

Qualitative Archiving and Data Sharing: Extending the reach and impact of qualitative data

Introduction

Archived qualitative data are a rich and unique, yet too often unexploited, source of research material. They offer information that can be reanalysed, reworked, and compared with contemporary data. In time, too, archived research materials can prove to be a significant part of our cultural heritage and become resources for historical as well as contemporary research.

But while there is a well-established tradition in social science of reanalysing quantitative data, there is not yet a well developed paradigm, nor a pervasive research culture of sharing or secondary analysis of qualitative data. The lack of discussion in the current literature on the benefits and limitations of such approaches is evident. In the UK, Finland and France some interesting debate has begun, but it is still early days. Further internationalisation of some of the arguments would certainly be welcomed, and I would encourage IASSIST members who are considering archiving qualitative data to consider kick starting some written discussion in their own countries¹.

This contribution provides an overview of some of the perceived barriers to re-use and highlights some of the positive pragmatic measures that are being taken to enable both sharing and re-use of qualitative data. I draw on the recent experiences of ESDS Qualidata and a new research council funded programme intended to investigate innovative ways of extending the reach and impact of qualitative data.

A brief history

Readers will be well aware that research data archiving has been around for some years. The data archiving movement began in the 1960s within a number of key social science departments in the United States who stored original data of survey interviews. The movement spread across Europe and in 1967 a UK data archive (UKDA) was established by the UK Social Science Research Council (SSRC). But the emphasis was strictly quantitative. The SSRC's successor, the ESRC introduced a formalised Datasets Policy in 1996 that contracted all award holders to submit data for possible accession to the UK Data Archive. The word 'data'; was typically, and perhaps conveniently, taken by researchers to refer purely to numeric data.

by Louise Corti¹

Pressure from a small insistent minority fought for qualitative data to be explicitly embedded into the ESRC portfolio of data resources. The Qualidata Centre set up in 1994 in the Sociology department at Essex complemented the UK Data Archive with a joint mission to actively acquire, curate, disseminate and promote the raw data from social science research.

From 2003, Qualidata became an integral part of a larger joined up one-stop-shop for data sharing, archiving and dissemination, under the ESRC/JISC supported Economic and Social Data Service (ESDS).

Over the past ten years ESDS Qualidata has contributed to the elucidation of some of the key perceived barriers to re-using data - through extensive contact with 2000 or more qualitative researchers and through the experiences of handing many disparate data collections. These arguments have been rehearsed in a number of publications by Qualidata staff (e.g. Corti and Thompson, 2004; Corti, L., Witzel, A. and Bishop, L. 2005; Bishop 2005; Corti 2000).

While ESDS Qualidata has conquered some of the 'mainstream' methods for archiving and sharing qualitative data, it has made significant efforts to spark more general academic debate. However, there is still a significant under-use of archived qualitative data when compared with survey data. Moreover, there is still a noticeable imbalance in attitudes towards sharing and re-using data across disciplines and types of methodological approaches.

The key research issues facing re-use of qualitative data

There are some insistent voices who suggest there is a widespread reluctance to deposit qualitative data with a research archive. While this was partially true some ten years ago, today we see a new generation of qualitative researchers who are more inclined to either embrace or gracefully accept the ESRC's Datasets Policy and its efforts to promote the value of sharing data (ESRC 2005). At the UK Data Archive, where some 150 qualitative datasets are catalogued, user figures have soared, particularly for use in research methods teaching.

Nevertheless, there are still barriers. The six key main perceived barriers that have been identified through contact

with researchers in the UK over past ten years can be summarised as follows:

- The practice of secondary analysis of qualitative data is not yet a common place research activity. The literature is not forthcoming on methodological guidance on how to approach the revisiting of data. Corti and Thompson (2004) provide the first inclusive and state-of-the-art chapter on the topic invited for a high profile methods reader by Seale et al. Progress is also hindered by preconceptions and sometimes less than innovative approaches to qualitative research. A cultural shift is required and we believe that this has been progressively happening since 1994.
- Problems of the implicit nature of qualitative data collection and analysis, of context and reflexivity, which are sometimes proclaimed to be indefinable. What are needed here then are practical strategies. Indeed, for research conducted in teams, data and fieldwork experiences are commonly shared, and for Principal Investigators who remain one step away from the field, it is imperative that they rely on their research staff on the ground to capture, document and communicate the nuances of the research process. It is vital to capture better and more systematically the context and the interrelationships among data and between data and other academic products, like analyses and write ups.
- Lack of time to get fully acquainted with research materials created by someone else. Social historians have been more forthcoming in revisiting data sources because of their willingness to embrace the slow and rigorous but commonly accepted practice of document analysis and the need to evaluate methodically the very sources they are revisiting. However it can be terribly time-consuming to locate suitable data sources, and to locate, for example, paper materials that may reside in traditional archival locations with limited access. New ways and tools that more efficiently expose the content and context of digital data sources need to be developed, in order to reduce such researcher burden.
- Constraints of informed consent. Informed consent is an ethical and legal requirement of the research process. It must be thought through at the time of research proposal planning and writing and be tailored towards the specific research questions and the sample. Often consent is not addressed until late in the research process by many researchers, and verbal consent alone is typically no sufficient for longer-term sharing and for effective use of research findings by the original researcher. Failure to realise the need to gain informed consent means that research efforts and the opportunities for archiving and secondary analysis are jeopardised from the start. But researchers require

more guidance on this area to better understand the nature and implications of consent and confidentiality. Additionally, pragmatic strategies are also required to aid the commonly accepted practice of anonymisation or pseudonymisation. Bishop provides a succinct reply to some of the recent scepticism of the possibilities of re-use (Bishop 2005)

- Insecurity about exposure of one's research practice, IPR or threat of misinterpretation. This may be relevant in some specific cases (e.g. an anthropologist's life work), but for the sake of data quality or auditing as Hammersley describes it (Hammersley 1997), exposure of data and methods is no bad thing. Capturing evidence, or specific reasons, as to why data cannot be shared is valuable.
- Finally, lack of a wide range of publicly available catalogued research data. While in the UK, the Economic and Social Data Service (ESDS) has done much to facilitate common resource discovery points of access through the use of standards at the study description level, ways to delve deeper into the qualitative data resource have not been as forthcoming as they have for survey data. The NESSTAR system is a good example of how data can be browsed online through the use of detailed data description down to the survey question (variable) level (Nesstar 2005). As the pool of rich and diverse shareable data expands, the greater the need also for interoperable and standardised description that will allow searching and location of key data across distributed sources. The means of enabling this information stock and flow to reach fruition needs to be investigated and common community methods agreed.

Prerequisites for making data shareable

There are two major issues that appear to be at the heart of making data fully shareable. The first is producing rich and full documentation about the data and the research processes used to conceptualise, collect, manage, process and analyse data. Full documentation enables effective resource discovery (i.e., catalogues) of distributed data sources and enables more informed re-use. The second challenge for sharing data is that of exposing data in the most flexible way possible so as to enable multiple methods of accessibility and innovative uses, for example, combine and link: activities that are the very core of some of the initial considerations of e-social scientists.

Both challenges require that:

- data are collected to a high standard using appropriate sampling strategies, rigorous data gathering methods and, where appropriate, systematic interview transcription

- research methods and practices (including the consent process) are fully documented
- the context of the data collection and analysis is captured
- the richness of the structure and features of data and are made available (use of mark-up)
- the interrelationships between data and analyses (intra-project) are made available (issues of representation)
- data are disseminated in sensitive ways that satisfy the ethical and legal requirements to which they are bound.
- data are represented in appealing and digestible ways, such presenting academic findings alongside evidence from the raw data (that is more than anecdotal quotes)

Enabling these requirements entails practical as well as conceptual challenges. And fundamentally, the underlying need is for the formulation, adoption and community subscription to commonly agreed methods, standards, and ontologies for data description and exposure. Previous work in this area has been spearheaded by ESDS Qualidata and the UK Data Archive in social science archival documentation and data processing for qualitative data (Corti, 2002).

Creatively exploring the barriers and looking forward

ESDS has found that data creation workshops have been very useful in helping unpack some of the specific issues and problems arising in the course of projects that are considering data sharing. The Medical Research Council (MRC) has also taken up an interest in data sharing of primary data from population and clinical trials data, with qualitative data firmly on the agenda. (Corti and Wright 2003). More recently the Natural Environment Research Council (NERC) has also implemented a formalised data management policy for a joint Council programme on Rural Economy and Land Use (RELU) that specifically includes all the qualitative data from the Programme. Context and consent are the two words that crop up most frequently in the debates.

For researchers, like myself, who have been seeking academic funding opportunities to confront such consent, context and technical issues, 2003 – 2004 saw a bumper harvest of such prospects. In the past it has been unusual for ESRC to fund research and development or consider dedicated methodological initiatives. But, over the past three years we have seen a much welcomed move towards dedicated funding for methods. These strands of money have enabled some innovative investigations

to be undertaken, particularly for qualitative data. Five main pots of ESRC funding appeared on the scene, thanks to a number of champions to the cause of methods and data analysis: E-social science; the Research Methods Programme; the National Centre for Social Research; the Qualitative Longitudinal Study and the QUADS scheme.

Innovation: The QUADS scheme



QUADS is the ESRC Qualitative Archiving and Data Sharing Scheme, running from April 2005 until October 2006. The aim of the scheme is to develop and promote innovative methodological approaches to the archiving, sharing, re-use 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 will be explored. The scheme also aims to disseminate good practice in qualitative data sharing and research archiving. This is part of the ESRC's initiative 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 a small initiative (some £500,000 over 18 months) but is dedicated to the mission of learning more about sharing, representation and re-use of qualitative data, in all of its disparate shape and forms. Five small exploratory projects have been funded together with a Co-ordination Role. The Co-ordination team based at ESDS Qualidata have been charged with the task of providing a pivotal role in fostering communication and understanding between the five demonstrator projects. Communication of the Scheme's innovative efforts to the broader spectrum of qualitative researchers is much needed. But equally it is must be appreciated that there exist various communities of practice with different data needs and methodological approaches to sharing and secondary analysis of qualitative research and data. Fruitful collaboration is required which can be achieved through guided discourse to inform and help guide the progress of QUADS demonstrators, and to encourage the broader acceptance and take up of data sharing and re-use.

Key areas for QUADS projects

Four key areas of needs and commonality identified across all the QUADS projects point to: defining and capturing data context, audio-visual archiving; consent, confidentiality and IPR; and web and metadata standards.

The debate on capturing context has been around for some time now on the qualitative data archiving scene. QUADS aims 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. Regarding audio-visual data, they are being handled by many of the projects and the scheme is providing an opportunity to share expertise on presenting and re-using such sources. On the hot topic of consent, confidentiality and copyright, while ESDS Qualidata maintain up-to-date detailed information many the QUADS projects do have specific consent and copyright issues, and it will be invaluable to see how these are confronted by the different projects during the demonstrator period. They will afford unique case studies that can be used in the future.

QUADS Coordination will hold an end of scheme hands-on demonstrator workshop, where projects will be able to talk about their investigations and developments and demonstrate any working QUADS products to an open invitation audience. Furthermore, QUADS Coordination

through its web site will mount papers, tools and training materials arising during the course of the projects in an easy-to-navigate manner. A session at the NCRM Summer School is currently being planned. QUADS is an exciting pilot initiative that is exploring experimental methods. The team is therefore very happy to hear from anyone who is already working in this area or who would like to contribute to these exciting projects.

But what about these standards?

In order to approach primary data now and in the future in years, we need that data to be accurately, richly and contextually described. And in turn, re-presentation of original data, methods and analytic interpretation and their interweaving requires agreed and exemplary standards and procedures. Fielding's scoping study that examined issues for the role of qualitative data in e-social science (Fielding 2003) aptly confirmed that that 'it is timely to anticipate emerging innovations in qualitative methods, including new data forms, sources, possibilities for research archiving and data mining and the potential for increased participation and access'.

Representation can be viewed across a spectrum starting from the simple publishing of anonymised digital qualitative data sources or banks (which are typically not present) through to the ability to link qualitative data to other distributed data sources (e.g. audio-visual or geo-

The screenshot shows the ESDS Qualidata Online search interface. At the top, there is a navigation bar with links for 'Home', 'Data Collections', and 'Explore Multiple Collections'. The main heading is 'ESDS Qualidata Online' with the ESDS logo and 'Economic and Social Data Service' below it. A sidebar on the left contains a menu with items: 'Introduction', 'Interview transcripts Search Browse', 'Interview Summaries', 'Thematic Coding', and 'Multimedia'. The main content area is titled 'Search interview transcripts across multiple data collections'. It includes a search box, a 'Filter by:' dropdown menu, and several search filters: 'Year of Birth', 'Geographical Region', and 'Do you want to search for any (OR) or all (AND) of your terms?'. A 'Submit' button is located at the bottom of the search area. The footer contains logos for WSC, JISC, and other partners.

coded data sources) and to creative and exciting ways of visualizing data. However, it is important to take a step back and see what is currently exposed. The researcher will find very little qualitative data even exposed to the web in any meaningful way. While there are archives of qualitative data to be found across the world, the majority is not even in digital format and “digitizing” these collections is often seen as merely providing an online catalogue of digitised metadata. The issue of how to make these data resources accessible to users has hitherto been a central concern for ESDS Qualidata who has continually been seeking ways to meet users’ requirements.

Standards of relevance are those for: building sustainable web sites; harmonious data descriptions to enable rich resource discovery (metadata); and marking-up data content. ESDS Qualidata recognised the need for standards and tools back in 2000 – tools that allow data to be published to the Web and support online interrogation of data via standard Web browsers. In 2000, Qualidata undertook pioneering work in this area through developing the Qualidata Online system and a methodology for sharing data (Corti and Barker, 2002, 2003).

The need to keep pace with the development numeric data browsing systems, that are now quite far advanced, is important not only for the UK Data Archive but also for other groups who wish to publish and share qualitative data.

Community efforts

QUADS Co-ordination is very aware that there are an increasing number of projects in the world that are looking at sharing qualitative data, typically via the web, and particularly in the wake of the e-science rush. But they are not linked up in any formal way. QUADS Coordination will be building an interactive map will be built to show the location of such initiatives and the key contacts. Additionally training and advice is being given on best practice in metadata creation and web standards for qualitative data. It is hoped that many of the e-science and methods research groups will be amenable to agreeing on some basic sets of standards.

*For more details about ESDS Qualidata see www.esds.ac.uk/qualidata and QUADS see: <http://quads.esds.ac.uk/> or contact Louise Corti, UK Data Archive, University of Essex, Colchester CO4 3SQ corti@essex.ac.uk.

An earlier version of this article appeared in Issue 1 of *Qualiti*, the Newsletter of the National Research Methods Centre, University of Cardiff. www.cardiff.ac.uk/socsi/qualiti/newsletter.html

References

Bishop, L. (2005). ‘Protecting Respondents and Enabling

Data Sharing: Reply to Parry and Mauthner’, *Sociology*, 39(2) pp. 333-336, London: Sage Publications

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). <http://qualitative-research.net/fqs-texte/3-00/3-00corti-e.htm>

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/inhalt1-05-e.htm>

Corti, L. and Barker, E. (2002) ‘Edwardians Online’ *IASSIST QUARTERLY* vol. 26 (2002)No. 4, pp. 5-8

Corti, L. and Barker, E. (2003) ‘Edwardians Online: An XML Application for Qualitative Data’ *ASSIGNation*, Vol 120, No 2, January 2003.

Corti, Louise (2002). *Qualitative Data Processing Guidelines*. Qualidata, UK Data Archive, University of Essex, Colchester

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). <http://qualitative-research.net/fqs-texte/3-00/3-00corti-e.htm>

ESDS www.esds.ac.uk

ESDS Qualidata www.esds.ac.uk/qualidata

ESDS Qualidata Online <http://www.esds.ac.uk/qualidata/online/>

ESRC Datasets Policy www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/Annex%20C_tcm6-9738.pdf

Fielding, N.(2003), *Qualitative Research and e-social science: appraising the potential*, University of Surrey, http://www.ncess.ac.uk/docs/qualitative_research_and_e_soc_sci.pdf

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

NCeSS Hub and Nodes www.ncess.ac.uk

NCRM Hub and Nodes www.ncrm.ac.uk

Nesstar (2005) <http://nesstar.esds.ac.uk/webview>

QUADS <http://quads.esds.ac.uk>

Footnotes

¹ See www.esds.ac.uk/qualidata/access/internationaldata.asp for an overview of progress on national data archives acquiring qualitative data