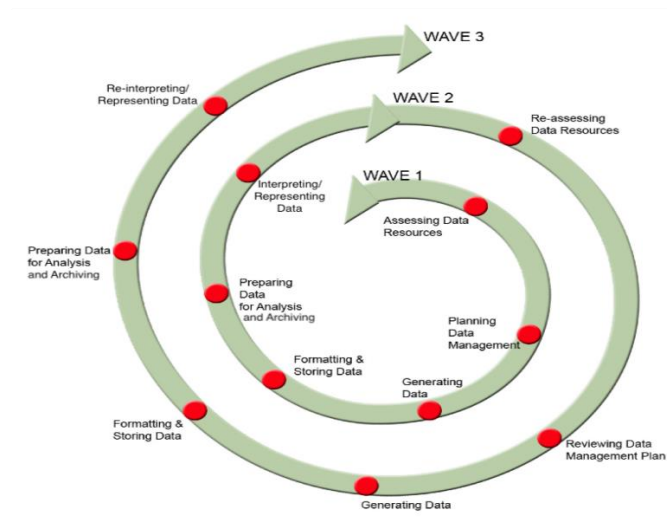timescapes

Data Management Planning:

A Practical Guide for Qualitative Longitudinal Researchers

Bren Neale with Kahryn Hughes[i]

University of Leeds Institutional Repository

2020 updated edition

Contents

Appendices

Anyone who has carried out longitudinal field research … will know that the central problem is dealing with complexity. … There is no release from the overwhelming weight of information, from the task of structuring and clarifying, from the requirements of … conceptualisation. The result is 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. (Pettigrew 1995: 111)

Now all this talk of filing may sound very boring – a long way from the humanistic concerns this book champions! But the fact is that filing – as C. Wright Mills once observed – is indeed an 'intellectual production.' … [A] good filing system harbours your intellectual life. (Plummer 2001: 152)

Introduction

As our opening quotation from Andrew Pettigrew graphically describes, and as each generation of researchers discovers anew, the challenges of managing and working with Qualitative Longitudinal (QL) data are acute. QL datasets are likely to include data transcripts, audio and/or video files, and participatory data (e.g. life maps, diaries, drawings, photographs), along with field notes and analytical case and thematic files that are likely to run over several waves of data generation (detailed in Neale 2021). While QL studies may be 'small' in terms of sample size, the resulting datasets are extensive, unwieldy, complex and continually open to change and growth.

Van de Ven and Huber (1995: xiii) wryly observe that, 'over time, data mount astronomically and overload the information processing capacity of even a trained mind'. For example, a study of the changing citizenship experiences of 65 young people, gathered over three waves of fieldwork, yielded over 4,000 pages of interview transcripts (Smith 2003), while Ferguson et al. (2019) estimate that each of the larger case files produced for their ethnographic study amounts to around 200,000 words. For larger-scale qualitative panel studies (e.g. the *Welfare Conditionality* study, which

generated over 1,000 interviews over three waves of interviews), the size of a QL dataset will be all the greater. Keeping track of data in waves, cases, themes and formats is all the more challenging when working with data of such scale and complexity. While the danger of data deluge or asphyxiation should not discourage researchers, it is a timely reminder that good data management is vital. It relies, as Plummer notes above, on the disciplined intellectual task of 'filing' along with a range of associated data activities. For QL research, diligent data management is the necessary grounding upon which the analysis and longer-term use of data can be built. It is the vital counterpart to the flexibility and creativity of the research process.

This practical guide sets out six steps to good data management. These steps will enable researchers to produce a well-organised, sustainable QL dataset for their own future use and for use by others. While the focus is on managing QL datasets, the guidance will also have relevance for researchers who are working with qualitative data more generally. The discussion is based on recent guidance and good practice in the UK, provided by the UK Data Service (Corti et al. 2020). It is also grounded in the work of the *Timescapes Archive*, hosted at the University of Leeds in the UK. This is a specialist repository for QL research data, and an international resource with scope for further growth and development. It is also a relatively new resource that embodies recent principles of good practice.

It is worth noting here that research funders in the UK, the USA and across Europe are gradually implementing data management and sharing policies in order to maximise openness of data, and transparency and accountability in research design and practice (Corti et al. 2020). While the guidance provided here is rooted in UK practice, it reflects

broader, international efforts to promote and implement good data management for qualitative data resources.

Working with Qualitative Longitudinal Data

For QL researchers, good data management serves a dual purpose. First, it facilitates the analysis and re-analysis of a dataset that takes time to grow, and which the original research team may wish to visit and revisit over many years. Clearly if a longitudinal dataset is to be of value it needs to have a long life. Second, good data management enables the archiving, sharing and re-use of a dataset for the benefit of new generations of researchers (Bishop and Neale 2012; Neale 2013). Preserving QL data creates legacy data that has considerable historical worth. This benefit has become more apparent over the past two decades as the practice of re-using qualitative data sources has begun to flourish (Hughes and Tarrant 2020; Neale 2021).

The fact that qualitative and QL datasets are usually gathered with public funds, and are relatively expensive to produce, adds further weight to the need to preserve them. They are best regarded as an academic and policy related resource, rather than the preserve of the original researchers (Neale 2013; Hughes and Tarrant 2020). 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. Two decades ago, debates about the use and re-use of qualitative datasets were in danger of becoming polarised (Moore 2007). However, attitudes are changing. The concern with whether or not

legacy datasets *should* be used has given way to a more productive concern with *how* they should be used (Neale 2017; 2021).

The broader context and rationale for data sharing and reuse, outlined above, makes it imperative for researchers to build in sufficient time and resources to manage their project data effectively. The process is best attended to in a systematic and co-ordinated way, as an integral part of an evolving study, rather than as an administrative task that is simply 'tacked on' at the end. Indeed, while data management has an important practical dimension, it also has important ethical, epistemological and legal dimensions. In other words, it is an interpretive process that requires careful planning and appraisal (McLeod and Thomson 2009: 133; Neale 2013).

The importance of managing QL research data ethically is sharpened by the fact that these data are inherently personal. The accumulation of biographical data through the waves of fieldwork creates a unique fingerprint for each participant that is increasingly revealing of their identities. Moreover, personal contact details, that directly identify participants, are vital parts of a QL dataset; they are needed to maintain contact with participants over the course of a study, and to facilitate longer-term follow-ups that may run for decades (Neale 2021). The highly personal, biographical nature of QL data places particular responsibilities on researchers to adopt robust ethical protocols for their safe preservation. Following the European Union *General Data Protection Regulation* (GDPR), this imperative has been integrated into UK law in the Data Protection Act 2018. As Corti et al. (2020: 160) note, the inappropriate sharing of data, exposed in the recent Facebook and Cambridge Analytica scandal, reinforces the need for robust safeguards for the protection of personal research data. It is reassuring to

know that the ethical and legal sharing of such data is perfectly possible where such safeguards are in place (Corti et al. 2020: 160; for a more detailed discussion of data leaks, see Hughes and Tarrant 2020).

Overall, the management and preservation of a dataset is an important aspect of research quality, a task that requires foresight and imagination. QL researchers need to work out in advance how data can best be managed through the cyclical waves of a study, while also maintaining some flexibility to respond to changes in the research and data landscapes. Six steps to good data management are set out below. They include three key areas where future planning is vital: gaining consent for archiving, ethically representing data, and preparing data for archiving.

Managing QL Data: Six Steps

The six steps to good data management outlined here draw on the detailed and comprehensive guidance provided for social scientists by the UK Data Service (Corti et al. 2020; see also the complementary resources on the UK Data Service website, www.ukdataservice.ac.uk/manage-data.aspx).These sources give invaluable advice on a range of legal, ethical and technical issues, including anonymising and confidentiality, file formats, and the secure storage and transfer of data files. Since this is a fast-moving field, with new technologies, practices and tools continually coming on stream, websites are likely to offer the most up-to-date advice. For advice and resources beyond the UK, researchers are advised to check local guidance, policies and protocols.

The guidance provided here is tailored to the particular challenges faced by researchers who are working with QL data, although, as noted above, it will have wider

relevance for qualitative researchers in general. However, this advice is not set in stone. Good practice will evolve in a QL study and will, in any case, vary according to the nature of the project, its disciplinary roots and any protocols laid down by funders, institutions and archives. Since every dataset is unique, the recommendations set out below are not intended to be overly prescriptive, but to highlight the broad dimensions of the process and the decisions that researchers need to make, along with examples of good practice that can be adapted for individual projects.

The varied tasks involved in good data management follow the data life cycle (Bishop and Neale 2012; Corti et al. 2020). While they are presented here as a series of steps, in practice these tasks do not unfold in neat chronological order; they occur in parallel and are interwoven with the processes of generating and analysing data (Neale 2021). For QL research, the data life cycle is best imagined as a spiral, a continuously spreading, cumulatively increasing process, rather than as a cycle of activity that is relatively stable and repetitive.

Figure 1: The Qualitative Longitudinal Data Spiral
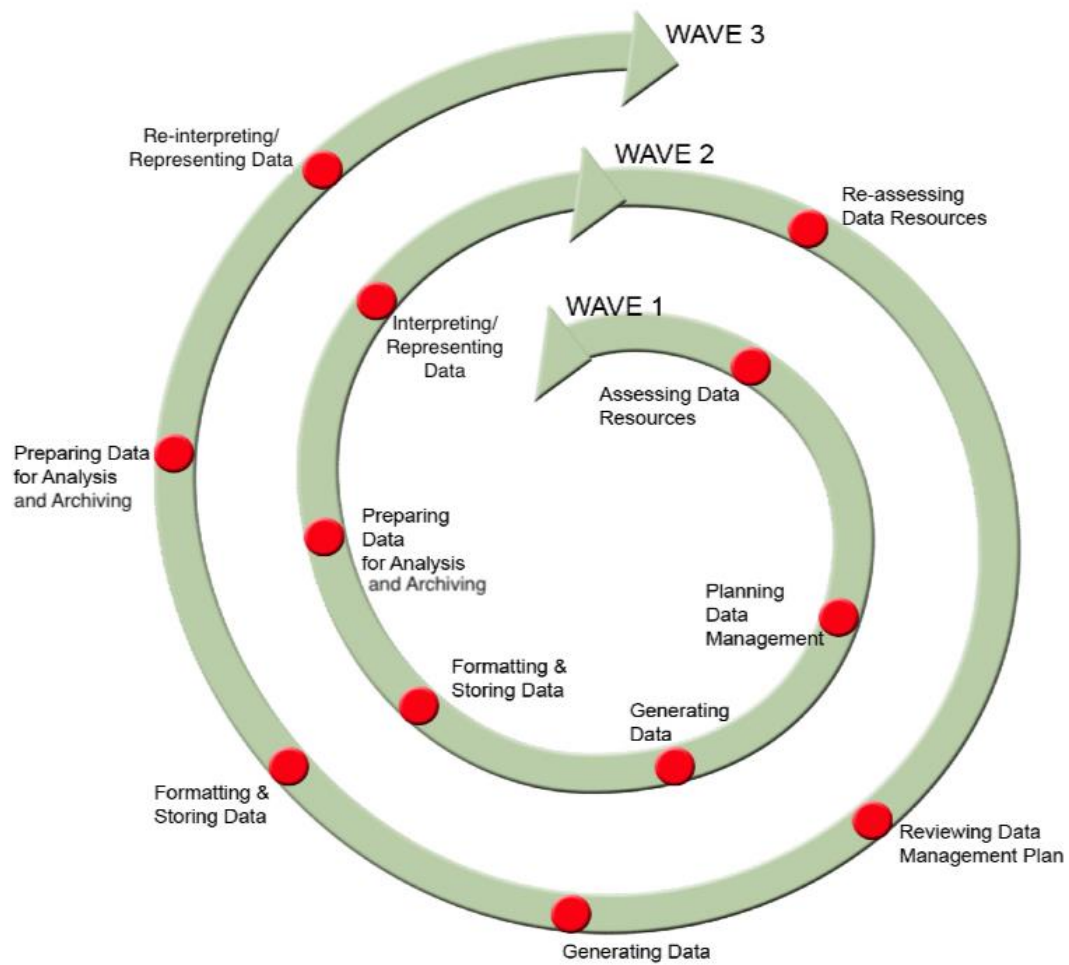
(Source: Adapted from Bishop and Neale 2012)

Step 1: Data Management Planning

The planning process helps researchers to develop a strategy for managing a dataset over the lifetime of a project, and being clear about what the process is aiming to achieve.

❖ At the design stage, produce a **data management plan** (see **Appendix 1** for examples). While this is usually a requirement laid down by funders (with templates provided), it is good practice for all researchers. The UK Data Service data management checklist is a valuable planning tool (see [www.ukdataservice /manage-data/plan/checklist](www.ukdataservice/manage-data/plan/checklist); Corti et al. 2020: 49–50). The plan can be reviewed at regular intervals and refined where needed. It should include:

- **A data-planning chart** that sets out the forms of data to be generated, from what sources and over which waves and timescales. This will help to gauge the extent and complexity of the dataset.
- **An assessment of pre-existing data resources**. Is there a need for new data? How will a new dataset fit with existing data resources, and what is the scope for bringing datasets together to broaden and historicise the evidence base?
- **An appraisal of data management resources**. Identify who will carry out data management tasks, and set a timetable and budget for this ongoing work. The UK Data Service planning and costing tools are likely to be very helpful here (www.ukdataservice.ac.uk/manage-data/plan.aspx). For a cyclical and cumulative QL project, time needs to be factored in at various points in the project schedule; it is easy to underestimate the resources needed.

❖ If working as part of a team, build data management and co-ordination tasks into the **remit of a skilled researcher or research manager** who understands the data needs of a project. Make this a standing item on the agenda of project team meetings.

❖ **Identify and liaise with potential data repositories**, who will advise on good practice for the management of a dataset, and on the steps needed for depositing data.

❖ In order to keep track of a burgeoning dataset, **set up a data log,** a record-keeping system that can document the dataset as it grows. The log should record all forms of data, including contextual data such as interview schedules and consent forms, and descriptive analytical files such as field notes and pen portraits. The log should indicate when, where and from whom files were generated, and who produced them, and should distinguish between abridged (anonymised) and unabridged (raw) files (for detailed guidance on documenting data, see www.ukdataservice.ac.uk/manage-data/document) . Such metadata (data about data) is vitally important for both the originating team and for future users of a dataset. The *Timescapes Archive* provides a spreadsheet as a data log, which is also used to ingest data into the archive (See **Appendix 2**).

Step 2: Informed Consent for Archiving

Gaining consent for data archiving requires advance planning; gaining consent of any kind can be time-consuming and difficult, particularly if contact with participants has been interrupted or lost. For this reason, it is now regarded as standard practice to seek consent for archiving alongside the general consent process prior to fieldwork (Corti et al. 2020). In a QL research context, consent can be seen as an ongoing process rather than a one-off event. It can be revisited with participants at key moments in an unfolding project, and before data are finally deposited in an archive (see Neale 2021, Chapter 5).

At this point, it is worth clarifying exactly what participants are being asked to consent to in relation to the future use of their data. Can consent for archiving be properly informed when future uses of a dataset (lines of enquiry, interpretations, contexts for

re-use) are unknown and cannot be anticipated? For an evolving QL study, where the themes under investigation may radically shift over the course of a study, the notion of informed consent is problematic in any case. It can't be specified with any certainty for the work of the original researchers, let alone for the work of others, who may subsequently use a dataset for a different purpose (see Neale 2021, Chapter 8 for a more detailed discussion).

One way forward is to reframe the consent process in terms of **enduring consent**. This generic or 'blanket' model of consent clarifies that data may be used in new research contexts, in relation to unanticipated themes, over unspecified timescales, and without the need to re-contact participants in future. **Procedural forms of consent**, i.e. consent for the way that data are to be stored, managed, accessed and generally protected, can be specified in advance and then built into enduring consent. Enduring consent is commonly used in medical research, and has obvious utility for longitudinal enquiry and for the future use of archival data (Neale and Bishop 2012b; Neale 2013; ESRC 2016; Hughes and Tarrant 2020).

It is also worth noting here that, for QL studies at least, the vast majority of participants readily agree to the archiving of their data (Neale 2013). They wish to have their accounts on record as part of social history. It is worth clarifying with participants that, unlike popular archives, research archives are not publicly available resources. Access can be carefully controlled and restricted to those working in officially recognised research institutions, who will pledge to uphold high ethical standards. In these circumstances, participants see little distinction between sharing their accounts with one bona-fide research team or several such teams through the medium of the

archive (Kuula 2010–11; Weller 2012; Neale and Bishop 2012a). Kuula's research on participant perceptions of archiving revealed that, 'participants perceive open access to research data for other researchers as self-evident' (Kuula 2010–11: 15). This was certainly our experience within the *Timescapes Study*, where over 95% of participants consented to archiving. Even where data are virtually impossible to anonymise, or are highly sensitive, consent is not usually withheld. Classic examples include data on the foot-and-mouth outbreak in selected areas of Cumbria, and Seymour's research on end-of-life care (www.ukdataservice.ac.uk/use-data/guides/dataset/foot-and-mouth.aspx; www.ukdataservice.ac.uk/deposit-data/stories/seymour.aspx). Re-thinking the ethics of consent for re-use in these ways may help researchers to move away from an unnecessarily protective stance in relation to their participants.

- ❖ Before entering the field, **devise a leaflet and consent form for archiving in an approved data repository** (See **Appendix 3** for an example; see also Corti et al. 2020 and www.ukdataservice.ac.uk/consentfordatasharing/. Consent should be audio or video recorded if written consent is difficult for participants.
- ❖ **Assess the requirement for confidentiality**. It is important to ascertain if participants wish for their personal information to be obscured in publications or archives, or whether they wish to be represented in their own right (for the ethics of this, see Neale 2021, Chapter 5). Both options can be built into the consent form. For disciplines such as anthropology and oral history, which share important affinities with QL enquiry, it is customary to seek and secure consent for an open sharing of personal information (Corti et al. 2020: 196).
- ❖ **The wording of consent should be as user-friendly, open and flexible as possible**, in order to comply with data protection legislation, and to meet the requirements of enduring and **procedural consent**. This is particularly important in QL research because studies are continually evolving, future data

uses are unknown, and some finessing of data protocols and practices may well be needed down the line. The consent leaflet should:

- Clarify how confidentiality will be addressed, if this is required (i.e. through archive access controls and anonymising data), bearing in mind that complete confidentiality cannot be guaranteed and should not be promised;
- State how/where research data will be stored and preserved for the longer term. As for all longitudinal enquiry, it is important to seek consent to store personal contact details, and to clarify the extra safeguards that will be used for these data (encryption, embargo, and storage in a dark archive);
- Make clear that data may be shared for future research or teaching purposes;
- State that participants have the right to seek to access their data in future, and/or to alter or erase their data from a study and/or archive;
- Give the name and details of a person that the participant may contact with queries about their data. This may be the researcher as data producer/depositor; the institution as data controller (with legal responsibilities for data use); and/or the archive as data curator/steward.

❖ A related task at this point is to consider copyright issues. 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 works. 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 holder control over the use of a dataset, including 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 or publishing of a dataset.

For data produced in interviews that are recorded and transcribed, the researcher holds the copyright in the transcripts and recordings, while participants hold copyright in their recorded words (Corti et al. 2020: 223). This creates a somewhat grey area in terms of protecting data. Archives generally recommend transferring copyright to the research team, so they can control how a dataset is used. However, Haaker (personal communication) suggests the value of seeking joint copyright for researcher and participant. This still

enables researchers to quote freely from a dataset, and to transfer it to an archive with a licence arrangement. But, in the spirit of participatory research and shared authority, participants can also use and control the use of their data, and publish their diaries and memoirs in their own right.

Step 3: Producing High Quality Data Files

This is the initial phase in the production of a polished, well-documented dataset. Audio and video recordings produced in the field need to be of high technical quality, for both immediate and longer-term use. High quality recordings will not only aid the transcription process; they are of great value for researchers who wish to get a feel for the conduct of an interview, its interactive dimensions, the oral dimensions of people's speech and identities, and the tenor of their accounts. Researchers who work only with transcripts (an increasingly common practice, since this is an easier and faster option) are likely to miss these nuanced and revealing dimensions of the data, for they are commonly ironed out or obscured in the process of committing an interview to a two-dimensional page (Alldred and Gillies 2012; Bornat 2020). The technical quality of a recording therefore goes hand in hand with and enhances the intellectual quality of the content.

The long-term accessibility of data depends on the quality and durability of the storage medium, the availability of data-reading equipment for that medium, clear labelling of the files, and high quality transcriptions that produce a written version of an audio file (Corti et al. 2020). All the tasks outlined here can help to generate high quality, well-documented data that are 'future proofed' as far as possible.

Producing, Labelling and Transcribing Data Files

❖ Review methods and equipment for **producing audio/video recordings and photographs of high technical quality**. This may entail seeking expert advice on sound or video recordings, or photographic techniques; reviewing recording equipment; and building in fieldworker training. It is good practice to quality-check recordings as they are produced.

❖ Consider the quality and future proofing of your chosen **storage media** (optical, magnetic or solid state media, or non-digital paper-based media) and how to create conditions that can optimise their longevity. (For a review of these different media and their nature of their vulnerabilities, see Corti et al. 2020, Chapter 6 and www.ukdataservice.ac.uk/manage-data/store/storage.aspx).

❖ It is good practice to preserve data files in more than one medium (for example, on a hard drive and a CD), and to copy or migrate data to new media every 2–5 years. Different institutions are likely to have their own policies for the safe storage of data, which researchers will need to follow. Since the storage of paper archives is becoming increasingly difficult for institutions and repositories, aim to create a fully digitised version of a dataset, including digitising handwritten, paper-based files such as drawings and letters.

❖ Consider what **digital file formats** you will use for your files and apply these consistently. They may be based on widely used proprietary software, such as MS Office, which is likely to have a long shelf life. But open, non-proprietary software (e.g. PDF/A, TIFF, XML) are less likely to become obsolete and should be considered for longer-term storage. File formats suitable for long-term preservation are listed at: www.ukdataservice.ac.uk/manage-data/format/recommended-formats.

❖ **A file-naming system** should clearly identify and distinguish between files, and between different versions of files, including abridged and unabridged versions of a dataset. A working copy of a dataset should also be clearly distinguished from backup versions. Careful labelling will ensure that different versions do not proliferate and cause instability in a dataset.

▪ In the *Following Young Fathers* study, a **unique digital identifier** was created for each data file (e.g. FYFAdamW2Interview06.doc; FYFAdamW1selfportrait03.tif). These file names start with the initials of the

project, followed by the case identifier (the name/pseudonym for the case); the wave of fieldwork in which the file was generated; the type of data (interview, life map, diary, field notes, and so on); a unique number assigned to each file in a particular case, followed by the digital format in which the file is saved (doc = word document; tif = tagged image format). It is worth checking protocols with your data repository; in this instance, the file names are suitable for ingesting in the *Timescapes Archive*.

❖ Create a system for systematically labelling and identifying data files (recordings, transcripts and other field data) and flagging up their content. For recordings, label the media themselves, as well as the packaging (by allowing space to record an audio or video introduction at the start of a recording). For transcripts and other documents, this labelling can be done by creating a one-page template for a front sheet which can be filled in and attached to each file as it is produced. The sheet should specify the project name, names of interviewer and interviewee (the case), date and location of interview, wave of fieldwork (temporal data) and file type (e.g. life map, interview transcript, diary). Assigning key words to the front sheet gives a useful indication of the thematic content of the file, ensuring that it can be accessed by theme, as well as by case and wave (see **Appendix 4** for an example).

❖ **Consider how to transcribe audio data files and who will undertake this**. Styles of transcription are likely to vary across disciplines and country contexts. Producing transcripts engages the researcher in a close and systematic reading of the data. Achieving such familiarity with a dataset is an important phase in its analysis. If you are using a transcription service, this same familiarity can be achieved by quality checking the transcripts against the recordings. The amount of time needed will vary, but some studies allow up to an hour of researcher time to check a transcript (this is considerably less time-consuming than the 2–3 hours needed to transcribe a one-hour interview; Dawson 2003).

❖ Develop **transcription guidelines and a template**, setting out the rules to be followed in translating audio files into text files (see **Appendix 5** for an example; see also Corti et al. 2020 and www.ukdataservice/manage-data/format/transcription.aspx; www.ukdataservice.ac.uk/media/622355/UKDA-example-transcription-instructions.pdf). Aim for a rigorous, orthographic

reading that preserves as much as possible of the original speech and the nuances of the interaction. Like all data management tasks, transcribing is not neutral. It is an interpretive act that may iron out and lose some of the richness and meaning of the original speech (Alldred and Gillies 2012). Transcription templates need to dovetail with anonymisation protocols (see below) or, if required, flag up sensitive information that needs to be anonymised. They also need to be compatible with qualitative data analysis (QDA) software, if this is being used.

❖ Produce a **Transcriber Confidentiality Agreement**, including protocols for the safe deletion of all files held by the transcriber once their work is completed (see **Appendix 5** for an example).

❖ Alongside a system for **quality checking transcripts** as they are produced, develop **protocols for the safe transfer of data files** to transcribers and archivists. File encryption (scrambling data to make it unreadable, except to those who have a key to decode it) is advisable for data on the move, although care is needed to ensure that the key is secure and that data are not lost through this process. Recorders with file encryption facilities are worth considering. A review of encryption practices and software, including *Pretty Good Privacy* (PGP), can be found in Corti et al. (2020: 144–8). Since encryption conforms to the standards laid down by the GDPR for the safe handling of personal data, it is also recommended for the safe storage of personal contact details, both prior to and after archiving.

Step 4: Organising Data – File Systems & Structures

The aim here is to create an online storage system that enables the easy identification and retrieval of digital files in a dataset. It is worth considering if data storage and documentation may be facilitated via a QDA software package, if one is being used. In

future, it may be possible to transfers files directly from selected QDA packages to a data repository (Corti et al. 2020).

- ❖ **Review where you will locate your digital data, and how you will organise the files**. While there are different options (portable devices, cloud storage, local desktop computers, networked drives, reviewed in Corti et al. 2020: 137–8), a networked drive is likely to be the most robust and secure for personal forms of data. Aim to organise your data files on **a password-protected digital drive** (on a secure server that has daily backups), with access restricted to the researchers.

- ❖ Consider how much storage space you will need for your dataset. If sharing data across teams, set up a system for the safe sharing of files (e.g. *Sharepoint Workspace* software or similar).

- ❖ For QL researchers, **the file structure should reflect the temporal logic of working across cases, themes and processes**, thereby supporting multi-dimensional analysis (see Neale 2021, Chapters 10 and 11). One strategy is to arrange files alphabetically by case name/pseudonym, or numeral identifier (see **Appendix 6** for an example). This brings together all files relating to a particular case, which are then nested by wave of fieldwork and finally by data type. This is a good foundation for building holistic analyses of unfolding biographies, and generating case portraits, histories and analyses.

- ❖ The alternative, recommended by Saldaña (2003), is to organise files chronologically by wave of data generation, within which files are nested by case and then by file type. Whichever structure is adopted, it should allow for the cross-referencing of files by case and wave. It should also complement the thematic organisation of analytical files, for example, in a QDA package or in Framework grids (See Neale 2021, Chapter 10).

- ❖ The file structure should include **contextual and analytical files** as integral components of a dataset. Separate areas can be allocated for files that run across cases and/or waves (e.g. interview schedules, leaflets about the study, sample consent leaflets and forms, and analytical files).

- ❖ **Abridged and unabridged versions of the dataset** should be kept in separate folders. The unabridged version will include all 'raw' data files; any identifying or confidential data, such as the personal details of participants, organisations and practitioners and their contact sheets; and any written communications received from participants, their consent forms and the anonymisation code book that links real to assigned identities (see Step 5). This unabridged version should be encrypted to give the additional security needed for personal information.
- ❖ It is worth checking local protocols for the safe storage of unabridged files while a project is underway. Some institutions may recommend that these are not held digitally but only in paper format, although paper storage for the longer-term preservation of datasets is becoming increasingly untenable for institutions and archives.

Step 5: Ethically Representing QL Data

The ethical representation of data in public displays and forums, publications and in archived datasets needs to be considered at an early stage of a project, and outlined in participant consent leaflets. This is a vital stage in fashioning a dataset that addresses a fundamental issue: how to balance the ethical concern to respect confidentiality with the epistemological drive to produce an authentic, fully contextualised dataset that reflects real lives (Neale 2013; Neale 2021). Anonymising (altering or substituting names of people, places, organisations, employment or other identifying features) is the most commonly used strategy, although it is a far from perfect tool. There is a growing appreciation that identities cannot be fully protected (Corti et al. 2020). Nor is anonymising the only route to protecting confidentiality: depositing data in a controlled-access archive also safeguards data and may enable different levels of access for different purposes (Neale and Bishop 2012a; Neale 2013; Corti et al. 2020).

As indicated above, data confidentiality is a particular issue in QL research, given the highly personal and biographically rich nature of the data. In the *Welfare Conditionality* study, for example*,* an early decision to carry out a light-touch alteration of data was reviewed during the last year of the project, when the politically sensitive nature of the data had become apparent. A decision was taken at this point to fully anonymise the full dataset (1,082 interview transcripts), for deposit in the *Timescapes Archive*. For this study, each interview had taken, on average, 50 minutes to conduct, and transcribers had been asked to flag up any identifying material in the interview texts. This helped to speed up the anonymising process. The team allowed up to an hour per transcript for anonymising (i.e. 1,000 hours, or approximately 29 weeks of work). Two part-time, trained workers were employed to carry out this task intensively over a six-month period (Fleur Hughes, personal communication; Dwyer and Patrick 2020).

Adopting a clear, systematic and proactive approach to the task of representing data can help to avoid the problems of inconsistency and lack of integrity that may arise if data are altered on a piecemeal basis (for a salutary lesson on this, see Taylor 2015).

Protocols for representing data and safeguarding confidentiality will vary across country settings, institutions and data repositories. Some details relating to the *Timescapes Archive* are set out below and in Step 6, as examples of how things might be organised, and to highlight the issues that researchers might look out for when negotiating the archiving of their data.

❖ Review varied ways to **safeguard data confidentiality**. Anonymising is usually strongly recommended or mandated by funding bodies, ethics committees, and archives. In the UK, these follow the strictures laid down by the EU GDPR and the

Data Protection Act 2018. As noted above, archiving a dataset also provides some protection. Archives control and regulate access to data via a user licence agreement. This places a contractual obligation on new analysts to use data only for research purposes, and to protect identities and preserve confidentiality.

❖ The *Timescapes Archive* offers different levels of access (public, registered, restricted) for different categories of data (Neale and Bishop 2012a and b; also Corti et al. 2020: 206–9). This granularity can be very helpful for researchers seeking to fine-tune access to their datasets. Placing personal contact details and other data that have not been anonymised on restricted access adds a further level of security to these data. Researchers can specify that highly sensitive data can only be accessed with the permission of the original researchers, or, in their absence, by an archive access committee that is authorised to make such decisions. There are issues of resourcing and scalability here. In days of stretched budgets, archives are increasingly seeking to automate their procedures. Nevertheless, for personal and unabridged data, a streamlined facility for discretional permissions, such as that provided by the *Timescapes Archive,* should be possible.

❖ **Anonymising** remains the standard tool for protecting identities, although, as noted above, it is not possible to fully anonymise rich, narrative data (for a useful web resource, see www.ukanon.net and Elliot et al 2016). Anonymising is a skilled and time-consuming task. It involves substituting the names of people, organisations or localities, or otherwise aggregating or generalising information (e.g. on a participant's employment or family composition) that may identify them, or make them distinctive within a sample (see [www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation)](www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation).

❖ Decide who will undertake this task, when anonymization will occur, and how far data will be altered (taking into account the ethical balancing act that also seeks to preserve data authentically and to acknowledge the data source, see Neale 2021). Anonymising is best seen as an ongoing process that follows on from transcribing. This creates an abridged version of a dataset that is ready for public outputs. For longitudinal data, some retrospective anonysmising may be needed, as sensitive

or biographical details that may identify a participant may accrue and become more evident over time. Audio transcribers may carry out a 'light touch' anonymisation (i.e. changing names and localities), leaving further finessing, or a more thorough 'belt and braces' alteration of data, to the research team. It is worth consulting with your chosen repository and reaching agreement on how a dataset will be altered.

❖ 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 preserved as part of the unabridged version of the dataset.

❖ Establish **anonymising guidelines**, and a securely held **code book** that links real to assigned identities and ensures consistency of approach across a dataset (see **Appendix 7** for examples; also Corti et al. 2020: 123-4). Using pseudonyms, or replacing text with different or more generalised descriptions, is preferable to simply removing or blanking out data. In this way, the intellectual coherence and integrity of the narrative is less compromised. Words that have been replaced need to be flagged up in the text, ideally in the most unobtrusive way possible, for example, by marking changed text with square [ ] brackets that are unlikely to appear elsewhere in the text (see examples in the transcript in **Appendix 5**).

❖ Proprietary online tools and services for anonymising text are available, and can automate some of the more straightforward changes. These include the useful QualAnon anonymization tool (ICPSR 2012), and a similar tool developed by the Irish Qualitative Data Archive. A basic *text anonymisation helper* tool has also been developed by the UK Data Service (www.ukdataservice.ac.uk/ manage-data/legal-ethical/ anonymisation/qualitative.aspx). This identifies capitalised words and numerals in the text, which may help to locate proper nouns, or people's ages or dates of birth. Great care is needed in using general search-and-replace tools, such as those in Microsoft Office Word, since they may unintentionally corrupt a text. Sophisticated tools for technically assisted anonymisation are currently under development (see, for example, the Washington-based *National Institute for*

Health-funded project, *Sharing Qualitative Research Data*, led by James Dubois, which aims to produce a new tool kit for ethical data sharing).

❖ Anonymising creates **two versions of a digital dataset**. As shown in Step 4, a complete, unabridged dataset, including personal contact details, should be digitised, encrypted and stored separately from the abridged (anonymised) version, in its own password-protected folder. Both are integral elements of a dataset that will have value at different historical times.

Step 6: Archiving a Dataset

Once the elements of good data management outlined above are addressed, it is a relatively quick and easy task to deposit data in an archive. The timing of this step, and rationales for taking it, are worth considering here. As shown above, over the past two decades, the preservation of research data through archiving has come to be viewed as a fundamental part of good data management. From a European perspective, at least, it is no longer seen as an optional extra (Bishop and Kuula-Luumi 2017; Neale 2017). This is underpinned by the idea that, since research datasets are produced with public funds, they should be regarded as an academic resource, rather than something that is 'owned' by the originating team for their exclusive use. In the UK, for example, archiving is mandated for research funded by the major research councils, and recommended by other major funders (e.g. the Joseph Rowntree Foundation), who are increasingly prepared to support the costs of this process. While there are no longer any requirements for research council-funded projects to archive with the UK Data Service, researchers are required to preserve their data for re-use within their own institutional repositories or a suitable specialist archive. In an international context, this same ethos of preserving data for archiving and sharing is also growing, but at

different rates and in widely different contexts of infrastructure and funding (see

Hughes and Tarrant 2020 for a review of international archiving resources).

The timing for archiving is a delicate ethical issue and is best left to the discretion of

the researchers. Given the open-ended nature of QL research, there is often no clear

point at which a study can be said to be completed. Moreover, researchers worry that

their contribution will be 'scooped' by others if they 'let go' of their data too soon.

Whatever the commitment to archiving and sharing data in principle, the push to

archive for the common good may be compromised by the pull to retain data for the

good of the research team.

These fears can be easily allayed once researchers become aware of the many

different ways in which data may be revisited, and the multiple insights that may be

produced (Maureen Haaker, personal communication; see Neale 2021, Chapter 8 for

examples). Dean et al. (2018) provide a convincing demonstration of this: when six

researchers, working in different disciplinary/methodological traditions, were asked to

analyse the transcripts from a series of radio broadcasts, they produced six varied and

equally plausible interpretations. An interpretive framework for analysis, with its

abductive underpinnings, allows for – indeed, welcomes – such diversity.

Even so, given these sensitivities about the 'ownership' of data and findings, there are

generally no time constraints laid down by funders or archivists for the deposit of data.

Archiving may occur at the end of a funded project, or later if researchers are

intending to continue their studies or extend their analysis. For QL researchers, who

are engaged in longer-term analysis, flexible time frames for 'letting go' of data are

essential. At the same time, however, there are solid professional and practical reasons to archive sooner rather than later.

First, archiving creates a published dataset as an important output from a study that can be cited alongside the publication of findings. For UK researchers, datasets can be included in the Research Excellence Framework (REF), which assesses the quality of research across the academy. Publishing a dataset through an archive increases the visibility of a study, establishes the quality and transparency of its evidence, and promotes it in ways that significantly increase its impact (Corti et al. 2020; for guidance on UK REF submissions, see www.ref.ac.uk/media/1092/ref-2019/01-guidance-on-submissions.pdf).

Archiving also transforms a dataset into legacy data, with value for future generations of researchers (see Neale 2021, Chapter 8). Where a dataset is revisited for research or teaching purposes, this is evidence of its quality and utility for the academy, giving added value to the funders' investment and enhancing the reputation of the researchers. Moreover, the anticipation of sharing and future use can provide the impetus to produce high quality, verifiable data files. This investment in the data is beneficial in raising standards.

There are also important practical and ethical reasons for depositing a dataset as near as possible to the conclusion of a study, rather than leaving this task for later. While data can be archived retrospectively, few researchers have the time and resources to manage this beyond the end of a project. Once staff resources and budgets are depleted, researchers may move on to new projects, change office or institution,

upgrade their computing facilities or equipment, or otherwise lose track of their project and the tacit knowledge that goes with it. The momentum to archive and to preserve and publish a dataset may be lost or severely compromised, unless it is built in as the culmination of the data management process. This would be a problem not only for the original researchers, who may wish to resurrect a project in the future, but would also mean the loss of an important historical resource for posterity (for a discussion on this issue, see Neale 2021, Chapter 8).

Finally, QL datasets are preserved more securely in a professionally curated, password-protected, access-controlled archive, than in a filing cabinet, or on a computer in a researcher's office, where they are more likely to get lost, destroyed, corrupted or inadvertently disclosed. The conditions under which data can be used (issues of copyright and licensing, see Step 2 above)) are much less clear when data are held outside the archive. Archiving, then, is not so much about 'giving away' a dataset, as transferring the responsibility for its longer-term preservation and stewardship to a data repository, where the data still remain available to the original researchers. For all these reasons, there are strong incentives to archive sooner rather than later, to plan for archiving at the outset of a project, and to prepare for it as an ongoing process. Overall, the timing of archiving can be a sensitive matter and is best left to the discretion of the researcher. But, as we have seen, there are sound reasons to archive as near as possible to the conclusion of a study, even if this involves placing the whole dataset on embargo for a specified period of time.

Archives vary in their procedures for depositing a dataset and facilitating their access by others; researchers will need to check the protocols used by their chosen

repository. The protocols used by the *Timescapes Archive* are set out below as an example of what the process might entail (see www.timescapes-archive.leeds.ac.uk and for a demonstration of the search and retrieval features of the archive, see www.timescapes-archive.leeds.ac.uk/archive-demonstration/. QL researchers will want to explore the options for archiving locally, in their institutional or national repositories. But they may also opt to deposit their datasets in the *Timescapes Archive*, which has been developed as an international resource. While datasets cannot be translated through this resource, they can be deposited in any language.

- ❖ Researchers complete a spreadsheet that logs the cases and files that make up the dataset as they are created (see **Appendix 2**). They also submit a depositors' licence, create a short landing page for the project (see **Appendix 8** for an example), and transfer the digital dataset to the repository.

- ❖ Once received, the archive team will assign a unique identifier to a dataset (a digital object identifier, or DOI), which new researchers will use to cite the dataset in their publications. This same identifier is then logged at the UK Data Service, which will then create a catalogue record for the dataset to aid its discovery by a wider pool of researchers.

- ❖ Any embargoes on the use of research data in the short and longer term need to be agreed with your archive. It is worth seeking to store the full, unabridged, encrypted version of the dataset under long-term embargo in a 'dark' area of the archive, until such time that it can be released as historically valuable data (this may be beyond the lifetime of the participant, when GDPR strictures on the protection of personal data cease to apply). This strategy was adopted by several *Timescapes* projects; researchers in the *Oldest Generation* study, for example, used the 'dark' archive function of the *Timescapes Archive* for their unabridged photographs. These are embargoed until 2040 (Bytheway and Bornat 2012).

- ❖ The final stage in the archiving process is the production of a **study guide** for long-term use. This should be archived alongside the abridged dataset. If resources

allow, aim for a 'gold standard' guide (e.g. Neale et al. 2015). This will give a detailed account of the study methodology and its development over time, along with details of the participants and how the dataset has been produced. The production of such metadata (data about data) is of vital importance for legacy data researchers. But it is also an essential part of the transparency needed to enhance the veracity of a study (Neale 2021). The guide is an important way to promote a study and provide the context for its production, ensuring, as far as possible, that a dataset can be independently and appropriately understood and interpreted by new researchers, and that it has some integrity as the evidence base for a study. This documentation can also form the basis for methods discussions in published outputs. A checklist of contents (adapted from Bishop and Neale 2012) includes:

- A description of the project, including title, research aims and questions, theoretical framing, overall design, funders, and details of the research team, institutions, and start and finish dates;
- A description of and rationale for the methods used to generate data, including any ethical or practical issues that have arisen; fieldwork experiences; sampling strategy; any changes in study design or conduct over time; and any protocols used in the management and production of the dataset. Any limitations or challenges that arose in the conduct of the study (e.g. relating to sampling or sample attrition, ethical issues or limitations in data generation or analysis) can be included here, enabling a wider audience to gauge the quality of the study;
- A description of the dataset: numbers and types of data files; numbers of participants, cases and waves; the timing of fieldwork; characteristics of the sample; and themes covered in the interviews;
- Field documents: details of interview schedules, field notes, research diaries, analytical files and thematic key words assigned to the data;
- A list of outputs and publications from the project and links to the project website.

Closing Reflections

This guide has explored the intricate process of managing a QL dataset and the dual purpose that this process serves: facilitating the ongoing use and analysis of the data

by the originating team, and creating legacy data that can be preserved and shared with others. Six areas of good practice in data management have been described here. While these have been presented as a series of steps, in practice they overlap and intersect as a project develops. The timing of these processes is a crucial consideration. A key message is that data management is not a simple administrative task that is tacked on to the end of fieldwork. It is an interpretive process, with ethical and epistemological implications that need to be carefully considered at the outset and addressed as an integral part of a project.

It will be abundantly clear by now that managing QL data for prolonged use over time is not a quick and easy task. It requires considerable foresight and planning. But, for a growing dataset that researchers may wish to return to over many years, good data management is essential. A high quality, well-organised, ethically attuned, fully documented and safely preserved dataset provides the foundation for temporal analysis, and for extending a study into the future. It is also an important output from a study that will enhance its credibility and accrue historical value for future generations of researchers.

References

Alldred, P. and Gillies, V. (2012) 'Eliciting research accounts: Re/producing modern subjects?', in T. Miller, M. Birch, M. Mauthner and J. Jessop (eds), *Ethics in Qualitative Research,* (2nd edition). London: Sage, pp. 140-156.

Bishop, L. and Kuula-Luumi, A. (2017) 'Revisiting qualitative data reuse: A decade on', *SAGE Open,* Jan-March: 1-15.

Bishop, L. and Neale, B. (2012) *Data Management for Qualitative Longitudinal Researchers.* Timescapes Methods Guide Series, no. 17, https://timescapes-archive.leeds.ac.uk/publications-and-outputs/

Bornat, J. (2020) 'Looking back, looking forward: Working with archived oral history interviews' in K. Hughes and A. Tarrant (eds) *Qualitative Secondary Anal*ysis. London: Sage, pp. 137-153.

Bytheway, W. and Bornat, J. (2012) 'The oldest generation as displayed in family photographs', in V. Ylanne (ed.), *Representing Ageing: Images and Identities.* London: Palgrave Macmillan, pp. 169-188.

Corti, L., Ven den Eynden, V., Bishop, L. and Woollard, M. (2020) *Managing and Sharing Research Data: A Guide to Good Practice*, 2nd edn. London: Sage.

Dawson, P. (2003) *Reshaping Change: A Processual Perspective.* London: Routledge.

Dean, J., Furness, P., Verrier, D. et al (2018) 'Desert island data: An investigation into researcher positionality' *Qualitative Research*, 18 (3): 273-289.

Elliot, M., Mackey, E., O'Hara, K. and Tudor, E. (2016) *The Anonymisation Decision-Making Framework* available at: www.ukanon.net/ukan-resources/ukan-decision-making-framework

ESRC (2016) *Framework for Research Ethics* Economic and Social Research Council Online, available at: http://www. esrc.ac.uk/about-esrc/information/research-ethics.aspx

Ferguson, H., Leigh, J., Cooner, T., Beddoe, L., Disney, T. et al (2019) 'From snapshots of practice to a movie: Researching long-term social work and child protection by getting as close as possible to practice and organisational life', *British Journal of Social Work*, 0: 1.18.

Hughes, K. and Tarrant, A. (eds) (2020) *Qualitative Secondary Analysis*. London: Sage.

ICPSR (2012) *QualAnon Tool: DSDR Qualitative Data Anonymiser*. Available at: www.icpsr.umich.edu//icpsrweb/DSRC/tools/anonymize.jsp

Kuula, A. (2010-11) 'Methodological and Ethical Dilemmas of Archiving Qualitative Data', *IASSIST Quarterly*: 34 (3-4): 12-17.

McLeod, J. and Thomson, R. (2009) *Researching Social Change.* London: Sage.

Moore, N. (2007) '(Re) using qualitative data' *Sociological Research Online,* 12 (3): 1.

Neale, B. (2013) 'Adding time into the mix: Stakeholder ethics in qualitative longitudinal research, *Methodological Innovations Online*, 8 (2): 6-20.

Neale, B. (2017) *Research Data as Documents of Life.* Guest-Post no. 13, www.bigglr.ncrm.ac.uk.

Neale, B. (2019) *What is Qualitative Longitudinal Research?* London: Bloomsbury

Neale, B. (2021) *The Craft of Qualitative Longitudinal Research.* London: Sage.

Neale, B. and Bishop, L. (2012a) 'The *Timescapes Archive*: A stakeholder approach to archiving qualitative longitudinal data', *Qualitative Research,* 12 (1): 53-65.

Neale, B. and Bishop, L. (2012b) *The Ethics of Archiving and Re-Using Qualitative Longitudinal Data: A Stakeholder Approach.* Timescapes Methods Guide no. 18. https://timescapes-archive.leeds.ac.uk/publications-and-outputs/

Neale, B. Hughes, K. Proudfoot, R. et al. (2015 edition) *Managing Qualitative Longitiudinal data: A Practical Guide for Depositors .* www.timescapes-archive.leeds.ac.uk. [replaced by 2020 edition]

Neale, B., Lau-Clayton, C., Davies, L. and Ladlow, L. (2015) *Researching the Lives of Young Fathers: The Following Young Fathers Study and Dataset.* Briefing Paper no. 8. www. followingfathers.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.* London: Sage, pp. 91-125.

Plummer, K. (2001) *Documents of Life 2.* London: Sage.

Saldaña, J. (2003) *Longitudinal Qualitative Research: Analyzing Change through Time.* Walnut Creek CA: AltaMira Press.

Smith, N. (2003) 'Cross-sectional profiling and longitudinal analysis', *International Journal of Social Research Methodology,* 6 (3): 273-277.

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 de Ven, A. and Huber, G. (1995) 'Introduction', in G. Huber and A. Van de Ven (eds), *Longitudinal Field Research Methods.* London: Sage, pp. vii-xiv.

Weller, S. (2012) 'Evolving creativity in qualitative longitudinal research with children & teenagers', *International Journal of Social Research Methodology*, 15 (2): 119-133.

Acknowledgements

Appendices

Appendix 1: Sample Data Management Plans

Following Young Fathers Further Study
(UKRI, Tarrant, University of Lincoln, 2020-2027):
Data Management Plan

1. Description of the data

Type of study: the proposed study will generate dynamic qualitative longitudinal data in order to understand the lived experiences and support needs of marginalised young fathers in international, comparative contexts.

Types of data: The dataset will include audio recordings of interviews, written transcripts of interviews in English and Swedish; text-based documents based on practitioner diaries; and YDC event feedback and participant meta-data.

Format and scale of the data: In strand 1, semi-structured, audio-recorded qualitative interviews, conducted with 10 young fathers in Leeds, and 20 young fathers (aged under 25) in Lincolnshire across two waves, will produce approx. 90-100 hours of audio and text-based files. In strand 2, 70-80 hours of new audio-recorded data will be generated with 20 young fathers across 3 waves of interviews (both models) and with 10 practitioners. In strand 3, an additional 10 young fathers will be interviewed across two data sweeps in the Nordic states producing 20 hours of audio. All interviews will be recorded in MP3 format and converted to Free Lossless Audio Codec (FLAC) (.flac) format, ready for ingestion into the Timescapes Archive. The UK Data Service recommends these as optimal formats for ensuring data quality and legacy required for data storage and re-use. Participants will use A4 notebooks for their reflections on the implementation process in strand 2. These will be scanned and digitised by PRDA2 and saved in rich-text format (.rtf).

2. Data collection / generation

Methodologies for data collection/generation: All interviews will be undertaken by experienced researchers with support and training provided by the PI and the University of Lincoln. Where the young fathers are trained as co-researchers for strand 2, a key focus of training will be data quality and technical use of the audio recorders. The practitioners who fill in the reflexive diaries will also be advised by PDRA1 about writing clearly and Dr Karen Bateson will monitor progress and consult with them about the diaries regularly to ensure clarity. The ongoing relationships and ethos of co-production supported by the QL design of the study will underpin continued consultation with participants. Experienced researchers will also conduct the interviews with young fathers in Sweden.

Data quality and standards: Quality assurance procedures for data generation will be determined by the research team early in the research process in order to generate data files that are of high intellectual and technical quality. Guidelines will be produced that ensure high quality, futureproofed data recordings and outputs. Data management and archiving procedures will also be a standing item on the agenda of research meetings. The Project Administrator will be tasked to report on, discuss and co-ordinate progress. Key issues include ensuring that the diary data is legible and that all interview data (audio recordings and transcripts) are complete and usable. All researchers will be expected to record a short, oral introduction at the start of interviews to provide vital contextual information, including who is being recorded, by whom, where, when and for what project. With the support of project partner Dr Karen Bateson, PDRA2 will oversee the development of the research diaries and their digitisation. Participants completing the diaries will be directed to write clearly and legibly and the research team will remain in contact with them as they are digitised to ensure data accuracy. The PDRAs will also be trained in the technical use of the recording equipment for interviews and with the methodological skills needed to generate high quality interview data.

Research interviews will take place in a quiet office location provided by the project partners to ensure the clarity of audio recording for the transcription process. As above, the PI and PDRA2 will train selected young fathers as co-researchers for the final wave of interviews in strand 2. A key aspect of this training will be to explain the importance of effective data recording to the young fathers, including the practicalities of using audio-recording equipment.

Transcription protocols and a guide for transcribers will also be developed by the research team prior to the transcription of the audio recordings to ensure consistency and data quality across the dataset. A reputable translator will be hired to translate the Swedish interviews and produce transcripts. They will be asked to provide detailed notes that explain the context in which particular words or phrases have been translated, In order to achieve 'conceptual equivalence' i.e. comparability of meaning between the two languages and cultures. The guidelines will state that transcribers and translators must adhere to UK Data Archive standards. The PDRAs and international project partners for each strand will quality check all transcripts once complete. The chosen optimal file formats (described in section 1.3) are based on recommendations by the UK Data Service and are most likely to ensure longer-term data access and usable data when archived. All data (from interviews, diaries and feedback) will be checked for errors and/or inaccuracies by the research team following conversion/digitisation to ensure data integrity is maintained during the conversion process.

3. Data management, documentation and curation

Managing, storing and curating data: An effective data management strategy will be established at an early stage of the project, attending to it as an integral part of the study (Neale et al. 2016). Given the time consuming and complex nature of

preparing, organising and documenting data, this early planning is essential and will be conducted by the PDRAs with oversight by the PI throughout the study. A robust system of data storage will be designed and established to ensure that all project data are securely stored and effectively managed. Digital files, including participant metadata and audio and textual data from diaries and interview transcripts, will initially be stored on a shared, password-protected digital drive on One Drive for Business at the University of Lincoln. All anonymised, non-personal, non-sensitive data will be backed up continually to this system, which will be accessible only to the research team. An additional password will be required for files that link participants to data. The information and consent forms for each participant will also be scanned and stored digitally as PDFs on this drive. The naming system for all documents will be organised to ensure that different versions of files are easily distinguishable e.g. raw data files, anonymised files, back up files. A secure and confidential paper archive will be stored in a locked filing cabinet, and hold the full, unabridged dataset, accessible only to the PI and PDRAs. This will include personal details of participants, organisations and practitioners, their contact details, their consent forms and the code book that links real to assigned identities. 'Raw' transcripts and other files and media that are not anonymised will also be stored here.

Metadata standards and data documentation: Data documentation and contextualisation processes will be mapped in a data log that provides a 'road map' to the dataset and a methodological guide to the study. This guide will include a detailed account of the project (aims, objectives, rationale), the methodology, details of the sample and dataset, key outputs and suggestions for emerging themes that are ripe for further analysis. Unique digital identifiers will be created for all files and consistently applied to each fieldwork file. This will include the initials of the project, a case identifier (the pseudonym for the case), the wave of fieldwork in which the file was generated, the type of data (interview, reflexive diary), a unique number assigned to each case file, and the digital format in which the file is saved.

Data preservation strategy and standards: All of the data produced will be ingested into the Timescapes Archive; a permanent and specialist resource at the University of Leeds that is officially endorsed and supported by the national UK Data Archive (Neale et al. 2016). This will ensure longer term data re-use by other social researchers both in the UK and internationally.

4. Data security and confidentiality of potentially disclosive information

Formal information/data security standards: The study will be fully compliant with the new EU General Data Protection Regulation (GDPR) and Data Protection Act (2018), to which the University of Lincoln is also compliant.

Main risks to data security: Inadvertent identification of participants via storage of their personal data is the main data risk in this study. Procedures to ensure data security will therefore be established from the outset of the study and will include

data anonymisation; confidentiality agreements with external project stakeholders; and effective storage and back up procedures for paper and digital files. To ensure anonymisation, pseudonyms will be assigned to participants and study areas during the transcription process. Care will be taken to ensure that no direct identifiers relating to participants (e.g. names and contact details) are included in printed transcripts. As far as possible, other identifying data will be removed with due attention to ensuring that the integrity of the data remains intact. The external transcriber and translator will be required to sign a non-disclosure agreement to ensure protection of sensitive information in the files and to adhere to the same levels of security and back-up as required at the University of Lincoln. They will also be asked to confirm that they have permanently deleted all files from all computers at the end of the study. No third parties will have access to participant contact information. The anonymisation process will be approached as an interpretive task, paying attention to the coherence and integrity of the dataset and to how the data will be represented in the study: including for analysis, publication and display, and for wider sharing and re-use (Neale et al. 2016). Pseudonyms will be used for all cases that are archived in the Timescapes Archive, to protect personal data but also to safeguard confidentiality.

5. Data sharing and access

Suitability for sharing: QL research yields rich, fine-grained data about social processes that is both of enduring historical value and a significant academic and policy relevant resource (Neale 2019). The dataset generated for this study expands and extends the ESRC funded Following Young Fathers baseline study that is already stored in the Timescapes Archive and increases the scope of longer-term cumulative and comparative analysis, both by the research team and other users.

Discovery by potential users of the research/innovation data: The data generated from the study will be stored in the Timescapes Archive. A short catalogue record for the UK Data Service will also be produced to record the project metadata (data about data). The Service will then promote the dataset so that secondary users are alerted to its existence, location and access options. The data sharing policy and project approach to data sharing will also be published on the study website and advertised via an open access briefing paper on the Timescapes Archive website.

Governance of access: The data produced will be sensitive in nature and potentially difficult to anonymise as the young fathers may become increasingly identifiable as the project progresses. Once ingested into the Timescapes Archive, access to the data will therefore be regulated and protected by 'restricted access'. Potential data re-users will need to apply for access via the PI, who will be responsible for decision-making. Re-users will be required to explain how they intend to use the data and for what purposes. If the PI is no longer reachable, then the decision about access will be devolved to the remaining research team. If the research team are not contactable or available, decision authority would devolve to the institutions holding the data (Lincoln and Leeds).

The study team's exclusive use of the data: While there are strong incentives to archive data as close to the completion of the study as possible, an embargo on data sharing and re-use for a period of 12 months will be applied following the completion of all fieldwork in year 4 of the study. This will give the originating research team time to complete their own analyses and develop research outputs. During this 12-month period, preparations will be made to deposit the dataset in the Timescapes Archive, including the submission of a depositor license and the creation of a short catalogue record for the UK Data Service.

Restrictions or delays to sharing, with planned actions to limit such restrictions: Informed consent for archiving will be sought by all participants at the same time as seeking consent for participation in the study. This will be done on an ongoing basis across data sweeps. A leaflet and archive consent form will be devised for use with participants. This leaflet will explain that copyright in the data will be held jointly by the participant and the research team. Participants will be asked to sign a form that establishes joint copyright with the researchers.

Regulation of responsibilities of users: Access to data stored in the Timescapes Archive is monitored and controlled through password controlled registration and a user licence system. Only researchers with institutional affiliations can register as users. They are contractually obliged to use archived data for research purposes only, and in ways that protect participant confidentiality.

6. Responsibilities

University of Lincoln: Naomi Timings/Anne Jolly (Data Protection Officers)

7. Relevant institutional, departmental or study policies on data sharing and data security

Research Data Management Policy and Procedured: www.secretariat.blogs. lincoln.ac.uk/ files/2013/08/Research-Data-Management-Policy.pdf.

Data Protection Policy: www.secretariat.blogs.lincoln.ac.uk/files/ 2018/05/20180508-Data-Protection-Policy-v1.5GDPR.pdf.

Institutional Repository Policies: https://eprints.lincoln.ac.uk/policies.html

Life on a Low Income: A Secondary Analysis Research Proposal
(ESRC submission: Neale, Hughes et al, University of Leeds, 2015)

Data Management Plan

Existing Data Sources: This project entails the generation of a small practitioner dataset, based on consultations with up to 8 policy and practice experts, which we will audio record and transcribe. The transcriptions will then be archived in the Timescapes Archive (see below). The prime focus of this study will be the collation and re-use of existing data sources across a network of thematically and methodologically related projects (The Life on a Low Income Research Network). Data from the 8 projects in the network will be aggregated to form the *Life on a Low Income* collection, and a comprehensive secondary analysis of sampled data from the collection will be conducted to address the themes of this research. The datasets will be prepared for archiving by the research teams, with support from Dr. Hughes, in her role as academic liaison for the Timescapes Archive. The composite datasets will be available for longer term scientific and social historical use through the Timescapes Archive.

Creating Data Resources for Re-use: Bringing a collection of thematically related datasets together, and enabling data linkages between them, is an innovative research strategy that will greatly enhance their value for re-use. We have included in this process datasets that are not mandated to be archived through the funding councils: four datasets funded by the Joseph Rowntree Foundation, and one funded by the Department for Work and Pensions (subject to DWP permission, which is currently being sought).

The infrastructure for this process is the ***Timescapes Archive,*** a specialist resource of Qualitative Longitudinal (QL) data that was developed under the ESRC Timescapes Initiative (2007-12). Specialist curation is needed for QL data because of the extended time frames for enquiry. Developed with support from the UK Data Service, the Timescapes Archive complements the work of the UK Data Archive. In line with the ESRC National Data Policy, the Timescapes resource collates QL data to aid understanding of micro-dynamic processes and mechanisms of social change. It complements large scale datasets that chart broad social trends and measure the extent of change. Under recent funding (ESRC *Changing Landscapes for the Third Sector, Kahryn Hughes)* the Timescapes Archive has been migrated to a new open source software platform (EPrints), and fully incorporated into the Institutional repository service at the University of Leeds. This has given the resource a sustainable technical platform and scope for further development. The resource provides a test case for the technical and scientific stewardship of datasets at institutional level, in a

climate where this is likely to become a growing requirement for institutions (RCUK, EPSRC, and ESRC National Data Policies; http://researchdata.leeds.ac.uk/).

The Timescapes Archive is founded on a stakeholder model of archiving (Neale and Bishop 2012; Neale 2013), that supports the needs of primary as well as secondary data users, and ensures that data and tacit knowledge about data are not lost through the extensive time frames of QL enquiry. Our strategic aim is to build collections of thematically related QL datasets to enable linkages and searching across (as well as within) datasets. The *Life on a Low Income* collection will complement existing collections in the Timescapes resource – the founding collection on the dynamics of personal and family lives (9 datasets); and a collection on the dynamics of third sector organisations (3 datasets, currently under construction).

Information on Datasets and Metadata: We will be working with a sample of 200 cases, each comprising multiple waves of data. These will be drawn from an overall pool of 500+ cases across the 8 projects. Datasets for re-use will take the form of verbatim transcripts of qualitative interviews, created as .docx files created in Microsoft Word. Docx is more open than older 'doc' formats, and therefore preferable for archiving and preservation purposes. Some participatory visual data may also be collated, e.g. diagrams and drawings. Data will be quality assured and formats checked as part of the curation process. Prior to ingest into the Archive, text documents will be converted into PDF-A; audio files will be converted to FLAC format and drawings will be saved as JPEG-2000 or TIFF Uncompressed Level 6 to ensure long term sustainability in line with guidance from the UK Data Archive. Appropriate metadata, to an agreed schema based on experience gained in previous projects, will be created and saved with raw data. An agreed set of core metadata will be applied across all projects to facilitate searching, browsing and to reflect the case and wave structure of longitudinal data. This will include a project guide for each dataset, covering aims and methods, sampling frame, research questions, fieldwork materials, and links to outputs, as well as descriptions of the data themselves. We will assign key words to transcripts to aid search and retrieval. Guidance on the production of gold standard metadata (Neale et al 2015 edition) will be provided through the Timescapes Archive website (www.timescapes-archive.leeds.ac.uk). This will update our existing guidance on data management planning (Bishop and Neale 2012). A demonstration of the features of the archive is available here (www.timescapes-archive.ac.uk/archive-demonstration). Costs for data preparation work across the network have been incorporated into this proposal. JRF has agreed to contribute £12,000 to support the archiving costs for the JRF funded projects.

Planned quality assurance and back-up procedures (security/storage): During the life of the project, data will be transferred to the University using secure file transfer procedures, and stored on the University Storage Area Network (SAN), which

comprises enterprise level file servers in physically secure data centres with appropriate fire suppression equipment. Snapshots are taken every day at 10pm (and accessible for 1 month). A second level of snapshots is taken every month and these are kept for 11 months. Snapshots are user recoverable from the desktop. An incremental copy to backup tape is taken every night (and kept for 28 days) and a full copy is taken every month. Every quarter, the full dump tapes are moved to a long term storage facility where they are kept for 12 months. Tapes are initially stored in on-campus fireproof safes and then moved to off-campus secure locations. The data will then be converted into an appropriate format and ingested into the Timescapes archive. The active storage for the Timescapes resource is also part of the University SAN, so is subject to the same backup / recovery procedures outlined above. In addition, immutable data objects in the archive will be copied to a long-term archive store to provide a further level of data preservation.

Access restrictions and data security during the project life time: Data will initially be generated using audio recorders which save to removable SD card. The data will then either be copied directly to the SAN or, if this is not possible within a reasonable timeframe, to a laptop encrypted to FIPS 140-2 standard. Once data is on the SAN, then the SD card will be securely erased using a utility such as CyberShredder (http://www.cylog.org/utilities/cybershredder.jsp) Access to electronic data on the SAN is controlled by Active Directory (AD) Group membership. The Faculty IT Manager will set up a dedicated folder for this research project and create read-only and read/write AD groups. The PI will determine which users require read-only and read-write access. Off-campus access is via the secure Citrix portal. External users who need access to the data will apply for a University username and then be assigned to the appropriate AD group. Once data have been deposited into the Timescapes Archive, access control will be managed through an application and registration process. Under Timescapes we addressed the safeguarding of data to make it shareable via informed consent, anonymisation, and regulating access via the archive. Legal licensing agreements are in place for depositors and users of the resource, and access is password protected.

Expected difficulties in data sharing, and possible solutions: For QL research there is no clear cut off point at which primary data use is completed and secondary use begins. Timings of archiving are therefore sensitive and flexibility is required. For this study, the 8 project teams in the network have made a firm commitment to prepare and share data as the basis for the secondary analysis work of the project. For all these data, permissions to archive have either already been sought from participants, or will be sought retrospectively by the teams as part of their commitment to the study. Longer term, mechanisms have been built into the Timescapes Archive to encourage archiving by primary teams: data can be securely stored for primary as well as

secondary use, and additional controls on access to data can be put in place if needed. We will abide by ethical guidelines for sharing and re-using QL data (Neale and Bishop 2012, Neale 2013; Neale et al 2015 edition). For this study we have devised new ethical principles for authorship of outputs, which support the members of the network and facilitate and acknowledge their involvement and contribution. We aim to maximise confidentiality of re-used data as far as possible, for example, by offering guidance on double anonymising by secondary users.

Copyright and intellectual property ownership of the data Where informed consent for archiving is needed – i.e. for the new transcripts relating to expert consultations, and where datasets are being archived retrospectively, consent for archiving will be sought using templates developed in Timescapes. These will also ensure that copyright is shared with the researchers, to facilitate protection and stewardship of the data. We do not envisage any problems in obtaining consent from participants; the researchers have remained or are currently in touch with their participants and our experience in Timescapes, and in other studies, is that consent for archiving is readily obtained once it is understood that data are not publically available but used by professional researchers under controlled circumstances (Neale 2013). We are working with an overall pool of 500 cases, from which we intend to sample 200 for our analysis. This allows for some attrition, if some permissions cannot be secured. The Timescapes Archive deploys a non-exclusive deposit licence which grants the archive rights to store, preserve and disseminate data, but without requiring transfer of copyright.

Responsibilities for data management and curation within research teams: Dr. Kahryn Hughes (leader of the archiving strand) will have overall responsibility for implementing this data management plan, working in collaboration with the Head of the Research Data Management team at the University of Leeds (Graham Blyth) and the research co-ordinator. Network members will be responsible for data management and preparation using guidance provided by the Timescapes Archive; obtaining permissions where needed; and facilitating data sharing across the datasets. The research co-ordinator will support network members in data preparation and quality checking, arrange for depositor licences to be signed, and for datasets to be safely transferred for sharing and for deposit in the Archive. The Data Management team at the University of Leeds will ingest the datasets into the Timescapes Archive, copy the files to the UK Data Archive for long term preservation, facilitate ongoing access to the datasets, and provide technical advice where needed. The Faculty IT Manager will be responsible for ensuring that electronic file permissions have been correctly assigned and for advising on other aspects of data storage and security.

Appendix 2: Data Log for the Timescapes Archive

Repository Data Ingest Spreadsheet - 2015

			Project (project name)	Case ID (participant pseudonym)	Gender (M / F)	Age group [0-9, 10-19, 20-29, 30-39 etc etc]	Ethnicity	Employment	Relationship Status	Location	Name of Project + Case ID	Description	Subject (keywords)	Fieldworker ID	Fieldwork date [Date or Date range]	Item type - case or context
								TIMESCAPES PROJECTS - CASES and Contextual Information									
1			Friends and Acquaintances	Trevor	M	10-19	White	Student	Single	Southern town	Friends and acquaintances: Trevor	Friends and acquaintances: Trevor - Waves 1, 2 and 3	Youth; young people; relationships; identity; life changes, aspiration	Bren Neale	2005-2007	case
2			Friends and Acquaintances	Tricia	F	10-19	White	Student	Single	Northern town	Friends and acquaintances: Tricia	Friends and acquaintances: Tricia - Waves 1, 2 and 3	Youth; young people; relationships; identity; life changes, aspiration	Bren Neale	2005-2007	case
3			Friends and Acquaintances	FAA_context												context

The above entry has been inputted for example purposes only…. Please do not alter, vary or abbreviate your entry in a way that differs to the style, format of the example given

TIMESCAPES PROJECTS - DOCUMENTS

				Project	Case ID	Gender	Age group	Year of Birth	Ethnicity	Employment	Relationship Status	Location	Name of Project + Case ID	Name of Project + Case ID + Data type and Year data collected	Subject (keywords)	Fieldworker ID	Fieldwork date	Wave	File Type (text, image)	Data Type (interview, timeline, focus group, photo)	File name [no brackets, hyphens, spaces or other]	Data Anonymised	Consent for archiving evidence available?
eprintid	rowid	documents docid	documents rowid	documents tsmd_project	documents tsmd_caseID	documents tsmd_gender	documents tsmd_ageGroup	documents tsmd_yearOfBirth	documents tsmd_ethnicity	documents tsmd_employment	documents tsmd_relationshipStatus	documents tsmd_location	documents tsmd_title	documents tsmd_description	documents tsmd_subject	documents tsmd_fieldworkerID	documents tsmd_fieldworkDate		documents tsmd_dataType	documents.main	documents security		
1	1			Friends and Acquaintances	Trevor	M	10-19	1995	Irish	Student	Single	Southern town	Friends and acquaintances: Trevor	Friends and acquaintances: Trevor - Interview W1 2010	Youth; young people; relationships; identity; life changes; aspiration	Bren Neale	15 May 2010	1	Text	Interview	FAATrevorW1Interview05.rtf	Y	Y
1	2			Friends and Acquaintances	Trevor	M	10-19	1995	Irish	Student	Partner	Southern town	Friends and acquaintances: Trevor	Friends and acquaintances: Trevor - Interview W2 2012	youth; young people, relationships identity, life changes, education	Bren Neale	12 June 2012	2	Text	Interview	FAATrevorW2Interview06.rtf	Y	Y
1	3			Friends and Acquaintances	Trevor	M	10-19	1995	Irish	u/e	Single	Northern town	Friends and acquaintances: Trevor	Friends and acquaintances: Trevor - Interview W3 2014	Youth; young people, relationships identity, life changes, employment; benefits	Bren Neale	13 April 2014	3	Text	Interview	FAATrevorW3Interview07.rtf	Y	Y
2	1			Friends and Acquaintances	Tricia	M	10-19	1997	Irish	Student	Single	Northern town	Friends and acquaintances: Tricia	Friends and acquaintances: Tricia - Interview W1 2010	Youth; young people; relationships; identity; life changes, aspiration	Bren Neale	15 June 2010	1	Text	Interview	FAATriciaW1Interview05.rtf	Y	Y
2	2			Friends and Acquaintances	Tricia	M	10-19	1997	Irish	Student	Single	Northern town	Friends and acquaintances: Tricia	Friends and acquaintances: Tricia - Interview W2 2011	Youth; young people; relationships; identity; life changes; aspiration	Bren Neale	18 September 2011	2	Text	Interview	FAATriciaW2Interview06.rtf	Y	Y
2	3			Friends and Acquaintances	Tricia	M	10-19	1997	Irish	Student	Single	Northern town	Friends and acquaintances: Tricia	Friends and acquaintances: Tricia - Interview W3 2012	Youth; young people; relationships; identity; life changes; aspiration	Bren Neale	13 April 2012	3	Text	Interview	FAATriciaW3Interview07.rtf	Y	Y
3	1			Friends and Acquaintances	FAA_context								Friends and acquaintances: Consent form						Text	form	FAAConsentForm.pdf	Y	Y
3	2			Friends and Acquaintances	FAA_context								Friends and acquaintances: Dataset Guide						Text	Guide	FAAGuide.pdf	Y	Y
3	3			Friends and Acquaintances	FAA_context								Friends and acquaintances: Interview schedule wave 1						Text	form	FAAinterviewschedule1.pdf	Y	Y
3	4			Friends and Acquaintances	FAA_context								Friends and acquaintances: Leaflet						Text	form	FAALeaflet.pdf	Y	Y

Timescapes Archive: Depositor Spread Sheet: Guidance

All depositors need to fill in an excel spread sheet for their datasets, arranged through two tabs. This records metadata – data about data – for each of the cases and contextual information in your dataset, and for each document in your dataset. The archive team will use these spread sheet to ingest your data into the archive and to organise it to facilitate search and retrieval of the files. The spread sheet functions as a Data Log for projects to record their growing dataset, as well as aiding the process of ingesting data into the Timescapes Archive.

The two fields in the spread sheet can be accessed by clicking on the tabs at the bottom of the excel file. To guide you in filling in the document we have provided an example (see excel tables above), supplemented by the advice set out below. You will

need to follow the style and format of the entries given in the example – please do not vary or abbreviate your entries from the formats shown, or it will not be possible to ingest your data.

We suggest that you set up this data log at an early stage of your fieldwork, and record information on your documents as they are generated or become available. This will help you to create a record of your dataset as it grows. It is best to start with Tab 2, which records the details for each document in the dataset. The summary case and contextual information in Tab 1 is best completed at the end of the fieldwork, once the whole dataset is in place.

Tab 1: CASES and Contextual information.

The first tab on the spread sheet enables you to record summary details about the various cases that make up your dataset, and to flag up the availability of contextual files that relate to the dataset as a whole.

Contextual Information. Please use one of the lines on this tab to flag up the existence of contextual information (e.g. interview documents, guide to your dataset, and so on) and to assign a summary tag to identify such documents in your dataset. For this line there are just three entries to make.

The first column should record your project ID. In the example sheet, this is Friends and Acquaintances. In the next column (case ID) you should enter the initials for your project followed by underscore and context. e.g. FAA _context. In the last column (Item type (case or context)) enter the word context.

We suggest you create this line at the top or bottom of the spread sheet, so it is separate from the information about the cases in your study.

Cases. Please use the spread sheet to record summary information about each case in your dataset. One line should be devoted to each case. The example sheet gives information for two cases in the Friends and Acquaintances dataset: Trevor and Tricia. The entries show the base data for each case, followed by summary descriptions of the documents held for each case.

Base data – here you can record details of the gender, age group, year of birth, ethnicity, employment status, relationships status and place of residence (location) for each participant or organisation (case) in your dataset. Your descriptions may be based on self-reported or researcher-reported characteristics. If you have not gathered such data or it is not relevant for a particular case, please enter ND (no data) in the appropriate box. If you are gathering longitudinal data, some of the information recorded here will no doubt change over time, in which case please record the characteristics of your case at the start (wave 1) of fieldwork.

Project name: a short, unique narrative identifier for your project.

Case ID (identifier): This must be a unique identifier for this case within your dataset. It is most usually a narrative tag - a name or pseudonym given to a participant or organisation. If you look at data in the archive, you will see how the case ID appears in the listings.

Gender: enter M or F.

Age Group: give the age group for your case at start of fieldwork.

Ethnicity: short narrative description e.g. White, White British, Black British.

Employment: short narrative description e.g. student, unemployed, retired, or job title (e.g. Waiter, Teacher).

Relationship status: short narrative description to capture nature of personal relationship, e.g. single, partnered, co-habiting, married.

Location: place of residence of the participant – this may be a specific town or city or a more general location, e.g. North of England.

Summary descriptions of data held on each case. All of these entries are needed, even though there is some repetition of information already entered.

Project name and case ID: repeat this basic information here.

Description: record here the project name and case ID, followed by the waves of data held.

Subject (key words): these describe the overall subject matter of the documents available for each particular case. While there is no limit to the number of key words that you can use, we suggest you enter a maximum of five key words here.

Fieldworker ID: this is usually a narrative identifier- the name of the fieldworker. You may enter more than one ID if more than one person has collected data on this case over time.

Fieldwork date (date or date range): record here the date on which data were collected for this case, or a range of dates if collected over time.

Item Type (case, context): In this last column, enter the word case.

Tab 2: Projects: Documents.

The second tab on the spread sheet allows the researcher to record details of each document that makes up the composite dataset; these entries can be made as the dataset grows. Each document has its own line on the sheet. It therefore complements, expands and elaborates upon the summary information given in Tab 1.

Contextual Documents. Please use this sheet to record details of the contextual documents that relate to the dataset as a whole – e.g. guide to the project, interview schedules and so on. We have added lines on the example sheet for a blank consent form, a dataset guide, a wave 1 interview schedule and a recruitment leaflet. More lines would need to be added for additional contextual files – e.g. wave 2 or 3 interview schedules, focus group brief, archive consent form and so on. Please provide as many contextual documents as possible to help users to orientate themselves to your dataset. Please enter the following information for each of these contextual documents:

The project ID: in the example this is *Friends and Acquaintances*

The case ID: for these documents please enter the initials of the project ID followed by underscore and context e.g. FAA_context.

Name of project + case ID + data type: Under this heading, please enter the name of the project, and type of contextual document e.g. Friends and Acquaintances: Study Guide

File type: text or image

Data type: please repeat here the type of document e.g. interview, audio file, video.

File name: you need to assign a unique file name to each document in the dataset – without brackets, hyphens or spaces. The format to use for contextual files is the initials of the project ID, data type, and format of the file – rtf (rich text format or a tag for images or other file formats) e.g. FAAGuide.pdf.

You may wish to add further information on these contextual documents –e.g. to assign key words for their content, or to specify the wave of fieldwork in which a particular document (e.g. an interview schedule) was used.

Data documents. These are the data files at the heart of your dataset, including interview transcripts, field notes, life maps, case histories or other descriptive or analytical files. You should list all documents available for each case in your study, devoting one line of the spread sheet to each document. The example sheet show the entries for the three interview transcripts deposited for Trevor and Tricia, which were generated over three waves of data collection. Each of these documents has its own dedicated line on this sheet. Further lines would need to be added for Trevor's life maps, pen portraits, and other documents held for him.

The entries for each document record base data for the cases – this closely resembles the entries in Tab 1 - followed by summary descriptions of each document:

1. Base Data – these entries are similar to those used in the case files in Tab 1, but here you are listing this information for every document in the study. While this is a repetitive process, it is necessary to allow the documents to be described and to

enable the search and retrieval functions of the archive to be set up at document level. It is important to note that, unlike the summary case files, these entries do a valuable job in capturing how the base data changes over time. The first line of Trevor's entry records his circumstances at the time of his first interview. At that time Trevor was single and a student. The second line records details of his interview two years later, when he was still a student but was partnered. By the third wave, a further two years on, he was single again and unemployed. His residence also changed over time.

2. Summary descriptions of each document - these follow on from the baseline entries. They are similar but not identical to the entries in the cases spread sheet. Again, the information is repetitive, but it is necessary to enter this information at this point to enable the archive to describe each document in your dataset. The example shows the documents for the three interviews undertaken with Trevor and Tricia. These are set out on separate lines, and the dates, times, key words and so on are specific to each interview. The entries to be recorded include:

Project name and case ID: your chosen shorthand name for your project and case.

Name of project, case ID, Data Type and year data collected: e.g. Friends and Acquaintances: Trevor-Interview W1 2010.

Subject (keywords): These describe the subject matter of the particular document – we suggest using a maximum of 5 key words to describe the content of each file.

Fieldworker ID: The narrative tag used for the person who collected this document.

Fieldwork date: e.g. 15 May 2010

Wave: enter the wave at which the data file was generated – in the examples above, the three interview documents were generated at waves 1, 2, and 3.

File type (text, image): Record here the format of the data: i.e. text or image

Data type: Record here whether this file is an interview transcript, life map, focus group transcript and so on.

File name: you need to assign a unique file name to each document – without brackets, hyphens or spaces. The usual format, as shown in the example, is the initials of the project title, case ID, wave of interview and a numerical identifier – 05, 06, 07, followed by the format of the file – rtf (rich text format or a tag for images or other file formats). e.g. FAATriciaW2interview06.rtf

Data Anonymised: Specify yes or no here.

We hope you find this guidance useful. If you have any queries about the spread sheets please contact Kahryn Hughes (k.a.hughes@leeds.ac.uk) in the first instance. Advice can also be obtained from researchdataenquries@leeds.ac.uk

Appendix 3: Sample Consent Leaflet and Form

FYFF

Archiving Information Sheet

Following Young Fathers Further (or FYFF for short!) is a four-year research study based at the University of Lincoln. It is funded by the UKRI Future Leaders Fellowship scheme. We would like to learn more about the lives of young fathers and how we might better support young men in their parenting journeys. To do this, we are gathering information via interviews and other activities through the course of the study.

1. Archiving
We would like to store all the interviews and information we generate with you in an archive. This is the space where your data will 'live'. It means that other people will be able to look at them in future and will help them to understand what young fatherhood is like today.

2. What is an archive?
An archive is a bit like a library where the books are all digital only. We would like to store your interviews in an archive based at the University of Leeds, called the Timescapes Archive. It will hold copies of all of the information from you and others who take part in the research. This includes recordings (like sound files) as well as written versions of interviews (usually called transcripts).

Everything will be stored in a digital form in the archive, not as paper in a building.

The archive allows people like researchers and historians to look at the material that we have gathered in our research project now and into the future.

Because we are learning so much and hearing so many interesting things we want to store the interviews and study information to give other researchers a chance to look at them too.

Your information will become an important part of history. What you tell us now will also help researchers to understand young fatherhood in the future too!

3. Protecting your identity
The archive that your interviews are being stored in will have restricted access. Your interviews, recordings, and personal data will not be available to just anyone. We will make sure that the people who look at your material promise to do so in a responsible manner and protect your identity.

We will make sure that any details that could identify you or anyone you talk about in the interviews will be changed before the material goes in the archive.

Your recordings personal details, such as address, telephone number and email, will be stored in the archive with extra protections. They will not be made available to anyone during your lifetime. If you want us to take them out of the archive you can do so at any time by contacting the research team or the Timescapes Archive Director. Contact details for the Archive Director are on the Archive website and are listed below.

4. Agreement to archive
To make sure that you agree that we can archive any data we produce with you, we would like you to sign a consent form. The FYFF team will also sign the form, and give you a copy to keep.

The agreement covers ALL interviews and activities that we do with you as part of the FYFF project.

The form also asks you to agree to share 'copyright', or ownership, of the interviews with our research team.

We ask for shared copyright with you because this means that both you and the FYFF team have control over the data. You can still make use of, or withdraw, any data produced should you wish to. However, shared copyright means that the FYFF team can continue to use your data for research purposes.

Because we take our responsibilities very seriously and do not wish your right to privacy to be affected by helping us with our research, we will make sure that nobody else can look at your interview material unless they have our approval and tell us why they want to see it.

UNIVERSITY OF LINCOLN

UK Research and Innovation

timescapes
An ESRC Qualitative Longitudinal Study

Director of FYFF:
Dr Anna Tarrant atarrant@lincoln.ac.uk

Timescapes Archive:
https://timescapes-archive.leeds.ac.uk/

Timescapes Director:
Dr Kahryn Hughes k.a.hughes@leeds.ac.uk

FYFF

RESEARCH PARTICIPANT:

• I have read and understood the information leaflet that outlines how my interviews and other data will be archived, and I have had the opportunity to ask questions about it.

• I agree that the material can be included in an archive.

• I understand that my personal details will be stored with protected status in the archive but that I am able to withdraw those details in future by contacting the archive.

• I agree to share copyright, or ownership, of my interviews and activity sheets with Anna Tarrant (Director of 'FYFF' project).

• I accept that including my research data in the archive will mean that, in the future, other researchers may also use my words in their reports, books and magazine articles.

NAME:………………………………………………………………………………………….

SIGNATURE:…………………………………………………………………………………….

DATE:……………………………………………………………………………………………

RESEARCHER:

o I have discussed with the 'FYFF' research participant how their interviews and other data will be archived, and given them the opportunity to ask questions about it.

o The 'FYFF' research team will make sure that recordings, personal contact, and identifying details are archived with extra protections, and know their responsibility to ensure that participants benefit from taking part in the research.

NAME:………………………………………………………………………………………….

SIGNATURE:……………………………………………………………………………………

DATE:……………………………………………………………………………………………

Appendix 4:
Sample Front Sheet for Data Files/Transcripts

Pioneers Oral History Study (Paul Thompson, UK Data Service)

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

(Source: Libby Bishop, UK Data Service).

Appendix 5: Sample Transcription Guidelines/Template

(main source: Corti et al. 2020).

Transcription is a process of translating data from one format (commonly an audio file) to another (commonly a text file). Good data management requires the production of high quality and consistent transcriptions across a project. Rather than improvising, a standard transcription template should be developed that ensures a uniform approach (see below). This is particularly important given that transcription is time consuming and commonly outsourced to external transcribers. Transcribers need guidelines that specify how a transcript should be translated and formatted, how it should be anonymised or marked for anonymization, and how to handle files securely to maintain confidentiality (see below for examples).

Styles of transcription are likely to vary, depending on the nature of a study and its purpose; there are no cast-iron rules. Projects that are concerned with emergent themes in people's lives are likely to produce narrative based transcripts. A project using conversation analysis, on the other hand, may specify a range of symbols to represent particular features of speech, such as length of pauses, laugher and other non-verbal utterances, overlapping speech or intonation, background noises, accents and so on (Corti et al. 2020: 63). Either way, it is generally agreed that transcriptions should reflect the speech of a participant, and should not correct grammar or vocabulary. As a general rule, Corti et al. (2020: 64) recommend that each transcription should have:

- ❖ A unique identifier, such as a name or number;
- ❖ A uniform layout for all transcripts, which is compatible with the import features of a QDA software package, if one is being used (formatting such as text in bold or italics, use of colons, or double columns may not be importable);
- ❖ Speaker tags (initials or names) to indicate who is speaking in turn-taking or question/answer sequences;
- ❖ Line breaks between speakers; single spacing for speaker's text, double spacing between each speaker;
- ❖ Numbered pages;
- ❖ A front sheet that contextualises the transcription (see below and Appendix 4);
- ❖ Markers to show where any amendments have been made e.g. where real names have been changed, or how sensitive text has been flagged up for later anonymization (see Appendix 7). Brackets [ ] can be used to identify anonymised data; & ellipses … to denote gaps or edits in the transcript;

❖ For studies conducted in a foreign language, a translation/summary translation in the original language if the file has been reproduced in English; or an English translation/summary translation if the original language is retained.

Standard Transcription Template: Following Young Fathers (Neale 2011)

Project Title	**Following Young Fathers**
File Type	**Interview**
Date of Interview	9th Feb 2011
Location of Interview	University of Leeds
Interviewer	Carmen Lau Clayton
Transcriber	Caroline Maloney
Anonymiser	Laura Davies
Recording Information	Olympus LS11
Consent for Archiving	Yes
Access Levels	Abridged data accessible via Archive; unabridged data embargoed in dark archive
Participant Information	
Unique Digital Identifer (File ID)	Jason01; FYFJasonW1Interview01.doc
Pseudonym	Jason
Wave no.	1
Age at Wave 1	22
Description of Data	**1 Tape, recorded over 80 minutes**
Key Words	Fatherhood; Life History; Education; Employment; Birth of child; Pregnancy; Relationships with co-parent; Relationships with Parents; Custodial experiences; Professional Support

(Start of side A)

Carmen: Well thank you again Jason for coming today. So can you just tell me a little bit about yourself, like how old you are, where you live and this sort of thing?

Jason: I'm twenty two and I don't really do much except go to the gym and look for work every day. And when I'm not doing that I'm just looking after [son] … And try to get him things. ... But there's not much you can do cause he's only [eight] weeks old so, basically you can only feed, feed him and change his nappies. … But I've just like, I've opened a savings account and just started putting money away for him because … he's got hundreds of clothes

Carmen: So [son] is your only child then? He's your first child then?

Jason: Yeah.

Carmen: So what's it like being a father to a very new baby then?

Jason: Well if I'm honest, before he were born I were a bit negative. I didn't want to be a dad, 'cause for starters I'm unemployed. So I can't give him best possible life. But obviously I can do something about that. But it's quite crazy. Once he were born and I seen him, like I didn't even want to go to watch him being born 'cause I were nervous and scared. … And I thought because I didn't like his mother that I wouldn't like him. 'Cause me and his mum fell out … . But once he were born, it's crazy. You just, nothing else matters. … Everything you do is for him. … It's impossible to describe, I think. It's just overwhelming. You are responsible for something that's … that can't be independent and needs help. I don't know. Just, you have to be there for him don't you. You have to sacrifice things to make their life better. But I suppose, obviously I, well, I've got a crap dad so obviously I want to be total opposite and be a good example to him. … Like I used to, and it's bad, I used to smoke weed. Not too often but since he's been born I just stopped it straight away. … I realised that's losing money or things I could do for him or, and setting a bad example. So I just stopped. … …

(end of side A)

(start of side B)

Example of Transcriber Guidance

(Adapted from Alice Jackson for a doctoral study of Bulimia, University of London, 2010; showcased in Corti et al. 2020: 81 -83)

Background to the research: Bulimia (an eating disorder) is very difficult to overcome. How and why some people manage to recover is not well understood. This doctoral study explores this experience through in-depth interviews with people who have recovered from bulimia. The research is funded by … and is being conducted in the Department of … at the University of … . Ethical approval has been granted by … .

Confidentiality: Due to the sensitive nature of the study you will be required to sign a confidentiality agreement which confirms that you will adhere to the principles of anonymity and confidentiality and will not discuss or share the content or nature of the research, or details of the participants, to anyone apart from the researchers.

Conceptual Approach: Verbatim Transcription: Transcription of speech is always a compromise. Greater detail gives more scope for interpretation, but too much detail can slow up the reading of the text and disrupt its flow. This project requires full verbatim transcription. As well as preserving the actual words which were spoken, extra verbal material captured in the recording also needs to be preserved, including

intonation, pauses, rhythm and hesitation. This gives insight into the manner in which the words were spoken (the feeling-tone of the original speech), and records the conversational exchange between interviewer and interviewee.

General Notes:

- ❖ Documents should include a header on every page with the ID number of the interview on the left hand side, and your name on the right hand side;
- ❖ Insert page numbers at the bottom of every page, in the centre;
- ❖ Denote start and finish points of each side of the tape (s) at appropriate points in the transcript (start of tape, side A; end of tape side A; start of tape side B; end of tape side B).
- ❖ Use Times New Roman, font size 12, and type what the interviewer says **in bold;** justify the text;
- ❖ Identify interviewer and interviewee and give line breaks between text;
- ❖ While a verbatim record is needed, this should not include phatic communication/back channel utterances by the researcher (e.g. 'encouraging' noises such as yeah, right, I see, uh huh, mmhhmm), which would interrupt the flow of a participant's narrative;
- ❖ Use punctuation as for normal written prose. Grammar should not be altered or 'tidied up'. Do not use 'eye' spellings (e.g. enuff for enough).

Things to Include in Full:

- ❖ Unfinished questions or statements that trail off. Indicate these with ellipses e.g. I never did understand what it might … it's difficult to express how … ;
- ❖ Repeated phrases, words, statements, questions;
- ❖ Informal discussion after the interview is formally completed;
- ❖ Non lexical utterances from the participant – umms, errs, uhs;
- ❖ Hesitations and pauses – indicate with ellipses, e.g. it means to me, err …
- ❖ Use an exclamation mark to denote surprise, shock or dismay;
- ❖ Use italics to indicate any word or phrase that the participant emphasises;
- ❖ Use a yellow highlight to flag up real names and places, or other identifying or sensitive material in the text, such as work place, job title, age or date of birth. This will help the researcher to anonymise the transcript.

Things to include in brackets:

- ❖ Noises in background (loud banging) (door slams) (muffled voices)
- ❖ The tone of the participant: including comments on mood, feeling, passion, (laughs loudly) (sighs) (mumbles slowly) (whispers) (sound angry) (falters)
- ❖ Use **(inaudible)** in bold, for unclear words/meanings - don't guess meaning.

After Transcription:

Please email completed transcripts to the researcher using [encryption package] and recorded delivery. After receipt and checking of the transcript, I will confirm when you can delete and destroy the recording and transcript from your files.

Thank you for contributing to this project. A sample interview is attached to show how the layout should look. If there are large parts of the interview that cannot be deciphered, please contact me immediately. Likewise, please get in touch if anything is not clear or you have any queries.

Example of a Transcriber Confidentiality Agreement

(source: Alice Jackson, *Bulimia Study*, showcased in Corti et al. 2020: 83-4)

As a transcriber of data from the Bulimia study, I will be hearing and working with recordings of confidential interviews. The participants agreed to take part in this study on the understanding that their material would remain strictly confidential. I understand that I am responsible for honouring this confidentiality agreement.

I pledge not to share any information on these recordings, about any party, with anyone except the researcher on this project. Any violation of this and the terms detailed below would constitute a serious breach of ethical standards and I confirm that I will uphold these standards and adhere to the agreement in full.

I ……………… …………… agree to keep all the research information that has been shared with me confidential by not discussing or sharing the content of the interviews or details of participants in any form (e.g. WAV files, CDs, transcripts) with anyone other than the researcher.

I agree to keep all research information entrusted to me, in any format (e.g. WAV files, CDs, transcripts) secure while it is in my possession (specify security measures).

I agree to return all research information in any form or format to the researcher when I have completed the transcription tasks, using file encryption and recorded delivery.

Following consultation and 'sign off' with the researcher, I agree to erase or destroy all research information in any form/format relating to this project that is not returnable to the researcher (e.g. held on CDs or stored on my computer hard drive).

Transcriber ……………………….. (print name) …………………. (signature) …………… (date)
Researcher ……………………… (print name) …………………. (signature) …………… (date)

Appendix 6: Sample File Structure (Case/Wave)

The case-led file structure set out below was used for the *Following Young Fathers* filing system. The structure distinguishes between unabridged and abridged versions of the dataset. Within each of these versions, the system distinguishes between fatherhood case files, practitioner case files, contextual files (e.g. examples of interview schedules and consent leaflets that cut across cases in the study), analytical files that cut across cases, and (for the unabridged version) confidential files.

The fatherhood files are nested alphabetically by case name, and then by wave and case document (see below). The researchers found it useful to bring together all files relating to a particular case. This brought greater granularity to the grouping of files, and supported case-led temporal analysis (see Neale 2021). As shown above (step 4) the alternative is to list files by wave and then by case.

Following Young Fathers: Abridged Dataset

Following Young Fathers: Unabridged ('Raw' Dataset)

Unabridged Dataset

Fatherhood Case Files

Practitioner Case Files

Contextual Files (recruitment and consent leaflets, Interview schedules)

Analytical files (cross case) (e.g. summary tables, framework grids)

Confidential files (e.g. name and contact details, consent forms, code book).

Fatherhood Case Files (arranged by case, wave, case document)

Adam

Wave 1

Interview audio file

Interview transcript

Life Map

Field notes

Pen Portrait (case analysis updated after each wave)

Case History (case analysis updated after each wave)

Wave 2

Interview audio file

Interview transcript

Life Map (updated, waves 1 and 2)

Field notes

Andrew

Wave 1

Wave 2

Ben

Wave 1

Wave 2

Appendix 7: Anonymisation Guidelines/Code Book

(Adapted from the *Timescapes Study and Archive* (Libby Bishop); and Corti et al. 2020)

Anonymisation guidelines are designed to create consistency and high standards in the way that data are altered. The code book creates a record of how data have been altered, and provides a key that links real with altered data for each transcript. All replacements, aggregations and deletions should be listed. The code book should be kept securely as part of the unabridged dataset. It is worth considering using an anonymization tool to help with light touch anonymising (see Step 5 for details).

File handling procedures

- ❖ Make a copy of the unabridged file and put the original in a secure location;
- ❖ Begin anonymisation on the copy;
- ❖ Tag the files so that abridged and unabridged versions are easily distinguished;
- ❖ Note the name of the anonymiser.
- ❖ Clearly denote altered text, e.g. by placing it in square brackets [ ]. The original *Timescapes* researchers were advised to use a transparent system: @@ at the start of the altered text and ## at the end. Check protocols recommended by your archive; strike a balance between transparency and intrusion in the text.
- ❖ Do not use global *search and replace* facilities, provided, for example, by Microsoft Word. This can lead to unintended corruptions of the text.
- ❖ Transcriber highlights may help to identify text that needs anonymising; this requires guidance to transcribers on what kind of text to mark up.

What to Anonymise

- ❖ Determine what needs to be anonymised for your project, in consultation with your chosen archive. The most common areas are listed below. Strategies include aggregating or generalising from specific information, as well as substituting with a pseudonym.
- ❖ **Names of people**: describe according to significance to the respondent: 'female / male friend', 'mother', 'father', 'teacher' etc. When different friends are mentioned it is best to assign each person a pseudonym, so that they can be followed throughout the waves of interviews. If participants choose their own pseudonym, this can engage them and avoid feelings of mis-identification;
- ❖ **Names of towns/cities/villages**: describe according to the significance of the place to the respondent's life: 'city she grew up in', 'town he moved to', 'small neighbouring village'. In some circumstances, 'London' could remain (e.g., when they are thinking of going to university there, or having a holiday there). Places or streets should be changed according to their relevance in the text. For

example the name of a housing estate may be changed to 'local housing estate' but may need more context, e.g. 'local loyalist housing estate';

- ❖ **Name of country**: describe according to its location in a part of the world or other suitable alternative, e.g. 'country regularly visited by backpackers';
- ❖ **Nationalities:** in close knit communities, talk of a different and specific nationality may easily disclose identity e.g. 'the Americans who took over the pub'. Use the generic 'people';
- ❖ **Name of school/college: e.g.** 'high school she attended' or pseudonym;
- ❖ **Name of workplace:** describe more generally e.g. 'fast food restaurant', 'pub', 'shop', 'factory' etc. Use 'department' or 'section' to denote a particular department within a workplace or company;
- ❖ **School subjects –** to be changed at the discretion of the anonymiser and possibly in consultation with respondents, especially young persons.
- ❖ Developing a guidance chart, such as the one below, can be very helpful.

Original	Example of possible change
Street names and names of local areas, places visited	Either change to a fictional name or describe as 'local area' or 'city centre street' or 'housing estate' or 'housing development'
School names	Change to 'local secondary school' or change to pseudonym
Parent's occupation	Change to a similar/generic description.
Specific businesses, places of work e.g. McDonalds	Change to 'fast food outlet/restaurant'
Places travelled to/ visited/ worked in.	e.g. Italy change to 'Northern European country'
Names of family and friends	Either change to a pseudonym or refer to 'younger brother' 'female friend'
Names of football clubs	Change for example, to 'English football club' or 'local football club'
Name of youth club	Use a pseudonym or 'local youth club'
Change in job, housing, locality over time	Generalise/alter details to obscure unique biographical identifiers

When to anonymise

As indicated above (step 5) it is best to anonymise files as they are generated. However, a full interview should be listened to before decisions are made about what to anonymise. Information may be revealed later in an interview that impacts on how earlier material is treated. In like fashion, some retrospective anonymising may be needed for longitudinal data. A biographical picture is gradually created over time that can be revealing of identities, while the significance of particular data may only be revealed after a number of interviews. A staged approach may therefore be needed, with a later refinement of earlier anonymising protocols. One strategy is to adopt a light touch anonymising at the outset, and develop a more comprehensive and retrospective approach once data for a case are fully assembled.

Example of Code Book
(Source: UK Data Service; Corti et al. 2020: 123)

Transcript /Page no.	Original	Changed to
Interview 2		
2/1	Age 27	Age range 20-30 (in their 20s)
2/1	Spain	European Country
2/2	Manchester	Metropolitan City
2/3	20th June	June
2/5	Amy (real name)	Moira (pseudonym)
2/6	Station Road Primary School	Local primary school
2/7	Head Buyer, Produce, Sainsbury	Senior executive, leading retailer
2/8	Francis	My friend
2/9	Aunty Jean	Aunty
2/11	North Colchester	Northern part of town.

Appendix 8:

Sample Project Landing Pages for the Timescapes Archive

Changing Landscapes: Pathways Through Participation Project

Timescapes

HOME | BROWSE | CONTACT

Browse by Project

🔺 Up a level

Export as [ASCII Citation ▾] [Export]   🔶 Atom 🔶 RSS 1.0 🔶 RSS 2.0

Young Lives and Times

This study followed an age cohort of young people to explore the unfolding of their personal and family relationships over time, and their changing educational aspirations and achievements. We aimed to shed light on how young people practice and 'work out' their personal relationships and identities at a time of intense change through their teenage years, and in relation to their values and aspirations for the future. There is evidence to suggest that the life chances of young people - their ability to achieve stable relationships in adult life, and to achieve well educationally - are causally linked to their earlier family experiences. However, longitudinal data that can shed light on how these influences operate are limited. We aimed to investigate these patterns and processes through a broadly constructed sample of young people, and more intensively through a subsample of young people entering early parenthood.

- **Title:** Young Lives and Times
- **Dataset Publisher:** University of Leeds, UK

More details about this project can be found here: Young Lives and Times.

Number of items: **28**.

Young Lives and Times: Alicia. [Case]

Young Lives and Times: Amy. [Case]

Young Lives and Times: Andy. [Case]

Young Lives and Times: Ben. [Case]

Bren Neale is Emeritus Professor of Life Course and Family Research in the School of Sociology and Social Policy at the University of Leeds. She directed the ESRC *Timescapes* Study from 2007 to 2012, as part of which the Timescapes Archive was established. She has published extensively on Qualitative Longitudinal Research and contributed to advances in this methodology across academia, government and the voluntary sector. She continues to provide training and support in QL methodology for new and established researchers.

Kahryn Hughes is Director of the Timescapes Archive and an Associate Professor in Sociology in the School of Sociology and Social Policy at the University of Leeds. She has played a leading role in the development of secondary analysis for qualitative archival data.