Speakers: Dr. Libby Bishop (LB) and Patrick Brindle, Publisher, SAGE Publications, Ltd., interviewer (I)

I: Libby Bishop, thank you very much for talking to me. The first thing I wanted to ask you is what is secondary analysis of qualitative data?

LB: It can sometimes be a bit difficult to define, but I use an informal definition, that secondary analysis is using data for any purpose that is different than the original purpose for which the data was collected, and it can be done in a lot of different ways.

It's a relatively new area, particularly for qualitative data. It's quite a bit more established for quantitative data, but it can be done in ways such as taking an existing dataset and asking a new question about it, which is something I did a small example of, working with an existing dataset where the original purpose was an investigation of health effects, generally a health study, and I used it to look at the examples of occurrences of conversation about food, which is a part of health, but the original researchers had not dealt in any detail with the food topic.

Most people tend to reuse data for topics or themes that are broadly similar to how the data were reused; not always, but if the original study had to do with collecting information about class or something, then perhaps a related topic might come up. In other cases it really again can be completely a different topic.

Other sorts of uses of the data are it can be very constructive to do comparative work, so you might have comparative work cross-sectionally, geography, it might be collecting your own data in one geography here in the UK and want to do comparative work internationally, or in comparison to another region in the UK; and increasingly of course with the idea of temporality, in time across periods, you can also do longitudinal comparisons. So some people have collected huge quantities, both in the US and in the UK, of very valuable data collected in the sixties and seventies, and now people are revisiting that data and trying to make it more contemporary. It's also an example that sometimes people revisit their own data, so it doesn't have to be a new researcher. People go back and look at their own data from decades ago and sometimes find very new things in it.

Increasingly, as well as the research aspects of it, you can reuse data to study methodology. It's a fantastic way to get below the surface of what's happened in the study, so you might read a published account and you'll get the official version, if you will, but sometimes reading the actual data, reading the flow of an interview, you'll get the sense of how the data collection actually went and get a much better and deeper understanding of the methodology. I sometimes make a

comparison with doing a literature review but it's a much deeper result than you can get with a literature review.

And the final important area to remember is it's very valuable research for teaching. So being able, sometimes new researchers in particular, new researchers taking up teaching activities, only have a small set of data, they might only have their own area of data to share with students. Using archives like the UK data archives or new qualitative archives in Ireland and in Europe even, there's a wealth of resources that can be used for teaching both substantive areas and methodology.

I: Can I just ask at this stage, how is what you do, in terms of secondary data analysis, different to what, say, historians are doing in revisiting data?

LB: I actually don't think that there is a great deal of difference. It's not surprising that a couple of the pioneers of doing secondary analysis, there are many names I could pick but Paul Thompson's major Edwardian study was one of the founding studies for qualitative analysis, qualitative data. Mike Savage of course was a key one who's redone data. These people come out of oral history or historical traditions. So it's maybe a bit unkind, but sociologists sometimes think they're discovering brand new things that in fact historians are well practiced in already doing. So it's a highly enriching cross-fertilisation, I think, to see the collaboration across the social science and historical methodology.

I: OK, so it's nothing really to do with the types of data that you''e working on that might distinguish?

LB: The distinction would be much more data in social sciences is actively collected; so, much more rarely for historians will data have been produced for the historical record. Now you an argue when someone famous is writing a diary, is that really a document for their own personal use or do they actually have in mind that it might become a significant historical document, but most of the kinds of materials that historians use have not been explicitly, actively generated by a research process like an interview elicitation process.

I: So who does secondary data analysis?

LB: Well, increasingly birds do it, bees do it, pretty much everyone is doing it in one form or another! I'll mention a couple of names but some of our stars of doing secondary analysis, people who've been doing it quite a while now, one is Mike Savage, who's also here, and in fact just released his book here at the Research Methods Festival, that is an example of using secondary analysis. So his work I found particularly innovative because he drew on a very large set of multiple collections of data; so oftentimes a reuse project might focus on only a single collection, whereas Mike drew across a really wide range of periods and

times and different kinds of data, actually drawing on even across the qualitative-quantitative boundaries. So he used affluent worker data, which is housed in Qualidata, he also used the mass observation directives, which are a particular kind of data where people are asked to write on themes and produce data, and he's investigated these issues of identity, and particularly people's class identity, over quite a long historical stretch now, by being able to draw on these historical perspectives. Another good example, I'm working with this team quite closely now, also people here at the Research Festival, Graham Crow and Dawn Lyon, and they're reusing data collected by Ray Pahl in the 1960s and '70s I believe. Ray did a large number of projects but one of them was essentially a community study done on the Isle of Sheppey, which is an island in the southeast of the UK, so it investigated many aspects of community life in Sheppey but one of the bits of data collection that Ray did was draw on essays that young people were writing about what they might imagine their future lives to be, and parts of Sheppey anyway were regarded as somewhat impoverished, socially excluded areas, so part of this investigation had to do with how do people growing up in these circumstances imagine their futures? So Graham and Dawn are coming along and doing many different kinds of work with this, but one of the exercises that they're doing is reintegrating themselves into the Sheppey community and into the schools and indeed having current young people write an almost identical kind of essay so that they will now have a very nice comparison of how young people have imagined their futures across this timeframe.

The Timescapes project, which is a qualitative longitudinal project based at Leeds, one that I also work with, is projects that are studying the life course and they're a series of projects and there's a pair of projects that are studying parenthood, so one on motherhood, one on fatherhood, a couple of other pairs of projects are focussed on order generation and grand-parenting issues. So another way, and it's a bit different, but that secondary analysis is being done, is even within the auspices of this ongoing live project, if you will, there's cross-fertilisation within these teams. So you might call it primary but on the other hand these people are using, the motherhood project is drawing on data from the fatherhood project, and vice versa, so it's kind of a contemporaneous form of secondary analysis.

But probably the most exciting thing about people doing it is that now increasingly I'm opening up conference programmes and looking at particular sessions and finding things on secondary analysis where I don't even know any of the people or any of the topics, because it is proliferating quickly. It used to be that I knew most of the players, and now it's quite exciting to see that it's growing so quickly that it's getting a bit difficult to stay on top of it all.

I: What might the problems of using secondary data?

LB: I think actually one of the problems, and I would say this in particular to younger students or post grads thinking about exploring this as a method, I actually think it's an image problem as opposed to a real problem. So it's what I call the poor relation problem, and that is that somehow, maybe because of the name secondary analysis, it's also thought of as second class, second tier in some way, and there is a little bit, I suppose still I think, of a bias that would say that somehow primary is privileged, that primary data is always better. I just don't think it's true. I think certainly talk to any historian, as you mentioned before, how could any historian think that somehow versions of secondary data would not be equally valid or rich to use?

I'm personally trying to even get away from the term secondary analysis, although it's very hard to kill a term unless you have a new term to replace it with, but I increasingly start trying to use language like reusing data (which is a bit clunky), revisiting data may be better. I tried recycling data—it sounds warm and friendly and climate-change friendly and whatnot, but typically reusing data is the term I will use. So it is a bit of an image problem with trying to get away from that.

The biggest challenge, I think, that probably comes up with reusing data usually focuses on the issue of context, and by context what is typically meant is the idea that there is something precious, invaluable and unreproducible about the original context and situation of the data collection moment, if you will, and there are hot debates about this in terms of various sub-branches of methodology and so forth, but the idea is that . . . even take our example, there are important contextual information that is happening in this particular interview: the room we're in, the temperature, body language, the fact that another person is watching this interview, these are all things that somebody would need to know, even analysing this as an example of a formal interview situation, something like that. Now within a secondary analysis, even if you have the transcript, quite frankly even if you have the audio, even if you were to have a video of this video session, it wouldn't be the same. It wouldn't represent . . . things it might have missed would be your briefing of the original setup for the interview, it would lack context about why you're doing this whole series of interviews in the first place, any previous context that we might have had, early images you might have had of me, me of you, these sorts of things. So all these things, it is quite true, cannot be fully reproduced as context in doing secondary analysis, but then the question is do they need to be? Must that information be present for anybody to make good use of the data?

So I have two answers for that. One is we're getting better actually about collecting contextual information, so that's one answer. OK, if the context is missing, try to fill it in, and all the qualitative data archives work quite hard to do that, so they would collect multiple genres of the data, audio, image, photos, etc. They would collect background information of how the project came into being,

how it's funded, reports of the setting, extensive kinds of contextual information. So that is helpful and I think that that's a very useful way forward. I still though would agree with the critic that says it doesn't fully answer the question, it doesn't solve the problem of fully reproducing the context. But I also think you don't have to fully reproduce context for data to be fully reusable and useful in very different ways, because in fact any time a researcher comes to new data, that is part of their job, to gather as much context as they effectively can, use good judgement and make interpretations in re-contextualising that data for the current setting. And that's a term from Niamh Moore, who's written on exactly this issue. So in some sense you can't analyse data without context because simply in the process of doing the analysis you are re-creating the context, and that is part of the job of somebody doing secondary analysis, reusing data, rather than primary. And again this is of course a whole experience and kind of technique and toolset with which historians are quite expert and very familiar, and I think we can learn a great deal from that in terms of bringing our tools to data reuse.

I: Can I ask when secondary data might be preferred to original data?

LB: There are several situations when that might be the case. One is, think about any sorts of instances where data already exist and there is anything burdensome about collecting the data again. So think about vulnerable populations of various kinds, so elderly, ill, people who have been over-researched—and I'm afraid increasingly, listening to again some of the talks around the festival today and yesterday, we know this is the case. You're seeing response rates drop on surveys, difficulty in accessing certain populations who are deemed challenging, so this idea of burden is important, and any time you can find data that might fit your requirements without burdening a new group, that is quite useful.

I think sometimes people don't appreciate that another whole value of reusing data is although reusing data isn't always faster and easier, I'll come back to that point in a minute, but somehow it does allow your mind to shift and pay attention to other aspects of research design and research methodology. So as opposed to having to focus a huge amount of effort on the particulars of recruiting, for example, you still of course have to think deeply about sampling and why you're choosing some people and not others, but you can focus on design research questions, you can focus on sampling design. And again I will just point out Mike Savage's work I think is brilliant on this, because he had to think quite deeply about not only which archives to pick and which collections to pick amongst those archives, but there was a vastly greater quantity of material than he could use, and so within any archive or collection, he had to figure out effective ways of sampling. And I think that is an important thing for us to pay more attention to. We sometimes, I think, shirk that a little bit as qualitative researchers.

And also just getting down to the deeper layers of what I would call research analysis, which are things like thinking through implications for theoretical development and thinking through things like causal mechanisms. I will personally go on record as saying I don't think we have done enough of this and I don't think we're doing it well enough. And again other people at the conference, Julia Brannen made these points yesterday, that she thinks that there actually will be an increase, a growth in reuse of data, and that although there are some downsides of that, there are some key benefits, and one is exploring these areas. And that raises a point, that I think it is a bit tricky right now because one of the reasons of course there might be more pressure to reuse data in the near future is that the money for collecting new data may be scarce and very hard to find. This was the point that Julia made. So I think that's true, and I will sadly enjoy the benefit that archives might get if in fact people come to us and need to reuse our data; I will take advantage of that, but there are a couple of important points to make. One is that it's still very important to remember that reusing data is not always faster and cheaper. You still have to do your homework, you have to investigate that data, you have to read data, you have to explore not only the context that you're given but try to find the additional context, you still have to pay attention to ethical issues; some of them will have been somewhat addressed, but you may well have to address them again or think through them also. So anybody who's thinking that somehow using existing data is the budget model should think again, 'cause I'm not positive that that is necessarily the answer.

I also, of course, do not want to see the availability of existing data used as a justification for cutting primary data collection. We need both, and that is not going to be an easy argument, I know, to push in the near future, but nonetheless they cannot be pitted against each other; they're very complementary and they need to continue to be done that way.