timescapes

Managing Qualitative Longitudinal Data:

A Practical Guide

Bren Neale, Rachel Proudfoot Graham Blyth, Kahryn Hughes,

and Brenda Phillips[i]

University of Leeds Institutional Repository

2016.

www.timescapesarchive.leeds.ac.uk

Contents

Introduction

Over the past decade attitudes towards data preservation, sharing and re-use have changed significantly. It is well established that Qualitative Longitudinal (QL) research yields important data of enduring historical value. The resulting datasets are usually gathered with public funds and are expensive to produce: they are best regarded as an academic and policy related resource, and there is a growing consensus that they should be carefully preserved and made available for use by future generations of researchers. This process increases the scope for building comparative research across datasets, and historicizing our knowledge and insights into the social world (evidenced, for example, through the growth in re-studies that combine secondary analysis of past datasets with the generation and analysis of new data, often some decades later). Preparing datasets for extended use also represents a transparent approach to evidence building in research, a willingness to make data available in order that scientific findings can be verified through a scrutiny of the quality, integrity and robustness of the evidence base. The need for such verification is beyond doubt.

There have been long standing debates among qualitative researchers about the value of archiving and sharing data, although concerns that others cannot understand the context and will misinterpret the data, or that archiving will compromise the confidentiality of research participants are slowly being dispelled. Even so, qualitative researchers have been slow to share data. The main barriers seem to be a misconception that they 'own' the data they have generated, and a fear that they will lose publication opportunities and be 'scooped' by others (Corti et al 2012: 9; and see wider literature cited here). But attitudes are changing and these fears and misconceptions are being allayed as the culture and ethics of sharing and longer term use becomes more widely established and accepted (see Neale 2017 for an overview of the key issues).

Archiving a Dataset.

Archiving is mandatory for research council funded empirical research in the UK, and recommended by other major funders (e.g. the Joseph Rowntree Foundation), who are increasingly prepared to provide resources to support this process. While there are no longer any requirements for research council-funded projects to archive with the UK Data Service (hereafter, UKDS), these projects are required to preserve their data for re-use within their own institutional repositories or a suitable specialist archive. They are also encouraged to contribute to a central catalogue of datasets that is co-ordinated by the UKDS.

It is worth reviewing archiving options at the outset of a project. One option for QL projects is to deposit with the Timescapes Archive. Operated by the Institutional

Repository at the University of Leeds, this is a permanent, specialist resource for the preservation and re-use of Qualitative Longitudinal (QL) data, regardless of where the data were generated and who has funded their generation. The archive accepts all qualitative longitudinal datasets that can be deposited in the English language, that are of sufficiently good quality to be of value for secondary analysis and for teaching purposes, and which are prepared appropriately for sharing and re-use. There is no charge for archiving with Timescapes. However, researchers need to document their dataset using a template provided by the Archive team. This is best undertaken as an ongoing process, building up a record of the dataset as it grows. Further information about archiving with Timescapes is provided in our companion guide for depositors (Neale et al 2017).

When to Archive?

There are no time constraints for the deposit of data and the publication of a dataset in the Timescapes Archive. As long as the checklist for good data management (see below) has been attended to, depositing data in itself is not a huge task, and researchers are free to deposit when they are ready to share their data. This may be towards the end of their funded research, or later if they are intending to continue their research or extend their analysis. This discretion is particularly important for QL researchers, who are likely to have a longer term commitment to the research process.

However, it is well worth planning for archiving as an integral part of a project, and taking the following considerations into account in deciding when to deposit data. Firstly, a dataset is an important output from a study. Archiving creates a published dataset, which is kite marked for citation by re-users and is eligible for entry in the UK Research Excellence Framework (the national process for judging the quality of funded research). Archiving enables a dataset and indeed a project to be more visible, the quality of the evidence verified, and the study promoted in ways that will significantly increase its academic impact. A re-used dataset is evidence of its quality and utility to future researchers, giving added value to the funder's investment. This process may occur after the originating team has completed their initial research, but it may also occur concurrently. In the ESRC Following Young Fathers project, for example, two affiliated researchers (Linzi Ladlow, a doctoral researcher and Anna Tarrant, a post-doctoral Leverhulme fellow) were engaged in secondary analysis of the emerging data during the life time of the project, enabling new kinds of insight to emerge and benefitting both the secondary analysts and the originating team (see www.followingfathers.leeds.ac.uk). It is worth remembering, too, that researchers are not 'giving away' their data, but gaining prestige for generating and publishing a dataset. Moreover the data are safely stored in the archive for the re-use of the originating team, as well as for sharing with others.

It is also worth considering the resources available to attend to data management and archiving, which may disappear at the end of a funded project. While data can be archived retrospectively, few researchers have the time and resources to manage this process beyond the end of their funding. For example, seeking consent from participants retrospectively is a time consuming and often difficult process if contact has been interrupted or lost; it is much better to seek consent at the outset, at the same time as seeking consent for participation in the study (and revisiting this, if necessary, for both processes as the study progresses). As a project comes to an end, researchers are likely to be involved in new research activities; they may have moved office or institution, upgraded their computing facilities or equipment, or otherwise lost track of their project and the tacit knowledge that goes with it. In these circumstances, the momentum to archive and to preserve and publish a dataset may be lost.

Finally, at the bottom line, data are more secure in a password protected archive, with additional levels of control on re-use, than in a filing cabinet, or on a computer in a research office, where they are more likely to get lost, destroyed, corrupted or inadvertently disclosed. Copyright for data held outside the archive may also be less clear; there are notorious cases of sensitive research data being subpoenaed or confiscated, for example by legal or government institutions, where they have not been protected through the licensing and registration system for archive users. Taken together, these considerations provide strong incentives to archive sooner rather than later. If data are still actively being analysed for publication, one way forward is to deposit a dataset with an embargoed period of time (usually 12 months) before it is fully published and activated for re-use.

Managing QL (Qualitative Longitudinal) Data

Good data management is essential not only for archiving and sharing, but to facilitate the longer term cumulative analysis of QL data within a research team. It is therefore worth building this into the research process at an early stage of a project, and attending to it as an integral part of a study, rather than seeing this task as an administrative one that is 'tacked on' at the end. This is also likely to be a cost effective strategy, since much of the time devoted to data management will support the longer term, cumulative use of the data by the research team, as well the process of eventual archive deposit and sharing: in other words, it serves both purposes simultaneously.

Below we set out six key stages in the successful management of QL data. It is worth bearing in mind, however, that these processes do not unfold in a neat chronological order; they are interlinked and overlapping. It is advisable, therefore, to read through all the stages at the outset, to gain an appreciation of the overall process. Note too that some of these stages will need to be repeated for each wave of data generation in

a QL study; it is a cyclical process (for more details on this and a visual representation of the process, see Bishop and Neale (2012) Data Management for Qualitative Longitudinal Researchers, Timescapes Methods Guide Series no. 17, available on the Timescapes Archive website). The process of archiving data with the Timescapes archive involves additional stages. These are set out in the companion guide to depositing data in the Timescapes Archive (Neale et al 2016).

There are no cast iron prescriptions for managing and documenting a dataset. Researchers need to use their discretion to decide how best to organise and represent their data to support their own analysis and longer term re-use. Even so, there are two broad sets of considerations that underpin these processes. The first is the need to ensure that ethical issues have been carefully addressed, so that data are protected, and that the confidentiality of study participants is respected. The second is the need to consider the authenticity and integrity of a dataset – and thereby to address important epistemological issues concerning the status and veracity of the evidence contained within the data. How researchers balance data protection and data integrity is likely to vary from team to team. Practices and protocols are likely to depend on the nature of the project, and the disciplinary traditions within which the research is conducted. The key point is that these issues need to be carefully addressed in the way that a dataset is produced (and, where possible, documented for future users – see separate guide to archiving). The Timescapes Archive recommends, as a minimum, that the cases in a study are assigned a pseudonym and that anonymised versions of interview transcripts and other files, based on these pseudonyms, are produced for deposit in the archive. Alongside the controls on access and use of data laid down by the archive, this creates a more secure system of data protection, while also facilitating re-use.

The guidance on data management set out below is not intended to be prescriptive, but to highlight the necessary dimensions of the data management process, the decisions that researchers need to make, and to offer examples of good practice. Since every dataset is unique, we have aimed here for a broad coverage of the process, and recommendations that can be adapted for the specifics of individual projects. The guidance given below should be read in conjunction with the more extensive advice provided by the UK Data Service, in particular, Corti, L. et al (2014) **Managing and Sharing Research Data: A guide to good practice.** London: Sage; and the practical guide produced by Van den Eynden and colleagues (2011). These give invaluable advice on a range of issues including confidentiality, anonymity, copyright, file formats and the secure storage and transfer of data files. Essential data management resources to guide researchers, and templates for the documents listed in the appendices here, are available on the UK Data Service Website: http://ukdataservice.ac.uk/manage-data.aspx. Ethical issues around data re-use, and issues of epistemology (the status

and veracity of data) are the subject of extensive debate (for a brief overview see Neale 2013).

Six Stages in the Management of QL Data

STAGE ONE: Planning for Data Management

Building data management into research design at the planning stage is an important first stage, involving the following steps.

1. **Assess existing data resources,** ascertain the need for new data, and appraise how a new dataset will complement existing data resources (Bishop and Neale 2012). Since not all Qualitative Longitudinal datasets are archived, it may be worth contacting researchers directly, to explore the possibilities for sharing or building on such data out-with the archive. Seeking advice from archives on existing datasets, and the protocols and possibilities for eventual deposit is a valuable part of the planning stage.
2. **Produce a Data Management Plan**. Funders may require researchers to do this as part of their application, but even if they do not, developing one at this stage is good practice. **A sample Data Management plan is provided in Appendix 1.**
3. **Appraise the resources needed for QL Data Management,** and set budgets for these resources. The UK Data Service Data Management check list is a valuable planning tool (http://www.data-service.ac.uk/create-manage/planning-for-sharing/data-management-checklist). Given the complexity of QL data and their generation over successive waves, the resources needed are likely to be higher than anticipated. Depending on the scale and complexity of the data, allowing for 40% of a dedicated role for these tasks over the life time of a project is likely to be the minimum requirement. The balance of time devoted to the task may also shift over the life time of a project, with up to six months of dedicated time needed in the final year. The UK Data Archive's Data Management Costing tool is likely to be very helpful at this early stage (http://www.data-service.ac.uk/create-manage/planning-for-sharing/costing).
4. **Identify who will take responsibility for data management.** It is good practice to build these tasks into the remit of a researcher or research manager who understands the complex data needs of a project (rather than a clerical officer, who may not), or, funding permitted, to have a designated data officer on the team who will take responsibility for this and related tasks. Whatever systems are set up, it is worth reviewing them at regular intervals; this can be done by including data management and archiving as a standing item on the agenda of research meetings, with the designated researcher/manager tasked to report on, discuss and co-ordinate progress.

STAGE TWO: Informed Consent for Archiving.

This is an essential process that needs to be thought through as part of the preparations for entering the field.

1. **Devise a leaflet and archive consent form** for use with participants. These should explain what an archive is, who has access to the archive, and what steps are being taken to protect confidentiality, including clarifying any alterations to data (see stage four, below). It is important to bear in mind that complete confidentiality can never be guaranteed, and should not be promised (Corti et al 2014). **Sample leaflets and consent forms can be found in Appendix 2. Sample leaflet needs to say in an approved data repository.**
2. **Consider issues of copyright.** Copyright is an intellectual property right that applies to research data (in written or recorded form), in the same way that it applies to literary or artistic work. Unlike other forms of intellectual property rights (patents, for example), copyright does not need to be registered, but accrues automatically once data are created. This gives the copyright owner control over the exploitation of a dataset, such as rights to copy, adapt or loan out the data, the right to communicate the data to the public, and the right to license and distribute it; it is therefore a powerful means to prevent unauthorised copying and/or publishing of a dataset. The copyright owner is the data creator (unless there is a contract that assigns copyright differently). In University settings, intellectual property rights are formally held by the institutions but, in practice, many institutions assign copyright for research materials to the researchers. For data generated in interviews that are recorded and transcribed, the researcher holds the copyright in the recordings and transcripts but each speaker is the author of his or her recorded words in the interview. This creates a somewhat grey area in terms of protecting data. In order to ensure that researchers can publish extended extracts of data from an interview and fully exercise their discretion over the use of the dataset, it is advisable to obtain a transfer of copyright from interviewees as part of the consent for archiving form. The data can then be held securely in the archive on license, with copyright retained by the researchers (for a fuller account see Corti et al 2012 chapter 8).
3. **Seek written consent from participants** for archiving at the outset of a study, and revisit this as appropriate as the research unfolds

In considering issues of consent for archiving, it is worth noting that the overwhelming majority of participants readily agree to this process, particularly when they are informed about archiving and the safeguards on confidentiality and re-use. They wish

to have their accounts on record and be part of social history, and they see little distinction between sharing their accounts with one bona fide research team or across several such teams (Neale 2018).

STAGE THREE: Generating and Transcribing Data

The potential sharing and longer term use of a dataset, including audio files, creates added pressure to ensure high technical quality and the 'future proofing' of files (e.g. through clear labelling and documentation and attention to file formats). From the outset it is worth considering how best to generate data files that are of high quality, both intellectually and technically, in order to aid transcription, analysis and re-use. Researchers often wish to listen to, or go back to an aural or video recording to get a feel for the conduct of an interview, the important aural or visual dimensions of people's speech, conversations and identities and the tenor and emotional content of their accounts or of events. The technical quality of a recording goes hand in hand with, and enhances the intellectual quality of the content. Here are the detailed steps:

1. **Review methods and technologies to produce high quality data recordings.** For the Timescapes programme we drew on standards and recommendations for equipment from the British Library Sound Archive. Fieldworkers may need training in the technical use of new equipment, as well as in methodologies for generating high quality content. Audio and video files should be quality checked as they are generated, particularly in the early stages of generation.
2. **Decide on optimal file formats for recording data.** These should allow for future proofing. File formats suitable for long-term preservation are given here: http://www.data-service.ac.uk/create-manage/format/formats-table.
3. **Create a system for labelling recordings.** Both the boxes and the media should be labelled, but it is also good practice to provide a short, aural introduction at the start of the recording itself, documenting who is being recorded, by whom, where, when and for what project. It is surprising how often recordings are preserved without this vital contextual information, making their subsequent identification very difficult, and compromising their value for long term use.
4. **Develop transcription guidelines and a template; and mechanisms for the safe transfer of data files.** Transcribers will need guidance on how researchers wish their audio files to be translated into text format i.e. what rules need to be followed to ensure preservation of the original speech. Styles of transcript are likely to vary across different disciplinary traditions (see Corti et al p 63 for details and templates). The transcription template needs to be compatible with CAQDAS analytical software packages, where these are being used. Producing a template for transcript layout, and introducing a process of quality checking on transcripts as

they are returned, are good ways to ensure consistency and a high standard. Ensure too that data security is attended to in the process of transferring files between different researchers in a team, or between the team and the transcription service. Encryption of files (the process of scrambling data to make it unreadable except to those who have a key to decode it) is one option, although it needs to be done with great care to ensure data are not lost through the process. Recorders with file encryption facilities are worth considering. **Sample transcription guidelines and templates are provided in Appendix 3.**

STAGE FOUR: Representing Data

This is another vital dimension of the process of fashioning a dataset from the data generated in the field. It addresses a fundamental issue in the use and re-use of data: how to balance ethical concerns to respect confidentiality with the epistemological drive to produce an authentic, fully contextualised dataset. **Anonymising data** – i.e. altering or substituting names of people, places, organisations employment or other identifying features, is a commonly used strategy to protect confidentiality. This can be a valuable tool that enables the contemporary sharing of otherwise sensitive or potentially revealing data, But in deciding how far to anonymise, it is worth bearing in mind that this is just one route to protecting identities, and it needs to be done in ways that do not strip data of their context and meaning. For example, a study that seeks to explore transport links across and in and out of a city would not be able to strip out identifying information about localities without destroying the value of the data. Anonymising, then, is an interpretive rather than an administrative task. It requires skill to ensure that any damage to the coherence and integrity of a dataset is minimised, particularly where a 'belt and braces' approach to the task is adopted. Early decisions therefore need to be taken about how data will be represented in a study: for analysis, for publication and display, and for wider sharing and re-use**.**

As shown above, the Timescapes Archive provides a number of safeguards on data access and re-use that are designed to protect the confidentiality of participants. However it also recommends, as a minimum, that cases in a sample are assigned pseudonyms that can be systematically applied to all fieldwork and analytical files that will make up the digital dataset for archiving. Part of the reason for this recommendation is that case-based data that are cumulatively gathered as part of a QL dataset are particularly revealing of identities and localities. The Archive also recommends that the full unadulterated dataset should be archived in digital format, for eventual release when the dataset has accrued in historical value. In effect, two digital versions of data – 'raw' (abridged) and anonymised (unabridged) - are likely to

be created, and both have value for re-use at different times and in different ways. Representing a dataset involves the following tasks:

1. **Review varied ways to safeguard confidentiality** and balance these considerations with the drive to produce an authentic, contextualised dataset. Whatever decisions are taken need to be informed and justified by a careful appraisal of ethical and epistemological considerations.
2. **Where data are being anonymised**, decide **how and when** to manage this task. The process may entail a 'light touch' approach that anonymises case names only (the minimum standard recommended by the Timescapes Archive), to a more 'belt and braces' approach that alters all or a greater range of names, places and other identifying features in a dataset. Similarly, practices vary in terms of the types of files that are anonymised, ranging from transcripts only, to a range of file types, including descriptive analytical files. Multi-media files present particular challenges for anonymising. Audio-visual data such as recordings, photos or videos are very costly, difficult, or may be impossible to anonymise, and are best stored as part of the unabridged version of the digital dataset. Decisions here depend on what files are to be included in the dataset for contemporary use, bearing in mind that descriptive analytical files such as field notes, pen portraits and case histories are likely to be of great value for historical use. A belt and braces approach to anonymisng requires a great deal of time, care and skill. Depending on the size of the dataset and the extent of anonymising, it can be a time consuming and costly task; a full scale alteration of data may be beyond the resources of many projects.

   The Welfare Conditionality project for example, gathered over 1,000 interviews across three waves of data generation. The length of the audio interviews varied – on average they were 50 minutes long. The data were initially anonymised with a light touch approach, but given the sensitivities of the data, each interview was subsequently fully anonymised. The team estimated that this process took an average of 40 minutes per interview. The task was undertaken intensively by two trained individuals over a six month period at a cost of £12,000 to the project. Identifying material had previously been marked by the transcriber with astericks, which speeded up the process somewhat.

3. A lighter touch anonymising of case names and selected identifying features, on the other hand, may be more easily managed, and can be applied across a greater range of files.

   In terms of timing, practices also vary. Some researchers anonymise systematically, as an ongoing task, following on from the creation and transcription of files, or as a bulk process at the end of fieldwork, when a full set

of transcripts are in place and the contexts in which sensitive data appear and their significance is more fully understood. The more information gathered, the easier it is to make appropriate anonymising decisions. Descriptive files such as field notes and pen portraits can be anonymised as they are created (although two versions, 'raw' and abridged, should always be produced). An 'upfront' approach saves valuable time later, when the pressures to produce findings are likely to override other concerns in a study. The files are then ready for use in publications and archiving. An alternative practice is for researchers to work solely with unabridged and fully contextualised files throughout the research process, and only create abridged versions of data later on, on a piece meal basis for publication purposes, and when preparing data for archiving. There are pros and cons with both approaches, but an upfront, integrated and systematic approach to altering data has undoubted benefits. It ensures greater consistency of approach, particularly in team-based research, while the investment of time earlier in the process pays dividends later (for examples and some salutary lessons on the implications of a piece meal approach, see Taylor 2015). Whatever decisions are taken, it is worth establishing protocols for anonymising files at the outset of a study, and building in sufficient resources to enable the process. It is also important that consent forms reflect these decisions.

4. **Identify skilled individuals or other means to anonymise data, and produce anonymization guidelines** that set out how and to what extent data will be altered. It is good practice to devise a system for changing data and flagging up such changes in transcripts. Simply removing or blanking out sensitive data is to be avoided. Pseudonyms, replacement terms, or more generalised or aggregated descriptions should be used to retain the narrative and intellectual content of the data file. For example, a person's name could be replaced with their role, or relationship with the participant (e.g. paternal grandmother, council employee). These changes then need to be applied consistently throughout the project and for subsequent transcripts. Words that have been replaced need to be flagged up in the text, ideally in the most unobtrusive way possible to preserve the integrity of the narrative. A simple bracket < > to mark text, that is not likely to be used for other purposes in a text document may suffice.

   In terms of resources to carry out anonymising, there are varied options. As well as, or instead of, utilising skilled resources in the research team, it may be possible to arrange a 'light touch' anonymising of case names through a transcription service. Using *search and replace* tools through the Microsoft office word programme is another option, but this may create unintended changes in other areas of a data file and should only be used with the utmost care. Online tools for anonymising text are also available, and can automate some of the more

straightforward changes. ICPSR, for example, hosts a web based tool called QualAnon, which is simple to use, requiring the user to upload two files: the original text, and a name key that links names to be changed with their replacement pseudonym. The tool returns the anonymised files along with a report (ICPSR 2012; see Corti et al, p. 124 for further details and references). Whatever system is chosen, anonynised files should be quality checked upon completion, to ensure minimal disruption to the integrity, flow and meaning of the content.

5. **Store 'raw' (unabridged) data files securely, along with a 'code' book** that logs any changes made to the data and provides a check for consistency across the dataset. This can be established at the start of the anonymization process and developed as the dataset grows. It is worth checking local protocols for the safe storage of sensitive files; some institutions may recommend that such data are not held digitally, but only in paper format. Either way, a complete unabridged dataset, giving the original sense of the data and the context in which it was generated, should be preserved and archived for longer term historical use. **Sample anonymisation protocols and code book are given in Appendix 4.**

STAGE FIVE: Organising Data: File Systems and Structures

This is the technical end of the data representation process. Creating a well organised filing system is important, whether for lone researchers, or collaborative research teams whose membership may change over time. It involves deciding how data files are to be stored, labelled and structured for identification, retrieval, analysis and cumulative use and re-use over time. Creating a file structure may seem, at first glance, to be a simple matter, but it entails making a range of complex decisions about a dataset, how it is constituted and how it is to be used. For example, will descriptive analytical files be included in the digital dataset? (as we have seen, they are likely to have value for longer term use). How are the two digital versions of the dataset (abridged and unabridged) to be distinguished? This is an important question, for different versions can proliferate and create instability in a dataset. It is surprisingly easy to lose track of QL data, even after a relatively short space of time, unless a well-organised system is in place. There are a number of steps in the process.

1. **Set up a shared, password-protected digital drive on a secure server** for the exclusive use of the researcher/research team, and with daily back up for data security. Create two digital versions of all files in the dataset, and clearly label

and distinguish between the original, unabridged dataset and the abridged version. The unabridged version will include all data of a confidential nature, for example, all personal details of the participants, organisations and practitioners, their contact details, their consent forms and the code book that links real to assigned identities. As indicated above, it is worth checking institutional protocols to ascertain if digital storage is permissible for sensitive files. Whatever system is set up, it needs to allow for clear labelling and differentiation between different versions of flies.

2. **Systematically label files and their contents.** A unique digital identifier can be created for each fieldwork file. In the Following Young Fathers study, for example, these show the initials of the project, followed by the case identifier (the pseudonym for the case), the wave of fieldwork in which the file was generated, the type of data (interview, life map and so on), a unique number assigned to each case file, followed by the digital format in which the file is saved (doc = word document, tif = tagged image format). The labels appear as follows: FYFAdamW2Interview06.doc; FYFAdamW1Timeline03.tif. These were compatible with the file names needed for ingesting in the Timescapes Archive; it is worth checking protocols for your chosen data repository. Whatever file names, codes and abbreviations are devised these should be consistently applied, and future proofed to guard against systems becoming obsolete with changing technology. File formats suitable for long-term preservation can be found at: http://www.data-service.ac.uk/create-manage/format/formats-table.

   As well as assigning a unique identifier to each digital file, documents such as transcripts of interviews, time maps, and other generated data can be individually labelled to give important identifying information and indication of thematic content. The easiest way to do this is to create a simple **template for a front sheet** which can be filled in and attached to each file as soon as possible after a transcript is prepared. This should specify the project name, names of interviewer and interviewee (the case); wave and location of fieldwork, and file type (e.g. time map, self-portrait, interview transcript, fieldwork note, focus group and so on). Assigning key words to the front sheet to describe the content is also to be recommended, for this enables the researcher to search these files thematically. **Sample front sheets are given in Appendix Five.**

3. **Structure files** in ways that facilitate retrieval and analysis of data by case and time (wave of data) and that allow for thematic analysis (an example is given in Appendix Six). This is a vital stage for Qualitative Longitudinal data; it is the foundation upon which to build three-dimensional analysis across cases, waves and themes. A simple system, for example, orders files alphabetically by case name/pseudonym. While a numerical identifier may be used for cases, and files

organised by number, researchers often prefer to use a narrative system that sits comfortably with the narrative orientation of qualitative enquiry. Within each case, files are then organised by wave of data collection. This brings together all files (e.g. transcripts, time maps, field notes, observations) generated for a particular case, and nests together case flies gathered at a particular moment in time. This helps to facilitate case-based analysis and the generation of chronological case histories. The same structure can be used for files stored electronically, and in the spread sheets used to ingest data into the Timescapes Archive. Alongside the organisation of data by case and wave, thematic analysis can be facilitated through the front sheets assigned to each file. The file structure is also likely to be complemented by a CAQDAS package such as NVivo, into which digital fieldwork files are entered and their contents coded and retrieved by theme.

The file structure should also include an area for descriptive analysis and fieldwork files that run across cases and/or waves – e.g. interview schedules, sample consent leaflets and forms, tables of sample characteristics, framework grids that condense data by theme, case and wave; pen portraits of participants; field notes, and cumulative case history files. As indicated above, these fieldwork and analytic files form an important and valuable part of a dataset.

STAGE SIX: Documenting and Contextualising a Study and Dataset

Alongside the system for organising data files, a good record keeping system (a data log) that provides a 'road map' to a growing QL dataset, is essential. As in all QL data management processes, the data log serves the needs of the research team as well as facilitating deposit in an Archive, for it enables the team to keep track of the burgeoning data generated through successive waves of fieldwork.

The other key task at this stage is to compile a guide to a study and dataset, which documents the research process and methodology, and how a dataset was constructed. This provides essential contextual information for longer term use and re-use. This too serves the needs of the research team, for it is surprising how quickly tacit knowledge about research methodology and experiences in the field are lost unless it is documented. The guide can also be used as the basis for describing the methodology of a project in published writings. The steps in the data documentation and contextualisation process are outlined below.

1. **Set up a data log** at the outset of your project to record and document a developing dataset. Filling in the data log is best done as an ongoing process, as a project grows, recording what data have been generated, when, where, by whom, and in what formats. Different tabs are provided for summary case

data records and the more detailed file records). (An example of a data log is provided by the Timescapes Archive – see the Guide for Depositors to the Timescapes Archive

2. **Compile a methodological guide to the study** that can be archived alongside a dataset and help to provide much needed context for researchers using the findings or data from a study. If resources allow, aim for a 'gold standard' guide, which will give a detailed account of the methodology for a study, as well as details of the sample and dataset (see, for example, Neale et al 2016, *Researching the Lives of Young Fathers: A Guide to the Following Young Fathers Study and Dataset,* Briefing paper no. 8 available at www.followingfathers.leeds.ac.uk The production of such a guide is an important way to promote a study and ensure, as far as possible, that data are not misrepresented but can be independently and appropriately understood and interpreted by re-users. A checklist of contents (adapted from Bishop and Neale 2012) includes:
    - **The project**, including title, rationale – i.e. relating to the broader socio-economic, historical, geographical, demographic and/or policy contexts in which the study was conceived; research aims, questions and design; funders and funding; and details of the research team, institutional affiliations and start and finish dates;
    - **The methodology,** including fieldwork tools (e.g. interview schedules), and templates; field work settings, timings and duration of field work; sampling decisions, recruitment, size of sample and sample characteristics; information on collaborations with practice partners and gatekeepers;
    - **The dataset,** including number of cases and waves generated and over what time scales; and decisions concerning the representation of the data, and protocols for transcribing and anonymising;
    - **Outputs,** a list of both published and web based outputs (e.g. presentations, reports and links to the project website);
    - **Emerging themes** that would be ripe for further analysis.

Concluding Reflections

The six steps outlined above are best seen as an integral part of a study. Ideally they should be attended to as a study unfolds, in a cyclical process following each wave of data generation. Once these stages are completed, it is a relatively quick and easy task to deposit data in an archive, following the guidance offered by individual facilities.

It will be abundantly clear by now that managing QL data for longer term use and re-use it is not a quick and easy task, but requires considerable forethought and planning. The challenges are acute for QL researchers, as each new generation discovers afresh. While sample sizes may be small, the accumulation of data over time creates extensive, unwieldy and complex data. The danger has been described by Pettigrew (1995: 111) as 'death by data asphyxiation: the slow and inexorable sinking into the swimming pool that started so cool, clear and inviting and now has become a clinging mass of maple syrup.' Avoiding this scenario necessitates good data management and this, in turn, relies on building sufficient dedicated resources into a project and budgeting for these at the outset. Overall, the rewards are well worth the effort. Researchers who have reached this point can congratulate themselves on the completion of their study (or this stage of its funding), and the production and publication of a high quality, well organised and contextualised dataset. This is an important output, which has been suitably prepared to demonstrate the value and robustness of the evidence base; to protect confidentiality; to preserve and sustain an important resource for posterity; to enable finely grained search and discovery of cases, waves and themes within the data; and to facilitate analysis and re-use for teaching purposes and for future generations of researchers.

References

Bishop, L. and Neale, B. *(2012) Data Management for Qualitative Longitudinal Researchers*. Timescapes Methods Guide Series, no. 17. Available on the Timescapes Archive website

Corti, L., Van den Eynden, V., Bishop, L. and Woolley, M. (2014) *Managing and Sharing Research Data: A guide to good practice*. London: Sage.

ICPSR (2012) *QualAnon tool. DSDR Qualitative Data Anonymiser. Data Sharing for Demographic Research.* http://www.icpsr.umich.edu/icpsrweb/DSDR/tools/anonymise.jsp

Neale, B et al (2016) Researching the Lives of Young Fathers: A guide to the Following Young Fathers Study and Dataset. Following Young Fathers Briefing Paper no. 8, www.followingfathers.leeds.ac.uk/findings

Neale, B (2017 forthcoming) *What is Qualitative Longitudinal Research?* London: Bloomsbury

Neale, B., et al (2017) *Archiving with Timescapes: A guide for Depositors.* Leeds, Institutional Repository, University of Leeds. (available at www.timescapesarchive.leeds.ac.uk).

Pettigrew, A (1995) Longitudinal field research on change: Theory and practice. In G. Huber and A. Van de Ven (eds.) *Longitudinal Field Research Methods: Studying Processes of Organisational Change,* London: Sage 91-125

Taylor, R. (2015) 'Beyond Anonymity: Temporality and the production of knowledge in a qualitative longitudinal study', *International Journal of Social Research Methodology* 18 (3): 281-292

Van den Eynden, V., Corti, L., Bishop, L. and Horton, L. (2011) *Managing and Sharing Data: Best practice for researchers,* UK Data Archive, University of Essex. www.data-archive.ac.uk/media/2894/managingsharing.pdf.

Appendix 1: Data Management Plan

Following Young Fathers Data Management Plan 2012.

This research benefits from its close association with the Timescapes initiative, which has been developing data management planning for Qualitative Longitudinal Research. Our plan has been developed in consultation with Dr. Libby Bishop (Timescapes and UK Data Archive/ UKDA). We will keep the plan under review as the research progresses but envisage this it will encompass a number of elements as described below.

Assessment of Existing Data. We have reviewed existing datasets available via the UKDA and the Timescapes Archive. While there are materials that can inform our study in general terms, there are no existing datasets suitable in the archives for our research questions. We have identified some existing small scale datasets held by researchers in the UK that we plan to share collaboratively as part of our secondary analysis strand of work. We plan to archive data from this study with the Timescapes Archive, as well as the UKDA, to build up resources of Qualitative Longitudinal (QL) research for re-use and to enable more refined access to and discovery of the dataset.

Information on New Data. Data will include audio files, transcripts of interviews and focus groups, and visual materials such as timelines, relational maps, and photographic records. These will be generated via in-depth interviews and focus groups and will be in open formats (.doc files formatted in word; .wav or MP3 audio files; and open format photographic files).

Quality Assurance of Data. Data sharing is integral to this research; we are committed to providing gold standard data and metadata as far as this is possible. The data will be checked, audio files transcribed, and the dataset anonymised ready for archiving. For this we will use templates developed within Timescapes that conform to UKDA standards. This will be facilitated through a data 'roadmap' - a chart which logs the data trail for each case from generation in the field through to deposit in the archive. Facilitating re-use by others as well as aiding longitudinal analysis by the primary team requires detailed documentation of the data, which is organised in two dimensions - case by case and wave by wave. This enables the building of both case analyses and temporal cross sectional analyses. In a QL context it is important that these data management tasks are done following each wave of data collection rather than at the end of the project. The metadata includes: guide to the aims and objectives of the project, including research questions; fieldwork materials, interview documents, letters and information leaflets, research diaries and fieldnotes; keywords assigned to data files and transcripts to denote content and aid data discovery. We will also consider the viability of incorporating thematic coding files (generated through NVivo Framework) and other descriptive files (e.g. pen portraits of participants) within our dataset.

Back-up and Security of Data. The data files will be stored securely and safely in well labelled and logically organised electronic files. We will develop robust procedures for version control - currently we use a system of dating files. The files will also be password protected. Secure back up files will be created using an external hard drive and the University N Drive. The data will be transferred to transcribers and for deposit in the archive using safe sticks.

Expected Difficulties in Data Sharing. Ethical considerations are of vital importance to the research and underpin our data management plan. Consent forms will include consent for archiving. To ensure consent is properly informed our information leaflets for participants will

include archiving details and potential reuse. We do not envisage any difficulty with this: evidence from our current research and that of others indicates that participants are positive about the preservation of their accounts for archiving and re-use. We will ensure confidentiality of data as far as this is possible, e.g. by providing guidance on double anonymising for secondary users. We will abide by ethical guidelines for sharing and re-using QL data that have been developed within Timescapes (in its Methods Guides Series).

Copyright/Intellectual Property Right. We will ensure that our protocols are in accordance with current copyright and data protection legislation. Consent forms will include transfer of copyright to the research team.

Responsibilities. The Research Officer will take overall responsibility for the data management tasks, including anonymisation, and will be supported by the Clerical Support Officer in this process.

Preparation of Data for Sharing and Archiving. We intend to share data from this study with other researchers through data sharing workshops during the lifetime of the project (years 2 and 3) and envisage depositing the data for preservation and re-use with both the Timescapes Archive and the UKDA during the third year of the project. We will review our data management plan at regular intervals – it will be a standing item on the agenda of our project meetings.

Appendix 1: Consent Leaflet and Form

To make sure that you agree that we can archive your interviews, we ask you to sign a consent form that says that we can do this. We will also sign the form, and we will give you a copy to keep.
The agreement covers ALL the interviews that we have already done with you, and any future interviews as part of the 'Your Space' project.
We take our responsibility to protect you from any harm as a result of taking part in our research very seriously. The form also says that you agree to give 'copyright', or ownership, of the interviews to our research team.
We are asking you to give us the copyright because this means that nobody will be able to look at your interview material without our approval and telling us why they want to see it.
The Economic and Social Research Council (ESRC) is paying for the research and the archive. They are an independent organization that funds research and training in social and economic issues, and receive most of their funding from the Government. They do not tell researchers what they should ask in the interviews.
The 'Your Space' research team is based at London South Bank University and is independent of the ESRC.
The 'Your Space' project is part of a larger research study called 'Timescapes' that is collecting interviews from a whole range of people in Britain over a five year period to look at how their lives change over time.
As you know, our particular project follows changes in young people's relationships with their sisters, brothers and friends as they grow older ~ including you!
As well as writing reports, books and magazine articles, we would like to store all the interviews and activity sheets in an archive so that in the future other people will be able to look at them. It will help them to understand what life was like for people at the start of the 21st century.
This leaflet gives you information about archiving so that you will understand what is involved. If you think of any more questions about the archive after you've read this, then do get in touch. Our contact details are on the back of this leaflet.

YOUR SPACE! SISTERS, BROTHERS AND FRIENDS PROJECT

A LEAFLET EXPLAINING HOW WE PLAN TO ARCHIVE YOUR INTERVIEWS

YOUR SPACE! SISTERS, BROTHERS AND FRIENDS PROJECT

We would like YOUR permission
to store your interviews and
activity sheets.

FAMILIES & SOCIAL CAPITAL RESEARCH GROUP
London South Bank University
103 Borough Road
LONDON
SE1 0AA
www.lsbu.ac.uk/families/yourspace
ROSALIND EDWARDS
Tel: 020 7815 5795
Email: edwardra@lsbu.ac.uk
SUSIE WELLER
Tel: 020 7815 5811
Email:wellers@lsbu.ac.uk

ement to archive

sure that you agree that we can archive your interviews, we
o sign a consent form that says that we can do this. We will
the form, and we will give you a copy to keep.

ement covers ALL the interviews that we have already done
, and any future interviews as part of the 'Your Space'

our responsibility to protect you from any harm as a result
part in our research very seriously. The form also says
agree to give 'copyright', or ownership, of the interviews to
rch team.

asking you to give us the copyright because this means that
will be able to look at your interview material without our
and telling us why they want to see it.

funding the research and archive?

nomic and Social Research Council (ESRC) is paying for the
and the archive. They are an independent organization that
search and training in social and economic issues, and receive
their funding from the Government. They do not tell
ers what they should ask in the interviews.

r Space' research team is based at London South Bank
y and is independent of the ESRC.

ct us:

ARDS
5795
a@lsbu.ac.uk

5811
lsbu.ac.uk

FAMILIES & SOCIAL CAPITAL RESEARCH GROUP
London South Bank University
103 Borough Road
LONDON
SE1 0AA
www.lsbu.ac.uk/families/yourspace

WHAT'S GOING TO
TO ALL THOSE INTE

A LEAFLET EXPLAINING HOW WE PLAN TO ARCHIVE YO

The 'Your Space' project is part of a larger research
'Timescapes' that is collecting interviews from a whole rang
Britain over a five year period to look at how their lives chan

As you know, our particular project follows changes in
relationships with their sisters, brothers and friends as the
including you!

As well as writing reports, books and magazine articles, we
store all the interviews and activity sheets in an archive
future other people will be able to look at them. It will
understand what life was like for people at the start of the 2

This leaflet gives you information about archiving so
understand what is involved. If you think of any more questi
archive after you've read this, then do get in touch. Our d
are on the back of this leaflet.

timescapes

LONDON S
UNIVERSITY

timescapes

An ESRC Qualitative Longitudinal Study

Consent Form for [name of project]

Appendix 2: Transcription guidelines and template

Project Information:

Programme name:	Pioneers
Project name:	Family Life and Work Experience Before 1918
Depositor name:	Paul Thompson

Interviewee Information:

ID number:	2000int002
Title:	Mr
First name:	John
Surname:	Stinchombe
Date of birth:	1886
Gender:	Male
Marital status:	Married
Occupation:	Farm Labourer

Interview Information

Interviewer first name:	Paul
Interviewer surname:	Thompson
Transcriber first name:	Jane
Transcriber surname:	Smith

Anonymisation:	Permission granted to use real names-P. Thompson 2007
Place of interview:	Colchester
Date of interview:	23 February 1976
Number of tapes/files:	2
Type of recorder:	Marantz Model 007 cassette audio recorder
Consent status:	Consent form signed for participation and archiving
Keywords:	fatherhood, siblings, turning points, historical events

[Start of Tape 1, Side A]

I: If I could ask you when you were born?

S: 1896.

I: Whereabouts was this?

S: Over at Horton, Woodley Arton.

I: Is this a village or is it -?

S: Oh no, I was out on the wilds as you might say. I had the farthest for to go to school. Two mile to walk.

I: Had you any other houses round about?

S: Well the farmer as I worked for thirty years. I was born in a cottage about 300 yards away as the crow flies. And I went to live in the old house where, as he bought afterwards. When the {ILLEGIBLE} estate was sold up. The farmers, they worked hard before that see. One of his brothers lived in that house and he was born there. And his uncle lived there before that. And then there was another brother down at his mill farm. Well he down at mill farm he was more boss than the one that lived in this old thatched house.

I: What was your father's job?

S: Farm labourer. Like meself.

I: It was the farmer's cottage that you had then?

S: That's only a cottage on the farm. That's the last cottage in the lane. In them days.

I: How large was it?

S: What the cottage? Well this was two up and two down. Then where I went to live, after the war, before I got married, cos I was married in 1915, and I went to this other old farm house in 1918. And then we stopped in there until 1948 when I shifted over to Old Sodbury on account of the wife cos if she wanted anything extra we had to come to Sodbury for to get it. And that was no unthankful office especially on a bike, a push bike. When you get out on Sodbury Common you know which way the wind is blowing. And I used to come across there of an evening if we wanted summat, extra. Sometimes I have had to stand on the pedals and I had to get off and walk when the wind was so rough all against me.

I: Yes, bad isn't it.

S: Oh it is.

I: Had you a garden at your first cottage where you were brought up?

S: Oh that was two gardens as you might say. One was one side of the road a three cornered piece and the other was adjoining the cottage. That wasn't what, {blank} oh I suppose all together it wouldn't be half an acre. Then when I went to live in this old thatched farm house well that was half an acre of garden to say nothing about the lawn and all that. And twelve month before we left over there I had the first prize for the garden. And he was twelve perch long. And wasn't the full length of the garden as was dug up see. And that was 96 rows of 'taters, two feet apart. And sometimes when the onions and all that during the war, well I didn't plant, the biggest part were onions and the rest of em were shallots. And sell 'em. For to get the extra shilling.

I: Did your father grow vegetables as well?

S: Oh yes.

I: What sort of things did he grow?

S: Some of them all sorts.

I: I see. Had you any fruit?

S: In this other, in this other old thatched house, it used to be a farmhouse in years gone by. Well I put it down to 300 year old but I was a long way out as it was more than 1600.

I: Yes, very old.

S: That was six rooms on him, three up and three down. Terrible great beams they were anyone wouldn't realize it.

S: And when we went into it, I had to put some thatch on first because the estate lot see let it go. And I put some on there because the thatcher that used to do all the houses thatching he was away at timber cutting. When he came home well the farmer had it thatched all over see then. He done it once afterwards. While I was there. When I left there, the fellow that took my job see, both of them was a little bit touched, I am sorry to say, and the old doctor said afterwards when the master, he came and see the missus, Mrs Stinchcombe, {BLANK} imagine that, and {ILLEGIBLE} they come from the beans. It was a darned great big whitened passage and {BLANK} just inside the front door and then there was the baking oven adjoining that see.

[End of Tape 1, Side A]

Appendix 4: Anonymisation guidelines and Code Book

Code Book (taken from the Timescapes Making the Long View Project)

Interview/Page	Original	Changed to
Int3 (ff2)/page		
3/1	Spain	European country
3/1	Salou	Holiday clubbing resort
3/1	20th June	June
3/1	Kenny	Ian
3/1	Julie	Mandy

Sample anonymisation log (source: UK Data Service Website: Guidance on anonymization).

Interview and page number	Original	Changed to
Int1		
p1	Age 27	Age range 20-30
p1	Spain	European country
p3	Manchester	Northern metropolitan city or English provincial city
p2	20th June	June
p2	Amy (real name)	Moira (pseudonym)
Int2		
p1	Francis	my friend
p8	Station Road primary school	a primary school
p10	Head Buyer, Produce, Sainsburys	Senior Executive with leading supermarket chain

Appendix 5: Front sheet for Data Files

Appendix 6: File Structure (used for the Following Young Fathers filing system).

1. **Following Young Fathers: Unabridged Dataset**
2. **Following Young Fathers: Abridged Dataset**

Unabridged Dataset

Fatherhood Case Files

Practitioner Case Files

Research Instruments (recruitment and consent leaflets, Interview schedules)

Analytical files (pen portraits, case profiles, sample characteristics)

Confidential files (Participants lists, contact details, Consent forms).

Fatherhood Case Files (arranged by case, then wave, then case documents)

Adam

Wave One

Interview

Time Map

Fieldnotes

Wave Two

Interview

Time Map

Fieldnotes

Andrew

Wave One

Wave Two

Ben

Wave One

Wave Two

An example of a case-led folder structure

Abridged Files

- **Adam**
  - **Wave 1**
    - Interview
    - Life map
    - Fieldnotes
  - **Wave 2**
    - Interview
    - Life map
  - **Analytical files**
    - Fieldnotes
    - Pen Portrait
- **Anthea**
  - **Wave 1**
    - Interview
    - Life map
    - Fieldnotes
  - **Wave 2**
    - Interview
    - Life map
    - Fieldnotes
  - **Contextual Files**
    -

- **Unabridged**
  - **Adam**
    - **Wave 1**
      - Interview
      - Life map
      - Fieldnotes
      - Interview audio-file
      - Consent form
      - Contact details
    - **Wave 2**
      - Interview
      - Life map
      - Fieldnotes
      - Interview audio file
      -

Wave 1

- Interviews
- Life maps
- fieldnotes

Wave 2

- Interviews
- Life maps
- Fieldnotes
  - Anthea

[i] **Bren Neale** is emeritus professor of life course and family research in the School of Sociology and Social Policy at the University of Leeds and formerly director of the ESRC Timescapes Initiative. **Dr. Kahryn Hughes** is senior research fellow in SPS at Leeds. She was a co-investigator with Timescapes and director of the Changing Landscapes for the Third Sector study, as part of which she played a leading role in the phase two development of the Archive. She has recently taken over the directorship of the Archive. Graham Blyth and Rachel Proudfoot (Data Management team at the University of Leeds) have played a key role in the development of this resource and its management as part of the University of Leeds repository. **Brenda Phillips** was initially employed as the Data officer for Timescapes and subsequently joined the University of Leeds Institutional Repository team, where she supports a range of data management functions, including the work of the Timescapes Archive.

We are grateful to Dr. Libby Bishop, UK Data Service and formerly Timescapes Archivist, for her continued support for the Archive and her advice on the compilation of this guide.