which means we take data that researchers have collected in the process of their work and we publish it and make it available to other researchers. That may be for starting a new research project, that may be for often teaching in a class, that may be for checking on the original researcher's work. There are all sorts of different reasons and what's specific about QDR as the name suggests is that we focus on qualitative data. So traditionally in the social sciences, most data that were collected, that were shared, were quantitative data. Think rows and columns of numbers mainly. And qualitative data was merely shared and was kind of like the odd one out. Some people also saw it as kind of soft or perhaps less scientific and so in the early 2010's a number of political scientists got together and said we need a place where people understand qualitative data and qualitative research, but that's also dedicated to sharing that material that's quite rarely shared. And so, in 2014, QDR went live and we've been working with researchers, initially very heavy in the political science area, but now we're very broadly placed within all of the social sciences and neighboring disciplines. So, for example, we get a fair amount of Public Health Data, too. And yeah, so we now have almost a hundred data projects of all different sorts. Our bread-and-butter I would say are interview transcripts, interviews and focus groups. We have a fair amount of historical data, so people would go to the archives, take scans and bring those back and deposit those with us and then we have all sorts of unusual and fun data. We have some video recordings from presidential ads in Latin America. We have some audio recordings from social linguists from a Scottish fishing village. We have some, we have some translated Constitutions [00:03:00] that are marked up thematically so that every sentence has kind of a mark for a what that refers to. So that I would say is kind of a fifth of our data is kind of unusual stuff that make qualitative data so diverse and so much fun.

Kasha Ely: That's great. That's so interesting. I definitely appreciate you mentioning the unusual types of data because I do think a lot of times when people say qualitative everyone just goes straight to interviews and focus groups but obviously it's a lot broader than that. I want to go back to something you said about how people traditionally saw qualitative research data as perhaps less scientific. And I'm curious, what would you say the value of qualitative research is versus traditional quantitative methods?

Sebastian Karcher: I think that really varies and I think you will and you should get very different answers by different qualitative researchers. So, within qualitative research, I would say perhaps in the more positivistic school. So, people who think of themselves more as social scientists would focus on the things that qualitative data just does better than quantitative data in explaining a social realities. So, for example, qualitative data may just be better at getting me a good measurement of a difficult to capture concept. When before we recorded, we talked a little bit about my dissertation research which was about the informal economy, right, and a lot about these are hidden structures and I can get at some official data about it, but it's going to be at best unreliable and at worst completely useless. Whereas when I begin and understand how our informal, how do informal workers work, how our companies are organized, when do companies hire informal workers? I just get more reliable information. So that's one reason you may use qualitative data. There's also a large part of qualitative social scientists who believe that the enterprise of social science should not just be about causes and effects but should really be about understanding - it's social science, right? So, we understand our counterparts and it's about almost stepping in their shoes, right? That's often referred to as interpretive social science. And there of course, if you want to see and make the world understandable through someone else's eyes, you need to use their words. They don't speak in numbers. They don't speak in rows and columns, right? And so, for that you can use statistics to kind of situate your research, but it's really the qualitative evidence where, where the act of interpretation happens [00:06:00]

Kasha Ely: That's really interesting, thank you for sharing your take. So, I've been talking to Dr. Cheryl Thompson. Obviously, she is an archivist here at Odum who connected us, which was great. And I'm curious, what does it mean that UNC is now an institutional member of QDR?

Sebastian Karcher: Yeah. So, we're very excited to have UNC as institutional member. The way QDR is organized is a researcher comes to us and they want to deposit data. There are obviously cost involved for us and every data repository has to think about how are those costs paid, that's storage costs, that's salaries. And what QDR said initially, well, the researchers who are depositing the data, they will have to bear some of that cost. And for some projects that's actually relatively easy to do. If you have a large NSF Grant you can just say and for data archiving, I need, say, two thousand dollars or so and NSF says that's great, we've always wanted people to archive their data and that's no problem.

A lot of projects don't have a large grant, especially in qualitative research, right? And in information science that is sometimes referred to as the long tail of research. All those small little things - maybe it's a dissertation, which by all means isn't small and little, maybe it's a side project, it can be all sorts of different things. And the great things about institutional membership is that it allows us to say if a researcher comes to us from UNC or from any of our other 25 institutional members, UNC has kind of covered the cost that it we, need to curate and store their data so we can do that for free.

So that's one of the biggest and I think most tangible benefits. There are other smaller things involved. So, for example, we will often work with data archivists and data librarians at member institutions to advise researchers on the early stages of data management. So that kind of the process. How do we organize data? How do we store it? What sort of things do we need to think about so that we can maybe later share it? And so, we often, when qualitative data are concerned, we will interface with whoever is advising that researchers, that may be Cheryl at Odum, that may be, may be Michelle at the at the UNC library or someone else entirely. So that's another thing we do I, and especially now since we're all virtual, that has become easier. I sometimes just join librarians or data archivists in workshops they give to talk a little bit about issues specific to qualitative data. And that's also kind of part of what institutional membership [00:09:00] entails

Kasha Ely: Great. I'm curious, what are the data services or tools that you're most proud of?

Sebastian Karcher: That's a good question. I'm proud of all them we do, which child do you love best? So, there's a number of things. So, one thing that I think we're very excited about, like the Odom Institute, we use the data repository software called Dataverse that's essentially developed at Harvard and it's developed at Harvard and it's the Institute for Quantitative Social Science. And as you can tell by the name, they really think about data principally in quantitative terms. So, one of the things that we have been doing and we continue to as we think about how do we use data versus the software at QDR, the data repository, how can we make that work better for qualitative data? And so, to give you some examples the first thing I think that we did was, one of the really nice things about Dataverse is you give it a spreadsheet and then it extracts all the columns, so all the variables, and you can search by individual variables.

Now that doesn't really have an equivalent in qualitative data, but we have all this text. So, the first thing that we did was build a full text search. So, you can now search through all of our holdings to the extent that they have text that's searchable, some archival stuff doesn't, most of it does. And then we first tested this out ourselves. There were some phony issues. How do you do deal with sensitive data those sorts of things, and then the part that I'm particularly proud of is that we then contributed that back to Dataverse. So other data repositories that maybe just have a small sliver of qualitative data are now also able to take advantage of that and there are a couple of other examples of things like that that we kind of build with qualitative researchers in mind, but then contributed to the software that's used by, I think 60 or so different data repositories worldwide. So that's number one. And if I get one more...

Kasha Ely: Oh, yeah, absolutely.

Sebastian Karcher: The, the other thing that we've built in terms of technology is what we call annotation for transparent inquiry. This is a little bit difficult and, in a podcast, because it's a very visual, visual tool. But essentially the idea is that qualitative research is often used evidence in a very different way as they talk about it. It's kind of dispersed throughout the article. You'll say something, and then you'll provide a bit of evidence, a link to an interview or you base that on an archival source or something like that, that may happen 30, 40 times in [00:12:00] an article. And so, if you think about transparency, that's really challenging, right? If you do transparency with quant work, we write some computer code and we, and we have the data, right, rows, columns. And then that together gives us transparency how our tables and our figures were derived, right. And your colleagues at Odum are kind of cutting-edge and pushing that forward, how we can be do that better.

And so, we are thinking about how we, can we do something similar for qualitative data. And so, what ATI does is it puts annotation sort of like a layer on top of an article, so just the way that in the very old days you would put one of those plastic slides on top of an article, right? You see highlights appearing on top of an article and then if you click on and highlight, a little bubble pops up and it says gives you a little additional background by an author, but then also links to the underlying data source that would typically be on QDR. And we're very excited about that. We have somewhere around 20 or so such projects. We've got an NSF Grant to make that easier to use for authors, and get a lot of enthusiasm from both leaders and, and authors on that. And it's also exciting because again, it's built all on open standards and open technology. So, it's all reusable by other data repositories and other projects.

Kasha Ely: That is so exciting. Thank you for sharing and we're going to share links down below in the description so people can explore all of these tools, it's a little difficult to describe it here, but I'm sure a lot of people want to check that out. That's awesome. What was the inspiration for that?

Sebastian Karcher: Yeah, so the inspiration for that is there's been some long term work, historical political scientist at Princeton Andy Moravcsik, has been talking about something similar since I think the late 2000s. He called it active citation. And that's, that's a similar idea. But Andy was, is very much kind of a conceptual visionary but didn't really think about the technology very much beyond, you know, this really should be possible with hyperlinks. And so, when I started at QDR I had interacted with some people thinking along different lines. So, the main technology that we use is a really amazing tool called Hypothesis, which is a company that provides open source web annotation. You can essentially annotate any web page anywhere online and then other people can see it and even interact with it. And so that's kind of the software we built this on top of and it's really crucial for making this visible and easy to use and happening not just on a given web page where I have full control, but I can actually show these annotations at the publisher, right? So, people don't have to come to QDR to view the article. They [00:15:00] come to Cambridge University Press, to Sage, to MIT Press, etc. And we are able to display that overlay and then they can click through and get to the source that's on QDR.

So, so that's, so the original information was Andy's work on active citation. And I think the inspiration there was really this idea, this frustration really off of social scientists working with qualitative works that you read something and something maybe doesn't seem quite right or something seems like oh that would be really interesting, I wonder what else he said in that document, right? And then you have no way of getting, there unless you know, you travel to the same archive and then you wait a day until the archivist is able to get you the box from the archive and then, you know, one month later you figure out that after all the document wasn't all that interesting. And that's obviously not viable and the fact that we're still kind of at a place where a lot of qualitative research requires you to do that if you want to dig deeper is really, really frustrating. And so, I think that's where the initial idea of active text citation and then our implementation, our annotation for transparent inquiry came from in that we want to make this easier, we want to bring the sources that researchers based their research on closer to the reader. Both to increase trust, so that readers can say, okay, I've now read the whole document and the author really provides an accurate description of what it says, but also to allow readers to go deeper into the tanks, right? If something really interests you, you can kind of follow go through the rabbit hole and follow a trail off of primary documents and what really happened there. and that's actually perhaps the part that I'm even more excited about.

Kasha Ely: I love that and all of these tools you are talking about and QDR, it seems like there's a huge emphasis on not just your group but the scientific community as a whole, kind of an all ships rise with the tide mentality. At Odum we talk a lot about data sharing and reuse, reproducibility. I'm curious, what does this mean to you? And how does it or does it not relate to qualitative and/or mixed methods research?

Sebastian Karcher: Yeah. Can I hook into that -- all boats rise with the tide first because I think that's really important in another way, too, and that's kind of a special situation that we have in the U.S., where social science data repositories have been collaborating very closely for a number of years now. So, Odum, ICPSR of course, the Roper Center for Public Opinion Research, QDR, a number of other archives have, are joined together in a group that's called Data [00:18:00] Pass: The Data Preservation Alliance for the Social Sciences is I think what that stands for, and we think about how can we help each other, how can we promote data sharing, we've had several workshops with journal editors and that will get us to reproducibility in a minute. But so we've had several workshops with journal editors together where we tell them, look, here's how your journal can work with data archives to make data more accessible, to make reproducibility materials more accessible, and, and that's really very much in the spirit that if we strengthen the community of institutions that help data sharing, that provide the infrastructure, that really helps all of us, and in a field like academia where a lot of it is about competition, it's intellectual competition, but it's still competition, I think it's a somewhat unique situation and its really great to work in a space like that. So, that's my aside on all boats rise with the tide. For reproducibility, I want to caveat that, that I don't really believe reproducibility is the goal of most, of most qualitative research. So, if you think of what I said earlier starting with your interpretive lists, right? Interpretation is always an act with two people, right? You are interpreting someone else, but you are also bringing your own personality, your own lived experience in it, right? And so, if I, Sebastian Karcher, interview someone and then interpret that through my lens, it's going to be different than if you, Kasha Ely, do the same thing, right? Because what you bring into the interview is different, and so it's not going to be the same. It's not going to be reproducible in the quantitative sense. And, and that's good, and it's not the goal of what many, most qualitative researchers do. There are some exceptions, right? There are some people who use qualitative research as kind of a building block in a very kind of quantitative framework where qualitative research does kind of build some block, maybe establishes the conditions for a natural experiment or something like that. There, you probably want reproducibility. A lot of time, you do not. So, at QDR, we almost never talk about it, we always talk about transparency. So, we always talk about research transparency. For quantitative work, transparency and reproducibility are very similar, right? Because if you are, if you provide your code and your data, that's very transparent. You show people exactly what your data are and how you analyze them. **So, you are transparent, but, and I always think that's by accident.** You're also reproducible [00:21:00] because someone else can then take that code and run it, but the core value I think really is transparency there, and so that's what we, that's the shared standard between qualitative and quantitative social science or general scientific work, and so research transparency for me is about making your data and your methods understandable and visible for a broad audience to the extent that ethics permit it.

Kasha Ely: I think that's very important distinction. So, thank you for going through that with us. How do you promote that transparency with the methods? Is that something that you list with the projects that you feature? Is that something that you promote when people are submitting data to you? Do you have requirements for that?

]

Sebastian Karcher: We don't have strict requirements, and that's again by the nature of the diversity of our participants, right? They may understand what they do and what they want to do in all sorts of different ways, but we certainly encourage people to be as transparent as possible. So, a lot of the time when we get data in there, we know relatively little about the data, right? We have like one paragraph about these are focus groups, interviews from X,Y, Z, say and so one thing that we do in our curation process is we have a conversation, either by Zoom or by email, and asks, can you tell us a little bit more about that, right? How did you select the focus groups? Can you give us the guide that you used to facilitate the focus groups? Who was there at the focus groups? Who was maybe not there at the focus group, which is often an interesting question in if you think about selection, those sorts of things, when were they... as much information as possible about how the data were collected that enables people who then access those data to understand. Okay, this is how these came about. This may be special about them. This is what I need to keep in mind as I try to read through this transcript and understand what was actually going on in that conversation as part of our curation process, and that's one of the key things I think that we do is we try to get as much information related to transparency, especially the production of the transparency as it relates to the production of the data into our documentation and then into structured metadata as we catalog it.

Kasha Ely: And do you have any advice for a person wishing to archive their qualitative data, maybe throughout the project and at the end of project?

Sebastian Karcher: Yeah, so the first thing, and I think that's no different from quantitative data is that you want to think about this from the start. One of the big objections we get to data sharing requirements [00:24:00] request for data sharing is that it takes a lot of work, and there is some work involved, right? You need to probably clean up stuff some odd, and you need to deposit it and then answer our questions, but if you plan ahead and think about okay, these are the things that I'm likely going to have to or want to provide; these are the formats in which I'm going to submit it; those sorts of things, you can save yourself a lot of trouble and more importantly perhaps time. There's also, beyond this, there are some things that can if you misplan, so to speak, can prevent you from sharing data all together, and the most critical component there and something that my colleagues at QDR have done a lot of work on is research ethics and specifically notions of informed consent. So, the first thing that I always look at when I get data from a researcher doing human participant research, my first question is always: could we please see a copy of the informed consent script that you used? And in a fair number of cases, these informed consent script that have been passed on from generation to generation contains stipulations like only I and my research team will ever see the data, or we will destroy any information you provide to us three years after the conclusion of the research project, those sorts of things, and at that point, I have typically have to say, well, we can't really share these data. In some very rare cases, people can go back to the IRB, file an amendment and then ask people, so I'd actually like to share this more widely. In most research settings, that's logistically just not viable, and so those data don't get to share it. So, the biggest advice I think would be plan ahead, but the second biggest advice plan ahead particularly in terms of thinking about human participant research and informed consent.

Kasha Ely: I'm also curious, I think it might be a common myth that people feel like no one's interested in my data. So, maybe they won't plan a plan for it because I think at the end it might not be of use to people. How would you respond to that?

Sebastian Karcher: I think especially because in qualitative data, what's interesting is that there just isn't that much out there. So, the examples that are out there particularly now get a fair amount of attention. The biggest draw for qualitative data tends to be teaching. Faculty are desperate to have real life examples in their field of usable data, collections and qualitative research. We often get requests. We get reports back from people who find data that they like on QDR, and I think that's amazing, right? Like, the thought [00:27:00] that you deposit your data and you helped train the next generation of social scientists maybe in your discipline, maybe even in a discipline beyond that, it's super cool. And again, because there is relatively little out there, if you deposit your qualitative data, there's a good chance your data is going to get picked up in that way. There is for larger data collections people definitely just poke around, and there is not a ton of cases of studies based on secondary qualitative data, but we are seeing some of that, so colleagues of mine are just guest editing issue of the journal, the qualitative report, where they took one of our data sets and gave it to, I think six or so different groups of qualitative researchers who then approached it from different qualitative traditions to see kind of what, how does it differ? How do we make different things out of the same data, which also you know gets you to the problems with reproducibility. That's not really our goal, but right? We get these different insights out of the same data sets from different perspectives. So, I think we're going to see more of that as people understand better that qualitative data are out there and that qualitative data can be used. The last thing I would say though is that the people who are going to look at your data are going to be fewer than the people who read your publications, but they are going to be your most dedicated readers, right? And so, you're probably not going to have a thousand people browsing through your data set, but you might have five people reading through it who have really, really engaged in your work, kind of your super fans or the people who really get a lot out of your work, and in a way, those are going to be the people who really push the science forward because they get engaged deeply enough to maybe build on your work, you know, maybe critique it, it happens, and that's how science makes progress.

Kasha Ely: That's a great perspective. Thank you. What do you see as some interesting challenges in working with qualitative data?

Sebastian Karcher: Yeah. So, research ethics is probably the biggest challenge that we work with and qualitative, all social scientists must be and are concerned about ethics and the well-being of the participants. There is something specific about qualitative research in that interaction between researcher and participants tends to be more direct and often tends to have a personal note, right? Some people I just interview once, and it's kind of an elite interview, and it's more transactional, but some people I interview multiple times, if I do ethnographic research, and I live with them, often become friends. So, the relationship is closer [00:30:00] which means there is more to reveal, but there's it also means there's perhaps an additional responsibility to think about and look out for people's well-being, and so we need to be very mindful of that as we share data. I think informed consent is definitely part of that, like being very straightforward with people with, this is what I planned with your data. Also giving people perhaps the choice to say, look, you can write about this, but you can't share our interview transcripts. I think that's perfectly fine, right? Like informed consent really has to mean that there is agency on part of not just a researcher, but also the participant, and we think about research ethics and what that means for data sharing at QDR, really every day, and I think probably more so and more deeply so than many people working at quantitative data archives, with exceptions, right? There are very deep, rich quantitative data sets where similar issues arise, but that's probably the most interesting and the biggest challenge for sharing qualitative data. How do we communicate that? How do we do it safely? How do we do it responsibly? That entire complex, and I think we'll be working on this for a good number more years, although I also think we've made a fair amount of progress in this.

Kasha Ely: Just to touch on my progress, and maybe just the field in general, I'm curious how it has changed in your time in this field. How has qualitative research changed in your mind?

Sebastian Karcher: I don't know how much qualitative research itself has changed. There is, I think, some more uptake of technology, and there is probably more awareness of multi-method research among many, so though certainly not all qualitative researchers. So, in political science, I would say, for example, which I know best. we see more people using qualitative data analysis software. Poli sci's a late comer there. When I did my research dissertation, I finished in 2014, it was very rare to see political scientists using something like Atlas or in vivo, and I would say, it's still not anywhere close to where you'd see say, sociologists or public health qualitative researchers, but we're seeing much more of that, and I think there is perhaps more researchers who're trying to interact with qualitative data, both in a quantitative and a qualitative way, right, and that goes from the digital humanities, right, who do computer assisted reading of corpora to qualitative social scientists who read, who have deep area knowledge, but then also say, [00:33:00] used some sort of text mining algorithms to find out more about the online foray in which they immersed themselves, those sorts of things. So, that's not super new. Right? Like, I don't think we're talking like, last two years here, but I do think that's probably areas that have been changing over the last 10 years, certainly. I mean, social media is probably the biggest there, right? I always forget. I think, Facebook is not even 20 years old, right? So, does the role of social media and understanding how we study social media, how we study it ethically, what it tells us both quantitatively and qualitatively, is a huge topic, and that's just some random thoughts.

Kasha Ely: That's great. I'm curious, how does qualitative research into social media? What does that look like?

Sebastian Karcher: Well, you can do things like participatory research, right? Like, people would become part of a Reddit group, say, and interact with the people of foreign bonds, read deeply, right? So, that would be a common example. Same thing with say, Facebook groups, if you follow specific Twitter hashtags. Yes, you can, you know, scrape them all automatically and then throw text analysis at them, and that might tell you something, but if you actually read through them and look at the interactions that are behind that and maybe are able to then also follow up with some of the participants in those online dialogues in some cases, they're anonymous; in some cases, they're not, but even the anonymous ones might be willing to talk to you through their direct messages, all those sorts of things. So, yeah, it's a rich field. There's a lot going on.

Kasha Ely: Stepping back a little bit into your research and your experience, you mentioned you did your dissertation in political science. What is it that drew you to political science originally?

Sebastian Karcher: What I liked about political science was kind of the big picture stuff, right? I want to figure out how things work and what kind of bigger picture than you know, figure out how power and politics shape a society, can you go, so I think that in the broad scheme of things really was what drew me to studying political science, and then as I looked further, what I ended up doing in my dissertation was that I was looking at labor markets, and that, again was driven by a similar idea that as I went to countries, and I talk to people, the kind of, the overarching importance of where do I work? How do I work? How am I getting paid? Am I getting paid? Am I going to get to keep my job? It's kind of, for many people, one of the most [00:36:00] important things that they think about in their life, and so I wanted to better understand how differences of that are shaped. Turns out, that's very difficult. So, I don't know if I would call my dissertation a roaring success, but it was fun trying to get at it.

Kasha Ely: There you go. That's all that matters. What was difficult about it?

Sebastian Karcher: What's difficult about it is that if you ask large questions, if you get great answers, you're great, but it also, you can get stuck and then you're left with not, you know, some interesting description, but not really a compelling story to tell that would surprise people or that would really add anything new to what people already know, and I wouldn't say that I didn't do anything new in my dissertation, but that certainly was probably the weak part. It also sometimes makes it hard to know where exactly to look. There are so many things that are potentially relevant. There may be too many things, and I may have gotten, you know, drawn into many different directions. I get easily distracted by other things that are interesting, and that's perhaps not always the best characteristic of someone writing a dissertation, so I don't recommend it. It's great if you work in a data repository though because I do 10 different things every day. So, there is that.

Kasha Ely: Just to wrap up here, what is one thing that you'd like to tell our listeners about QDR. If there's you know one takeaway you want them to have, what is it?

Sebastian Karcher: I think if I wanted one takeaway, I, is that we really like working with researchers, and we really like working with researchers through kind of their research process, and so if you do work with qualitative or multi-method data, and you are thinking well, maybe I want to share my data. I want to share my data, or I need to write data sharing into my grant application, drop us a line and, and talk to us, and we may end up deciding your project doesn't work, or we may end up, you know, putting in the building blocks for then sharing your data 3, 4 years down the line, and both are great outcomes.

Kasha Ely: Well, thank you so much. I really appreciate you being a guest today and chatting with us about Qualitative Data Repository. Again, we're going to drop the link down below. So, everyone can check that out. Stay safe and well.

Sebastian Karcher: Thank you so much, Kasha!