

Getting the most from archived qualitative data: epistemological, practical and professional obstacles

NIGEL FIELDING

Introduction: patterns in the contemporary use of qualitative methods

There are signs that qualitative research is currently enjoying something of an improvement in its fortunes. In several European countries and in North America, qualitative research is increasingly used in applied research and evaluation research, attracting sponsors such as government departments. Qualitative methods appear to have gained increased legitimacy, even in US social science, for long a bastion of quantitative research. Several new journals in the field have been launched in recent years, and events such as the International Sociological Association Research Methodology conferences include an increasing number of sections relating to aspects of qualitative method. In applied qualitative research especially, the popularity of focus group methodology has done much to increase the use and legitimacy of qualitative research. There is also evidence that the emergence of qualitative data analysis software has served to legitimate qualitative research amongst audiences which had customarily entertained reservations about qualitative methods.

The most regular and frequent practitioners of qualitative research are probably found in applied and market research. This is one pattern among several which account for limited progress in secondary analysis of qualitative data. Applied and market researchers are much less likely to archive their research data, for several reasons, including commercial confidentiality and the relative superficiality of some, at least, of the analyses they produce. Another pattern is that qualitative methods increasingly attract practitioners with little social science background but whose work has presented them with a pragmatic requirement to analyse some corpus of qualitative data (Fielding and Lee 2000). While this group may well practise secondary analysis, the results are unlikely to be documented in the literature and their work may proceed without reference to conventional methodological canons or accepted processes of peer review.

These emergent patterns relate to the increasing autonomy of applied social research from its former base in social science disciplines. Among examples cited by Williams is that, increasingly, 'not just social scientists require research training but also G.P.s [medical doctors], nurses, mid-wives and health policy analysts are encouraged to become at least

research literate' (Williams 2000: 160). Researchers and practitioners in fields like health and criminal justice are being called on to use qualitative methods with little background in the field, because of the turn of governments and many social agencies towards 'evidence-based policy' and the increasing legitimacy of qualitative research in such work. This group, and particularly the practitioner-researcher, may not even recognize that the data they have are 'qualitative' but instead regard the data simply as text, requiring no greater skill to analyse than to write an adequate summary. Such views are encountered, for example, among medical doctors who use qualitative techniques to analyse patient records. Researchers with limited backgrounds in qualitative method are unlikely to grasp the criteria of analytic adequacy which customarily apply, or, indeed, their increasingly contested nature.

Archiving and re-using qualitative data sets

Secondary analysis is a well-established practice in quantitative social research. Re-analysis of key data sets informs many academic debates, much policy analysis, and, though largely unpublished, the business decisions of many companies. The same is not true of the secondary analysis of qualitative data. It is a far more modest, indeed, an almost invisible enterprise in social research.

However, commentators on secondary analysis of qualitative data seem to agree that its purposes are similar at the broadest level to those of secondary analysis of quantitative data. These purposes are: to perform additional in-depth analysis; additional analysis of a sub-set of the original data; or to apply a new perspective or a new conceptual focus (Heaton 2000). Work of the last kind seems most frequent, where the original data are re-analysed from a new point of view. Examples include Bloor and Mackintosh (1990), Mauthner *et al.* (1998) and Fielding and Fielding (2000). From the archivist's perspective—with a view to the value of archived qualitative data for future generations, particularly with respect to historians—Qualidata have noted a range of applications. These include 'describing the contemporary and historical attributes and behaviour of individuals, societies, groups or organizations', providing case material for teaching, and methodological development, where researchers' own diaries, logs, memos and notes can offer insight into the process of the fieldwork in a way which is seldom forthcoming from methods textbooks.

Elaborating on the possible uses of secondary analysis of qualitative data, Hammersley (1997) argues that the activity may be useful in evaluating the generalizability of findings from qualitative research by different researchers on similar populations. If so, it would help qualitative research to address one of the charges most frequently made against it, its lack of a cumulative character and the limited generalizability of its findings (or to put it another way, the specificity of its insights). Hammersley takes a broadly positive stance towards the activity, but those with reservations about it are probably in the majority.

A major line of criticism has been epistemological, taking the view that, because the context in which the data were originally produced cannot be recovered, the normal criteria by which qualitative analysis is evaluated cannot be applied. While several writers have put forward this position, we might note particularly the approach of Mauthner *et al.* (1998), since their criticism is based on attempting to conduct secondary analysis of qualitative data from research which they themselves had conducted in the first place. They maintain that, because qualitative data 'are the product of the reflexive relationship between researcher and researched, constrained and informed by biographical, historical, political, theoretical and epistemological contingencies', secondary analysis of archived data is valid only if limited to methodological exploration. Attempts to go beyond this, such as for the purpose of establishing generalizability, or to demonstrate the warrant for an additional analytic theme, are 'incompatible with an interpretive and reflexive epistemology' (Mauthner *et al.* 1998: 743).

Against this position one might argue that, since an essential part of qualitative research has always involved monitoring the effects of reflexivity and taking account of these effects in the analysis, there is no logical incompatibility between assessing the influence of contextual features in primary data analysis or in secondary data analysis. Rather, it is a practical matter. Qualitative researchers have always been in the position of having to weigh the evidence, and often have to deal with incomplete information or speculate about what may have happened if a researcher had not been there. The difficulty is not, therefore, epistemological but practical. Information regarded as vital in providing evidence for a given analytic point may well be missing from the archived data. But that happens in primary data analysis too—the tape runs out 'just when things get interesting', or the respondent withdraws their remark, or the observer leaves the police station just before the suspect gets violent, or any number of other contingencies. One might, and should, expect the professional researcher to respond to such a contingency in exactly the same way regardless of whether the data source is primary or secondary—by saying 'that is too bad but I cannot evidence this point' and moving on to what can be evidenced by the material available. Since one of the attractions of qualitative research is the richness of the data it can produce, this is not such a terrible problem.

Let us now take a brief look at the kinds of context problems with which qualitative researchers already struggle, some of the things that can happen when we're in the field. We may have a change in the focus of the study during fieldwork. It is often said that no PhD turns out precisely as it was proposed. We can get changes as a result of the growing familiarity of ourselves with the setting. We literally become blind to some things that were very prominent when we began. We will have changes from one fieldwork session to another, in the fieldworker's attentiveness. There can be changes in the setting occasioned by external factors, or by internal factors; effects from the exertion of pressures of various kinds by members of the setting on the fieldworker, or similarly, effects from the drift into familiarity and increasing disinterest of research subjects towards the researcher. Such context effects can undoubtedly affect the data.

They all have one, broad outcome—they make the data uneven. A topic which comes up on two occasions may be covered by depth documentation in one case but not in the other. In an interviewing context, two respondents presented with the same question may contrast strongly in the extent and depth of their response. In short, one can hardly argue that, as distinct from secondary analysis, the background information necessary to monitor and take account of such effects is always present in primary data analysis.

Since no one has ever argued that fieldwork can adequately provide data as evidence for *every* potential analytic theme applicable to the data in *primary data analysis*, we can challenge the idea that the context effects make *secondary data analysis* an epistemologically distinct activity.

If the position is accepted that the issue is not an epistemological but a practical one, Hammersley's claim regarding generalizability, along with some other potential advantages, can be realized. For example, secondary analysis may be useful in research on issues which participants find sensitive or where the relevant research population is elusive. Even the most elusive populations sometimes consent to research and secondary analysis enables us to fully exploit those relatively rare cases where researchers do gain access. Further, where research topics are sensitive, perhaps eliciting intense emotional responses, others can be protected from similar upset if the data from existing studies are fully exploited before approaching new participants. As well as protecting the sensitivities of research participants by avoiding the likelihood of their being over-researched, secondary analysis can help subsequent research to position itself so that it is cumulative rather than just repeating the same enquiries made before.

One might identify another virtue of secondary analysis. Primary data analysis is always subject to the problem that researchers will have entered the field and collected their data with particular interests in mind. There are many methodological discussions of the distorting effects this may have. This is probably more often an unwitting process, but this actually makes the problem worse, since the primary researcher may sincerely believe that such processes have not been at work and so may be blind to their effects. We generally regard data as more convincing the less the researcher has had to intervene directly in order to elicit them. For example, a volunteered statement is generally regarded as more reliable than a response to a direct question.

Secondary analysis may have a claim to greater plausibility since it is less likely that the analytic interests employed will have played a part in the interactional field from which the data were derived. There is a parallel here to the practice in qualitative evaluation research where, to overcome affinities developed in the field, the fieldworker hands over the data to a second team member who will carry out the analysis.

There are, then, several respects in which secondary analysis may be a desirable practice in qualitative research.

Since I try to avoid indulging in counsels of perfection it is perhaps time to acknowledge what follow-up researchers may be struggling with if archived material is prepared without secondary analysis in mind. Just short of 30 years ago I carried out a qualitative study of the National Front (NF). It was the first academic study of the NF, although my publisher's libel lawyers ensured it was not the first such study in print (Fielding 1981). As

17

6. Open air meeting, Lincoln's Inn Fields, 4.10 - 5.15

We walked to the Southeast end of the park and stood in the street opposite a row of houses. There was a stand with a banner - "Britons unite with the NATIONAL FRONT". After about 5 minutes I felt a poke in the back and someone said "Stick your hands up." It was Dave McCalden with Christine. He said "Are you here in an objective capacity?" and I said "No, I'm here for my own reasons, I decided to come with the Brighton branch and we've had a lot of fun." He asked if I'd liked the march and I said I had. Christine said she didn't like it and I said I didn't normally enjoy being in a crowd. We discussed the ceremony at the Cenotaph and I commented about the workmen, which David and Christine had also noticed. Christine said they were probably told by their union to make a noise and David said they would have been on double time too. David said had I seen the linked arms of the police in the street before the turning into the park "cause we were near the LSE." I said "Yes I know we're near there" and he said "oh, of course," and smiled. I had been standing with Butler and the flagbearer and several other Brighton branch members; they remained in front of us throughout the meeting and I stood slightly behind with the McCaldens on my left.

Shortly after our comments on proximity to the LSE Martin Webster began the meeting. There was a small amplification system. He said that the NF's opponents had tried to stop the Cenotaph service by a phone hoax where they called in the fire brigade. He said he thought "they could have tried something more original." They had also phoned in a warning that there was "a bomb on one of our coaches" but "this has not stopped us. The police checked and there was nothing." This was the kind of trickery that was used but it couldn't stop us from remembering our glorious dead. He then introduced John Harrison-Broadley, using his military title, as a man who had run 2 distinguished campaigns for the NF and had a distinguished war record.

John Harrison-Broadley spoke with an upper class tone and a chest full of enormous medals. He spoke with an easy assurance. His theme was his wartime experience. He recounted how he had been trained as a young pilot during 1939. "It was summer, the days were long and we lived life to the full." By 1944 all but 2 of the training squadron were dead or P.O.W.s. He himself was a P.O.W. The country had sent its men into battle without weapons, without training but they had fought valiantly. Had anyone told him that these men would be forgotten he wouldn't have thought it possible. In 1962 he'd heard a sermon in which forgiveness was preached and the sentiment was expressed that the enemy ought to be forgotten, without one word being said of the bravery of England's fighters. He hoped this would never happen again. Britons must now fight again, and they must not forget again the glorious dead. (see transcript, Spear 80 p14)

(ret.)

John Harrison-Broadley, Wing Commander, returned the microphone to Martin Webster. Before Webster introduced the next speaker he began raising funds. He began at the figure of £ 50, and decided to up it to £ 50. During this David and Garry did not seem very interested, and Garry made comments ridiculing the meeting. I kept quiet. People gave fairly freely, especially of course towards the lower amounts. Stewards passed buckets around for donations, and people also dropped in change. The amounts went from £50 to £30, £25, £10, £7, £5, £1. This took about 10 minutes though I felt it went slowly.

Figure 1. Fieldnote extract, NF open air meeting, Lincoln's Inn Fields, 1974.

well as interviews and analysis of party documents, the study involved a very uneasy mix of covert and overt observation. That uneasiness is very apparent in the first paragraph of figure 1.

Let us imagine that we are a sociologically-informed alien with no knowledge at all of the original context. What might our alien glean from this page of data? It might first think that this page is rather a mess. But it might also notice that quite a bit of the context is tracked in at-the-time coding and other aspects of context are clear from the look of the page. For example of the latter, it is clear that I had a problem getting the names of my contacts right. I had to go through the data correcting the name of my key informant's partner.

What about the hieroglyphics down the left margin? Sharp-eyed readers will notice that one of them says ‘My role’—some of the data is about me as a researcher—in this instance, what I did when my key informant, Dave, unexpectedly encountered me with the group holding the branch banner of the Brighton & Hove NF, when he had formerly known me only as someone who had been interviewing headquarters staff. I doubt that our alien would have much difficulty de-coding the sense of the data coded ‘My role’ here. Assuming the alien has sharp eyesight, it will notice several other codes relating to ‘My role’: personal relations, concealed identity, and *entrée*.

It isn’t hard to link the code labels to the text. I am not making a case that my coding is especially accessible. Since Armstrong *et al.*’s important study (1997) of inter-rater reliability in qualitative coding, we know that most experienced qualitative researchers code data in rather similar ways. Armstrong looked at a group of medical sociologists, given the same piece of data and asked to code it. They had no communication with each other, but they were in the same field, and they came up with substantially similar codes. So I do not have to make a case about the special accessibility of my coding, to say that most people eyeballing the data will come up with some broadly similar themes.

The other codes down the margins are substantive and relate to the analysis I was doing. Since I was developing a rather stock interactionist organizational analysis, many of the themes could be predicted and checked against my book. Again, these would be accessible to follow-up researchers. The codes used in the study included: NF as an organization; relations with other organizations; overt action; security; violence at public activities; hoaxes; the NF and the military establishment; nationalism and ideology of members; getting funds; the NF as a bureaucracy.

I would not argue that these are the only themes our alien should recognize in these data. I would argue, however, that the fit between the data and the codes is quite close and works in some predictable ways. Equally there are themes in the data that I did not develop but our alien might notice. Of course it would be easier for our alien if I had prepared the data with secondary analysis in mind from the outset, but the context effects do not rule out re-analysis *per se*.

Mauthner *et al.*’s case against secondary analysis was that we normally evaluate qualitative analysis by reference to the context in which the data were originally produced. That approach to evaluation has been widely challenged in recent years. The early criteria of validity were indeed oriented to data collection issues, such as how long one had spent in the field, but latterly other criteria have emerged, some of which do not relate to context, such as empowerment and what Denzin and Lincoln (1995) call ‘transferability’, ‘credibility’ and ‘dependability’. We will never find a universally accepted and incontrovertible criterion of validity, but that does not mean that our field has no standards of quality at all. The field judges quality not by embracing standardized criteria of validity, but by critical peer review, liberally-interpreted community standards of analytic adequacy, and learning from experience. Problems of context *are* material to such judgements but are not the sole means to evaluate an analysis and much depends on exactly what analytic claims are based on the data.

Getting the benefit of secondary analysis

If the debate over epistemological issues relating to secondary analysis tells us anything, it is that it is very important that archived materials include as much information about the context of the original data collection as possible. Because secondary analysis of qualitative data is complicated by the contextual issue, contemporary qualitative researchers need to design their research with archiving in mind from the outset. Since our previous patterns of professional practice lead us to associate the archiving of data only with the eminent, not to mention the deceased, it may be that we need something of a change in our own culture to accept that someone else may later take an interest in our work. It is important that research design, instrument design and fieldwork decisions are fully reported. It is helpful, but not essential, to include material that gives some insight into the way the original analysis was done. But Qualidata have found that archive users most often want 'raw data', such as interview transcripts, rather than already-coded data sets.

Apart from the benefits of secondary analysis already mentioned, a further benefit arises from the archiving process itself. In order to archive material, the data set has to be kept in an organized way. This may well be useful for the original researchers, who may want to re-use the data set later. Another benefit is that, for archived material to be useful, it has to meet minimum standards of access: tape recordings have to be clearly audible, documents have to be legible, and there should not be large gaps where important material is missing. A third benefit is a matter of increasing importance, where qualitative researchers may have been rather lax in the past: obtaining research subjects' consent to participate in a way that meets legal standards on copyright and confidentiality.

Let me suggest a dream scenario for secondary qualitative data analysis. A website exists which holds an index of every substantial qualitative data set publicly available in a particular country. By clicking on a link to the data set of interest, one can go to a repository site and download the data. The data set is organized and formatted in such a way that it can be imported into a qualitative software package. This is not an impossible dream, and the work to make it generally possible could be done now (and in some cases, it has been done), if we were to accept some modest change in established research practices.

Laws protecting the interests of research subjects, and researchers' efforts to honour commitments made to research subjects, do present substantial obstacles to the free transfer of qualitative data sets via the Internet. In most cases, researchers place access restrictions on archived data. By providing a new technology of access, the Internet has exposed an interesting conflict in the archival mission. One might suppose that archives are all about access and are uniformly delighted to have their holdings used. But archives also have duties to depositors and worry about relinquishing carefully-negotiated control over the use of material. My naive assumptions about access were brought home to me when I followed up a press report that a new website offered access to rare illuminated manuscripts of medieval musical scores. Apart from a single tantalizing image, the site

consisted almost entirely of legal documents prescribing and justifying a highly restrictive access procedure. A conversation with the archivist revealed that the depositors, largely religious orders and private collectors, were extremely worried about the website, in case the material was copied and published without their oversight (and, perhaps, without a payment), and had responded to the creation of the academically-based website by imposing tighter restrictions than had applied before it had existed. So in that respect, at least, the 'dream scenario' will never come about in a general sense, because there have to be restrictions which will make access rather more than a quick 'double click' process.

'Re-using data' versus 'secondary analysis'

Discipline-based differences in the utility of archival data tend to be glossed over. The discipline of history appears to provide the guiding premises in respect of some archival centres. For historians, the necessity of archiving is particularly acute: without it, there would be no prospect of new insight or analysis which went beyond the existing literature. The situation is not the same in social and behavioural science. Behavioural science, where there is a tradition of verification based on the natural science model, may be somewhat more open to the value of secondary analysis than is sociology. Psychologists embarking on a study routinely seek out tests whose validity and reliability have been established, so there may be less resistance to using 'someone else's data' or research instrument.

Such considerations suggest that perhaps we need to make a distinction between 're-using archival data' and 'secondary data analysis'. The timeframe suggested by the first term is historical and implies that there may be no other data addressing a given issue. Archival data are self-evidently useful from this perspective, because the epistemological and methodological worries that mark sociology's discussions of secondary analysis are very simply countered by the rejoinder that there is no alternative data available. But the term 'secondary analysis of data' suggests that the data still have some currency and can be taken as having contemporary relevance. It then falls prey to a range of doubts over its ontological and epistemological status. Researchers may react to these doubts by concluding that it is safer to carry out an original study and concentrate their efforts on primary data analysis, where they will enjoy the advantages of having the first bite at the cherry. Indeed, one might observe that, for narrow but significant reasons of career advancement, the incentive in most social science disciplines is not to focus on previous knowledge but to document the 'new'. For historians, there is probably no need at all to rehearse the case for secondary analysis, whereas the social sciences have largely grown up with their face set towards the present and their back on the past.

Copyright of International Journal of Social Research Methodology is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.