

Libby Bishop
University of Essex

RE-USING QUALITATIVE DATA: A LITTLE EVIDENCE, ON-GOING ISSUES AND MODEST REFLECTIONS¹

This paper uses the occasion of the founding of the Archives of Qualitative Data at the Institute of Philosophy and Sociology in the Polish Academy of Sciences to consider the subject of re-using qualitative data. Drawing on limited, but still useful research about the UK Data Archive, I will consider what we know about how archived qualitative data are being used. Then I will briefly review several methodological and ethical debates that continue to be vigorously discussed. I close the paper with some reflections on where we have been, some lessons that might be useful for new archives, and speculations for the future.

Key words: qualitative data archives; reusing qualitative data; ethical and methodological issues of qualitative data archiving.

Introduction

The year 2014 is significant for the domain of re-using qualitative data. It marks the 20th anniversary of the founding of Qualidata, one of the earliest initiatives to archive qualitative data and make it available for re-use (Corti 2000). That single data collection in one department of one university in the UK multiplied, eventually integrated with the UK Data Archive (UKDA), and now holds over 300 datasets. Twenty years later, another exciting sign of success is the founding of the Archives of Qualitative Data at the Institute of Philosophy and Sociology in the Polish Academy of Sciences. Its inauguration was marked by a conference in Warsaw in 2013.

Welcoming a new member to the family of qualitative data archives affords an opportunity to consider the current situation facing all archives, new and old. First, I will consider what we know about how archived data are being used. Then I will briefly review key issues about re-using qualitative data—methodological and ethical—that continue to be vigorously debated. These issues have been present throughout the history of archiving, yet they also evolve and change, reflecting the context of their times, as is the case with all things qualitative. I will close with reflections and thoughts for the future.

UK Data Archive, e-mail: ebishop@essex.ac.uk

¹ Acknowledgements. This paper is based on a presentation I gave at the Polish Academy of Science on 6 December 2013. I am grateful to Piotr Filipkowski for providing me with the transcript and to Tony Wood for helpful editing.

How is qualitative data being used?

Not a great deal is known about how qualitative data are being re-used. There are several reasons for that: holdings are not that large; scarce resources in smaller centres and repositories are needed for locating, ingesting, and supporting new data; and tracking usage is difficult. It is costly, in time and labour, to obtain detailed information about re-use when such information must be manually collected. The growing practice of data citation and DOIs for data will automate this work (Brase 2014). The information below is based on very preliminary analysis of data generated at UKDA about use of its collections (Bishop 2013). The data consisted of instances of nearly 5000 downloads of qualitative and mixed methods collections from 1994 to 2013. The data show that nearly all (96%) of collections have been used at least once. This finding is noteworthy as it debunks the perception that the majority of data languishes in archives untouched. Even a single re-use, if that produces an insight, significant publication, or policy impact, may be an important contribution. This is not to deny that a very important debate needs to occur about the costs of archiving and how to assess data based on their potential re-use (Harvey 2008).

That said, there is little doubt that the pattern of re-use is highly unequal, with about 70% of collections used fewer than 20 times and only 5% used more than 100 times. In terms of who uses data, nearly two-thirds of downloads are by students and just over a quarter by staff members at educational institutions. Government, third sector and others make up the balance. While we knew that many of our materials were being used by students, this number was a little higher than we predicted; the biggest group are post-graduate users of our materials. Finally, over 60% of uses were for teaching, with the majority of cases in the context of teaching qualitative research methods. This may be explained by the growing diversity of methods in use.

Now, for any archive – a new archive, or a growing archive – it's important to understand what it is that makes a collection a good candidate for re-use. Could we see any patterns in the characteristics of the collections that are being re-used most often? The list of most used collections is highly varied, making patterns hard to assess, but some features stand out. Mixed methods collections and collections with multiple types of data are used, perhaps because of the convenience of learning about several genres of data within one study. Topics tend to be current and to be relevant for policy debates, e.g., crime, retail food sector, and gender. In addition, some subjects that might be seen to appeal to students did well: body image, youth crime, and cannabis. But in addition to specific focused studies, oral histories covering many topics, are also used heavily. (A caveat: these data do not control for how long the collection has been held at UKDA.)

The findings made one fact unambiguous: active promotion of a study, by either a particular faculty member or by the UKDA, dramatically increases the use of its data. This should not surprise, but the implications are not always fully acted upon. Namely, archives need not be passive agents, trying to fathom what users want. They can actively shape those needs and wants, ideally in an interactive and collaborative manner with re-users. Probably the most important thing we learned was from the

methods teachers: they observed that students learn qualitative methods best if they use real data that matches their interests. That is, being engaged with the specific topics the data describe enables students to learn methodology better than if they work with generic or artificial data.

This is a need that data centres are ideally positioned to meet. What archives can offer is a wide range of types of data, from diverse methods (or sub-methods) of qualitative research. Teachers, especially recent Ph.D.s, have the benefit of deep expertise in their subject area, but often, they will not have had the time or opportunity to build a broad range of materials with which they are well acquainted. Archives are perfectly suited to filling that gap.

The UKDA has created several resources to illustrate interview methods using its extensive data collections, and, in particular, to assist instructors who have limited research materials of their own (UKDA 2014b). The first resource provides brief summaries of several different interviewing techniques: semi-structured, unstructured, oral history, psycho-social, etc.; the second resource covers non-interview methods: focus groups, the written word, ethnography, visual methods, and online data collection. Each summary is accompanied by full transcripts or excerpts and the interview schedule (or guidance notes). It concludes with selected references and practical suggestions for how to use the materials for teaching. (Other teaching resources are described in more detail in Bishop (2012) and UKDA (2014b)).

Feedback from various sources has provided other guidance. Teachers have expressed a preference that at least some materials sit outside our registration process. While much data at UKDA are safeguarded and requires registration, some are open and publicly available. To be most useful to teachers (often assembling course material under tight time pressure), we opted to gain permission to make some data publicly available. In addition, we know our teachers also need small data sets, very small collections to work with, and they greatly appreciate teaching activities to go along with data; we've tried to meet some of those requirements (e.g., Haynes 2012).

The challenge in developing these resources is, as always, the time and cost to produce them. But even smaller data centres have used this strategy. The Irish Qualitative Data Archive produced a series of learning resources based on archived data from the Life Histories and Social Change Project. The resources consist of guided introductions to key sociological concepts, such as social class, using audio and text extracts from life history interviews. In a second example, Mass Observation is a key historical resource of popular accounts of everyday life in the UK produced by a panel of recorders who respond to thematic directives (e.g., life under austerity). English teachers in higher and further education use its resources for inspiration for creative writing or performance. All these examples demonstrate that providing archived data to teachers enables innovations in substantive and methods instruction. Although requiring an investment, there are two obvious benefits for new archives: usage numbers for the promoted data will rise, and a next generation of researchers will learn to see re-use of qualitative data as routine, just another part of their methods training. This would be a significant accomplishment.

The methodological issue of context

The debates in the literature about methodology of re-using data, or secondary analysis of qualitative data, have changed considerably over the past twenty years. As recently as ten years ago, there was relatively little published. Now, there is too much to keep up with (Barbour and Eley 2007, Edwards 2012, and Valles, et al. 2011). It is instructive to take just one issue – context—in order to appreciate that while it continues to be discussed and debated, there is clear development and deepening of the arguments through the years. (I have drawn on an earlier article (Bishop 2009) for the main point below.)

The role of qualitative data and its context concerns whether or not the primary researcher has some privileged position to the data by having “been there”, that is, by having unique knowledge of the original context (Bishop 2007; Mauthner et al. 1998; Gillies and Edwards 2005). Critics of re-use claim that it is impossible for another researcher to interpret the original work, or to interpret it differently, because the secondary researcher does not have the context of the original research. He or she was not there in the scene collecting data, and doesn’t have what are sometimes called “head notes” – the tacit experience of being in the face-to-face interaction in the interview or other data collection moment.

In 2007, Moore rebutted with the concept of ‘recontextualisation’ to make clear that all researchers, primary and secondary, engage in contextualisation. While this is true, it risked going too far in the direction of relativism, e.g., a stance that data are constructed, or made, not found. Hammersley (2010) usefully returns to definitions, restricting “data” to what is found or generated, and then using the term “evidence” for that which is reconstituted.

The access to the context does give the primary researcher a distinctive relationship to some of the data as given – the original data experience. But that understanding doesn’t apply to evidence. It does not apply to the analysis, to the interpretation, to the arguments that are derived from that data (Hammersley 2010; Winterton and Irwin 2012). Deriving insights from data depends more on researchers’ analytical capabilities, not merely their proximity to the data.

Primary and secondary research have a somewhat different positioning: primary analysts have a privileged relationship to the data they have generated, but do not necessarily have a privileged claim on the arguments which can be made from that data. With Hammersley (2010) we argue that sociological data will support different theoretical understandings, and researchers’ presence at data generation is not the final arbiter of the adequacy of such understandings (Irwin and Winterton 2012; 2.5).

Something quite fundamental about academic scholarship is lost if we deny that, in the hands of different researchers, data can produce different theoretical understandings.

Moreover, context is always dependent on the actual research question being asked. In many cases the entire context of the original work bears little relevance to the new enquiry if there is an entirely different question posited in the innovative research.

It is not the role of archives to rule in methodological debates, but it is equally naïve to deny their actions may affect those debates. Data archives encourage depositors to provide extensive “contextual” documentation (and re-users find it valuable), but doing so does not imply that the “original” context can—or should be—reproduced (Bishop 2006). At a minimum, anyone providing guidance to researchers depositing qualitative data can expect to be asked questions about context.

The methodological issue of multiple interpretations

The second methodological point I will explore is researchers’ concerns that archived data will be misused (Alderson 1998; Broom et al. 2009). One key fear is that archived data may be used to reach different, even opposing, conclusions from what the primary researcher intended, for reasons other than the context argument given above. But these concerns are not foreign to natural scientists, artists, and others. For historians, for example, the idea of multiple, and even unimaginable uses for data comes more easily.

However intelligent and well thought out his work, it is inconceivable that his will be the only selection of texts that could be made. The information which he brushes aside as irrelevant may be just the thing upon which a future researcher will seize—if he is given the chance. Research can never be a once- and- for- all affair, nor is there ever a single use to which evidence can be put (Samuel 1998: 392).

Some will then respond by claiming that the project of social science is different from art, or natural sciences, or even history, typically citing the interpretivist epistemological assumption that data are not objectively out there, but (inter-) subjectively constructed. If so, then the question remains of how to adjudicate between competing interpretations of co-constructed data. And this issue, as with context and consent, is not a matter of archiving but of epistemology and the nature of academic debate.

In the focus groups facilitated by Broom et al. some researchers described their relationships with data as “‘intuitive’, ‘organic’, ‘intimate’, and ‘personal’ . Ultimately, it was an encoded account only decipherable to the individual who collected it.” (Broom et al. 2009:1170). The logical extension of such a stance does, I suggest, leave researchers on flimsy footing. It is no different from saying ‘trust me, I was there’. This will not do; if a second researcher made an opposing argument, it is unlikely that the first researcher would accept it with such uncritical trust.

What is at issue here is not re-use at all, but scholarly procedures, including sharing data, for assessing validity. Some qualitative researchers find the term validity problematic, but I am not using the term in its narrow, positivistic sense. What I mean is, put simply: why should anyone believe my claims and not those of my counterpart? Researchers have to defend their analyses, interpretations and conclusions, and the presence of archived data may be relevant, but the issue—adjudicating competing claims—exists whether or not data are archived. Further ways to show validity can place equal emphasis on exposing procedures, the paths that led to an interpretation (Mason 2007). Sharing data is part of this process; it is not the

only way, and not suitable in all cases, but it should be the default, barring special circumstances.

The ethical issue of consent

There is vigorous debate about informed consent for all forms of research: medical, social and others (Wiles et al. 2005). Some question whether participants who have read information sheets and signed forms really understand what has been agreed. These are important matters, but not ones I can develop here. I will instead address two questions: do people willingly consent to share data, and what can informed consent mean when future uses of data are unknown?

While there is little data on the first point, what exists speaks volumes. The Finnish Data Archive undertook an ambitious project to contact research participants and ask for re-consent to archive their data. The original researchers had not asked for consent for archiving *because* they believed strongly that their participants would refuse. In total, 169 participants were successfully recontacted and 98% (165) of the original participants agreed to archive their data, even knowing that some of it was very sensitive and personal. Significantly, they gave as reasons for their agreement a general desire to advance science and learning (Kuula 2010/2011). And this is not an isolated case. Projects on sensitive topics such as family relationships, or those affecting known, small communities, have been successfully archived (Henwood 2012, Mort 2006).

However much we may believe we have a certain kind of rapport and relationships with our participants, we, as researchers, are not always the most accurate interpreters of the willingness of our participants to share their data. And as important, participants can, and nearly always do, have many, complex motivations for their actions. This includes participation in research, and it is not surprising that it extends into reasons for sharing data as well. The same respect and autonomy we grant them regarding their decisions to participate, refuse, or withdraw from research projects must apply equally to their decisions about sharing their data. Our role is to inform and advise of risks and benefits, not to assume, even with the best of intentions, that we know what is best for them.

Let me next turn to the second question of the meaning of informed consent. With respect to data archiving, the issue centres on the granting of consent to re-use data. Can consent for future research projects be regarded as “informed consent”? The problem is not unique to data re-use; researchers employing exploratory or emergent methods are unable to give detailed descriptions of research outcomes or procedures at the time when consent is sought. One resolution takes the approach of considering the meaning of “informed”. In such an approach, it is deemed reasonable to provide more general information about the nature of the research and the most likely procedures, as well as making clear the possibility of unforeseen outcomes. At the same time, limitations have to be acknowledged: it may not be possible to specify detailed research questions, and the possibility of interpretations different from, even opposed to, the original findings must be communicated. Similar strategies, called

generic or enduring consent, are growing more common in medical research (Miller 2010). In the case of consent for data archiving, researchers give participants information about the ways data may be used, who typical re-users may be, and the security measures in place protecting their confidentiality. Data archives can also provide resources and advice to researchers to prepare them to hold conversations with participants about consent.

This is an exceptionally challenging area in the area of data archiving now, affecting all data, not only qualitative. Administrative data, for example, is routinely collected without consent. Efforts are underway to explore how such data can be—safely and ethically—re-used for research purposes, including linkages with other data (UKDA 2014a). The challenges are immense—research will be prohibitively expensive or outright impossible should consent be needed for all cases. However, participants have rights regarding confidentiality and privacy that must not be breached. The debate regarding the extent of consent required for data re-use more generally is currently under debate in the European Parliament, with the possibility of greater restrictions and requirements for consent for re-use (European Commission 2014). There are no easy answers.

Reflections for the future

Let me close with some reflections on where we have been, some lessons learned, and speculations for the future. One of the capabilities made possible by institutionally based data archiving and sharing is that it frees depositors from the administrative burdens—which are not trivial—of sharing their data. They delegate and licence another entity, the archive, to attend to the responsibilities of data processing, dissemination and preservation. This has huge advantages, and without doubt, more data, and more high quality data, has been archived and shared from this institutional arrangement. However, there is a trade-off.

This “hands-off” model does not foster close relationships between depositors and re-users. As a consequence of consultations with depositors over the years, formal surveys and focus groups but also many informal conversations with depositors, it has become very clear that depositors want to know how their data is being used. In fact, our depositors can request usage reports at any time, but they don’t often know that. Several years ago, we experimented with “pushing” the reports out, unsolicited. We were relieved to receive no complaints; indeed, there were many gratifying responses such as, “Hello, this is fantastically helpful, thanks very much. I had no idea that my data was being used so much by such a range of people”. In the future, I think archives will increasingly use their position in the network between depositors and re-users to actively promote relationships (where interest is mutual and confidentiality protections can be maintained). This has been successful in some cases of with domain specific repositories, such as the Timescapes repository for qualitative longitudinal data (Neale and Bishop 2012). It is also the direction being taken at the UK Data Archive, where we are in the process of improving support for data producers by developing dedicated producer account webpages.

This focus on fostering relationships will continue to expand beyond the “designated community” users of archives, and this is happening because issues that are central for data centres are, increasing, in the public eye as well. Just a few examples will serve to make the point. The proposal to share medical data for research (care data) in the UK had to be postponed for six months because of a pathetically inadequate information roll-out (Triggle 2014). Core issues included: opt-in vs. opt-out, levels of de-identification of the data, and access to the data. The recent arrest of Gerry Adams in the UK was based, in part, on evidence obtained from interviews done by US researchers who promised their participants confidentiality. The interviews, held at Boston College, had to be surrendered in response to a US Supreme Court ruling (McDonald 2014). Finally, recent disclosures have shown that recommended uses of statins have been based on data that, when finally opened for independent scrutiny, did not support the original findings (Goldacre 2014). Each of the cases is hugely complex, and I am emphatically *not* saying that all are directly comparable to the archiving and re-use of qualitative data, or indeed any data. Issues that we, as researchers and archivists know to be complicated, nuanced, and in need of slow deliberation are also spread across tabloid headlines. We all have a collective responsibility to bring what knowledge, reason and sense we can muster to these debates.

References

- Barbour, Rosalin, and Eley, Susan (eds.). 2007. Refereed special section: reusing qualitative data, *Sociological Research Online* 12(3).
- Bishop, Libby. 2006. A proposal for archiving context for secondary analysis, *Methodological Innovations Online* [Online], 1(2), <http://sirius.soc.plymouth.ac.uk/~andyp/viewarticle.php?id=26>.
- Bishop, Libby. 2007. A Reflexive Account of Reusing Qualitative Data: Beyond primary/secondary dualism. *Sociological Research Online*, Special Section on Reusing Qualitative Data, 12(3) <http://www.socresonline.org.uk/12/3/2.html>.
- Bishop, Libby. 2009. Ethical Sharing and Reuse of Qualitative Data. *Australian Journal of Social Issues*, 44(3) 255–272.
- Bishop, Libby. 2012. Using archived qualitative data for teaching: practical and ethical considerations, *International Journal of Social Research Methodology*, 15(4) [Special Issue: Perspectives on working with archived textual and visual material in social research], pp.341–350. doi: 10.1080/13645579.2012.688335
- Bishop, Libby. 2013. Analyzing qualitative data usage: What can (and cannot) be learned from user registration. International Association for Social Science Information Service and Technology Annual Conference, Cologne, 28 May.
- Brase, Jan (2014) DataCite registers its 3 millionth DOI name! <https://www.datacite.org/node/120>.
- Broom, Alex, Cheshire, Lynda and Michel Emmison. 2009. Qualitative researchers understandings of their practice and the implications for data archiving and sharing, *Sociology*, 43(6), 1163–1180.
- Corti, Louise. 2000. Qualitative Data Archival Resource Centre, University of Essex, UK. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(3). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/1048/2269>.

- Edwards, Ros (ed.). 2012. Perspectives on working with archived textual and visual material in social research, *International Journal of Social Research Methodology* (Special Issue), 15(4) doi:10.1080/13645579.2012.688308
- European Commission (2014) http://europa.eu/rapid/press-release_MEMO-14-186_en.htm.
- Goldacre, Ben 2014. What the Tamiflu saga tells us about drug trials and big pharma. *The Guardian*. 10 April. <http://www.theguardian.com/business/2014/apr/10/tamiflu-saga-drug-trials-big-pharma>.
- Gillies, Val and Ros Edwards. 2005. Secondary analysis in exploring family and social change: addressing the issue of context. *Forum: Qualitative Social Research*, 6(1), Art. 44, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0501444>.
- Hammersley, Martin. 2010. Can we re-use qualitative data via secondary analysis? *Sociological Research Online*, 15(1)5, <http://www.socresonline.org.uk/15/1/5.html>.
- Harvey, R. 2008. Appraisal and Selection. DCC Briefing Papers: Introduction to Curation. Edinburgh: Digital Curation Centre. Handle: 1842/3325. Available online: <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation>.
- Haynes, Jo. 2012. British Migrants in Spain: the Extent and Nature of Social Integration, 2003–2005: Teaching Data [computer file]. O'Reilly, K., [original data producer(s)]. Colchester, Essex: UK Data Archive [distributor], March 2012. SN: 6963.
- Henwood, Karen. (2012) *Masculinities, Identities and Risk: Transition in the Lives of Men and Fathers, 2000–2009* [computer file]. Colchester, Essex: UK Data Archive [distributor], April. SN: 6984.
- Irwin, Sarah and Mandy Winterton. 2012. Qualitative Secondary Analysis and Social Explanation. *Sociological Research Online*, 17(2), <http://www.socresonline.org.uk/17/2/4.html>.
- Kuula, Arja. 2010/2011. Methodological and ethical dilemmas of archiving qualitative data. *IASSIST Quarterly*, 34(3–4) & 35(1–2), 12–17.
- McDonald, Henry. 2014. From IRA militant to Troubles historian, *The Guardian*, 3 May, <http://www.theguardian.com/uk-news/2014/may/03/gerry-adams-ira-anthony-mcintyre-hate-campaign>.
- Mason, Jennifer. 2007. “Re-using” qualitative data: on the merits of an investigative epistemology. *Sociological Research Online*, 12(3), <http://www.socresonline.org.uk/12/3/3.html>.
- Mauthner, Natasha S., Odette Parry, and Kathryn Backett-Milburn. 1998. The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology*, 32(4), 733–745.
- Miller, Franklin. 2010. Consent to clinical research. In F. G. Miller and Alan Wertheimer (Eds.), *The Ethics of Informed Consent* (375–404). Oxford: Oxford University Press.
- Moore, Nlamh. (2007) (Re)Using Qualitative Data? *Sociological Research Online*, 12(3) <http://www.socresonline.org.uk/12/3/1.html>.
- Mort, Maggie. 2006. Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001–2003 [computer file]. Colchester, Essex: UK Data Archive [distributor], November. SN: 5407, <http://dx.doi.org/10.5255/UKDA-SN-5407-1>
- Neale, Bren and Libby Bishop. 2012. The Timescapes Archive: a stakeholder approach to archiving qualitative longitudinal data' *Qualitative Research* 12(1), 53–65.
- Parry, Odette. and Natasha Mauthner. 2004. Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data. *Sociology*, 38(1), 139–152.
- Triggle, Nick. 2014. Giant NHS database roll-out delayed. *BBC News*. 18 Feb. <http://www.bbc.co.uk/news/health-26239532>.
- UKDA (2014a). Administrative Data Service. <http://ukdataservice.ac.uk/news-and-events/new-sitem/?id=3611>.

- UKDA (2014b). Teaching resources: interview and non-interview methods. <http://ukdataservice.ac.uk/use-data/teaching/teaching-resources/qualitative.aspx>.
- Valles, M., L. Corti, and M. Tamboukou (eds.). 2011. 'Qualitative Archives and Biographical Research Methods', *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 12(3), <http://www.qualitative-research.net/index.php/fqs/issue/view/38>.
- Wiles, R., S. Heath, G. Crow, and V. Charles. 2005. Informed consent in social research: a literature review. NCRM Methods Review Paper, 001, Southampton, eprints.ncrm.ac.uk/85/1/MethodsReviewPaperNCRM-001.pdf.

Wtórne wykorzystanie danych jakościowych: nieco aktualne problemy i skromne wnioski

Streszczenie

Korzystając z okazji powstania Archiwum Danych Jakościowych przy IFiS PAN w artykule poruszam problematykę dotyczącą powtórnego wykorzystania danych jakościowych. Opierając się na wciąż jeszcze nielicznych, ale pouczających badaniach na temat brytyjskiego UK Data Archive przedstawiam różne sposoby analizy danych jakościowych przechowywanych w archiwum. Następnie referuję pokrótce najważniejsze, żywo dziś dyskutowane, kwestie metodologiczne i etyczne związane z tą problematyką. Artykuł kończy refleksja nad naszym dotychczasowym doświadczeniem, które może być pomocne przy tworzeniu nowych archiwów oraz zastanowiam się nad przyszłością archiwizacji.

Główne pojęcia: archiwa danych jakościowych; wtórne wykorzystanie danych jakościowych; etyczne i metodologiczne problemy archiwizacji danych jakościowych.