Dealing with Data:
Using NVivo in the Qualitative Data Analysis Process

Elaine Welsh

Abstract: This paper will assess the way in which a qualitative data analysis software package—NVivo—can be used in the data analysis process. Computer assisted qualitative data analysis software (CAQDAS) has been seen as aiding the researcher in her or his search for an accurate and transparent picture of the data whilst also providing an audit of the data analysis process as a whole—something which has often been missing in accounts of qualitative research. This paper will compare manual techniques in the qualitative data analysis of interview transcripts with the use of NVivo. In particular, this paper will consider the difficulties surrounding interrogation of interview transcripts and will assess issues of reliability and validity in the data analysis process. The time investment required in order to make full use of NVivo's tools will also be discussed. It is shown that a combination of both manual and computer assisted methods is likely to achieve the best results. [1]

Table of Contents

1. Approaches to Qualitative Data Analysis

There are many different approaches to qualitative data analysis and these have been widely debated in the social sciences literature (BRYMAN & BURGESS, 1994; COFFEY & ATKINSON, 1996; DEY, 1993; MASON, 1996; MILES & HUBERMAN, 1994; SILVERMAN, 1993; STRAUSS, 1987). For example, MASON (1996, p.54) outlines three possible approaches labelling them "literal", "interpretive", and "reflexive". The first approach is an analysis process that focuses on, for example, the exact use of particular language or grammatical structure. The second approach is concerned with making sense of research participants' accounts, so that the researcher is attempting to interpret their meaning. Finally, the reflexive approach attempts to focus attention on the researcher and her or his contribution to the data creation and analysis process. Whichever of these three possible approaches is taken by researchers they face a choice of using either manual and/or computer assisted methods in their data analysis and this paper will discuss the advantages and disadvantages of both. [1]

MASON (1996) suggests that in practice many researchers would use a combination of the above approaches. To begin with, though, there is usually agreement that most researchers will "organise" the data (MILLER, 2000), which can be done by coding text and breaking it down into more manageable chunks. The point at which this happens differs, and some researchers are almost reluctant to break the data down. For example, MAUTHNER and DOUCET (1998, p.135), in their research into postnatal depression and motherhood respectively, say that "the decision to 'cut up' the transcripts was a difficult moment in our research". In addition, THOMPSON & BARRETT (1997, p.60) suggest that in their approach to qualitative data analysis, known as Summary Oral Reflective Analysis (SORA), the main aim is to retain the context of qualitative data, and this "facilitates actually 'hearing' what the data have to say *rather than* splicing them into arbitrary units before searching for topics, themes or meanings" (emphasis added). [2]

2. Data and Software

The data on which this paper are based are drawn from the author's study of gender and local politics which consisted of both qualitative and quantitative parts[1]. In the qualitative element thirty-five face-to-face interviews were carried out with women councillors elected to local government office in England. Local government in England consists of tiers of local councils each with different decision making powers. The top tier is made up of county and unitary councils which have the most power in local government, the second tier consists of district councils and the third tier is made up of small parish councils. The interviewees in this study were drawn from councils which fell into the top two tiers of local government and mainly from the three main UK political parties[2]. Interviews lasted between 1-3 hours and each interview was recorded and fully transcribed[3]. Six of the thirty five interviews formed a pilot study that was carried out in order to "test" the interview schedule and to gain a feel for the issues which were important to the women councillors. Following the pilot study the data were manually analysed and at the end of this process it was decided that it would be necessary to use a software package for the full study. This decision was initially made on the basis of volume of data, and the possible options were explored by attending day courses on different packages before the decision was made to use Nvivo[4]. Nvivo was chosen over other packages primarily because it was very new at the time and it had therefore addressed some of the earlier problems of other packages—particularly the need in programs like NUD.IST to determine minimum text units in advance of the analysis. Nvivo is relatively simple to use. It

---

1   This paper focuses exclusively on the qualitative part of the study. The quantitative element consisted of a postal questionnaire sent to over 800 men and women local councillors from thirteen local councils in England, a response rate of 71% was achieved.

2   The party breakdown of the interviewees was 10 Conservative, 12 Labour, 11 Liberal Democrat and 2 Green Party.

3   Comments made in this paper about the advantages and disadvantages of manual vs. software methods refer only to data as interview transcripts; other types of qualitative data are beyond the scope of this paper.

4   NVivo is a software package to aid qualitative data analysis designed by QSR. Its full title is NUD.IST Vivo. In this paper where NVivo is referred to it is to the first version of the software.

is possible to import documents directly from a word processing package and code[5] these documents easily on screen. Coding stripes can be made visible in the margins of documents so that the researcher can see, at a glance, which codes have been used where. In addition, it is possible to write memos about particular aspects of documents and link these to relevant pieces of text in different documents. Many social science researchers selecting software do not have the expertise to make informed assessments of the different software choices, thus, decisions made can be based on colleagues' recommendations or on the basis of trying out one package and finding it appropriately user-friendly. In addition the time required to become familiar with the package can be an important part of this decision making process, thus the availability of short courses and a support network are necessary so that the researcher can quickly become proficient in use of the package. [3]

3. Computer Assisted Qualitative Data Analysis

Much has been written about the use of computers in qualitative data analysis with some commentators expressing concern that the software may "guide" researchers in a particular direction (SEIDEL, 1991). Others have commented that using CAQDAS could serve to distance the researcher from the data, encourage quantitative analysis of qualitative data, and create a homogeneity in methods across the social sciences (BARRY, 1998; HINCHLIFFE, CRANG, REIMER & HUDSON, 1997). However proponents of CAQDAS argue that it serves to facilitate an accurate and transparent data analysis process whilst also providing a quick and simple way of counting who said what and when, which in turn, provides a reliable, general picture of the data (MORISON & MOIR, 1998; RICHARDS & RICHARDS, 1994). [4]

Qualitative data analysis software is often thought to be based on grounded theory approaches to data analysis in that theory will emerge from the data, and the software often has "memoing" tools which facilitate theory building from the data. Taking a grounded theory approach to data analysis means allowing the data to "speak for themselves" rather than approaching the data within, for example, existing theoretical frameworks. However, KELLE (1997, p.20) suggests that the manufacturers have jumped on the "grounded theory band-wagon" because it is "an established brand name" and that many researchers claim to be using grounded theory when in fact they are applying a "coding paradigm" which is neither inductive nor deductive, but a mixture of both. Whilst the "memoing" tools in NVivo do push the researcher to draw theory from the data, it is not necessary to follow the grounded theory guidelines when using this software. [5]

---

5   In NVivo data are "coded" at "nodes".

4. Validity and Reliability

Debate on the usefulness of the concepts of validity and reliability in qualitative research has been undertaken for many years (KELLE & LAURIE, 1995). Some researchers suggest that whilst these terms are inappropriate in qualitative research, preferring to use terms such as "trustworthiness", "rigorousness", or "quality" of the data, it is nevertheless important that qualitative research and data analysis are carried out in a thorough and transparent manner (CRAWFORD, LEYBOURNE & ARNOTT, 2000; CRESWELL, 1998; KIRK & MILLER, 1986; LINCOLN & GUBA, 1985; MILES & HUBERMAN, 1994; SEALE, 1999). However, in most published research it is unusual to find accounts of exactly how researchers analysed their data and it is partly because of this missing information that this research tradition has been open to allegations of "unthorough" research practices. KIRK & MILLER (1986, p.21) suggest that validity in qualitative research "is ... a question of whether the researcher sees what he or she thinks he or she sees" so that there is evidence in the data for the way in which data are interpreted. Qualitative data analysis has been regarded as akin to "impression analysis" because of the lack of detail and scrutiny on how the analysis process itself is carried out. One of the main benefits of the advent of software in this area is that the practice of qualitative data analysis has been open to debate and scrutiny and many of these methodological issues have been discussed resulting in close consideration of the contribution qualitative research can make across the social sciences. [6]

Using software in the data analysis process has been thought by some to add rigour to qualitative research (RICHARDS & RICHARDS, 1991). One way in which such accuracy could be achieved is by using the search facility in NVivo which is seen by the product designers as one of its main assets facilitating interrogation of the data. This is certainly true when the data are searched in terms of attributes[6], for example, how many women from the Labour party self-identified as feminist? Clearly, carrying out such a search electronically will yield more reliable results than doing it manually simply because human error is ruled out. This kind of interrogation of data is important in terms of gaining an overall impression of the data which has not been unduly influenced by particularly memorable accounts. However, in terms of interrogating text in more detail it is a little more difficult, as BROWN, TAYLOR, BALDY, EDWARDS and OPPENHEIMER (1990, p.136) suggest, "the existence of multiple synonyms would lead to partial retrieval of information", so that although it is possible to search for particular terms, and derivations of that term, the way in which respondents express similar ideas in completely different ways makes it difficult to recover all responses. For example, I tried to search (within the text coded at the node "orgopin" which was about councillors' views of their organisations) for councillors who had expressed negative sentiments about the way in which the committee system worked in their councils. I initially tried a search for the word "showpiece" and two "finds" were returned as only two people, unsurprisingly, actually used this word. However when I carried out a manual search I found

---

6   An attribute is a particular characteristic of the data, for example, age or political party of the interviewee. The researcher can create attributes for any documents in NVivo.

more instances of this kind of attitude, expressed in terms such as "set-piece", "full of people who just want to talk", "male show", and "waste of time". It would have been difficult to find other responses by using the search facility because of the different ways this idea was expressed. Thus, whilst the searching facilities in NVivo can add rigour to the analysis process by allowing the researcher to carry out quick and accurate searches of a particular type (the researcher may be reluctant to carry out these searches manually, especially if the data set is large), and can add to the validity of the results by ensuring that all instances of a particular usage are found, this searching needs to be married with manual scrutiny techniques so that the data are in fact thoroughly interrogated. [7]

Once the researcher has data collected together following a search the results are stored as another node by default. If the searching has been done manually and electronically then the data done manually would have to be individually added to the node. For example, all the responses that were negative about the committee system could be coded together and *this new node* could then be interrogated through the search facilities. One of the problems with this, though, is that every time the researcher asks a question of the data and gets a sub-set of data in response it is tempting to re-code this sub-set. The decision to stop coding and sit back and think about possible thematic connections across the data can be made at very different points in the analysis process. Because the electronic coding process is quick (compared to cutting and pasting pieces of text manually) it is possible that more coding will take place in a study which makes use of software than one that uses only manual methods, and it is not necessarily the case that this additional coding contributes much to an understanding of the data. Instead it may make the researcher *feel* as though she or he is being more rigorous and transparent than would be the case using manual methods, and hence data are interpreted more confidently. [8]

Often among qualitative researchers there are two camps, those who feel that software is central to the analysis process and those who feel that is unimportant and in fact can result in the "wrong" kind of analysis taking place. However, in order to achieve the best results it is important that researchers do not reify either electronic or manual methods and instead combine the best features of each. If the data set is relatively small it would be possible to use only manual methods, although the researcher would risk human error in searching for simple information on the whole data set. In their study of qualitative researchers who had used data analysis software, SMITH and HESSE-BIBER (1996) found that it was used mainly as an organising tool. Qualitative data analysis software is designed to carry out administrative tasks of organising the data more efficiently and should therefore be exploited to the full on this basis. For example, it is easier and quicker to code text on screen than it would be to manually cut and paste different pieces of text relevant to a single code onto pieces of paper and then store these in a file. Clearly, in this situation it makes more sense to use dedicated software. The extent to which the software is exploited beyond this basic use is related to the expertise of the analyst. For example, writing memos within the software rather than manually (by, perhaps, writing in a notebook) and linking different pieces of data together through electronic memos can be useful

when building up themes across the data. In order to make sense of these memos though, it is useful to return to "manual" methods. This involves going through coded text as well as memos and making notes on how all of these link together. At this point it is useful to think of the qualitative research project as a rich tapestry. The software is the loom that facilitates the knitting together of the tapestry, but the loom cannot determine the final picture on the tapestry. It can though, through its advanced technology, speed up the process of producing the tapestry and it may also limit the weaver's errors, but for the weaver to succeed in making the tapestry she or he needs to have an overview of what she or he is trying to produce. It is very possible, and quite legitimate, that different researchers would weave different tapestries from the same available material depending on the questions asked of the data. However, they would have to agree on the material they have to begin with. Software programs can be used to explore systematically this basic material creating broad agreement amongst researchers about what is being dealt with. Hence, the quality, rigour and trustworthiness of the research is enhanced. [9]

5. Making Sense of Themes

In order to understand how the different themes knit together to form a whole, it is first necessary to analyse individual themes. Using Nvivo to do this is difficult. Whilst it can be helpful in terms of counting "who said what" within a theme, in order to relate the theme to other ideas it is necessary to consider, for example, the memos written during the analysis process. The model explorer tool in Nvivo is useful at this point for mapping out diagrammatically how the themes relate to each other. However, because it is difficult to show the whole model on screen at once it is easier to do this on, for example, a very large piece of card so that the researcher can view the whole picture and the inter-relationships of the codes at a glance. When considering the memos and coded data together in order to pull out themes across the data I found it useful simply to write a short summary on each node. These summaries included details such as how many women from a particular political party appeared within text coded at each node, for example, how many women from the Conservative party self-identified as feminist? This information was placed alongside relevant memos and using this information, notes were made of possible themes within the nodes. For example, when considering the motivations of the councillors the relevant text from all interviews was coded at "motivations" and a coding report[7] was made of this node. The theme of "credibility" (which was related to the councillors' search for a professional identity) was identified from this node coding report and thus initial ideas were formulated for a discussion in the research of a "political investor" identity among councillors. [10]

The temptation, as mentioned above, when using software in the data analysis process to extend the coding—because it is relatively easy to do so—beyond any real benefit for understanding the data is a difficult issue. For example, once data have been gathered together under descriptive codes and thematic ideas have

---

7 A "node coding report" is all the pieces of text coded at one node (which is the location of coded text) drawn together and, if desired, printed out.

emerged from this process with the data connected together through memos, it is possible to begin coding again, with only thematic codes being applied. The purpose of this stage of analysis is to ensure that the theoretical ideas which have emerged in the first round of coding can be systematically evidenced in the data, thus addressing the validity of the research results and this may be "easier" to see if all data relevant to, for example, "professional identity" are coded electronically together rather than manually highlighted on paper. This would, in turn, make it possible to search this "professional identity" code to find out, for example, how many councillors from a particular political party fell into this category. However, the decision to code data at this stage, once it has been identified manually on paper, will be influenced by the size of the data set. In this research project into women local politicians, it was possible to see these codes easily on paper without electronically re-coding them simply because there were so few women in any one category. Thus, once again we see manual and electronic methods being combined whilst also taking account of the research reality of available time. [11]

6. Conclusion

The searching tools in NVivo allow the researcher to interrogate her or his data at a particular level. This can, in turn, improve the rigour of the analysis process by validating (or not) some of the researcher's own impressions of the data. However, the software is less useful in terms of addressing issues of validity and reliability in the thematic ideas that emerge during the data analysis process and this is due to the fluid and creative way in which these themes emerge. Of course, details can be checked on the content of particular nodes and this could affect the inter-relationships of the thematic ideas, but in terms of searching through the thematic ideas themselves in order to gain a deep understanding of the data, NVivo is less useful simply because of the type of searching it is capable of doing. It is important that researchers recognise the value of both manual and electronic tools in qualitative data analysis and management and do not reify one over the other but instead remain open to, and make use of, the advantages of each. [12]

References

Barry, Christine A. (1998). Choosing Qualitative Data Analysis Software: Atlas/ti and NUD.IST Compared. *Sociological Research Online, 3*(3). Available at: http://www.socresonline.org.uk/socresonline/3/3/4.html.

Brown, D.; Taylor, C.; Baldy, G.; Edwards, G. & Oppenheimer, E. (1990). Computers and QDA—can they help it? A report on a qualitative data analysis programme. *Sociological Review, 38*(1), 134-150.

Bryman, Alan & Burgess, Robert G. (Eds.) (1994). *Analysing Qualitative Data.* London: Routledge.

Coffey, Amanda & Atkinson, Paul (1996). *Making Sense of Qualitative Data.* California: Sage.

Crawford, Ken H.; Leybourne, Marnie L. & Arnott, Allan (2000). How we ensured rigour in a multi-site, multi-discipline, multi-researcher study. *Forum: Qualitative Social Research* [On-line Journal], *1*(1), Art. 12, Available at: http://www.qualitative-research.net/fqs-texte/1-00/1-00crawfordetal-e.htm.

Creswell, John W. (1998). *Qualitative Inquiry and Research Design: Choosing Among Five Traditions.* California: Sage.

Dey, Ian (1993). *Qualitative Data Analysis*. London: Routledge.

Hinchliffe, S.J.; Crang, M.A.; Reimer, S.M. & Hudson, A.C. (1997). Software for qualitative research: 2. Some thought on "aiding" analysis. *Environment and Planning A,* 29, 1109-1124.

Kelle, Udo & Laurie, Heather (1995). Computer Use in Qualitative Research and Issues of Validity In Udo Kelle (Ed.), *Computer-Aided Qualitative Data Analysis*: *Theory, Methods and Practice* (pp.19-28). London: Sage.

Kelle, Udo (1997) Theory Building in Qualitative Research and Computer Programmes for the Management of Textual Data. *Sociological Research Online, 2*(2). Available at: http://www.socresonline.org.uk/socresonline/2/2/1.html.

Kirk, Jerome & Miller, Marc L. (1986). *Reliability and Validity in Qualitative Research.* London: Sage.

Lincoln, Yvonna S. & Guba, Egon G. (1985). *Naturalistic Inquiry.* California: Sage.

Mason, Jennifer (1996). *Qualitative Researching.* London: Sage.

Mauthner, Natasha & Doucet, Andrea (1998). Reflections on a Voice-centred Relational Method: Analysing Maternal and Domestic Voices In Jane Ribbens & Rosalind Edwards (Eds.), *Feminist Dilemmas in Qualitative Research: Public Knowledge and Private Lives* (pp.119-146). London: Sage.

Miles, Matthew B. & Huberman, Michael A. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* (2nd edition). California: Sage.

Miller, Tina (2000). Exploration of First Time Motherhood: Narratives of Transition. Unpublished PhD Thesis, Sociology Department: University of Warwick, UK.

Morrison, Moya & Moir, Jim (1998). The role of computer software in the analysis of qualitative data: efficient clerk, research assistant or Trojan horse? *Journal of Advanced Nursing, 28*(1), 106-116.

Richards, Lyn & Richards, Tom (1991). The Transformation of Qualitative Method: Computational Paradigms and Research Processes. In Nigel G. Fielding, & Raymond M. Lee (Eds.), *Using Computers in Qualitative Research* (pp.38-53)*.* London: Sage.

Richards, Lyn & Richards, Tom (1994). From filing cabinet to computer. In Alan Bryman, & Robert G. Burgess (Eds.), *Analysing Qualitative Data* (pp.146-172). London: Routledge.

Seale, Clive (1999). *The Quality of Qualitative Research*. London: Sage.

Seidel, John (1991). Methods and Madness in the Application of Computer Technology to Qualitative Data Analysis. In Nigel G. Fielding, & Raymond M. Lee (Eds.), *Using Computers in Qualitative Research* (pp.107-116). London: Sage.

Silverman, David (1993). *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction*. London: Sage.

Smith, Beverly A. & Hesse-Biber, Sharlene (1996). Users' Experiences with Qualitative Data Analysis Software: Neither Frankenstein's Monster Nor Muse. *Social Science Computer Review, 14*(4), 423-432.

Strauss, Anselm L. (1987). *Qualitative Analysis for Social Scientists*. Cambridge: Cambridge University Press.

Thompson, Susan M. & Barrett, Penelope A. (1997). Summary Oral Reflective Analysis: Method for Interview Data Analysis in Feminist Qualitative Research. *Advances in Nursing Science*, *20*(2), 55-65.

Author

Dr. *Elaine WELSH*, research interests: sociology of gender; family sociology; qualitative research methodology

Contact:

Elaine Welsh

Department of Sociology and Social Policy
Oxford Brookes University
Headington
Oxford
OX3 0BP, UK

Phone: +44 (0) 1865 483763
Fax: +44 (0) 1865 483937

E-mail: emwelsh@brookes.ac.uk

Citation