

Progress and Problems of Preserving and Providing Access to Qualitative Data for Social Research —The International Picture of an Emerging Culture

Louise Corti

Key words:

qualitative data,
data archives,
Qualidata,
International
networking, data
preservation,
secondary
analysis

Abstract: In this paper, I hope to offer a global picture of what is happening in the world of qualitative data archiving. Qualidata is in a strong position to be able to offer this insight as it was the world's first initiative to pioneer preservation of qualitative social science data on a national scale. This was facilitated by the Economic and Social Science Research Council (ESRC), Britain's largest sponsor of social science research, implementing a mandatory policy for research grant holders to offer datasets of all kinds created in the course of their research. The policy has been met with both great support and animosity from the research community. In this paper I examine some of the reasons why the concept of sharing qualitative data generates such mixed feelings.

Qualidata's work has provided sparks of inspiration to a number of research groups across the world beginning to consider the systematic preservation of qualitative data. Over the past four years we have been approached by embryonic "qualidata" projects for advice on issues surrounding archiving and providing access to qualitative data. Many have used Qualidata procedures as a starting point for developing their own archiving procedures (which were devised initially from a cross-fertilisation of UK Data Archive and traditional archiving procedures). Typically these groups tend to be sociologists, and surprisingly have had little or no contact with the social science Data Archives in their own countries. Furthermore, we are still not aware of any other national funders of social research across the world who have realised the added value that archiving of qualitative data can bring.

I hope to provide a quick world tour of progress in the field and then suggest some of the key objectives that I think need to be met in order to achieve a respectable tradition and infrastructure for preserving and re-using qualitative data. I will touch on optimal and cost effective models for qualitative data archiving, discuss issues surrounding the documentation of data, and finally, address the need for meaningful collaboration at the international level, such as by creating a Network for Qualitative Data Archiving (INQUADA).

Table of Contents

- [1. Introduction](#)
- [2. Who is Archiving Qualitative Data and where Might you Typically Find these Data?](#)
- [3. Finding the Right Home for Data](#)
 - [3.1 The rise of electronic data in qualitative research](#)
 - [3.2 Can "traditional" repositories cope with electronic \(non-numerical\) data?](#)
- [4. What about the National Data Archives? Are they Ready to Accept Qualitative Data?](#)
 - [4.1 The UK position](#)
 - [4.2 The rest of the world](#)
 - [4.3 Data archives](#)
- [5. What are the Major Issues Surrounding Adopting a National Archival Policy for Qualitative Data](#)

[6. Researchers vs. Archivists](#)

[6.1 Researchers' worries](#)

[6.2 The data archivists answers](#)

[6.3 Changing attitudes: resignation?](#)

[7. Key Issues for Data Archives Acquiring Qualitative Data](#)

[8. Documenting Qualitative Data](#)

[9. Optimal Models for Qualitative Archiving](#)

[9.1 The distributed network model: Qualidata](#)

[9.2 The UK Qualidata funding situation](#)

[10. Re-use of Data](#)

[10.1 The million dollar question: How are people re-using qualitative data?](#)

[10.2 So, what do users want?](#)

[11. Connecting the World](#)

[12. Conclusion](#)

[References](#)

[Author](#)

[Citation](#)

1. Introduction

In the UK, the ESRC Qualitative Data Archival Resource Centre ([QUALIDATA](#)) is now regarded as one of the international centres of expertise for archiving qualitative data, on a national scale. The Centre has been established for 7 years and, in this lifetime, has had to create a niche for itself within the UK research environment. The Centre has been very fortunate in working within the framework of a national policy for archiving data, and as such offers a pioneering model to other countries. The aims and work of the Centre are discussed fully elsewhere [in this issue](#) and I would ask readers to browse this paper to gain an insight into the remit and operations of Qualidata. [1]

2. Who is Archiving Qualitative Data and where Might you Typically Find these Data?

In the first year of Qualidata's life, we did a great deal of exploratory research to find out where, and to what extent, qualitative data were being kept, stored, preserved and shared in the UK. One of Qualidata's key objectives is to identify both actual and potential sources of qualitative data across the UK and then publicise these sources. Qualidata is concerned with a wide range of qualitative data across the spectrum of social science disciplines—sociology, anthropology, social policy, criminology, political science, education, geography, social psychology, socio-linguistics, and management and business studies. [2]

Our surveys of university libraries, public records offices and museums across the UK suggested that a number of them held collections, ranging from materials

which were small and idiosyncratic in nature, to those which were substantial and coherent sets of raw research materials. Overall, we identified five main concentrations of collections across a range of materials and with a varying accessibility and visibility (see Table 1). The materials include personal papers of professors to sets of interview transcripts and tape recordings.

	Evi- dence of cCollec- tions	Examples of qualitative data	Format and reason for acquiring / storing	Accessibility / willingness to Share	Visibility (e.g. finding aids) and Promotion
Traditional Archives, Libraries	Rare	Personal papers of academics containing raw data, methods and papers; reports on methods, and substantive research, correspondence about research design	Strong research tradition, e.g. anthropology; for posterity; default deposit by retiring academic of the institution	High	Low
Records Offices and Museums	Medium	Thematic collections of interviews e.g. local oral history collections	To create collections based around displays and to add value to collections of artefacts etc.	High	Medium
Research Groups: academics engaged in qualitative research	Medium	Diverse collections of data from past research projects	For own organisational purposes. By default as staff leave	Very Low	Non-existent
Individual Researchers	Common	Solo collections, often based around similar topics, life's works	Personal History and possibility of own re-use	Low - High	Non-existent
Data Archives and Digital Libraries	Rare	Study-based research data accompanying survey data	For sharing and long term preservation	High	Low

Table 1: Qualitative Data Collections [3]

Qualidata's main aim is to increase the number of collections and raise the visibility and access to these disparately housed collections, as summarised in Table 2:

Qualidata	Common	Solo and multiple collections of data (level based on the Investigator)	For sharing and long term preservation	High	High
-----------	--------	---	--	------	------

Table 2: Aims for Qualitative Data Collections [4]

3. Finding the Right Home for Data

One of Qualidata's ongoing objectives is the selection of academic and public repositories suitable and willing to receive qualitative research data. Qualidata staff conducted a programme of visits to key national archives during the first six months of the project. One of our on-going activities is to liase with new repositories with specialist collecting priorities, to meet the needs of new kinds of data we encounter. Meeting with "traditional" archivists raised a number of interesting issues concerning the way in which these professionals view the acquisition and cataloguing of qualitative data collections, and about their relationships with traditional librarians. (I use the word "traditional" here to denote the older established style of archiving, largely concerned with collections of historical papers). [5]

In order to gain credibility with traditional archivists, we appointed, from the start, a professional archivist onto the team. However, it became clear that for the Centre to pursue its mission in dealing with data-holding academics, the needs were for a team of highly trained social scientists with experience of qualitative research. The Centre's approach to cataloguing and dataset description was adopted by a cross-fertilisation of data archiving and traditional archiving procedures. While the two traditions clearly share some similarities, there are also fundamental differences, primarily in the way that multiple collections are described, and also in the way in which, say individual transcripts, are documented. [6]

Repositories who are willing to acquire qualitative deposits from Qualidata include:

Digital repositories

- The Data Archive, University of Essex
- The Oxford Text Archive, University of Oxford

University archival repositories or Special Collections across Britain

- British Library of Political and Economic Science, London School of Economics
- The Modern Records Centre, University of Warwick
- National Social Policy and Social Change Archive, University of Essex

Specialist Institute Libraries

- Institute of Criminology, University of Cambridge
- Contemporary Medical Archives Centre, Wellcome Institute, London
- British Universities Film and Video Council, London
- Institute of Education, London

National Library and Museum Archives

- British Library (Sound Archive and Manuscripts)
- Imperial War Museum, London
- Labour History Archive, Manchester
- Science Museum, London [7]

Each repository has its own identity, mission, collecting policy and research specialisms. Although some institutions had not dealt with qualitative research data before, all were very keen to learn more and begin to acquire this new kind of material.¹ Furthermore, some have taken a proactive interest in working with Qualidata and offer support in helping publicise the value of collections through workshops and publicity materials. However, up to the late 1990s most of the repositories (with the exception of the "Electronic repositories") were only used to dealing with paper-based data, and in some cases audio materials. [8]

3.1 The rise of electronic data in qualitative research

For the past 10 years or so, the majority of qualitative researchers have used word-processors to transcribe their interviews. Furthermore, since around 1996 in the UK we have witnessed a huge growth in the use of computer-assisted qualitative data analysis software (CAQDAS) packages in qualitative research. CAQDAS software, such as ATLAS-ti, NUDIST and WINMAX are rapidly becoming the accepted tool for handling the description and interpretation of qualitative data². For Qualidata, issues about preservation of data from these packages is something we have had to address with some urgency. These are proprietary software packages and in the past it has not been possible to import and export data from one package to another. Qualidata has developed guidelines on what to keep for archival purposes—i.e. extracting or reducing the

1 CORTI, DAY and BACKHOUSE ([in this issue](#)) discuss the nature of the depositing process—from researcher to archive.

2 CAQDAS networking Group, University of Surrey, <http://caqdas.soc.surrey.ac.uk/>.

data to its simplest form—ASCII text or rtf. As expected, in the past year we have seen software developers taking steps to encourage sharing between packages, for example adding export and import facilities to their programmes. Qualidata is liaising with the developers to discuss extending the functionality of the packages to include archiving features. [9]

3.2 Can "traditional" repositories cope with electronic (non-numerical) data?

Well, in short, some can and some can't. Some of our host repositories have the facilities to provide copies of, say transcripts on disk, whereas others haven't had the infrastructure to provide that service. This is usually simply a case of under resourcing. It is not uncommon in the traditional British archive world to see one, or at best two, archivists responsible for sorting, cataloguing, housing, and providing access to archives. This leaves little time for digitisation programmes and resources may not stretch to obtaining high-powered computing equipment for storage. A survey carried out in 1998 of the prevalence of electronic documents in personal papers and organised records held by archival repositories in Britain, confirmed problems of staffing, software, hardware, expertise and dissemination. [10]

The other side of the picture, and of course an ironic one, is the increasing lack of physical storage space for paper-based archives. Many archives are approaching full capacity for paper documents, and those with inadequate storage facilities are using hot or damp basements for storage. Microfilming and digitising saves on storage space, but does not necessarily represent a cheaper option: filming and scanning are expensive operations and the maintenance of electronic records in the long-term involves periodic transfers of data to new media and software. Technological changes—and the ever-reducing cost of computer storage—will undoubtedly mean that digitisation becomes a more attractive option over time, not least because it allows the records themselves to be disseminated electronically. [11]

However, change is underway, spurred by the dawning of the Age of the e-Library (e.g. [JISC E-lib programme](#)) and other electronic archiving initiatives in the archives world ([H.E Archives Hub](#)). We are witnessing closer relationships being forged by academic libraries and archives with IT departments suggesting that it won't be long before traditional archives will be able to handle materials in any format. [12]

4. What about the National Data Archives? Are they Ready to Accept Qualitative Data?

4.1 The UK position

Qualidata's original strategy was to place "digital" data alongside paper-based materials in repositories or, where possible, to offer it to the UK Data Archive at Essex. The UK Data Archive is experienced in handling, storing and disseminating textual data, and presently have the advantage over some traditional repositories in being able to keep up with changing media and storage technologies. Whilst the Archive archives primarily numerical data, it also acquires textual data, image data and databases (see CORTI and AHMAD [in this issue](#) for a description of the archiving of an image based dataset). Documentation for datasets (such as questionnaires and interviewer instructions) is now stored in image format mostly in the form of PDF files; and more recently they have also begun to acquire image based datasets. [13]

However, in the early days Qualidata witnessed a number of barriers to establishing closer co-operation with the Data Archiving Community. In the 1990s the national Data Archives across the world were generally not ready to take on board qualitative data. The main problem was the lack of any prevalent demand for qualitative data. Demand for social science data is largely a reflection of what is already known to be out there, and how the goods are marketed. Because qualitative data were not available and because the culture of secondary analysis of qualitative data was not established, there was no pressing need for the Archives to consider them. Second, the nature and format of "soft" data was something data archive staff were not at all familiar with. Although the holdings of the Archives extended to image and tabular data, over and above their primary collections of numeric, format such as audio and video formats were not on the agenda. Finally, incorporating qualitative data would present apparently huge and insurmountable problems for safeguarding anonymity. [14]

In 1995, collaboration between Qualidata and the UK Data Archive based, also at Essex University, led to the Data Archive accepting qualitative datasets, on a case-by-case basis. For Qualidata, the Data Archive was used in the same way as any of the other repositories in the Qualidata network. It makes sense to hold data from mixed methods studies in the same place, for example, so that accompanying in-depth interview transcripts sit alongside the statistical dataset. For the Data Archive, accepting materials depended on whether their own strict acquisitions criteria (see <http://www.data-archive.ac.uk/depositingData/suitability.asp> [Broken link, FQS, 04/07/14]) were met, one of which required textual data had to be, as far as possible, completely anonymous. The debate about the reality of anonymisation for qualitative data is discussed in depth (CORTI, DAY & BACKHOUSE [in this issue](#)). [15]

In order to acquire "new" types of data, staff require specialist skills for evaluation, processing and documenting data and for user support for the data. The reason that the UK Data Archive is able to acquire qualitative material is that

Qualidata acts as the front-line, engaging in most of these key functions. Consequently, expertise and resources to deal with qualitative data (especially dealing with confidentiality, see CORTI et al. in this volume) are not required of the Data Archive's own personnel, who are busy enough with the demands of their own specialist roles. With this infrastructure in place, the UK Data Archive is able to provide access to a greater range of social science data. [16]

4.2 The rest of the world

One of my own personal missions has been to try to promote and persuade other National Data Archives across the world to consider the value of extending their data holdings above and beyond primarily quantitative research or administrative data. Responses to these efforts from the data archiving community (through the International Association of Social Science Information Service and Technology, IASSIST, <http://datalib.library.ualberta.ca/iassist/>) have been none too impressive. Perhaps, understandably, the IASSIST community had their own key agendas, driven by a pressing urge to improve resource discovery technology and develop and harmonise documentation standards (the Data Documentation Initiative, <http://www.icpsr.umich.edu/DDI/>; see also KUULA [in this issue](#)). [17]

I am pleased to report that in the last year or so we have seen dramatic progress in the willingness of the Data Archiving community to begin to evaluate the issue. In October 2000 the picture looks as in Table 3:

Status	How many	Who
Acquire qualitative data on a regular basis for some years	2	UK Data Archive, http://www.data-archive.ac.uk/ US Murray Research Centre*, http://www.radcliffe.edu/murray/
Recently acquiring qualitative data	1	Swiss SIDOS, http://www.sidos.ch/
Currently conducting feasibility studies for acquiring qualitative data	2	Finland FSD, http://www.fsd.uta.fi/english/index.html Denmark (DDA), http://www.dda.dk/

Table 3: National Data Archives which have qualitative data on their agenda (*The Murray Research Centre has a specialist and unique collection of largely longitudinal data, but is not strictly a "National" data archive that is, like others, supported by the social science funding councils; see JAMES & SORENSEN [in this issue](#)). [18]

In my view, the climate is changing fast. Along with global warming we are also seeing a warming climate within social science research methodology—interdisciplinary and mixed-method approaches are now more common and, in a most unexpected manner, even some hardcore economists are embracing

qualitative methods. Data Archives now need to adapt to this changing climate, by investing in a wider range of data sources and products. None of the barriers referred to earlier are insurmountable—there are strategies for overcoming these. The UK model works well—but, like everything else, the model is evolving too. [19]

4.3 Data archives: Particular problems for storing qualitative data

Audio and video recordings, images (such as photos) all present difficulties for archiving in two ways. First, it is essential to gain consent from participants to archive their recordings/pictures, and second, the media pose an issue for storage in digital data archives. [20]

Tape recordings of interviews are almost always used in qualitative studies—in-depth interviews, focus groups, many observations and naturally occurring conversation for discourse analysis, often rely on a record of the spoken word. For some projects, full transcription is essential, for others summaries may suffice. Methods of transcription also vary: sociologists generally want to capture the words, whereas conversation analysts and socio-linguistics are more concerned with recording other contextual features of the interview, such as pauses, laughter, tears etc. [21]

In terms of the potential of a qualitative dataset for re-use, the ideal is to retain the original tape recordings. There is really no substitute for listening to people's own words—a transcription is a subjective interpretation of the real-life conversation. However, in reality, it is often not possible to archive audio-tapes where the material is of a "sensitive" nature, without imposing either restricted access, a period of closure and/or retrospective permission from participants. For video, the added complexity of faces means that there is no way around seeking permission to archive video data. In the UK video methods are still not that popular, with only a few branches of social science having taken it on board—social anthropologists; socio-linguists and discourse analysts and educationalists. Anonymising tape recordings is vastly time-consuming and prohibitively costly. Blanking out identifying information on analogue media is also rather pointless as it distorts the data. New kinds of software are now available which enable researchers can edit, anonymise, label and copy their own digital data with far more ease. However the task is still labour-intensive. In archival terms, an additional problem is the lack of consensus (at least in the UK) about the relative longevity of these audio media for archival purposes. Whilst current technology favours CD-R and Minidisc, R-DAT is probably the most archivally sound (and most expensive). DVD is fast becoming a popular media and will, in time, replace audio and video CD. Since all Windows operating systems will be supporting it, it looks likely to dominate the market. Whilst it is still very expensive, inevitably costs will drop. [22]

As technology moves forward many Data Archives across the world may begin to consider the storage of digitised and indexed data from audio, video and multi-media data. The UK Data Archive has taken the decision not to investigate audio and video data at present. The UK has other digital data services that specialise

in the long-term preservation and provision of access to a range of electronic historical and contemporary humanities' data in many formats (Arts and Humanities Data Service, <http://www.ahds.ac.uk/>). It is wiser to work with these experts than attempt to duplicate efforts. Still the issue remains as to how we can provide the user of qualitative data with seamless access to the "whole" dataset when the components are dispersed. There are new technologies currently addressing the prospect of one-stop-shops for distributed electronic resources. [23]

5. What are the Major Issues Surrounding Adopting a National Archival Policy for Qualitative Data

No matter how much a Data Archive may want to embrace qualitative data there are still significant barriers—deriving principally from researchers. Qualidata has carried out numerous surveys of and interviews with qualitative researchers at all levels of seniority across the UK—from PhD to Principal Investigator. The feedback coming back points to a small number of key concerns. There are strong feelings out there, yet we have encountered a vastly diverse response to archiving: from overwhelming joy at the thought of rescuing someone's life's works from consignment to the skip, to vehement anger and displeasure at the thought of being asked to share a "possession" considered to be of personal value. Some of the negative feelings have not waned over the seven years, and we continue to confront a hardcore of sceptics. Just to set the UK scene, I would like to share some of the reasons for which I believe qualitative researchers harbour scepticism about sharing and re-using qualitative data. Whilst many of these relate to issues about confidentiality and concerns about agreements made at the time of fieldwork, others are more to do with academic arrogance, fear of criticism or lack of understanding about how data can be sensibly re-used. Researchers' worries about confidentiality are discussed in detail in our other paper in this issue (CORTI, DAY & BACKHOUSE 2000 [in this issue](#)). The Danish Data Archive (FINK in this issue) also speak about their discussions with qualitative researchers on this topic. [24]

6. Researchers vs. Archivists

6.1 Researchers' worries

Vulnerability: Fear of exposure

Generally, qualitative social "scientists" are just not used to making their findings accountable. They are worried about others seeing their data, and possibly picking holes in them. Some argue that certain approaches used in qualitative research, for example, grounded theory (GLASER & STRAUSS 1967) which opposes the scientific paradigm of testing hypotheses, do not lend themselves to verification. [25]

"Being there"

Sociologists are not used to consulting colleagues' data and the concept of "secondary analysis" is still viewed by most qualitative researchers as pertaining to "number crunching" activities. Some researchers are concerned that qualitative data cannot be used sensibly without the accumulated background knowledge which the original investigator acquired during its collection. This is particularly so with longitudinal studies of a group where the researcher feels that a special rapport has been developed without which the material may be meaningless. Thus the essential contextual experience of "being there" cannot be shared. [26]

6.2 The data archivists answers

Since there has not been an established culture in social science for secondary analysis of qualitative data there is not a mass of evidence of successful re-use of qualitative data, as there is for survey data or historical texts, for example. However, in spite of this historical deficit, the arguments for preservation and re-use of qualitative data have managed to persuade a sizeable portion of the qualitative research community to come to terms with the practice. [27]

Substantiation is good

First, if we are to accept the label "scientist", then we should adopt the scientific model of opening up our data to scrutiny, and the testing of reliability and validity. The quality of social research is highly variable, and in the UK there are no quality control standards for qualitative studies (the exception being for market research³). [28]

Avoiding duplication is cost effective—make use of existing data

Second, in order to avoid unnecessary replication of research, and to gain a more informed approach to a new topic, new research should make more attempts to delve into earlier related research and, where possible, try to include some comparative element. This can only be established if information about existing projects is available and data are easily accessible. In the UK a centralised database of current research does not exist, and the information available relies on sponsor's own databases of research they support. Built into the ESRC's research grant contract is an obligation to check for similar existing datasets to the one(s) grant applicants are proposing. This requirement relies on an established and expanding bedding of archives. Finally, generally speaking, many projects generate huge quantities of data which are rich and often, relatively unexploited. [29]

3 BS 7911 is the trademark for the standard adopted by the Market Research Society in 1988 for "Specification for organizations conducting market research". This came about partly as a result of the hugely varying quality of qualitative studies in this arena.

You don't need to have "been there"—good documentation can help

While qualitative research uses reflexivity relating to experience of fieldwork as a means of enhancing data collection and forming new hypotheses in the field, the secondary analysis of data should not be dismissed that easily. Indeed, there are instances where research data is, in a sense "re-used", by the investigators themselves. For example, some principal investigators who write the final articles resulting from a project, have employed research staff or a field force to collect the data. Similarly, for those working in research teams, sharing one's own experiences of the research are essential. Both rely on the fieldworkers and co-workers documenting detailed notes about the project and communicating them to each other. Of course, audio and video-tape recordings enhance the capacity to re-use data without having actually been there. For archives, documentation of the research process provides some degree of the 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. [30]

6.3 Changing attitudes: resignation?

Since 1998 we have found a damping down of resistance as both acceptance (perhaps grudgingly in some cases) and positive support for archiving has become commonplace. We are in no doubt that this cultural shift has been accomplished by two factors:

- The ESRC Datasets Policy implemented by Economic and Social Research Council (ESRC) in 1996 which contractually obliges holders of ESRC research grant awards to offer all kinds of research data for archiving at the end of their project. Research applications now ask proposers to verbalise plans for preparing and archiving any qualitative data they produce in the course of their research. Other UK research sponsors have followed these steps; the Wellcome Trust, the Joseph Rowntree Foundation and the Nuffield Foundation refer new grant holders to Qualidata for advice on archival strategies for data they intend to generate.
- The publicity and training work Qualidata has done in trying to promote the culture of sharing, preserving and re-using qualitative data. Centre staff try to attend key gatherings of qualitative researchers and either get the issues onto the agenda or at least into the conversation. [31]

7. Key Issues for Data Archives Acquiring Qualitative Data

Below I have summarised some of the key issues that data archivists should ask themselves:

- *Setting priorities for acquisition*: what kind of collections or themes for qualitative datasets will enhance the profile of the archive? Should collecting policies concentrate on studies which have produced both survey and qualitative data? Should the collections reflect contemporary themes or agendas which tie in with funding bodies' priorities? Do we focus on large and expensive studies such as longitudinal studies.
- *Procedures and standards for processing data*: what standards or practices should be adopted for processing data? How do these differ from preparing survey data? What about issues of confidentiality and copyright? Should existing staff be trained to process qualitative data or should staff undertaking this be experts in qualitative research?
- *Metadata—standards for documentation*: are the existing standards for study description for numerical datasets adequate? How do the emerging document type definition standards for data (e.g. the Data Documentation Initiative, DDI) suit qualitative data? Do they need to be extended or reworked? At the same time, how relevant are standards adopted by the "traditional" and library communities for more complex qualitative material?
- *Access procedures for safe-guarding data*: are existing data preparation procedures adequate for safeguarding participants? Should qualitative and survey data from the same study be provided together? Are the access control and vetting procedures adequate?
- *Format—digital, text, multimedia*: Can the archive cope with digital audio or video data? Is that part of their policy or could it be handled better by another specialist data service?
- *Researchers—sharing and training*: how can we encourage researchers to provide data and in a format that enable us to deposit it economically and rapidly? What mechanisms have we in place to train researchers to document their own datasets? How much hand-holding might they require?
- *Funding—taking responsibility*: how can we persuade sponsors of research to implement archival policies for qualitative research they fund? How do we get long-term commitment for archiving data onto the research resources agenda? How can we ensure that peer review considers the value of a dataset rather than just intended publications and media outputs? How do we forge both a partnership and a working relationship with funding bodies that allows for smooth and relatively seamless deposit of data? [32]

These issues are familiar to all Data Archives—the difference is that for qualitative data there is more groundwork to be done. The state of the art is relatively new and a period of feasibility testing needs to be undertaken to find out what works and what doesn't, for example, which datasets are possible to archive and, whether people want to use them. [33]

8. Documenting Qualitative Data

Qualidata developed standards for the documentation of qualitative data in liaison with the UK Data Archive. The Standard Study Description, used by many Data Archives since the 1980s, was adopted and tweaked to fit the slightly different characteristics of qualitative data. In terms of archival format, materials were required to be reduced to their simplest form, e.g. for text ASCII or Rich Text Format (rtf) and for images, TIFF4 or Adobe's Portable Document Format (pdf). [34]

Over the last couple of years, documentation standards have been at the forefront of Data Archives' activities. Aimed at the greatest possibility of interoperability, transferability and visibility, the Data Documentation Initiative (DDI) is seen to be the solution. Arja KUULA ([in this issue](#)) gives us an excellent introduction as to how this might suit the description of qualitative data. Furthermore, in March 2000 a meeting will take place at Essex to discuss the adequacy of the DDI for qualitative data. Five groups from the Cologne 2000 ISSM conference sessions on "Preserving and sharing Qualitative Data" will be contributing. [35]

9. Optimal Models for Qualitative Archiving

BODDY (2000) identifies two models of data storage and provision under which a national archive could operate: a centralised facility in a single location or a hub and spokes model (using the metaphor of a wheel); see Table 4. He sets out a number of characteristics of each approach:

Centralised archive model	Typical distributed model
<ul style="list-style-type: none"> • Acquisition of data from researchers and others, storage and distribution at a single location • Centralised control over the conditions of supply and use of data • Checking, cleaning and processing data according to standard criteria • Centralised support service, describing the contents of the data, the principles and practices governing the collection of data and other relevant properties of data • Cataloguing technical and substantive properties of data for information and retrieval and offering user support following the supply of data 	<ul style="list-style-type: none"> • Data holdings distributed over various sites • Data disseminated to users from each of the different sites, according to where the data is held • The various suppliers of data ideally networked in such a way that common standards and administrative procedure can be maintained, including agreements on the supply and use of data • A single point of entry into the network for users, together with some form of integrated cataloguing and ordering service

Table 4: Models of data storage and provision, BODDY (2000)⁴ [36]

⁴ This is taken from a recent consultation document commissioned by ESRC to establish the long-term future of resource provision for social science data archives in the UK, see BODDY

9.1 The distributed network model: Qualidata

Qualidata was established using the latter approach, with the Centre as the hub, or the focal point, bearing responsibility for evaluating, acquiring, preparing, documenting, setting access conditions, transferring and publicising data. The network of archives act as spokes which enable the long-term storage of data. Whilst a distributed or "clearing house" model has costs savings over a centralised one, for Qualidata, the hub and spokes model falls short when we consider monitoring access to and usage of data. [37]

9.1.1 Problem 1: Keeping tabs on re-use and re-users

All archiving initiatives will be more than familiar with the fact that preservation for posterity is not a convincing enough reason for funders of social science research to want to make long-term commitments to archiving. For Qualidata, collecting information and statistics about re-use is one of the key performance indicators demanded by our funders, ESRC. These annual figures are increasingly being seen as a measure of our success. How many have used each dataset this year? Who is using the data? What for? And so on. For a centralised model, because data are acquired from a single access point, details of users and reasons for use are readily logged. Nevertheless, even Data Archives still find it difficult to get users to tell them about publications some years down the line. [38]

Qualidata's on-line catalogue, [Qualicat](#), points people to data sources located in a given repository within the UK. Whilst we do ask those identifying data from Qualicat to enquire directly to us about datasets, we have little control over those approaching the "spokes" directly, via the host repository's own finding aids or indeed, through Qualicat itself. We gather our annual usage statistics by asking our network of host repositories to provide us with figures. While most archives are happy to feed back to us basic details about usage of "Qualidata collections", others do not have appropriate mechanisms in place to register usage of particular collections at all. Clearly this is an unsatisfactory situation, but one we have little control over, in spite of an agreement with repositories which asks them to provide Qualidata with re-use details. Furthermore, we are rarely supplied with names of users and are not therefore able to link up data enquiries or research grant holders with those actively using data. The model we want to move to is one where all potential users register with Qualidata first, and then, either we offer more support in helping them acquire data or, they inform us when they have acquired data from the host repositories. [39]

9.1.2 Problem 2: Providing users with timely access to data

We are finding, increasingly, that as knowledge about deposited sources of data increases, so do requests for help in finding and obtaining suitable datasets. This is particularly true for requests from those wanting datasets for teaching purposes. Researchers too want expert advice on datasets, and instant access to

data. As Qualidata does not physically hold all the data it publicises in its catalogue, (other than having a degree access control over the National Social Policy and Social Change Archive collection at Essex), users are often put off by the fact that they may have to travel to Scotland to access a single dataset based in a Scottish Repository. Moreover, Qualidata sometimes finds itself having to acquire data on the user's behalf, for example by arranging to get copies made for and dispatched to a user. User support at this intensive level needs to be resourced, but currently, for Qualidata, it is not. In the short-term, there is still no getting round users having to visit archives in person to access large paper-based collections, as repositories are in no position to digitise all their holdings. [40]

Is there an optimal model for providing qualitative data? I think we still have some way to go in discovering a perfect model for providing access to a wide range of qualitative data. Would it help to have all data in one place? This way we could ensure standards—in terms of data quality, preservation (particularly of digital data) and controlled access. Or should we trust the model Qualidata chose at the beginning—to invest in a network of high-class repositories? Perhaps we need both models, whereby data are both centrally and locally stored giving a double benefit to a host of dispersed and disparate user communities. I would be happy to choose this latter model as a future one, if we could ensure archival repositories get up to speed on a number of key functions:

1. being able to safely and respectably preserve digital data on a long-term basis,
2. enabling both secure and speedy access to qualitative data,
3. enabling provision of timely and accurate statistics about re-use of data. [41]

On a final note, offering access to qualitative data requires highly-trained user support staff, who can help make the process of finding, acquiring and re-using data less painful than it can sometimes be. [42]

9.2 The UK Qualidata funding situation

The 5 year ESRC Resource Centre award for Qualidata ended in September 2000, and we were awarded a reduced budget from Oct 2000 to Sept 2001 (see Table 6).

1994	5 year funding of £.7m from ESRC
1996	50k for Charitable Foundation (JRF)
1996	top up funding of £.4m from ESRC
1999	35k from Medical Research Council
2000	100k from ESRC for 1 year
2000	12k from Essex University
2001	? Dependent on current evaluation of UK archiving scene

Table 5: Financial profile of Qualidata 1994 - 2001 [43]

In the past year, ESRC have shown indecisiveness about their level of commitment to archiving social science data across the board. In order to help steer their strategy they have commissioned a consultancy aimed at gaining a cross-national picture of researchers' experiences of, views about and demand for data archives. This is a worrying time for archival enterprises like Qualidata and the UK Data Archive, at whom this review (one of many) is targeted. Qualidata has served not only as a test bed for the archiving of qualitative research on a national scale (for which the mission has proved possible) but also as a centre of advice for other budding projects across the world working in this field. We have ambitious plans, including the creation of significant electronic data resources for both research and teaching, and hope to be able to realise these with alternative funding. The prime aim at the moment is keeping the basic Qualidata machine going. [44]

In the mean time we are planning a longer-term strategy for merging Qualidata with the UK Data Archive. Most immediately this will benefit ESRC researchers and Qualidata itself who do not have access to technical staff or advanced resource discovery tools that are being developed within the UK Data Archive (e.g. NESSTAR, <http://www.faster-data.org/technology/nesstar/index.htm> [Broken link, FQS, December 2004]). [45]

While many of the particular problems of confidentiality remain for qualitative data, which subsequently requires careful thought about access, the earlier distinction between paper-based qualitative research and machine-readable survey material no longer applies. In this sense, a closer integration of archiving services is practicable, points to efficiency gains, and is equally welcomed by the UK Data Archive. Extending the breadth of data collections and expertise about archiving is good for the Data Archive's portfolio and equally, is beneficial for social scientists in terms of having a "one-stop-data-shop" discussed earlier. [46]

10. Re-use of Data

10.1 The million dollar question: How are people re-using qualitative data?

The ways in which qualitative data can be re-used have much in common with those applicable to the secondary analysis of survey data. Here I identify six distinct ways:

- *"new questions for old data"*⁵: approaching the data in ways that weren't originally addressed. The more in-depth the material, the more possible this is,
- *for research design*: using the sampling and data collection techniques and tools to design a new study, or investigating a study for methodology's sake,
- *as case material for teaching*: using the methodology, data and methods from, for example, classic studies for teaching research methods across a range of social science disciplines,

5 I have borrowed this neat phrase from the Murray Research Centre, Harvard, USA.

- *for comparative research*: data can be compared with new or other data sources, across time or region, social group etc.,
- *for verification*: for substantiating results. We have yet to see any evidence of this, apart from a couple of much earlier classic cases in anthropology and psychology,
- *historically*: inevitably data created now will become a historical resource. [47]

THOMPSON [in this issue](#) offers an insight into using in-depth interviews for approaching a new study—for both informing the design and for comparative purposes. [48]

10.2 So, what do users want?

To some extent, a definite answer to this question is not yet available. The patterns that Qualidata is seeing tend to vary year on year—from a relatively small base of about 100 catalogued datasets in 1995, to 460 in 2000, potential users have not had a great choice. In January 2000 we conducted a national survey of potential users of archived qualitative data. Researchers and teachers were asked to say whether they made use of/would make use of qualitative data archives, and what kinds of qualitative data archives might be most useful to them. Over 550 valid responses were received from a range of user communities, of which 99% wanted to see datasets, mostly in electronic form, available for both research and teaching, across a wide range of disciplines. Health, criminology and education data came out on top. [49]

11. Connecting the World

11.1 The need for international collaboration

Since October 1996, Qualidata has established a growing number of key international contacts through meetings and correspondence with archivists and researchers in Europe, North America and Asia. Not only has the Centre been actively consulted by a number of archival initiatives across the world but we are also seeing evidence of embryonic "qualidatas" spawning across Europe and beyond. Typically these initiatives have been small scale projects with narrow remits run by academics or researchers based in sociology departments, and have tended to have no links with their own country's Data Archiving community. Qualidata has helped to bring the "factions" together to promote a better understanding of the needs of both parties:

- for the small, isolated and largely underfunded groups to understand that there exists a national infrastructure set up specifically to manage the preservation and provision of access to social science data;
- for the national Data Archives to appreciate that there is life beyond numerical or administrative data, and that there are groups out there who want to open the door to sharing of qualitative data. [50]

In early 2000, I proposed a stream for the International Social Science Methodology Conference in Cologne in October 2000 on the Preservation and Re-use of Qualitative Data. Fifteen papers spanned over 3 sessions, and many of these are published in this volume. The three sessions focussed on different, but highly interconnected, aspects of sharing and preserving qualitative data:

- Key Developments and Problems in Qualitative Data Archiving,
- Digital Qualitative Data Archives,
- Providing Access to and Re-using Qualitative. [51]

For me, the most satisfying and promising aspect of this meeting was the presence of representatives from five of the national Data Archives—some planning the acquisition of qualitative data (and some hopefully feeling guilty about not making such plans!) [52]

11.2 The key issues for international collaboration

In summary, I see the most critical issues for qualitative data archives which need to be discussed on a global level as

- encouraging a culture of sharing in research practice,
- developing collection priorities and assessing re-usability of datasets,
- developing methods for deposit of and access to sensitive datasets,
- developing requirements for contextual data to provide suitable "background" to raw data,
- developing documentation standards for data (metadata, e.g. DDI),
- encouraging researchers across the whole social science spectrum to re-use qualitative data,
- creating digital resources for teaching and research,
- working with the major national funders of social research to implement archival policies,
- enabling greater access to qualitative datasets across national boundaries. [53]

11.3 Linking arms

One of my wishes, and perhaps dreams, is for a genuine two-way communication to take place between the smaller groups of researchers archiving qualitative data and their own country's national Data Archives who possess both influence and infrastructure. Finding secure funding is the most problematic issue for all the archive initiatives you will hear about in this volume. Qualidata needs peers across the world—it needs to participate in working groups to discuss all these issues and arrive at sets of international recommendations. Whilst we have our own procedures and standards, as do many of the groups beginning to take the tasks on board, these alone do not have the international standing or recognition I

feel they need. International reputation reflects well on prospects for national funding. [54]

Some key work with qualitative data is now being done from within established Data Archives, which are perhaps more secure than some of the smaller enterprises. However all have one thing in common—the need to persuade their own research communities that preserving and sharing qualitative data is sensible and a good use of public funds. The next year will be a time of great advancement in qualitative data archiving as the some of the National Data Archive across Western Europe step out of their evaluation phases into programmes which are implementing the acquisition and processing of qualitative datasets. [55]

INQUADA was established in October 2000 at the Cologne conference (see <http://www.essex.ac.uk/qualidata/current/inquada.htm>). The main impetus for establishing INQUADA was the international isolation felt by Qualidata and the need for a new space for groups wanting to gain advice and support on qualitative data archiving, as well as being able to share experiences and contribute to the knowledge pool. [56]

The overall aims of INQUADA are:

- provide a forum for professionals involved in the practice of preserving and providing access to qualitative data,
- promote the preservation, dissemination and re-use of qualitative data,
- foster the exchange of qualitative data and collaborative working,
- enable both individuals and organisations wanting to set up new initiatives to learn, adapt and share their knowledge about work practices,
- disseminate sets of Best Practice Guidelines and Recommendations covering a range of key issues relating to the preservation, dissemination and re-use of qualitative data including: research project management; confidentiality and copyright; methods of transcription; technical and recording issues; anonymisation; digitising and indexing paper-based research materials,
- develop methods of computer-assisted archiving of qualitative data,
- develop and refine standards for documenting a wide range of qualitative data (metadata),
- build and maintain a focal Web site. [57]

12. Conclusion

I hope that those reading this will have been persuaded that qualitative data have much to offer above and beyond the investigator's own analyses. While the responsibility for providing the infrastructure to enable this process should fall on the shoulders of those who sponsor the research, it may be uphill struggle for some time to come, to prove that preserving qualitative data is worthwhile. My hope is that I can look back in ten years time and see that it was worth all the hard slog! [58]

References

- Boddy, Martin (2000). *ESRC Green Paper on Data Policy and Data Archiving: Consultation Paper with Questions*, <http://www.ilrt.bris.ac.uk/ubris/esrc/p3.html> [Broken link, FQS, December 2004].
- Corti, Louise & Ahmad, Nadeem (2000). Digitising and Providing Access to Social-Medical Case Records: The Case of George Brown's Works [19 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 1(3), Art. 6. Available at: <http://www.qualitative-research.net/fqs-texte/3-00/3-00ahmadcorti-e.htm>.
- Corti, Louise; Day, Annette & Backhouse, Gill (2000). Confidentiality and Informed Consent: Issues for consideration in the preservation of and provision of access to qualitative data archives [50 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 1(3), Art. 7. Available at: <http://www.qualitative-research.net/fqs-texte/3-00/3-00cortietal-e.htm>.
- Fink, Anne Sofia (2000). The Role of the Researcher in Qualitative Research [69 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 1(3), Art. 4. Available at: <http://www.qualitative-research.net/fqs-texte/3-00/3-00fink-e.htm>.
- Glaser, Barney G. & Strauss, Anselm L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.
- James, Jacqueline & Sorensen, Annemette (2000). Archiving Longitudinal Data for Future Research. Why Qualitative Data Add to a Study's Usefulness [59 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 1(3), Art. 23. Available at: <http://www.qualitative-research.net/fqs-texte/3-00/3-00jamessorensen-e.htm>.
- Kuula, Arja (2000). Making Qualitative Research Material Reusable: Case in Finland [28 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 1(3), Art. 19. Available at: <http://www.qualitative-research.net/fqs-texte/3-00/3-00kuula-e.htm>.
- Thompson, Paul (2000). Experiences of Re-analysing Data in Qualitative Research [48 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 1(3), Art. 27. Available at: <http://www.qualitative-research.net/fqs-texte/3-00/3-00thompson-e.htm>.

Author

Louise CORTI is currently the Deputy Director and Manager of Qualidata, the ESRC Qualitative Data Archival Resource Centre, based at Essex. In January 2001 she will be taking up the post of Director of User Service of the UK Data Archive, where alongside the duties of that role, she will retain an overall responsibility for qualitative data archives. In the past she has taught sociology, social research methods and statistics, and spent six years working on the design, implementation and analysis of the British Household Panel Study at the University of Essex. She is interested in both qualitative and quantitative aspects of social research.

Contact:

Louise Corti

Qualidata
University of Essex
Colchester CO4 3SQ, UK

Tel.: +44 1206 873058

E-mail: quali@essex.ac.uk

URL: <http://www.essex.ac.uk/qualidata/>

Citation

Corti, Louise (2000). Progress and Problems of Preserving and Providing Access to Qualitative Data for Social Research—The International Picture of an Emerging Culture [58 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(3), Art. 2, <http://nbn-resolving.de/urn:nbn:de:0114-fqs000324>.