Sharing Qualitative and Qualitative Longitudinal Data in the UK:

Archiving Strategies and Development by Libby Bishop and Bren Neale

Abstract
Over the past two decades significant developments have occurred in the archiving of qualitative data in the UK. The first national archive for qualitative resources, Qualidata, was established in 1994. Since that time further scientific reviews have supported the expansion of data resources for qualitative and qualitative longitudinal (QL) research in the UK and fuelled the development of a new ethos of data sharing and re-use among qualitative researchers. These have included the Timescapes Study and Archive, an initiative funded from 2007 to scale up QL research and create a specialist resource of QL data for sharing and re-use. These trends are part of a wider movement to enhance the status of research data in all their diverse forms, inculcate an ethos of data sharing, and develop infrastructure to facilitate data discovery and re-use. In this paper we trace the history of these developments and provide an overview of data policy initiatives that have set out to advance data sharing in the UK. The paper reveals a mixed infrastructure for qualitative and QL data resources in the UK, and explores the value of this, along with the implications for managing and co-ordinating resources across a complex network. The paper concludes with some suggestions for developing this mixed infrastructure to further support data sharing and re-use in the UK and beyond.

Keywords: Qualitative Data, Qualitative Longitudinal Data, Data Archive, Data Policy, Secondary Analysis, Re-use, UK

1 Introduction
Over the past two decades significant developments have occurred in the archiving of qualitative and qualitative longitudinal (QL) data in the UK, supported by two major funders of social science research and of data archiving: the Economic and Social Research Council (ESRC) and the Joint Information Systems Committee (JISC). The first national archive for qualitative resources, Qualidata, was established in 1994. This followed feasibility studies that set out the case for gathering such resources together for preservation and encouraging data sharing and secondary use by enabling access to publically funded research data. Since that time further scientific reviews have supported the expansion of data resources for qualitative research
an overview of the main data policy initiatives and recommendations that have set out to advance qualitative data sharing across the social sciences. We provide an overview of the mixed infrastructure in place for qualitative and QL resources, focusing in particular on Qualidata as a key generic resource and Timescapes as a specialist distributed resource. The paper concludes with some suggestions for developing this mixed infrastructure to further support data sharing and re-use in the UK and beyond.

2 Development of Qualitative Archiving in the UK

Qualitative datasets are rich and varied in nature, based on in-depth interviews and a range of ethnographic methods (including participant observation and the generation of fieldnotes, case studies and aural and visual data) to capture the contexts and complexities of real life experiences (Hammersley and Atkinson 1995; Mason 2002). The establishment of Qualidata in 1994 as the national archive for the curation of such data was a major landmark in the UK. Much of the recent history of developments in qualitative archiving in the UK equates with the history of this initiative. At its inception, Qualidata was not a place of deposit itself, but acted as a clearing house by locating existing data collections and arranging for their deposit in suitable institutions such as archives, libraries, museums and other repositories. In 1997, the major funder of qualitative social science research in the UK, the ESRC, made it a condition of funding that researchers should deposit their datasets with Qualidata. This policy change was a critical factor in enabling Qualidata to accelerate the acquisition of its own holdings of qualitative data. In 2000, Qualidata was incorporated into the UK Data Archive, itself established in 1967 and curator of the largest collection of digital data in the social sciences and humanities in the UK. In 2003, the national infrastructure was further bolstered through the establishment of the Economic and Social Data Service (ESDS), a national data service of the UK Data Archive, which provides access to and support for an extensive range of key economic and social data, both quantitative and qualitative. ESDS Qualidata became one of the core components of the new service: The Qualidata Quinquennial Report 1994–1999 (Corti and Thompson 1999) provides a comprehensive summary of the early years of Qualidata and addresses existing archives, cataloguing procedures, dissemination, re-use, management and funding. More recent developments are documented in subsequent reports for the UK Data Archive and ESDS (ESDS 2009).

3 The Development of Qualitative Longitudinal (QL) Research and Archiving

In the UK and elsewhere QL research is well established. Qualitative researchers have a long history of engaging with time, through a wide range of methods and from different disciplinary perspectives, most notably, anthropology and oral history (Elder 1981). Time is built into these studies in a complex variety of ways. Retrospective studies capture change through biographical, historical or inter-generational accounts. Recently there has been growth in the number of projects that re-visit classic studies of communities, institutions or groups to understand changes and continuities and to re-interpret past findings. Prospective studies, on the other hand, track individuals or groups over time in order to capture changes in the making and to revisit changing perceptions of the past and future. Individuals may be tracked extensively through particular transitions, or extensively across different periods of their lives. Prospective tracking is valued because it captures the immediacy and complexity of real lives as they unfold (Neale and Flowerdew 2003; Saldana 2002; Thomson and Holland 2003). Over the past decade QL methods have begun to gain legitimacy as an integral part of the methodological canon. Researchers are increasingly seeking to incorporate QL methods into their research design and a range of studies are now being funded by government, the ESRC and the main UK charities (the Nuffield and Joseph Rowntree Foundations); e.g., on lone parenthood, families after divorce, the life trajectories of offenders and probationers, passages through primary school or the benefits system, and the life histories of migrants or people living in poverty. Until recently QL datasets tended to remain the preserve of the originating researchers. Archival collections that bring such data sets together to facilitate re-use remain scarce. Notable exceptions are the Oral History archives at the British Library and the Mass Observation Archive at the University of Sussex. Mass Observation is a key historical resource, a paper archive accessible in person or through an online catalogue, of popular accounts of every day life in the UK, produced by a panel of recorders who respond to thematic directives (e.g., rationing, family food, life during the war, birthdays). The archive is seeking funding to digitise parts of its considerable collection that date back to the 1930s, and a range of secondary analysis projects have been funded to use materials from the resource.

Recent developments have placed QL archiving more centrally on the map. This began in 2006 when ESRC funded the archiving of case studies from the Inventing Adulthoods Study, a nine year study tracking a sample of young people from different regions of the UK (Inventing Adulthoods 2010). The archiving of the case studies has continued with further funding under the Timescapes initiative (described below) and the data are held at both ESDS Qualidata and the Timescapes Archive. A feasibility study into the development and scaling up of QL research and resources (Holland, et al. 2005) led to funding for Timescapes under the ESRC Qualitative Longitudinal Initiative 2007-12. The Timescapes study is resource-led as well as having a strong substantive and conceptual focus. The Archive, which was launched in October 2009, is being developed as a resource of QL data, with the current collection focused primarily—although not exclusively—on studies of personal lives and relationships across the life course. As well as data from the Inventing Adulthoods Study, the archive is collating data from seven core projects that span the life course and from a growing number of separately funded QL projects that are affiliated to the Initiative. In this way, Timescapes aims to build up a range of QL data collections on life course themes across diverse substantive fields in the social sciences, as well as encouraging re-use through secondary analysis initiatives. Its holdings are primarily digital and are multi-media, including audio and written data, as well as still and moving images. Currently the archive is supported by an institutional repository (LUDOS: Leeds University Data Objects Store) which uses DigiTool proprietary software. Documentation about the technical and procedural development of the Timescapes resource is available on the website (Timescapes 2010a). These include consent forms, guidelines for interview transcription and anonymisation, user registration documents, a depositor licence and the multi-media metadata schema. Further documentation will be added as it becomes available. Priorities for the near future are to further develop the Timescapes resource as a working archive across a broader range of projects, and to encourage and assist secondary use of the data.

Timescapes is innovative in encouraging archiving as an integral part of the research process rather than an administrative task relegated to the closing phase of a project. This feature is important because of several characteristics of QL research. Firstly, since qualitative researchers usually generate their own data and, in the process, build up relationships with their participants, they have a uniquely personal affinity with and ‘feel’ for the data and the context within which it was generated. Secondly, QL research often involves the generation of highly sensitive data that is contextually rich, difficult to anonymise, and therefore runs higher risk of disclosing identities. Particular care is therefore needed to preserve confidentiality. Thirdly, QL projects are often the product
of individual or small team scholarship that can last over many decades. The originating teams control and have exclusive access to the population samples which make up a study, and determine how and when they are followed up over time. Such projects may have a continuously provisional feel; they are never quite finished, either in terms of the potential for further data generation or the endless possibilities for complex analysis and reworking data to produce new insights and interpretations. Finally, unlike quantitative longitudinal data, which is gathered solely for secondary use, QL data is generated, at least initially, by and for primary researchers to enable them to address particular research questions. Archiving for secondary use in this context must therefore run alongside the tasks of ongoing data gathering and analysis by the originating team. This has implications both for the resources needed to attend to these tasks and for the timing of archiving within the project life cycle.

Given these characteristics, QL data needs specialist curation to encourage deposit and sharing while a prospective study is ongoing. Timescapes has developed an innovative stakeholder model of data sharing that enables archiving to be seen as an integral part of the research process. Researchers who deposit their data with Timescapes are stakeholders in the resource and are encouraged to re-use data as well as depositing, thereby combining primary and secondary analysis to raise new questions and produce new insights. Researchers can thereby continue to use their data and link it to other related data as their research progresses. The commitment to good data management planning at the research design phase enables data to be generated and organised for archival as well as primary use and prepared according to international archiving standards along with appropriate documentation or metadata (i.e., data about data that provides important context for the resource). Depositors are in the best position to provide rich and descriptive metadata that is aligned with the requirements of temporal analysis. As users of the dataset, depositors have a vested interest in ensuring that the resource is fit for purpose, with accurate metadata and refined thematic search and retrieval functions (e.g., through assigning key words to interview transcripts).

Crucially, the Timescapes Archive enables finely granulated controls on the re-use of data by building in different levels of access: public, registered, case-by-case approval, and embargo. For example, permission to use highly sensitive or un-anonymised data can only be given by the originating researchers, and the proposed use needs to be specified in consultation with the originating team. In this way, the archive opens up the potential for the sharing of data that might otherwise remain unarchived (Bishop 2009a). In essence the archive builds the necessary infrastructure to enable a more personalised mode of data sharing, which (as will be shown below) has been the primary way in which researchers have chosen to share their data in practice.

Timescapes is not a stand-alone archive; it is a distributed satellite of the UK Data Archive which also holds the data for long-term preservation purposes. It is simultaneously a part of the canon of longitudinal resources in the UK, and of qualitative resources, and needs to develop in both directions. Part of the remit is to encourage the linking of Timescapes data with data from other longitudinal resources, both qualitative and quantitative. These have included, for example, Understanding Society (the UK Household Longitudinal study), the National Child Development Study, Mass Observation and the Oral History collections at the British Library. There is also evident scope for comparative research and secondary use projects with international collaborators. These are beginning to emerge through EQUALAN (see introduction to this issue). Distributed archives such as Timescapes can play a crucial role as brokers between specialist research communities (whether defined in terms of data genre, methodology or thematic content), and generic data centres with broader remits, such as the UK Data Archive (Bishop 2009b). The specialist infrastructure being developed in Timescapes has the potential to form a valuable bridge between the research and archiving fields, and between primary and secondary research, that would enable these to be seen as iterative and reciprocal processes.

4 Data Sharing: UK Policies, Practice and Ethos
The development of infrastructure to support the re-use of qualitative data goes hand in hand with the development of an ethos of data sharing, both are necessary if data are to be made available for sharing and valued as a resource for re-use. The process of enabling data sharing is developing in a wide variety of ways when viewed comparatively across Europe. This is shown clearly by Ruusalepp (2008) who comprehensively reviews developments across the 30 countries of the OECD. He shows that organisations such as the OECD, UNESCO, ESFRI (European Strategy Forum on Research Infrastructures) and CODATA (The Committee on Data for Science and Technology) have policies that promote or recommend data sharing, and that these policies have influenced the policies of numerous UK organisations (e.g. Office of Science and Innovation (e-Infrastructure), JISC Strategy 2007-2009, and the Research Information Network’s Strategic Plan). These policies stop short of recommending mandatory data sharing. To date there are no national policies across the countries of the OECD that mandate data sharing in this way, although there is an increase in recommendations for ‘data management plans’ which ask researchers to take into account data sharing and curation, most notably in 2011 by the Research Councils UK. Even so, the ethos of data sharing is strongly endorsed within these policies and is beginning to have a discernable impact at the organisational level. Since 2000, The ESRC Data Policy, for example, has required all award holders to offer for archiving and sharing copies of both digital and non-digital data, and similar obligations continue in the 2010 Policy (ESRC 2000, ESRC 2010b). The recently revised ESRC Framework for Research Ethics notes that data should be collected with the expectation that others will re-use it (ESRC 2010a). (See Research Information Network (2011) for a detailed review of funders’ policies and the funded Sherpa Juliet project (JISC 2009b) for an international inventory of such policies).

While at the level of UK policy there is a clear and growing commitment to data sharing, the extent to which this translates into practice among the wider community of researchers is less clear cut. One way of gauging the ethos of re-use in the UK is through academic debate on this issue which has taken place primarily among a small community of sociologists and archivists. Publications that promoted qualitative data sharing began to appear shortly after 1994 (Corti 1995, Corti and Thompson 1998). These inspired rejoinders about the value and ethics of re-use as a research strategy (Mauhther et al. 1998, Perry and Mauhther 2004) that were then taken up in special issues of a number of journals. These debates have done much to open up the issue of qualitative data re-use to the research community. In the introduction to a special issue in Sociological Research Online, a leading qualitative researcher in the UK reflects this changing ethos.

What is particularly refreshing and useful about the articles contained in this special issue is the way that they push past the more moralistic overtones of the ‘re-use’ debate to focus instead on what happens, what is involved, what can and cannot be achieved, when sociologists get on and do it. In the process the articles give grounded and finely grained insights into the challenges but also the potential for qualitative ‘secondary’ analysis. In their different ways, the articles are qualitatively analytical about ‘re-use’ and they are engagingly reflexive in their arguments. They make the case for
using any qualitative data carefully, revealingly, and reflexively, rather than arguing that a specific set of rules applies to so-called data re-use (Mason 2007: 1.3).

A further important metric has been the willingness of funders to underwrite projects that are fully or partially engaged in data sharing. These include continuing support for the UK Data Archive and ESDS, specific initiatives designed to encourage secondary analysis of micro data (e.g., Understanding Populations Trends and Processes and the Collaborative Analysis of Micro Data Resources), and funding for a number of qualitative data sharing projects (see the Data Exchange Tools and Conversion Utilities (DExT) 2006-8, and QUADS-Qualitative Archiving and Data Sharing Scheme 2005-06 (UK Data Archive 2010). These initiatives have taken place alongside the advent of core funding for research methods and infrastructure initiatives, in particular the National Centre for Research Methods. New depositors, especially major research centres, are also an important signal about attitudes toward data sharing. The National Centre for Social Research is the largest independent social research institute in Britain, with major holdings of public policy data. In 2009, it began discussions with the UK Data Archive to plan for depositing its qualitative data. As a further example, over the past three years a steady stream of QL researchers have sought to affiliate their research with Timescapes and pursue secondary analysis of the archival resources, or to combine their primary research with secondary analysis as a way of broadening the scope of their data and providing a more robust evidence base.

However, there remain many qualitative resources that are not archived centrally but held as independent datasets by the originating teams or institutions, and, it is also the case that much archived data remains under-utilised. This is evident from a number of surveys that have been conducted to attempt to gauge the level of support for data sharing, both in principle and practice. While these often have low response rates and attitudes toward data sharing are in any case difficult to discern from these sources, they do give some indication of prevailing trends. A recent feasibility study into the co-ordination of UK data resources (UK Research Data Service, 2008) found that:

- although only a minority of researchers share data via a data centre, almost half need to access others’ data and most share data by informal means, usually peer networks.”

The current picture is clearly mixed. Evidence of sharing includes the fact that approximately 1000 data sets are downloaded each year from ESDS Qualidata and this represents only a fraction of the re-use of qualitative data, much of which still takes place informally. Also, panels on re-use are becoming more frequent in mainstream academic events such as the biennial Research Methods Festival. However, criticisms continue to be voiced. For example, a recent special issue of the Australian Journal of Social Issues (2009) devoted to data archiving reports the views of Australian researchers, some of whom oppose data sharing, and such views continue to hold sway among some UK researchers as well.

Overall, the current picture reflects an uneasy tension between pressures to share data, for the benefit of the wider research community and public good, countered by requirements to protect data and confidentiality, for the benefit of the subjects of research and also for the qualitative researchers who both generated and analysed the data. While it is no longer seen as legitimate to protect data because the originating researchers wish to have exclusive use of it, valid concerns remain about how to protect sensitive and confidential data and how to accommodate the sometimes conflicting demands of conducting primary research with the production of archive ready datasets for secondary use. The close nature of the relationship between researchers and the qualitative or QL data they generate outlined above needs to be taken into account in the way such data is curated and its re-use facilitated. Key factors in the strong move towards data sharing include: making “unmined” data available, avoiding duplication, reduced burden on research participants, greater transparency of research procedures, alignment with open access principles, and recognising that outputs of publicly funded research are public assets (Fry et al. 2008). Equally important are the concerns for protection, codified in the UK Data Protection Act 1998 (and international laws) which intend, rightly, to assure that all data sharing is done ethically. Overall, then, the current environment is challenging and complex, with many general laws, little applied case law, and researchers often subject to contradictory advice (e.g., archives demanding data sharing and research ethics committees calling for data destruction). (For key reports in this debate, see Thomas and Walport 2008, Swan and Brown 2008).

5 Complex Infrastructures for Qualitative and QL resources

Mapping the field of qualitative and QL data resources in the UK is a complex and seemingly never-ending task. Given the wealth of resources and their scattered nature, it would take a dedicated project to provide a truly exhaustive inventory. Our mapping exercise is therefore highly selective. It is probably safe to say that ESDS Qualidata, as a national resource, is a central hub in this network, especially since its incorporation into the UK Data Archive; but it is by no means the only hub, and the network is vast. In part, this is because what might count as qualitative data is so diverse – ranging from open ended responses on otherwise quantitative surveys to large holdings of historical materials, to newly emerging blogs, Twitter and other “born digital” resources. The forms of these data are also highly diverse, ranging from written and other paper resources, visual and audio materials, film and photography, through to web-based and other digital materials. Furthermore, qualitative data for social research is available in a growing number of organisations in the UK. These span libraries, museums, funders’ archives, universities, government departments, broadcasting and media archives, independent institutions and organisations, and localised collections held by community or special interest groups (for a comprehensive review see Foster 2004).

One reason for the complexity of the network is its interdisciplinary nature and the obvious attraction of bringing thematically or methodologically linked data together in special collections to increase their visibility and enable specialist curation and ease of re-use. For example, oral historians have produced extensive resources of qualitative and QL data. The Oral History Society website provides information about archives in the UK, including regional collections. A related discipline, discourse analysis, produces its own collections e.g., CHILDES – Child Language Data Exchange and TalkBank. Two major resources with ESRC funding are regionally based, with a remit to develop holdings of locality data for re-use: the Wales Institute for Social and Economic Research and Data and ARK: Access Research Knowledge on Northern Ireland. As a further example, The British Film Institute holds an extensive collection of social historical documentary films that is international in scope. A dramatic change since 2000 is the proliferation of digital content held in institutional repositories, most often affiliated with universities (JISC 2009a). Timescapes, as a specialist resource of life course data, is one such example. It is currently part of the LUDOS repository at the University of Leeds, which also holds the extensive Disability Archive. Clearing houses for open source repository data are also emerging (e.g., Open DOAR). While such repositories offer immense potential for the curation of specialist or locally generated data, our
initial investigations indicate that, as yet, holdings of qualitative social scientific data in such repositories are limited, and that, where they are held, limited metadata and searching options make these data difficult to locate.

This brief overview reveals the great complexity and diversity of qualitative data in the UK and of the infrastructures in place to manage and facilitate access to these data. The picture is one of a mixed economy with centralised, generic, national level resources existing alongside distributed specialist resources. The latter are valuable in enabling the specialist curation, discovery and re-use of data that take particular forms, are generated through distinctive research methodologies, or that have a particular thematic or substantive focus.

The UK Research Data Service recently carried out a feasibility study (UKRDS 2009) for a national shared data service, as part of which they considered a number of options for the future management of UK’s research data outputs. These included, firstly a continuation of the current proliferation of data services and resource with little change in management or co-ordination; secondly, the creation of a highly centralised agency to provide and manage all new capacity; and thirdly a co-operative service by which the UKRDS would be an enabling framework, working across a range of UK stakeholders and acting as a catalyst for new services and partnerships. The report found substantial research infrastructures existing in ‘islands’, with limited coherence and communication among them. The authors recommended a co-operative model for future development that would enable good co-ordination of existing data resources and maximum value from infrastructures and services already in place. This recommendation recognises that data is held, and will continue to be held, at a variety of levels, including project level data sets held by the originating researchers, institutional repositories, specialist archives that focus on particular kinds of data (defined by format, methodology, or thematic content) and in national level data centres. Alongside this, however, the need for some rationalisation of existing data services has been recognised, and the development of a more integrated national data service, that could encompass and oversee both generic and distributed resources, is a likely next step in the UK (ESRC 2011).

6 Future Developments

The overview presented here suggests a number of priorities for the future development of qualitative and QL archiving and data sharing. Firstly, there are practical considerations in building capacity in this area in the UK. There is a skills shortage in data curation and management, which is particularly evident given the scale of the UK network. Standards for the management of data and the production of metadata across the network are currently lacking, with metadata remaining inadequate for easy resource discovery. However, a new working group for qualitative data exchange within the Data Documentation Initiative is a promising development. Technical challenges also arise, for example, in curating and organising complex forms of qualitative data, such as audio and video formats, and in developing adequate protection of confidentiality as part of the broader ethical challenges of data re-use.

More broadly, our review of the developing ethos of data sharing and the mixed infrastructure of qualitative and QL resources suggests a number of challenges. The inherent tensions between protecting and sharing data are evident with qualitative data, and become even more acute with QL data. Various strategies may help to further the ethos of data sharing. Further mandating by funders will help, but just as important may be the rise of a new status for data as bona fide citable research outputs, even enabling researchers to receive recognition in the Research Excellence Framework for the production of data sets for sharing (such moves are being explored in the JISC Managing Research Data Programme (2010) and in DataCite (2010), among other places). Perhaps just as critical in encouraging sharing and re-use is the further development of new approaches to archiving that are more closely integrated with research processes and that build on dialogue and collaborative models of sharing. This will depend on the specialist archiving of data with distinctive formats, content or modes of generation, to run alongside and complement generic archiving and to act as brokers between the research community and the national level facilities. Such a model has been developed in Timescapes but will require follow on funding to be properly realised and tested.

There are challenges, too, in working across the mixed infrastructure identified above, to ensure that distributed resources develop in consultation with centralised resources such as ESDS Qualidata, enabling special requirements to be met but without re-inventing the wheel. The development of effective co-ordination between generic and specialist resources and across the network of resources is important, and this needs to include the development of key portals so that data resources can be easily identified, described and located. Nonetheless, there are many reasons to be optimistic, even in the face of complex challenges. Common principles for managing and sharing data across all UK Research Councils (and other funders) signal a demonstrable shift toward an ethos of sharing. Nor are such signs only in high places. Recent workshops on managing and re-using data offered by both Timescapes and ESDS Qualidata attracted hundreds of participants. The potential for combining primary and secondary analysis to broaden the scale and historical reach of qualitative and QL research and produce robust evidence for policy and practices is an exciting development that is likely to flourish over the next decade. The provision of well co-ordinated generic and specialist infrastructure to support this development is a vital next step.

References


Notes
1. Dr. Libby Bishop e.l.bishop@leeds.ac.uk; Senior Research Archivist, Timescapes Study and Archive, University of Leeds; http://www.timescapes.leeds.ac.uk/
ESDS Qualidata, UK Data Archive, U. of Essex; http://www.esds.ac.uk/qualidata/
Professor Bren Neale; b.neale@leeds.ac.uk
Director Timescapes Initiative and Archive, University of Leeds UK
http://www.timescapes.leeds.ac.uk/
2. Documents are available on the ESDS and UK Data Archive websites that give details of archiving policies and procedures for these major services (e.g., preservation, back-up and storage information) and extensive information on managing and sharing data (confidence, ethics, consent, documentation, etc.) (ESDS 2009; UK Data Archive 2008).
3. Additional articles can be found at http://www.esds.ac.uk/qualidata/support/reusearticles.asp and for further publications see http://www.disc-uk.org/publications.html#data_sharing).
4. UK and international qualitative data providers are listed here: http://www.esds.ac.uk/qualidata/access/otherdata.asp and UK and international QL resources are here: http://www.timescapes.leeds.ac.uk/methods-ethics/international-qualitative-resources/. Links to these and other data providers can be found here: http://www.esds.ac.uk/qualidata/access/otherdata.asp.