

INTRODUCTION

The archiving, sharing and re-use of research data is a growing part of research culture in the UK, and is central to the Timescapes Initiative. Advances made in these fields under Timescapes have been the subject of several guides in this series (guides no. 16, 18 and 19). In this guide we focus on a dimension of research practice that is an important precursor to data sharing and re-use – the effective and ongoing management of complex data by primary research teams prior to archiving. In developing this guide we are drawing on some of the general principles of data management developed by the UK Data Archive (Van den Eynden *et al.* 2011a), and our experience within Timescapes of developing a specialist infrastructure for Qualitative Longitudinal (QL) archiving and re-use. This guide draws out the distinctive features of data management for QL research, and—used in combination with these existing guidelines—it provides a short overview of good practice for researchers who are gathering complex qualitative data over time. All data management guidance stresses the value of early planning for data management when undertaking QL research. Building this task into the research process at an early stage serves two purposes. It enables QL data and metadata (data about data) to be organised for cumulative analysis by the primary team, while at the same time creating archive-ready datasets that are available for re-use and which are acknowledged in their own right as important outputs

KEY POINTS

- Data management planning for QL research serves two important purposes: it aids longitudinal analysis of QL data by the primary analysts, and supports the sharing and re-use of the data by secondary analysts.
- The complex nature of QL data, the elongated time frames for enquiry and the cyclical nature of the research process create particular challenges in managing QL data; building data management into the research process from an early stage will help to meet these challenges.
- In assessing existing data sources, it may be beneficial to approach QL researchers directly; currently, existing datasets are scattered and many are not readily available in archives.
- Sufficient resources need to be built into QL research budgets to enable data management as an integral part of the research process.
- Consent for archiving is a process, rather than a one-off event: this needs to be borne in mind in data management record keeping.
- Attention to secure storage and labelling of data, both digital and non-digital, is essential. The temporal framing of QL research, and the implications for the continuity of the research team increases the potential for data loss; tacit knowledge by researchers cannot be relied upon in the long term.
- Data need to be organised by wave and case, as well as by file format and other categories. The iteration between waves and cases is essential to aid longitudinal analysis.
- Producing documentation for a QL dataset is integral to good research practice. Of most value are descriptive case profiles and longitudinal case histories that condense complex data into manageable forms for primary and secondary analysis.

ISSN 2049-9248 (Online)

TIMESCAPES METHODS GUIDES SERIES 2012

Guide No. 17

Data Management for Qualitative Longitudinal Researchers

*Libby Bishop &
Bren Neale*

SERIES EDITORS

*Bren Neale &
Karen Henwood*

from a study. It is also important to bear in mind that since the QL research process is cumulative and cyclical, rather than linear, the tasks of QL data management are also cumulative and cyclical. With each new wave of data collection, the data life cycle begins afresh, requiring careful, timely and ongoing attention to these tasks, to ensure they are synchronised within the research process. Drawing on our experiences under Timescapes we make a number of practical suggestions for managing QL data, while acknowledging the challenges that face QL researchers in dealing with complex data. Since every dataset is unique, we have aimed here for broad principles that can be adapted to the specifics of individual projects.

BACKGROUND

Data management planning involves anticipating prospective challenges and setting up systems to create high quality and sustainable data for cumulative analysis by the primary team, and for archiving and sharing with secondary analysts. Data sharing is increasingly expected by funders; indeed, a data management plan is now a required element of an application for ESRC funding. Building good data management into the QL research process at the early design stage is vital. QL datasets are complex; they grow through successive waves of data collection and the extensive volumes and types of data need to be organised, preserved and represented in waves, cases, file formats and themes, to facilitate their analytical use and re-use.

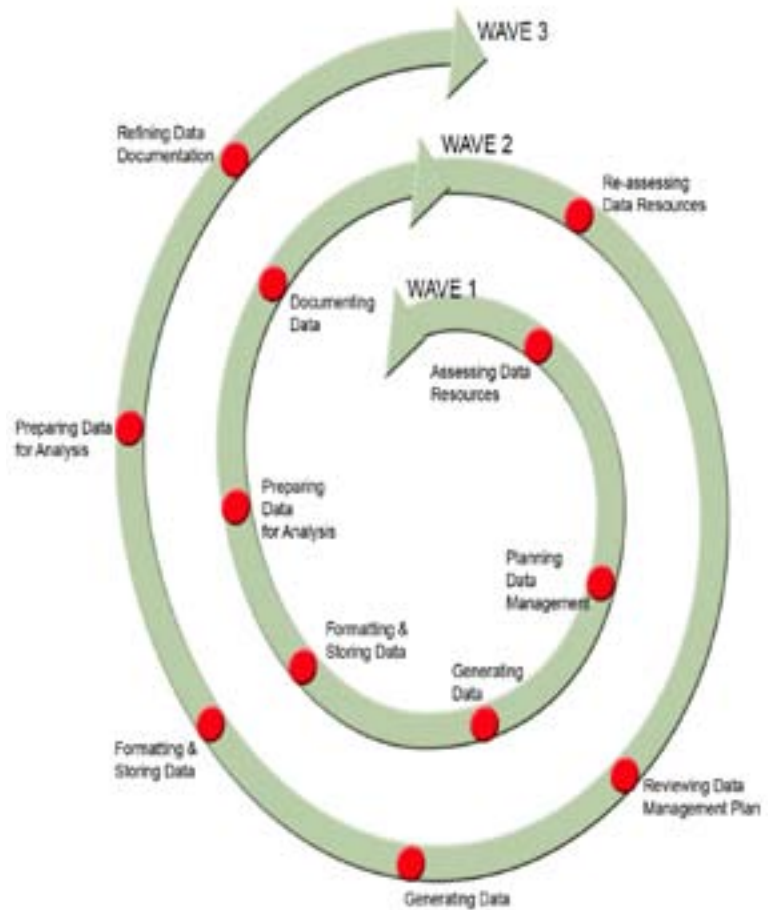
Below we outline a number of considerations to be taken into account in QL data management. We have grouped these under the following broad headings: Assessing existing data sources; Planning for data management, Generating data for primary and secondary use; Organising and storing data; Preparing a QL dataset for primary and secondary use; and Documenting data to aid analysis and re-use. Some of these considerations also appear in the UK Data Archive's data management life cycle (<http://www.data-archive.ac.uk/create-manage/life-cycle>) but have been adapted for QL data, where tasks occur in overlapping and iterative cycles, linked to successive waves of data collection (see diagram).

In what follows we address these different areas of the life cycle, in each case drawing out the specific considerations that arise in dealing with QL data.

RESEARCH DESIGN AND PRACTICE

1) Assessing Existing Data Resources

This is a crucial element at the early design stage of a new empirical study. What existing data resources exist in the proposed substantive field of study? Is there a need for new data? How will a new dataset fit in with extant data resources and what is the scope for bringing datasets



together for cross project comparison and analysis? It is worth bearing in mind that QL datasets are scattered and may continue to reside with individual research teams rather than in data archives. It may be worthwhile to contact QL researchers working in the same field to ascertain what data currently exist, and to explore possibilities for data sharing outside formal archives.

2) Planning for Data Management

Having ascertained the need for a new dataset, advance planning for data management is essential. Key issues to address early are obtaining informed consent and clear intellectual property rights for data sharing. A number of further questions arise about the resources and funding needed for managing the data: who will hold responsibility for ensuring data is managed and who will attend to data management and preparation tasks? What time is needed and what technical infrastructure should be located or created for safe storage, preservation and preparation of data for re-use? The UK Data Archive's Data Management Checklist is a valuable planning tool at this stage (<http://www.data-archive.ac.uk/create-manage/planning-for-sharing/data-management-checklist>).

Given the complexity of QL data and the gathering of these data over successive waves, the resources needed are likely to be significantly higher than anticipated; depending on the nature and complexity of the data, allocating 40 percent of a dedicated full time post for these tasks, and

ensuring an appropriate skill set in the post holder, is likely to be a minimum requirement. The UK Data Archive Data Management Costing tool is likely to help in this early stage planning (<http://www.data-archive.ac.uk/create-manage/planning-for-sharing/costing>). A crucial consideration in a QL context is that data management needs to run alongside the research process, rather than being seen as an administrative task that is 'tacked on' to the end of a project. It is also important to consider at this stage where a dataset is likely to be archived. Archives have different conventions and facilities for the storage and dissemination of data, which will need to be taken into account in the way that data management tasks are approached. The Timescapes Archive, for example, offers four levels of access for data and facilities for thematic searching, requiring the production of key words attached to transcripts (guide no. 16). Whatever systems are set up, it is worth reviewing them at regular intervals; data management planning and review should be a standing item on the agenda of research team meetings.

3) Generating Data for Primary and Secondary Use

An important consideration at this stage is to ensure that data is of a high quality, both scientifically and technically. Qualitative fieldwork is a highly skilled task that requires training to ensure data files contain high quality material for primary and secondary use. The potential sharing of a dataset creates added pressure to ensure quality. One example is the need for sufficient technical skills in sound recording to ensure that audio files are of a high standard for transcribing and also preserving as part of an enduring dataset. Acquiring high quality sound equipment is important here; in Timescapes we drew on standards adopted by the British Library Sound Archive. QL research often utilises life history interviewing, where audio files have particular value as sources of data for analysis (see guide no. 12). Decisions need to be made early about the file formats for gathering data and their future proofing.

Early planning for data sharing entails seeking consent for archiving from research participants, and consulting over how they wish data about their lives to be represented (e.g. choosing pseudonyms or retaining real names). In QL research in particular, informed consent is not a one off event but a longer term process (see guide no. 18). Some researchers keep consent for archiving separate from the process of attaining consent for participation in a project, while others combine the two processes. The nature of consent also varies: from written consent through to recording consent during an interview. A template for informed consent is available on the Timescapes website (<http://www.timescapes.leeds.ac.uk/data-archive/archiving-guidance.php>). The provisional status of consent therefore needs to be built into data management record keeping. A related task at this stage is to consider issues around copyright and intellectual property rights. Researchers should gain permission for copyright in the data to be transferred from the participant to the research team; the data is then held by the archive on license, with copyright

retained by the researchers. Preparing information sheets for participants about the project, about archiving and the potential uses of data will support these processes. (For more detail on Timescapes specific procedures, see Bishop 2009).

4) Organising and Storing Data for Security, Preservation and Retrieval

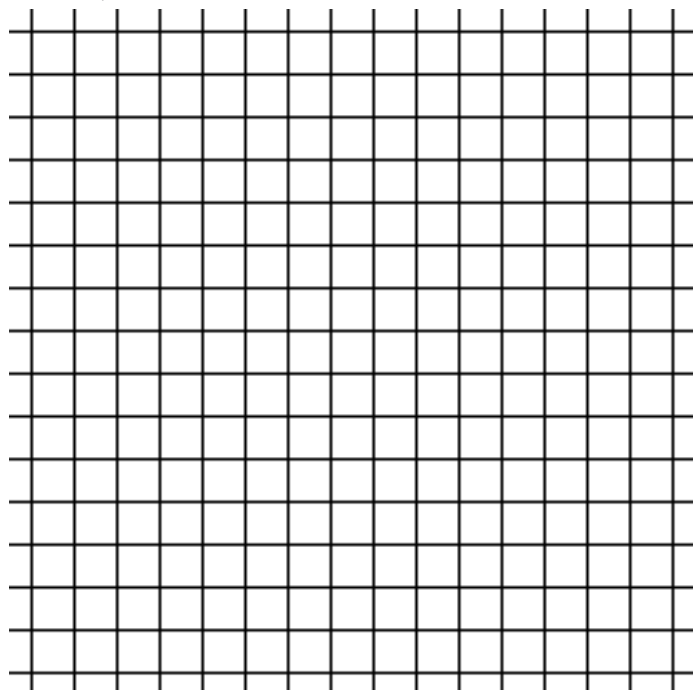
QL data are rich and detailed and gathered in waves (over time) as well as cases (cumulative data relating to each individual, group or organisation). They may appear in more than one format. QL researchers often combine in-depth interview techniques (generating audio files and transcripts) with focus groups and ethnographic methods that yield data such as participant generated diaries and written accounts, and visual and diagrammatic data. The temporal framing of QL research opens up the possibility to creatively combine different techniques of data gathering, resulting in complex datasets.

Decisions need to be made early in the research process about how best to store and label data, and in what file formats, to ensure they are securely and safely preserved and easily identified for retrieval. For example, digital files may be held in a shared drive that is password protected for team access only, and with daily back up to hold data securely. A well organised file structure, with clear file names, codes and abbreviations that are consistently labelled is particularly important for collaborative research involving teams with changing membership over time. Decisions also need to be made about storage of hard copies of data, including, for example, the safe and secure storage of consent forms for archiving, and the real names and contact details of participants. Whatever file formats and software are used, these need to be future proofed, to guard against systems becoming obsolete with changing technology. File formats suitable for long-term preservation can be found here: (<http://www.data-archive.ac.uk/create-manage/format/formats-table>).

The system needs to distinguish between, and control for, different versions of files (raw data files, anonymised files, back up files). These different versions can proliferate and create instability in a dataset. It is surprisingly easy to lose sight of data, even after a relatively short period of time, unless a well-organised system for storage, labelling and retrieval is in place. The tasks and the challenges are magnified in QL research, where data may accumulate over long periods of time and the likelihood of breaks in continuity of the research team is magnified. Tacit knowledge held by individual researchers cannot be relied upon and may eventually result in lost or unidentified data.

For QL research, data should be organised for retrieval by waves, as well as by cases and data formats. The iteration between waves and cases is essential to aid longitudinal analysis by primary and secondary researchers and reflects

the growth of the dataset in two critical dimensions – waves of cumulative cross sectional data, building up a picture of changes across the sample; and cumulative case-based data, building up case histories of particular individuals or groups over time. Data grids that allow for two dimensional documentation, with waves in one dimension and cases in the other, which can be built up thematically (e.g. the Framework programme, now incorporated into NVivo 9), may be useful here as an organising structure for data and providing a foundation for analysis.



Alongside this system, good record keeping is essential. A basic task at an early stage is to set up a template for a data inventory - an expanding chart or 'roadmap' that logs the data trail from generation in the field to deposit in the archive. This can be filled in as a project grows, recording what data has been gathered, when, where and by whom, in what formats, and where data files (both digital and hard copy) are located. The chart needs to reflect waves, cases and formats and can record progress against data processing tasks for each file.

5) Preparing a QL dataset for primary and secondary use

The basic considerations here concern transcribing and anonymising data for analysis and representation. These are the main elements of data processing and they need to be undertaken with care. Consistency and accuracy are important for transcripts and transcriber guidelines should be produced. The transcription template needs to be compatible with analytical software packages, where used. Anonymisation ranges from light touch to more extensive alteration of the data, with a balance to be struck between preserving confidentiality for the participants and ensuring that the data is not stripped of context and meaning (guide no. 18). We have produced guidelines for both of these tasks (<http://www.timescapes.leeds.ac.uk/data-archive/>

[archiving-guidance.php](#)), which conform to UK Data Archive standards. Where non-digital files are created, decisions are needed on whether to digitise the data for wider access.

An important task at this stage is the creation of descriptive, factual data about the sample and content of interviews. These descriptive files condense large volumes of narrative data held in transcripts and audio files and provide a bridge to the analytical process. The production of case histories (guide no.6) and case profiles that summarise data across waves of fieldwork is common practice for longitudinal analysis, and in Timescapes, proved equally useful for secondary analysis.

6) Documenting Data

An important task in preparing data for re-use is the provision of contextual information, both structured metadata and unstructured documentation, that can be used to situate the dataset and promote and aid its re-use. Researchers take a variety of approaches to documenting their datasets, from minimal guidance to the production of gold standard guides. Gold standard guides are valuable for the primary researchers in promoting a dataset and framing it as an important output of a study (Henwood *et al.* 2012; Weller and Edwards 2012; Inventing Adulthood 2011). They are also important in ensuring that data are not misrepresented but appropriately understood and interpreted by secondary users. However they take time to produce and need updating with each wave of data generation. Guidelines for the production of documentation for QL research are available and recommendations have been made by the Timescapes secondary analysis team (Irwin and Winterton 2011). The following checklist draws on these sources:

- A description of the project, including title, research aims and questions, research design and methods, funders, and details of research team, institutions and start and finish dates.
- A description of the dataset: the number of participants, cases, waves, gathered over what time scales. This can include the data inventory and details of how data has been anonymised and formatted.
- Field Documents: interview schedule, fieldnotes, research diaries, and details of samples, sampling frames and modifications over time.
- Descriptive data – e.g. key words assigned to data files to denote content and aid thematic search and retrieval. If descriptive analytical files have been produced as part of the dataset these can be documented here. A publication and presentation list linked to the project.
- Specific information for secondary analysts: for example, guidance on sampling from archives, and overview of themes with potential for further research.

CONCLUSION

Data management is essential for ensuring that the data produced can be shared, and planning for data management is best done early, hand-in-hand with research design. Due to the particular complexities of data generated during QL research—cases, waves, multiple file formats, versions, and so on—the tracking, organising, and general management of data is that much more vital, making it essential to lay out an overall data management strategy. That strategy needs to create a plan, and also the approach, specifically by determining which elements of management need more centralised planning and which benefit from a more devolved approach. Some activities, such as security and file naming, benefit greatly from high levels of standardisation and thus a more centralised model; others, such as consent procedures, must be customised to specific projects and therefore are best decentralised (even if, for example, a standard consent form is used as an initial template). Reassuringly, these are the same findings reached by the JISC/ESRC collaborative Data Management Planning Project (van den Eynden 2011b). However, finding the right balance is more art than science, and, like the doing of QL research itself, remains a continuous and always provisional undertaking.

REFERENCES

Bishop, L. (2009). 'Moving data into and out of an institutional repository: Off the map and into the territory,' *IASS/IST Quarterly IQ*, 31(3&4), <http://iassistdata.org/downloads/iqvol313bishop.pdf>.

Hadfield, L. (2010) *Balancing on the Edge of the Archive: The Researcher's Role in Collecting and Preparing Data for Deposit*. Timescapes Working Paper No. 2, 60-73. <http://www.timescapes.leeds.ac.uk>

Henwood, K., Shirani, F. and Coltart, C. (2012) *Masculinities, Identities & Risk: Transition in the lives of men and fathers* - Timescapes Project 4 – Project Guide. <http://www.timescapes.leeds.ac.uk>

Inventing Adulthoods (2011) *Making the Long View: Sharing the Inventing Adulthoods Project*, <http://www.lsbu.ac.uk/inventingadulthoods/archiving/mlv2/index.shtml>.

Irwin, S. and Winterton, M. (2011) *Debates in Qualitative Secondary Analysis: Critical Reflections*. Timescapes Working Paper Series, no. 4. <http://www.timescapes.leeds.ac.uk>

Van den Eynden, V., Corti, L., Woollard, M., Bishop, L. and Horton, L. (2011a) *Managing and Sharing Data: Best practice for researchers* Wivenhoe Park: UK Data Archive. ISBN: 1-904059-78-3.

Van den Eynden, V. Bishop, L., Horton, L. and Corti, L. (2011b) *Data Management Practices In The Social Sciences*, (August 10) UK Data Archive.

Weller, S. and Edwards, R. (2011) *Timescapes: 'Your Space! Siblings & Friends' Project Guide*. <http://www.timescapes.leeds.ac.uk>

RESEARCH TEAM CONTACT DETAILS

Libby Bishop | ebishop@essex.ac.uk

Bren Neale | b.neale@leeds.ac.uk