

Editorial

Louise Corti

UK Data Archive, University of Essex

Introduction

There is currently a growing interest in the methods, resources and tools for analysing or re-using existing qualitative data. In 1994, Qualidata was set up by the Economic and Social Research Council (ESRC) to provide a tested [noun missing] for archiving and disseminating UK-based researcher's data (ESDS Qualidata, 2006). Some ten years on, we have seen dramatic progress in defining a workable model for archiving, safeguarding and providing access. Additionally, and importantly, we have evidence of an emergence of new culture – secondary analysis of qualitative data. Not only are researchers routinely depositing data for sharing and are requesting access to other's data sources, but the body of literature devoted to debate surrounding the processes and methods is also starting to pile up.

In 2005, the ESRC supported an additional funding scheme to complement the UK Data Archive's national qualitative data service, ESDS Qualidata. The Qualitative Archiving and Data Sharing Scheme (QUADS), running from April 2005 until October 2006 aims to develop and promote innovative methodological approaches to the archiving, sharing, re-use and secondary analysis of qualitative research and data. This stems from ESRC's drive to increase the UK resource of highly skilled researchers, and to fully exploit the distinctive potential offered by qualitative research and data.

The QUADS is small in terms of cost and scale, but is dedicated to the mission of learning more about the sharing, representation and re-use of qualitative data, in all of its disparate shapes and forms. Five small exploratory projects have been funded together with a Co-ordination Role.

QUADS Co-ordination plays a pivotal role in fostering communication and understanding between the ESRC Qualitative Data Archiving and Dissemination Scheme (QUADS) projects. It facilitates promotion and publicity via a web site, printed materials, a discussion list for information communication and exchange, and by hosting forums to engage projects in debate, sharing developments and to show working demonstrators. It engages with stakeholders and gives presentations at key events and encourages publication and dissemination of project findings (QUADS 2006).

As the Scheme Coordinator, one of my primary objectives has been to facilitate communication of the Scheme's efforts to the broader spectrum of qualitative researchers, while appreciating that there exist various

communities of practice with different data needs and methodological approaches to sharing and secondary analysis of qualitative research and data. A range of new models for increasing access to qualitative data resources, and for extending the reach and impact of qualitative studies are being explored – in the hope that it will encourage the broader appreciation and take up of data sharing and re-use. Demonstrating Best Practice in qualitative data sharing and research archiving is also a key aim. Early on in the Scheme, I identified four generic areas across the projects that would benefit from Best Practice Guidance. These are: defining and capturing data context; the challenges of audio-visual archiving; consent, confidentiality and intellectual property rights (IPR) issues; and web and metadata standards.

The debate on **capturing context** typically arises in any mention of re-using qualitative data collected by someone other than the original researcher. **Audio-visual** data present particular description, representation and re-use problems. **Consent, confidentiality and copyright** continue to provide sometimes major challenges for data sharing and these projects are all exploring them in their own ways. Finally, **standards** for building and presenting sustainable data ‘products’ must not be ignored – consistent and prescribed web standards, data description and data mark-up enable rich resource discovery and cross-data usage and comparability.

This collection of papers in this second issue of *Methodological Innovations Online* focuses on the first of these issues – **defining and capturing context** of raw qualitative data – in relation to sharing and re-use rather than original interpretation. The papers arise out of a lively and productive workshop organised in as part of the QUADS Demonstrator scheme in Spring 2006 by ESDS Qualidata. The day fostered an opportunity for QUADS projects and three other groups working with large collections of previously collected qualitative data, to share experiences from work in progress, with a remit of addressing context. In my opinion, the depth of insight arising from these papers has produced some concrete and fairly pragmatic advice on how best to capture context – things that future qualitative researchers should consider when undertaking new studies. The papers reflect the stimulating presentations and debate encountered in this workshop, and are all good and easy reads.

A little bit of backdrop to context

Representation, coverage and context of research data are topics that have given rise to some heated debates within the qualitative data community. The fundamental issues of if and how someone else’s raw data can be used were addressed by the UK qualitative data archive in 1994 by Corti and Thompson (2004, 2006).

The basic argument lies with the belief that qualitative data cannot be used sensibly without the accumulated background knowledge and tacit understanding that the original investigator had acquired – understanding typically not written down formally, but held in the researcher’s head. There have been a couple of vociferous critiques in the literature that consider the act of secondary analysis of qualitative data both impractical and impossible. For example, Mauthner et al (1998) argue that ‘data are the product of the reflexive relationship between the researcher and researched, constrained and informed by biographical, historical political, theoretical and epistemological contingencies’ (1998: 742). The researcher’s own deep engagement in the fieldwork and ongoing reflexivity enhances the raw data gathered and stimulates the formulation of new hypotheses in the field. And, in the process of analysing and coding data, researchers do use their own personal knowledge and experiences as tools to make sense of the material, that cannot be easily be explicated nor documented.

Thus the original context can probably never be completely reconstructed. The complexity, quirks, and lack of adequate documentation of data may thus present difficulties in re-analysis, particularly when no input from the original investigating team is possible (Corti and Wright 2002). The loss of the essential contextual experience of 'being there' and the lack of being able to engage in reflexive interpretation may then be viewed by some as a significant barrier to re-use. Even when revisiting one's own data, the problem of loss of context can apply. Mauthner et al (1998) highlight how their own 'ability to interpret their own data may also decline over time as memories wane; changes in personal situation and new knowledge that they have gained since the primary study may also influence their re-interpretation of the data'.

A pertinent question we must consider then, is whether data can be effectively used by someone who has not been involved in the original study? How much of the jigsaw can be missing yet leave the puzzle still worth attempting?

From my own perspective and from our long-standing mission at Essex to build up a qualitative data sharing infrastructure, through ESDS Qualidata and coordinating the QUADS projects, the loss of context in archived data should not be seen as an insurmountable barrier to re-use. Indeed, there are very common and accepted instances where research data is used in a 'second hand' sense by investigators themselves. For example, principal investigators writing up their final analyses and reports may not have been directly engaged in fieldwork, having employed research staff to collect the data with which they are working. Similarly, those researchers working in teams rely upon sharing their own experiences of fieldwork and its context. In both instances, the analysers or authors must rely on fieldworkers and co-workers documenting detailed notes about the project and communicating them – through text, audio and video. Indeed, documentation of the research process can help recover a degree of context, and whilst it cannot compete with 'being there', field notes, letters and memos documenting the research can serve to help aid the original fieldwork experience. Audio-visual recordings of interviews can also significantly enhance the capacity to re-use data without having actually collected them. Representation of the interview is also significantly affected by the nature or method of transcription. Transcriptions are usually a subjective interpretation of the real-life original and ways of transcribing interviews can vary enormously between disciplines and individuals. While sociologists typically want to capture the words, conversation analysts and socio-linguists are more concerned with documenting the para-linguistic features of speech, such as pauses, laughter, tears and so on.

Thus even from these few examples, we can see that defining how to provide context for raw data to make it more 'usable' is a complex and contentious topic. ESDS Qualidata has spent twelve years working in the area of sharing qualitative data, and has done much to establish informal ways of documenting raw data – an example is providing advice on what research documents to keep, and how to gather contextual documentation, such as field notes, log books, award applications, reports and analyses. In addition to written materials arising out of the project, interviews with depositors have proved to be one of the most effective ways of quickly and succinctly capturing context (see ESDS Qualidata Online 2006).

In terms of maximising the potential of a qualitative dataset for re-use then, the ideal scenario is to retain original audio recordings. If an original interview transcription has been selectively edited, the ability to pursue a new line of enquiry that considered the nuances of for example, hesitancy and embarrassment previously unobserved, would require re-transcription of the data. This illustrates the value of retaining audio-visual material for archival collections of qualitative data.

Before I move on to introduce the contributions, I think that the current social research must broaden and revise its methods. Secondary analysis of qualitative data is quite a recent tradition, and we must look to

extend the usual bag of social science approaches. Grounded theory can be used to uncover patterns and themes in data and biographical methods can be employed to document lives and ideas by verifying personal documentary sources. But the bottom line is that a data 'reuser;' must first assemble and 'verify' the data sources they wish to analyse.

Hammersley notes that 'the data collected by different researchers will be structured by various purposes and conceptions of what is relevant. As a result, users of archives are likely to find that some of the data or information required for their purposes is not available' (Hammersley, 1997, p. 139). Hence re-users of data will need to use their own judgement in assessing the quality of the material. But I have always viewed this scrutinising discipline practice as taken for granted: let us note that the practice of research in other disciplines, such as history, is fully based on the critical interpretation of evidence created or documented by others.

I have always found it fascinating to be in a room where qualitative sociologists and historians are discussing the use of qualitative archives. Unlike the sociologist, the historian will not be daunted by the concept of re-use of material that is unfamiliar to them. Historians have had to deal with the challenges of assessing provenance and veracity for many hundreds of years – take the Dead Sea Scrolls, Testaments and many other critical texts. They appreciate that academic archives typically comprise the cultural and material residues of institutional and intellectual processes, for example, in the development of ideas within leading social science departments. The portions of a research collection that finds its way into an archive may also not represent the original collection in its entirety. Overly sensitive interviews may never have been recorded or transcribed or have been subsequently destroyed or destroyed at the time of fieldwork. In short, material is judged to be worthy of preservation by the originator as well as the archivist. Archives are thus a product of 'sedimentation' over the years - collections may be subject to erosion or fragmentation - by natural (accidental damage or loss) or man-made (selection or disposal policies) causes. Storage space may also have had a big impact on what was initially acquired or kept from a collection that was offered to an archive. It is therefore important for archivists to document, where possible, what data are missing and why. They have a responsibility to help with this part of providing context.

However, social scientists are still far more sceptical of 'other' sources, because many prefer to collect 'new' data instead of using 'old'. We must move beyond this scepticism and provide some practical exemplars on how to assess veracity. These papers provide exemplars of how to go about locating and assessing archival material. They place the onus on the original researchers to document context; the archivist to add formally to this background knowledge; and the re-user to consider context as part of their interpretation processes.

Once located and assessed, a researcher is at liberty to delve into the materials - to evaluate, review and reclassify data, to test out prior hypothesis or to uncover emerging patterns and themes. But, contextual needs for 'raw data' also depend on the particular intended usage: description; comparative research, restudy or follow-up study; augmenting new data collection; re-analysis or secondary analysis; discourse and linguistic analyses; verification; research design and methodological advancement; teaching and learning. Situating data in its context also requires both micro and macro level features to be considered including: how the research question was framed, the research application process, project progress, fieldwork situations, analyses processes and output/publication activities. For instance, when undertaking a replication or restudy, detailed information on sampling procedures, fieldwork approaches and question guides will be essential. Knowing that fieldwork strategies radically tack midway or a number of interviews were destroyed or could not be shared, might affect the way research findings are constructed or a project was finally archived. And in the survey literature there is published evidence to suggest that presence of third parties in an interview can

influence question answering. Providing information about macro factors, such as providing a political chronology might be critical when re-using focussed data from the foot and mouth project. There are many more instances that we can think of where particular aspects of context might matter to a reworking of data. Van de Berg (2005) provides an interesting piece on reanalysing qualitative interviews from different angles – cultural, discourse and linguistic, and the risks of decontextualisation.

While ESDS Qualidata staff and others have published in the area of context within more general papers on sharing data and in edited collections (ESDS publications 2006), these contributions have enabled us to develop these ideas about contextualisation further.

The QUADS contributions: adding insight to capturing context

All of the QUADS projects have, in some form or other, the analysis of context built into their remits. In my role as scheme coordinator I felt it necessary to move towards attempting to devise and recommend a minimum set of contextual constructs that would be necessary to document a collection of qualitative data to enable informed secondary use. This has been met with some criticism, mainly from the challenge that asking researchers to provide **structured** information moves us away from the openness and complexity of qualitative data. I do believe this is the case, and we should not be scared of quantifying some aspects of our research findings. A framework on which to hang study-specific context is useful. The insight the authors offer in the following papers have given some very useful input into consensus on mandatory elements for providing ‘necessary’ but maybe never ‘sufficient’ context.

Libby Bishop’s broadly focused paper starts the proceedings by considering what the objective is of recreating context – we perhaps cannot ‘recreate’ original context but we can ‘recontextualise’ data. She identifies multiple levels or layers of context and the processes of recontextualisation, from conversational context at the interview level to cultural context at the global level. The data collection level is already quite well handled by most archives at the study level - information on sampling, data collection methods and so on. Bishop’s article offers practical advice on how to build up context information at the ‘data unit’ level (e.g. a single interview), such as descriptions of participants and interrelationships. The institutional/cultural level is also important and too rarely taken into account in the archiving process – although ESDS Qualidata does provide chosen UK classic studies with as much published context as possible - newspaper clippings, article and book reviews and so on. Reviews can provide a good barometric reading of the cultural and political climate at the time of the original data collection. Finally, Bishop outlines work being undertaken and promoted by ESDS Qualidata to create a ‘standardised transcript’ with systematic header information. This is useful insofar as it enables a degree of consistency in capturing some basic aspects of context – for example, who is being interviewed by whom and who is speaking. Extensible Mark-up Language (XML) can provide a flexible way of recording this to enable multiple publishing outlets, such as to a word document, to the web, or to an archival format. This leveraging of meta information into constructs must be balanced by the appreciation that a single interview is highly situationally dependent.

Niamh Moore follows this paper by also querying the ‘re’ in ‘reusing data’, by asking ‘how does *reusing* qualitative data differ from *using* qualitative data?’ She suggests that we are obsessed with the context of the original project – yet context is not given but produced, and therefore contextualising is a process. Moore concludes that secondary analysis or re-use can be understood, not so much as the analysis of pre-existing data, but rather as involving a process of re-contextualising, and re-constructing, data. And, the complexities of how data are co-constructed come to light when approaching a new research project - new contexts and ‘data’ emerge through the contemporary production of the relationship between researcher and data. Equally,

attention should be paid to the construction of an archive, with researchers having debate and input into what kind of social realities we should be creating and what kinds of records should be generated.

Bella Dicks et al.'s paper focuses on context and hypermedia ethnography. With multi-media ethnographic data in mind, they discuss how it is often difficult to distinguish between data and context, and highlight some of the inherent problems in the notion of archiving ethnographic context. They describe the non-linear paths that ethnography typically takes and the resulting data that can be 'messy' with very rich description. They outline different kinds of contextual information that might be necessary to interpret data in different media forms. The originators of data and re-users have qualitatively different kinds of knowledge-bases, as data and the data-records presented in an archive are different. They propose that communicating context to re-users can be done using hypermedia representation (hypertext and hyperlinking) as a key contextualising tool. Appropriate metadata should also be attached to each 'data record' to aid capture of original complex context, for example, consent information, methods and facts about the 'activity' recorded. Like Bishop, Dicks et al. highlight the usefulness of XML as a standard language in facilitating common descriptions of diverse data sets. Set against this positive stance, they warn against any tendency to mould research data into regular bite sized chunks or to burden researchers with excessive documentation tasks.

Sheila Henderson et al's paper is based on their project on archiving, representing and sharing ten years of longitudinal qualitative data from The Inventing Adulthood study based at London South Bank University. The data collected from the young people is sizable and comprises a variety of multi media sources gathered over time. In exploring some of the more creative ways of overcoming ethical and practical problems involved in providing access to this dataset, the project has gathered contextual information not only from their *own* fieldnotes and reflections but also, innovatively, from those of the young participants. The paper suggests that research [such as?] historical and biographical timelines are important factors in helping frame and understand the complexity of the project and resulting data.

Robert Miller and Peter Macloughlin's aptly titled paper, 'Whatever you say, say nothing' takes us through the number of contextual concerns that, like Bishop's contribution, consider the micro-macro question, but also introduce the idea of moral context. They neatly classify context into basic questions that can be associated with any systematic interrogation of data – who, when, where and how the data has been collected. In the context of research on the Northern Ireland troubles, who did the research and how they gained access to a research situation or site can be political charged matters – the paper takes the theme of policing and the RUC. Time is an important factor, and a chronology of the NI conflict providing information about how the contemporary social/political climate affected research is critical. Capturing geographical context and the intricacies of methodological strategy is also beneficial. However, as the authors note with caution, recording 'cultural' context (for example a researcher's identity, religious affiliation, ethnicity, political leaning) is highly contentious and unethical - it may mean applying labels to both researchers and participants that the individual does not agree with.

Tanya Evans and Pat Thane consider issues of context in their detailed study of unmarried motherhood in England and Wales between the First World War and the mid-1990s. Using data from interviews with unmarried mothers carried out by Dennis Marsden in the 1960s, held at ESDS Qualidata, and other historical sources they aim to increase understanding of recent changes in demography and family structure, particularly the growth since the 1970s of unmarried parenthood. Older norms of serial partnerships, complex families, and late marriage ages, are explored though looking at various legal and cultural contexts across the various historical periods. The qualitative data captured in the mid-1960s study were reconstructed from notes and memory after the interview and contain subjective observations by the interviewer about the interviewees that

now days might be construed as controversial. How should such observations be interpreted some forty years later? Taking a social historian's perspective, the authors conclude that historical, cultural and political context should be taken in account when re-using older research material. The nature of this paper authored by social historians confirms my earlier argument about how assessing provenance is a practice that historians really do take for granted.

The final contribution by Maggie Mort and Cathy Bailey is a fascinating paper that tells of their experiences of archiving a large, sensitive, mixed method data set based on extensive qualitative research of the foot and mouth crisis of 2000. Their research aimed to understand the health and social consequences of the early twenty-first century UK foot and mouth disease epidemic for a rural population. Rich longitudinal qualitative data were collected from North Cumbria, the worst affected area in Britain, documenting the distress experienced across diverse groups well beyond the farming community. Through adopting a collaborative and democratic process for enabling the deposit of these data, they uncover some pertinent context issues that arise. The authors suggest that sensitivity in the data lies not just in the stories that the data conveys but also in the context of the storytelling. Government handling and public perception of the epidemic gained much press – some of it quite heated and controversial. In seeking to establish the key question of how an 'outsider' might make sense of the particular context of a data set, Mort et al. note that shared lexicons and shared linguistic terms, e.g. local agricultural terms around local farming and rural language can add additional meaning to the data, which might not be picked up by the secondary user. They discuss ways and means of sharing this language with the re-user. Equally, the authors point out that retrospective telling though in-depth? interviews and long-term diary-keeping needs to be understood in the context of shared experiences. Shared language may be used as a code and may only make sense to those who also share the experiences. Equally, narratives told may be an attempt to find meaning, to organise and make sense of extraordinary and often traumatic events. The contextual issues arising here are thus complex.

Conclusion

The papers arising out of the cutting edge QUADS demonstrator projects presented in this issue afford unique case studies and help elucidate models of sharing, archiving and re-using data that we can look to in the future. We have better insight into the definition and meaning of context as it applies to different kinds of research strategies and research settings.

For those who collect data, you will see from reading these papers that thinking about context at every key step of the research process is beneficial. Stepping back to provide a bird's eye view of the research and the data being gathered should be done – both during fieldwork and afterwards. It is even more helpful if this scrutinizing view can be captured verbally in a final report. I have long argued that research grant holders who have been awarded public funds to support fieldwork-based investigations should be asked to report more fully and explicitly on the pathways their research took. We know researchers rarely sail directly from Plymouth to Massachusetts but have to navigate choppy waters, may capsize, suffer seasickness, and succumb to mutiny or pirate attacks. Secondary users want to know about these events - as the conditions under which data are sampled, collected and analysed may have been driven or influenced by the events. Warts and all reporting can add significant context to a piece of research – tell us about research conceptualisation, methodological decisions envisaged and taken, design, sampling, fieldwork, consent, analysis and output, impact of the study and data sharing plans envisaged; give us a glossary of lay language.

Re-users must also play a role in giving something back to the original data – the archives. Feed back anything you have gleaned about context in the course of your reanalysis. Let's make data sharing and archiving an ongoing evolving and participative process rather a one-off static one.

Finally, it is clear that documentation of context can be resource intensive. If every research project constructed an historical, political and biographical chronology, day-by-day research logs, and full metadata - not only would it be incredibly expensive, it would most likely be an unattractive burden to be faced with. The resources required to document work also compete with the capacity to undertake the actual research at hand – limited pennies. Thus, I think we have to be careful about recognising what researchers can realistically do. But we can start with minimal mandatory standards and build up to rich, ideal, gold standard ones. A guide to best practice with advice on capturing context as it relates to different kinds of secondary analysis and data re-use scenarios is currently being drawn out of these findings and collated by ESDS.

Useful web sources

QUADS website, 2006 <http://quads.esds.ac.uk>

ESDS Qualidata, 2006 www.esds.ac.uk/qualidata

ESDS publications, 2006. Publications on data sharing and re-use
www.esds.ac.uk/qualidata/news/publications.asp and
www.esds.ac.uk/qualidata/support/reusearticles.asp

References

Corti, L. and Thompson, P. (2004) 'Secondary Analysis of Archive Data' in C. Seale et al (eds.), *Qualitative Research Practice*, London: Sage Publications.

Corti, L., Witzel, A. and Bishop, L. (2005, January). 'Secondary Analysis of Qualitative Data' *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* [Online Journal], 6(1), <http://qualitative-research.net/fqs/fqs-e/inhalt3-00-e.htm>

Corti, L. (2000, December). 'Progress and Problems of Preserving and Providing Access to Qualitative Data for Social Research - The International Picture of an Emerging Culture'. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* [Online Journal], 1(3).

Thompson, P. and Corti, L. (eds.) (2004) 'Special Issue - Celebrating Classic Sociology: Pioneers of Contemporary British Qualitative Research', *International Journal of Social Research Methodology, Theory and Practice*, Vol 7, No 1, February 2004.

Corti, L. Kluge, S. Mruck, K. & Opitz, D. 'Text . Archive . Re-Analysis', *FQS*, Volume 1, No. 3 – December 2000, URL: <http://qualitative-research.net/fqs/fqs-e/inhalt3-00-e.htm>

Corti, L. and Wright, M. (2002), *Consultants' Report to the Medical Research Council on the MRC Population Data Archiving and Access Project*, UK Data Archive, University of Essex.

ESDS Qualidata Online (2006) www.esds.ac.uk/qualidata/online

Hammersley, M. (1997), 'Qualitative data archiving: some reflections on its prospects and problems', *Sociology*, 31(1): 131-42.

Mauthner, N., Parry, O. and Backett-Milburn, K. (1998), 'The data are out there, or are they? Implications for archiving and revisiting qualitative data', *Sociology*, 32(4), 733-45.

Van den Berg, H (2005), 'Reanalyzing Qualitative Interviews from Different Angles: The Risk of Decontextualization and Other Problems of Sharing Qualitative Data', *Forum: Qualitative Social Research*, 6(1), Art. 30. www.qualitative-research.net/fqs-texte/1-05/05-1-30-e.htm