

The Breadth and Depth Approach to Big Qual Data Analysis: Transcript



[0:00:00]

Susie Weller: Welcome to our video on the breadth and depth approach to big qual analysis.

My name is Susie Weller and I'm a senior researcher in the Clinical Ethics, Law and Society Research Group at the University of Oxford. Emma?

Emma Davidson: And I'm Emma Davidson, I'm a senior lecturer in the social policy subject area at the University of Edinburgh. And Lynn?

Lynn Jamieson: I'm Lynn Jamieson. I'm also at the University of Edinburgh and I'm in sociology and I'm handing to Ros.

Ros Edwards: Hi there, I'm Ros Edwards and I'm Professor of Sociology at the University of Southampton.

Emma Davidson: Okay, so tell us about the types of big data that you work with and the challenges that it poses. I think the first thing to say is that when we think about big data, we're often thinking about very large volumes of digitally born data, and that often means that we're thinking about data sets that can be analysed computationally or using kind of quantitative approaches, and we're interested in understanding high level patterns and trends and associations.

But our work, both individually and as a team altogether, we've been really interested in this growing availability of big qualitative data sets and big qual. And our view is that this is really presenting new opportunities for social research. And I think partly this new opportunity has been facilitated by growing archiving movement and data sharing movement, but there are also lots of other types of big qual data that's out there.

For our project, we've been focused on developing a method that allows researchers to manage and analyse large volumes of qualitative data and Susie's going to tell you a little bit about the Timescapes Project that we've been using to do that project. Susie.

Susie Weller: Thanks, Emma. So Timescapes was the first large scale qualitative longitudinal study to be funded by the ESRC in the UK and it comprised a set of seven empirical projects looking at different life stages, but there was also a large focus on archiving and data sharing and thinking about what the implications might be for the data and for qualitative research teams, this fantastically rich, detailed set of qualitative projects that we could use to help develop this approach to big qual analysis.

And these archived datasets bailey comprise qualitative interviews and we've really been working with qualitative text based data. And maybe Ros can tell us about some of the other projects that we've been involved in.

Ros Edwards: Yes, I know that Susie's been working with large international data sets, merge data sets. Again, that was transcribed interviews and focus groups. But we've also tried to branch out a little bit. We've been working with a public engagement exercise, which is written responses to questions, and so we've moved in from working with transcribed data into textual based data as well.

I think maybe Lynn, you could say something about what's happening more generally in the qualitative archive data field.

Lynn Jamieson: Yes. I mean, archive data, of course, does require a certain amount of resources to make the archive possible, but I think increasingly, there are not only official archives held by institutions that are maybe publicly funded by governments or educational bodies, but also community archives that are building up and maybe oral history data about particular places or events, sometimes particular political projects. So, there's a growing body of archive data. Obviously, that is more dense in the rich world, but it exists, I think,

increasingly everywhere, as long as people have got access to some kind of electronic device.

[0:04:17]

Ros Edwards: So maybe we could have a think about what are the challenges that we face with using the particular data sets. One of the things that comes to mind is that you need to have funding opportunities to actually go and reuse data. Other things, Susie, has anything come to mind?

Susie Weller: I think whilst that availability of data offers qualitative research or all researchers great opportunities for asking new questions of existing material, I think we have to be really conscious that we could be reproducing inequities if we're just constantly relying on the same data sets that may not be made up of diverse populations and by constantly reusing them we may be missing out on the voices and experiences of some populations. And there's also that kind of reliance on archive material could exacerbate the dominance of research from European and North American context, I think. So, we have to be very mindful of that when reusing data. Of course, the breadth and depth method can be used for data that isn't held in large scale repositories as well.

Ros Edwards: And I suppose another challenge would be context, wouldn't it? It's really important for qualitative researchers to take account of the context and when you're working with large amounts of data, I think that our method does allow us to do that.

What is your approach to analysing big qual and what does it allow you to do? So, I think maybe, Emma, would you like to kick us off on that question?

Emma Davidson: Yeah, so our approach to big qual, as I say, as a team, we were creating this method and we felt that we had to create a method that would allow us to engage with big data, but do it in a way that had qualitative integrity. So often we found that when we looked at work that was using computational methods, it would end with the computational analysis and wouldn't go any

further. So, you were getting the breadth of understanding, but not necessarily the depth.

So, I guess our method was really designed, and that's why it's called the breadth and depth method, was to try and combine these two elements. So, combining computational techniques with the kind of in-depth analytical work that is more conventional forms of qualitative data analysis.

And just really importantly, we wanted to ensure that new innovations and big qual analysis were developed in a way that actually involved and engaged qualitative researchers because we think that's a really important thing to do. And as we say, we were working with big qual data to enable researchers to scope out questions that allow for comparison and for generalisability.

The method can be used with any big qual material, as Susie's mentioned. So, we worked with archive data, but we can also work with other forms of data, whether that's primary data, multiple projects that we bring together or multiple cited projects. Susie, do you want to pick up on the actual method here?

Susie Weller: Yeah, sure. So, the breadth and depth method is an iterative approach to managing and analysing large volumes of qualitative data. And you can apply it flexibly, whatever the theoretical logic, whatever your substantive topic or the nature of the qualitative data. And it's a good way of conveying this movement between, a good way of conveying this movement between breadth and depth using an archaeological metaphor.

So, Step 1 is the breadth element of the breadth and depth method, and that's very much like an archaeological aerial survey used to gain a broad overview of the landscape of available potential secondary data. And the purpose of this is to identify several sources of existing material that might be relevant for your research topic or questions.

[0:08:37]

And then Step 2 continues the breadth element, and archaeologically, that's a geophysical survey. So geophysical surveying maps the patterning of the landscape features and detects these subsurface areas that might be of interest to your project. And this step uses computational data mining approaches to identify areas of potential interest within the corpus, this big corpus of data, that you've created. from pooling lots of data sets to point towards areas of preliminary deeper investigations.

And then the third step is your preliminary analysis, and this is where you start to move from breadth to depth. And drawing on our archaeological metaphor, this is really like digging shallow test bits into the data landscape to see whether there's anything of interest or not to your research questions. And at this stage, you don't go into the data in great depth, you're just trying to see where these areas of interest might lie.

And then a fourth step moves us fully into the depth element, from examining extracts of data to deep excavation in whole cases. And deep excavation is part of the archaeological process of working with big qual that heralds the ability to bring back into conversation with breadth. So, it's a rigorous process. You can track back and forth. It's iterative. So, at different points in time, you might want to go back to a previous stage to look at the data again.

Ros Edwards: And that's the rigor of it, isn't it, really? That you can track backwards and forwards in that way, I think.

Lynn Jamieson: If I could just come in then and talk about the limitations of the approach. The first thing to say is that we are assuming that people are working from an idea, a theory, a question. We're not talking about randomly putting lots of text or qualitative data of some sort together and then just doing some kind of data sorting using quantitative techniques from data mining and then having a bit of a look. We're talking about bringing together data purposefully to answer particular questions or to explore in more depth some particular

issue. So, there's already a theoretical framework there at the very beginning. You know, theoretical might sound more elaborate than it is, but it's nevertheless informed by some questioning of some issue.

And it's not a fast process. I think people imagine, well, we can use quantitative techniques for large volumes of qualitative data, and this will be a quick fix sometimes, but that's not the case. One limitation of this is it's slow. It's got to be slow because it involves these different steps, and for some researchers, it would maybe involve learning techniques that they don't already have. And some of these techniques are not completely straightforward and easy to acquire if you've never done any statistical analysis or you're not sort of quantitatively minded, although there are proprietary softwares that you can use that are relatively easy as well. So maybe I'll stop there and hand over to Ros to develop that a little bit further.

Ros Edwards: Yeah, I think that those are really important points, Lynn, but the data doesn't need to be there. That's another limitation. If you're working with secondary data, if you're collecting massive amounts of primary data, then you'll have the data that you want. But another feature is that if you're working with secondary data, actually you have to be looking for that data, it has to be there for you to work with. You know that we have quite good archiving infrastructure in this country, but not everywhere does, and it's quite variable as well. We haven't actually used visual methods yet, so that could be a limitation that we can't handle it, but we're intending to try at least.

Lynn Jamieson: I was also just going to add that when the data have got some comparison, a comparative element, I think that's one of the strengths of the data as well as a challenge. It allows you to bring together data to make new comparisons and having a comparative element to the data is a big advantage.

[0:13:19]

Susie Weller: So, we have some questions that we'd like other researchers to think about, and I think I suppose the biggest question that we've always tried to hold in

mind is how can we ensure that big qual analysis is conducted with qualitative integrity. And I think our focus on combining breadth and depth and combining computational techniques with more conventional forms of qualitative analysis has really been key in trying to ensure that. Emma, are there any other questions that you'd like to pose to other researchers for them to think about?

Emma Davidson: Yeah, I suppose there still is this issue that the privileging really of more quantitative techniques. So, I guess I'd like us all to think about collectively why that's the case, why do we privilege that kind of data, and to really think more deeply about how this kind of method can benefit our understanding. You know big data is this great opportunity, can we think about it in a more qualitative way and what value can that bring?

I think in our training, we've been asked quite a bit about AI, haven't we guys, about how we might utilise or make use of AI and I think as social researchers, we're all slightly scared of the potential of AI being used for qualitative research. I personally think we should be, again, like computational techniques, not ignoring it, but trying to work out how we can play a role in that conversation and making sure that we're not kind of excluded from that. It feels like that's even more important now in the context that we're living in, that it's really important for qualitative researchers to be part of that conversation.

Yeah, and I think, I don't know if someone wants to pick up on just the ethics of care because often when we're working, particularly with archive data, we're assuming that, oh, right, we can use this data because it's been archived in an infrastructure that's already dealt with ethics. But we've spoken a lot as a team about the importance of ensuring that we work with the data that we have in a really careful, ethical way. So, yeah, does someone else want to pick up on that? Yeah, this is Susie. Yeah, go ahead.

Susie Weller: I think it's really important. I think it's often underestimated how much work, whether that's in time, in the intellectual contributions, in emotional labour

goes into developing and curating a qualitative data set. I think that's often underestimated and something we need to really bear in mind when working with data generated by other researchers and teams.

I think Emma's point as well about AI and the increasing discussion around AI and qualitative search is a really interesting one. And in developing the method for Steps 2 to 3, we experimented with a range of computational tools to conduct that breadth analysis and then sort of moving into the depth analysis. And some of those use the programming language R, but we also experimented with AI enabled tools like Leximancer. So, we have started to look into that and think about how that might fit within our own work.

Lynn Jamieson: I think one of the issues that is seen as a potential advantage of the method is increasing generalisability. Now, I think it's important to think about that quite carefully so we're not claiming that if you just add lots of stuff together, it becomes somehow more statistically generalisable. We're never, ever arguing that because what we're doing is drawing on large volumes of qualitative data that have usually been collected in quite small samples. But nevertheless, what you would maybe have are small samples with different characteristics of your research participants. So, the thing that is often said about a small qualitative sample is, well, how would you know that would be the same if it was people who were older or people who were in a different national context or people who were in different socioeconomic circumstances? You may be able to piece together different circumstances and say something about them, but you're never going to be trying to do a statistical generalisation.

[0:18:00]

Susie Weller: Thank you very much for listening to our video on the breadth and depth method. Please do check out resources that we've generated over the last few years. Thank you.

[End of Transcript]