

The Shared Fate of Two Innovations in Qualitative Methodology: The Relationship of Qualitative Software and Secondary Analysis of Archived Qualitative Data

Nigel Fielding

Key words:
secondary
analysis,
qualitative soft-
ware,
methodological
innovation

Abstract: This article considers the contribution that software to support qualitative data analysis can make in the secondary analysis of qualitative data. The article suggests some benefits of secondary analysis of qualitative data and addresses some of the methodological criticisms that have been made about secondary analysis in qualitative research. The article's focus is largely practical, but it also offers an account of why the apparent advantages of using qualitative software in the secondary analysis of qualitative data have not so far been fully exploited. It does so by reference to the social context of the research environment.

Table of Contents

- [1. Introduction](#)
- [2. Current Patterns in Adoption of Qualitative Software](#)
- [3. Archiving and Re-using Qualitative Data Sets](#)
- [4. The Related Fate of Two Methodological Innovations](#)

[Acknowledgement](#)

[References](#)

[Author](#)

[Citation](#)

1. Introduction

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. [1]

Associated with this trend, and contributing to it, are two developments which provide the focus of this article. One of these is the development of specialist computer software to support the analysis of qualitative data ([CAQDAS-Computer-Supported Qualitative Data Analysis](#)). The other is the development of an infrastructure for the archiving of qualitative materials, with a view to promoting the secondary analysis of qualitative data. This article concerns the relationship between these developments. The focus of the article is practical. It explores the

ways that the two developments can be exploited in the practice of qualitative research. One might expect that, because there are obvious attractions in using qualitative software to conduct secondary analysis of qualitative data, those proficient in the use of the former would feature in the latter activity. One might further expect that, because both qualitative software and the provision of facilities to support secondary analysis of archived qualitative data are relatively new developments, those most open to either innovation would be interested in the other. The article offers an account of why this has so far not proved to be the case. [2]

2. Current Patterns in Adoption of Qualitative Software

Software for qualitative data analysis began to be developed in the early 1980s and by the late 1980s several first generation packages were available. Perhaps a greater challenge than that of writing programs and negotiating the limited capacities of the personal computers then available was that of overcoming the suspicions of a field which had customarily shown limited enthusiasm for information technology and which perhaps contained a stronger streak of anti-technological Luddism than one might expect to find in other fields of social science methodology. Qualitative methods, with their emphasis on context, personal experience, staying close to the data, and their lack of documentation of how one actually goes about data analysis, seemed to be particularly stony ground for the introduction of new software, and qualitative software has indeed taken a long time to get established. In fact, if one were to survey practising qualitative researchers today it would not be surprising to find that such software is still not in general use, if the basis were a simple head-count of those who use it and those who do not. [3]

There is reason to believe this is particularly likely with respect to academic social research, as opposed to applied and market research. The most regular and frequent practitioners of qualitative research are probably found in the latter field, and use of qualitative software (sometimes created in-house and kept to a particular company or group of users) may well be most established amongst those who are the most prolific practitioners of qualitative analysis. This is one pattern among several which accounts for limited progress both in secondary analysis of qualitative data generally and its facilitation by the use of qualitative software in particular. 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 (there is no assertion here in respect of the quality of these analyses, merely a recognition that depth of analysis varies between academic, applied and market research). Although this may be a group of researchers who are particularly likely to adopt qualitative software we cannot look to them to provide an impetus to the development of qualitative software applications for secondary analysis. Another pattern is that qualitative software increasingly attracts users with little social science background but whose work has presented them with a pragmatic requirement to automate the analysis of some corpus of qualitative data (FIELDING & LEE 2000). For reasons I shall discuss, while this group may

well practice secondary analysis of qualitative data, 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. Turning away from applied and market research towards academic research, the current evidence is that of an adoption pattern in academic research where novice researchers are more likely to adopt than established researchers and methodologists, which is again a matter I will discuss further below. [4]

These emergent patterns relate to several developments (ABBOTT 1998). One is the growing application of CAQDAS in the analysis of focus group data in market research and social research. Another is the increasing use of multiple method studies, where an efficient way of analysing qualitative data is necessary to justify the place of qualitative methods in the overall research design. Another factor is internet-based research, where qualitative software is used to analyse downloaded data. But there is a wider trend than these factors alone, that of the increasing autonomy of applied social research from its former base in social science disciplines (WILLIAMS 2000). Among examples cited by Williams is that, increasingly, "not just social scientists require research training but also G.P.s [medical doctors], nurses, midwives and health policy analysts are encouraged to become at least research literate" (WILLIAMS 2000, p.160); these are indeed amongst the non-academic users we increasingly see in CAQDAS training. [5]

Like the secondary analysis of qualitative data, the use of qualitative software is not a mainstream interest among methodologists. Contemporary methodological literature could even be taken as suggesting that academic social scientists regard qualitative software as a separate kind of analysis, to put alongside analytic induction or grounded theory. The authoritative *Handbook of Qualitative Research* (DENZIN & LINCOLN 1994) lists "computer assisted analysis" as a "method of analysis" (table 1.1, p.12) and comments that "faced with large amounts of qualitative materials, the investigator seeks ways of managing and interpreting these documents, and here ... computer-assisted models of analysis may be of use" (DENZIN & LINCOLN 1994, p.14). This characterisation of qualitative software is unsatisfactory both because it exaggerates the coherence of a field which actually provides a variety of types of computer support for qualitative data analysis and because it confuses a technical resource with an analytic approach. Its effect is to sideline qualitative software as a special interest, which contributes to an adoption pattern where novices, e.g., postgraduates, are more likely to adopt than established researchers and methodologists. [6]

If one were to look at this pattern from the perspective of the social organisation of intellectual production one might emphasise that those with best access to publication outlets are those of established reputation and authority. This almost inevitably means they will be senior figures, who were educated in the craft of qualitative analysis before the advent of personal computing. When methodological innovation takes place, there is a danger that such figures may be uninformed and even actively hostile to it. Instances of such views in respect of qualitative software amongst authorities on methodology were documented in

FIELDING and LEE (1998), drawing on the testimony of a number of CAQDAS users. However, while this may characterise the first phase of response to this particular technically-based methodological innovation, we can expect more accurate accounts to emerge in the literature and the increasing acceptance of CAQDAS into the methodology curriculum will eventually produce a generation of academics with a clearer understanding of what CAQDAS is and is not. [7]

While methodologists may be coming to grips with CAQDAS, another pattern of use has appeared which poses its own problems. Qualitative software increasingly attracts users with limited or no social science background but whose work has presented them with a purely practical need to automate the analysis of qualitative data. 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 recognise that the data they have is "qualitative" but instead regard it as text, requiring no greater skill to analyse than to write an adequate summary. Such views are encountered, for example, among medical doctors who wish to use qualitative software to analyse patient records. [8]

In recent years, two significant user groups other than academics have emerged: applied researchers who have a social science background but whose involvement in applied research means that some, even most, of their work is not conducted in a disciplinary framework, and a second group comprising people who do not primarily work in a research role but in some other field of professional practice, such as medicine, who have no background in social science, and whose involvement in research is an adjunct to their normal field of work. Both groups challenge some conventional understandings of legitimate research practice but the second group is especially independent of the normative standards of social research. In the estimation of one of the principal providers of qualitative software training in the UK, the CAQDAS Networking Project, between 15-20% of participants in its qualitative software training programme are currently drawn from the non-academic group and have limited or no social science background (FIELDING & LEE 2000). [9]

These patterns of adoption may also account for the typical mode of use, which tends to exploit data management rather than conceptualising or analytic features (FIELDING & LEE 1998). While applied researchers face few problems justifying acquisition compared to academic adopters—whose supervisors and/or colleagues are often sceptical—applied research often involves tight deadlines and has relatively straightforward analytic requirements. In research with users, applied researchers complained that data entry and setting-up occupied a disproportionate time relative to the analysis the sponsor wanted (FIELDING & LEE 1998). Applied researchers found the pace of their work denied them time to exploit advanced features; sometimes they simply did not have time to code all the data, seriously limiting the kind of analytic work possible. To them, CAQDAS was valuable as an electronic "filing cabinet". A considerable proportion of users,

perhaps as much as 60%, testified to a pattern of use where CAQDAS was chiefly employed for data management and what one user branded "basic analysis" such as "very basic frequency counts". [10]

Under-utilisation of software features is not a worry just because some users are getting less than they could from qualitative software. Users with limited backgrounds in qualitative method are unlikely to grasp the criteria of analytic adequacy which customarily apply, or, indeed, their increasingly contested nature. Their appreciation of qualitative method is largely defined by the software. While academic qualitative research is oriented to the intellectual and processual safeguards which have been developed in the scholarly community, in applied research the CAQDAS user may be the only team member who knows anything about qualitative methods. [11]

There are other potential concerns, too. Automation of code assignment allows blanket re-coding, and in applied or practitioner research there may be especially strong pressures to skimp on careful inspection of each segment before codes are assigned. The complexity of some software means that users may sometimes be unclear about what particular operations have actually done. Neo-quantification of program output, and the provision of features borrowed from quantitative content analysis techniques, may encourage apparently precise numerical analyses which are not in fact justified by the data itself. Most packages can provide counts of "hits" from specified retrievals, and the inexperienced and those subject to time pressure may be tempted to trust the count rather than examine the data segments to check that what has been counted gives an adequate reflection of the data. Experience teaches us that inferences made from counts are often undermined when the data itself is examined. Such experiences alert users to the importance of precise coding and systematic retrieval strategies, but to see this one needs to be aware that there are interpretive principles other than simple counting of things seen as similar. A researcher who encounters conflicts between their initial analysis and retrievals from a coded data set may not know how to handle any contradictions or, where there is time pressure, may decide that there is not time to re-code the data and so will ignore the contradictions. [12]

Users lacking experience in qualitative method may also be more inclined to accept awkward or dubious procedures or to think they must be intrinsic characteristics of qualitative research. Those who learn software in isolation from an appreciation of qualitative method tend to regard the analytic features of their chosen package as 'qualitative analysis' and not be aware that there are other approaches which might, in fact, better suit the requirements of their particular project. [13]

If this analysis of the adoption of qualitative software is right, researchers in applied fields, and those with a need to analyse qualitative data as an adjunct to their professional occupation, join those newly entering academic social science as groups most likely to adopt qualitative software, but have distinctive needs and characteristics which bear implications for their practice of qualitative

methodology. I have gone into some detail to document the trends in adoption of qualitative software because of the implications they bear for the practice of another currently marginal concern, the secondary analysis of qualitative data. These observations provide one thread of an argument that the article will develop, after considering the second main development under discussion, the creation of an infrastructure to enable the secondary analysis of qualitative data. [14]

3. 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 (in a sense, the bulk of historical research involves secondary analysis of qualitative data, but falls on the "other side" of a humanities/social science distinction). [15]

However, the few 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. In HINDS, VOGEL and CLARKE-STEFFEN'S view (1997), these purposes may be to pursue interests distinct from those of the original analysis or to apply other perspectives to the original research issue. HEATON (1998) offers three analytic purposes which take us a little further. These are to perform: additional in-depth analysis; additional analysis of a subset of the original data; or to apply a new perspective or a new conceptual focus. Although examples (outside the current volume!) are sufficiently uncommon as to render generalisation hazardous, work of the last kind seems most frequent, where the original data is re-analysed from a new point of view. Examples include BLOOR and MACINTOSH (1990), MAUTHNER, PARRY and BACKETT-MILBURN (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—CORTI (1998) notes a range of applications. These include "describing the contemporary and historical attributes and behaviour of individuals, societies, groups or organisations", 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. [16]

Elaborating somewhat further 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 this proved to be the case, it would help qualitative research to address one of the charges most frequently made against it by its critics, 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. There may be echoes here of the resistance to qualitative software. If it was right to argue that such resistance was at least in

part a reflection of the qualitative researcher's preference to stay close to the data and to elevate the importance of context in understanding qualitatively-documented social action, we can indeed see an affinity between the suspicion of qualitative software and the arguments that have been put against the secondary analysis of qualitative data. [17]

A major line of criticism has been epistemological, taking the view that, because the context in which the data was originally produced cannot be recovered, the normal criteria with which qualitative analysis is evaluated cannot be applied. While several writers have put forward this position, we might attend particularly closely to the approach of MAUTHNER et al. (1998), since their criticism is based on their attempt to conduct secondary analysis of qualitative data from research which they themselves had conducted in the first place. These authors 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 suggested by HAMMERSLEY (1997), or for the purpose of demonstrating the warrant for an additional analytic theme as in FIELDING and FIELDING (2000), are "incompatible with an interpretive and reflexive epistemology" (MAUTHNER et al. 1998, p.743). [18]

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 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 or been thought or said if a researcher had not been there observing or prompting talk on a topic by conducting an interview. 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 for the toilet just as the arrestee gets violent and the police beat him up, 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. One might suspect that at least some of the resistance to secondary analysis is actually to do with resistance to change and discomfort over the emergence of a new technique for which one's training and experience has not equipped one with the necessary skills. This would be a further and troubling parallel with the case of qualitative software adoption. [19]

If the position is accepted that the issue is not an epistemological but a practical one, both HAMMERSLEY's (1997) claim regarding generalizability and the incremental utility of results from several qualitative studies of the same population, along with some other potential advantages of secondary qualitative analysis, can be realised. For example, secondary analysis may be useful in research concerning issues which research participants find sensitive or where the relevant research population is elusive. Even the most elusive populations sometimes consent to research (see, for example, the field of elite studies; even the eminent and/or powerful sometimes like to hear themselves talk) and secondary analysis enables us to fully exploit data from those rare cases where researchers do gain access to such populations. Further, where the topics the researcher wants to address are sensitive, perhaps eliciting intense emotional responses from research participants, others can be protected from undergoing similar upset if the data from those studies already in existence are fully exploited before approaching new research participants. As well as protecting the sensitivities of research participants (including gatekeepers) by avoiding the likelihood of their being over-researched, secondary analysis can help subsequent research to position itself so that its fieldwork is iterative rather than just repeating the same enquiries which have been made before. [20]

One might also 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 (even though these evolve, as BECKER described in his notion of "sequential analysis", BECKER 1970). There are many methodological discussions of the distorting effects this may have, where the data collected may be oriented to particular analytic purposes. This is probably more often an implicit or 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 one in response to a direct question from the researcher. Secondary analysis may have a legitimate claim to greater plausibility since it is less likely that the analytic interests which are employed will have played a part in the interactional field from which the data were derived. In fact, there is a parallel here to an established practice in qualitative evaluation research. To overcome affinities the fieldworker may have developed in the field, some evaluation research designs involve the fieldworker handing over the data they have collected to a second team member who will carry out the analysis. [21]

It seems, then that there are several respects in which secondary analysis may be a desirable practice in qualitative research, as it is in quantitative research, although it will probably never be the dominant activity in either kind of social research. The case for secondary analysis appears to have gained ground institutionally in several countries recently. Several universities and university-based institutions in North America have moved to create archives for qualitative material. Plans are underway to establish national-level qualitative archives in

Germany. In Britain, the government's funding agency for social science research has supported the [Qualidata Archival Resource Centre](#) at the University of Essex, which acts as an intermediary between potential data depositors and potential data repositories. Each of these developments has taken a somewhat different form, but that is not our concern here. [22]

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. HEATON (1998) offers advice consistent with this. 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 (because 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; after all, even the eminent were not born that way). It is important that research design, instrument design and fieldwork decisions are fully reported. Following the advice given in Qualidata (1995), HEATON (1998) suggests that, in producing a secondary analysis, follow-up researchers should include: an outline of the original study and the data collection procedure; a description of the processes involved in categorising and summarising the data for secondary analysis; and an account of how methodological and ethical considerations were addressed. [23]

We might work backwards from these points to see what kind of documentation we would want to provide for researchers following up a study of our own. In the course of the research reported in FIELDING and FIELDING (2000) we reviewed several archived qualitative data sets. One common problem was the superficial nature of the index provided to describe the material. Indexing tended to be at "box" level; if there were ten boxes, the index would have ten main headings, with few sub-headings. Inside individual boxes it was unusual to find the material organised in any way. It was rather as if one had just two or three file directories into which one assigned every word-processed document one had ever written. Qualidata has observed that the system of box listing is that employed by traditional archivists; Qualidata does not believe this to be adequate and typically indexes qualitative data by interview or observation. For interview-based material, Qualidata's system provides a list of major biographical details, from which the user can pick out interviewees of their choice. Because they are not regarded by Qualidata as primary data, press cuttings are indexed only with a basic heading referring to the topic and time period. As this example suggests, a practical obstacle is that most existing archived material is held in printed form on paper. It is only since the advent of word processors that qualitative data sets have been "machine readable", but word processors have now been in common use for some years and progress towards secondary analysis of qualitative data sets remains modest. There are very few sets of archived data held in electronic form. When I searched amongst British archives for data sets in two fields, crime, and work and organisations, I found just one data set held in electronic form. This means that, to use the material, one has to create photocopies, which inevitably

means a lengthy stay at the archive. Some archived material may well be handwritten rather than typed, too. If one wants to manipulate the data using software, one will have to word-process or scan the original documents. Mention of data manipulation using software brings us to the point where the main concern of this article can be considered. [24]

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 organised and formatted in such a way that it can be imported into a qualitative software package of one's choice. Secondary analysis can then proceed. This is not an impossible dream, and the work to make it generally possible could be done now (in some cases, it has been done). As we shall see, although there are elements of our dream which we will never be able to realise, most of our dream could be a reality now, if we were to accept some modest change in established research practices. [25]

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 organised way. This may well be useful for the original researchers, who may want to re-use the data set later. Let us look at the practicalities involved. Following the advice of Qualidata, machine-readable data sets should be held in ASCII without line breaks on High Density disks which are DOS-compatible and be in the form of "external files" rather than program-specific or "internal files". For images, the prevailing current standard should be used (e.g., TIFF 4). For machine readable audio, the recording should be on recordable CD-ROM; if machine readability is not required, the audio should be held on C60 audiocassettes (which are least likely to stretch or break over time). ASCII is recommended because we cannot forecast future word processor or CAQDAS developments. [26]

To be of most benefit in secondary analysis the archived material would consist of an external file of data in ASCII format and with the identity of the respondent or subject(s) of the material anonymised; a diagram showing the thematic codes applied to the data, and their relationship (e.g., the code hierarchy); a chronological, cumulative "log book" of memos generated in the study; an index and description of files; a file of fully-coded data in the package-specific format used in the original study (if CAQDAS was in fact used); any visual data held in digitised form; a file of scanned documents relating to the setting and which were used in the original analysis; a file in ASCII format offering a basic fieldwork summary (for example, in an interview-based study, this would contain dates of interviews, summary of topics covered, basic descriptive information about the respondents). The ASCII text version of the data set should adopt a common format to be applied at the beginning of each data source (e.g., in an interview-based study, at the beginning of each interview), to act as a Key to Files. For an example, see Figure 1. The data itself should include Interviewer and Respondent

markers in the case of interview data, and should be single-spaced with a single line between paragraphs.

rosie.txt	ascii	40kb	15pp.	Interview no. 1	Word 7
anonymised transcript for respondent Rosie—full					
Female, age 35, from Northwest, lawyer					

Figure 1: "Key to Files" [27]

Although there is only limited experience to go on, Qualidata has found that archive users most often want "raw data", such as interview transcripts, rather than already-coded data sets. The exception is very large data sets, where users request data sorted by codes/themes. An associated issue is how useful it may be to provide data and other documentation in program-specific ("internal files") format. Bearing in mind that archived material may be used decades after the original study it is likely that the qualitative software originally used will have long ago become obsolete. CAQDAS developers are aware of archiving as one of several reasons why a common exportable program format is desirable and some software has moved in this direction but this has yet to be generally put in place. For these reasons the approach using ASCII and providing code lists and thematic schema is the best to follow at present. Overall the implication is that a data set can be of use to others if the minimum requirement for archiving is satisfied, providing the data is in ASCII format. [28]

As noted earlier, archivists also request some documentation about projects. There should be a brief project description; a description of the research design and methodology; a list of publications associated with the project; a record of other sources that were consulted in the project (e.g., of other research projects with which information may have been exchanged); a copy of the research instruments used, such as interview schedules or topic guides; other contextual information, such as correspondence with research sponsors about the findings or, if the project was part of a programme of research, information about other projects in the programme. Also desirable is the transcript of an interview with the depositor about the project, in which, if the depositor was involved with the original research, an account is given of the project. [29]

There is probably some redundancy between these items and, as with the inclusion of coding schemes and coded versions of the data, it is no doubt possible to create a serviceable data set for archiving without including every item listed above. The main point researchers should keep in mind is that the deposited material should provide access to as much of the data as possible in a form that can be re-used. It is helpful, but not essential, to include material that gives some insight into the way the original analysis was done, but this can be achieved by several means—a logbook of analytic memos, the data set with codes applied, the diagram(s) of thematic codes. Since many researchers use memos as a kind of catch-all aide memoire, it may be worth considering writing a "second-level" memo just for archival purposes, but this is one of several ways in

which the eventual archival use of qualitative data sets would have to be designed-into the project from the outset and, to be blunt, most of us have our hands full doing the research without changing our procedure to accommodate later archival use. In other words, the lists above of material to be held electronically, and of the documentation it is useful to have, should probably be regarded as statements of the ideal rather than the norm. [30]

This said, there are several ways in which archiving can be a stimulus to good practice. We have already noted one—that archiving obliges the researcher to impose some degree of organisation on material from their project. Another way 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 way is a matter of increasing importance, where, one suspects, qualitative researchers have been rather lax in the past. This is the matter of obtaining consent from research subjects to participate in the research in such a way that it meets legal standards such as copyright and confidentiality law. Public archives cannot accept material that does not meet these standards, and in the main, these standards are benign. Where they are not, and writing purely in a western European context, it is because legislators have not taken into account the particular character of qualitative research. When this happens, researchers should participate in the legislative process to ensure their objections are taken account of, and if law is adopted which threatens the legitimate practice of independent social research, to lobby for change and provide professional support for researchers caught up in test cases. [31]

Laws protecting the interests of research subjects, and researchers' efforts to honour commitments made to research subjects, do provide substantial obstacles to the free transfer of qualitative data sets via the Internet. In most cases, researchers place access restrictions on archived data sets. Many data sets can only be used by "*bona fide* researchers" and a frequent access condition is that the primary researchers have to be consulted before access is granted. Because of the difficulty of achieving a level of anonymisation which one can be confident will not permit identities to be established and yet does not remove contextual details of likely analytic importance, this obstacle is inevitable. So in that respect, at least, the "dream scenario" suggested earlier will never come about in a general sense, because there have to be restrictions which will make Internet access rather more than a quick double click process. [32]

4. The Related Fate of Two Methodological Innovations

Having sketched-in some logical and practical characteristics of secondary analysis and discussed some contemporary patterns in adoption of qualitative software, we are in a position to explore the link between archived data and the software to be used for the re-analysis. There are many published discussions of data analysis using qualitative software, but does qualitative software have any special merit in a secondary analysis context? [33]

This article has argued that the concern over recovering and taking account of the context in which the data were originally collected is a complicating factor in secondary analysis of qualitative data. There are some important ways in which software can help the secondary analyst to take account of the complexity of the data. The context problems may include changes in the analytic focus of the study during fieldwork; the growing familiarity of the fieldworker with the setting; changes from one fieldwork session to another in the fieldworker's attentiveness; change in the setting occasioned by external factors; change in the setting occasioned 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. These are some of the ways in which context effects can come about and affect the data that are recorded. [34]

Since no one has ever argued that fieldwork can record and/or adequately provide data as evidence for *every* potential analytic theme applicable to the data in *primary data analysis*, we have already been able to discount the idea that the context effects make *secondary data analysis* an epistemologically distinct activity from other kinds of hermeneutic analysis. But the context effects all have one, broad outcome—they make the data uneven. By this I mean that a topic which comes up on two occasions may be covered by depth documentation in one case but not in the other (or, in an interviewing context, that two respondents presented with the same question may contrast strongly in the extent and depth of their response). How might software help here? [35]

One aid provided by software is the ease with which the coding process can be done. Nearly all qualitative software packages now enable users to assign codes by "dragging and dropping". It follows that, because codes are readily assigned, they can (in most cases, depending on package architecture) readily be re-assigned. Many packages provide automated procedures for re-coding. Changes can easily be made to the code which has been assigned, either to a single segment, or to a sub-set, or the complete set, of coded segments. Complexities introduced by context problems are less likely to corrupt the analysis. For example, a researcher working consecutively through the data from 20 interview transcripts, "filling up" a code category with segments which are instances of the code, may encounter in interview 18 a significant variation in the response. This variation may imply the need for a revision of the code. The easier it is to do that, the more likely the researcher is to accommodate the necessary revision. [36]

Perhaps more importantly for many secondary analyses which are directed at applying a new conceptual framework to existing data, the use of qualitative software encourages researchers to test the new analysis they are developing against the complete body of data rather than finding a few instances of data supporting their conceptualisation and focussing thereafter largely on those particular instances. If part of the activity involves weighing-up the evidence for particular interpretations, the content analysis capacities of most qualitative software packages can also help. Most packages provide a means to check the proportion of data to which a given code has been assigned, for example. [37]

We have noted that, for secondary analysis, the availability of the original coding frame can be a boon, as it is in quantitative secondary analysis. One needs to know what questions were asked and how they were coded. The advice from archivists is that documentation should include the original coding frame. What, apart from checking in a verificationist way, might the availability of the coding frame help a CAQDAS-based re-analysis do? One benefit might be an elaboration of the analysis, where, rather like the "system closure" idea of software like NUD*IST, the original and secondary analysis are related together as the basis of a more sophisticated conceptualisation (in "system closure", results of each round of retrievals of coded data are added to the body of project data, creating a "history" of work with the coding scheme). Another benefit might be that elements of the coding frame might be similar to those developed in another study of the same phenomenon, enabling the development of a meta-analysis. [38]

Before closing this point, we might also observe that there may be disciplinary differences between the applications of secondary analysis. Because archivists are, naturally, concerned to maximise the use of the resources they compile, the 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. [39]

One imagines that behavioural science, where there is a highly developed 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 psychometric tests whose validity and reliability have been established, so there may be less resistance to using "someone else's data" or research instrument. While qualitative research in psychology may be rather less informed by the verificationist canon, one can imagine that studies where there is some standardisation of instruments and constructs would offer a useful base on which subsequent studies could build. [40]

These considerations suggest that perhaps we need to make a distinction between "re-using archival data" and "secondary data analysis", too. The timeframe suggested by the former term is historical and bears the connotation that there may be no other data addressing a given issue. Archival data is 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 has some currency and can be taken as having contemporary relevance, or at least some value as the basis of a trend analysis. It then falls prey to a range of doubts over its ontological and epistemological status of the sort discussed earlier. 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. Such considerations especially apply to the younger researchers who, in an academic context, we have already noted, are the group most likely to have expertise in the use of qualitative software. 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. [41]

We have been examining current patterns in respect of two methodological innovations—qualitative software and the secondary analysis of qualitative data—which are at present relatively marginal elements of the qualitative methodology scene. One might assume that there should be a real affinity between the tool—qualitative software—and the technique—secondary analysis. However, as we have seen, there are a number of technical problems that obstruct the application of the tool and the practice of the technique. [42]

In previous research on user experiences with qualitative software it emerged that the actual use made of new research tools and techniques is as much a result of the social context of research as it is a matter of the intrinsic characteristics of the tool or the technique. This analysis can be applied to the use of qualitative software to conduct secondary analysis, too. Some, at least, of the purposes of secondary analysis can best be achieved by the use of qualitative software. For example, CAQDAS enables researchers to handle the large volumes of data associated with meta-analysis or to maintain differently-coded versions of a single data set with a view to comparing and assessing different coding schemes (such as the original coding scheme and alternative schemes). However, at present we have a situation where adoption of qualitative software is most likely to be by applied researchers, who have limited interest in secondary analysis (since sponsors usually want "the latest" knowledge), by practitioner-researchers, who do indeed carry out secondary analysis but are unlikely to attend to the wider utility of their analysis outside their own immediate practical concerns, and by the newest generation of social researchers, who need to establish their understanding of qualitative data analysis (and their reputation) by gaining experience of the whole research process—from fieldwork through to publication in the context of their own empirical study (typically for a doctorate)—before tackling the complexities of secondary data analysis. This last group also faces the problem that research careers are made by "discovering" the "new" rather than by extracting further analytic value from the "old". It seems that while there may be a useful affinity between the tool of qualitative software and the technique of secondary qualitative data analysis, those most likely to have expertise in the former are unlikely to apply it to the latter. [43]

Acknowledgement

With grateful acknowledgement to Louise CORTI, Qualidata for advice throughout and for the use of Figure 1.

References

- Abbott, A. (1998). The causal devolution. *Sociological Methods and Research*, 27(2), 148-81.
- Becker, H. (1970). *Sociological Work*. London: Penguin.
- Bloor, M. & Macintosh, J. (1990). Surveillance and concealment: a comparison of client resistance in therapeutic communities and health visiting. In S. Cunningham-Burley & P. McKeganey (Eds.), *Readings in medical sociology* (pp.159-181). London: Routledge.
- Corti, L. (1998). *Archiving qualitative data from CAQDAS*. Unpublished working paper, Qualidata, University of Essex.
- [Denzin, N.](#) & Lincoln, Y. (1994). *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage.
- [Fielding, N.](#) & Fielding, J. (2000). Resistance and adaptation to criminal identity: using secondary analysis to evaluate classic studies of crime and deviance. Forthcoming in *Sociology*.
- Fielding, N. & Lee, R.M. (1998). *Computer Analysis and Qualitative Research*. London: Sage.
- Fielding, N. & Lee, R.M. (2000) Patterns and potentials in the adoption of qualitative software: the implications of user experiences and software training, *Proceedings of the Fifth International Conference on Social Science Methodology*, International Sociological Association, Universität zu Köln.
- Hammersley, M. (1997). Qualitative data archiving: some reflections on its prospects and problems. *Sociology*, 31(1), 131-42.
- Heaton, J. (1998). Secondary analysis of qualitative data. *Social Research Update*, Autumn issue. Guildford: University of Surrey Institute of Social Research.
- Hinds, P.; Vogel, R. & Clarke-Steffen, L. (1997). The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. *Qualitative Health Research*, 7(3), 408-24.
- Mauthner, N.; Parry, O. & Backett-Milburn, K. (1998). The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology*, 32(4), 733-45.
- QUALIDATA (1995). *Guidelines for Depositing Qualitative Data*.
<http://www.essex.ac.uk/qualidata/forms/index.htm>.
- Williams, M. (2000). Social research—the emergence of a discipline? *International Journal of Social Research Methodology*, 3(2), 157-66.

Author

[Nigel FIELDING](#)

Contact:

Nigel Fielding

Institute of Social Research
University of Surrey
Guildford GU2 5XH
United Kingdom

E-mail: n.fielding@surrey.ac.uk

Citation

Fielding, Nigel (2000). The Shared Fate of Two Innovations in Qualitative Methodology: The Relationship of Qualitative Software and Secondary Analysis of Archived Qualitative Data [43 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(3), Art. 22, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0003224>.

Revised 2/2007