

INTRODUCTION

Qualitative data archiving for secondary use is a relatively new field of research practice, giving rise to a range of ethical considerations. A growing body of literature is devoted to addressing these issues, and there is growing expertise in applying ethical protocols for re-use, for example, through the UK Data Archive (<http://www.data-archive.ac.uk/create-manage/consent-ethics>), British Library (<http://sounds.bl.uk/information/legal-and-ethical-usage>) and Oral History Society (<http://www.ohs.org.uk/ethics/index.php>). In this guide we focus on the ethics of sharing and re-using Qualitative Longitudinal (QL) data. QL datasets, with their extended timeframes for data generation and analysis, present some distinctive ethical challenges. We draw here on our experience of archiving and sharing data under the Timescapes initiative, in which we developed a stakeholder approach to ethical practice. This guide complements guide no. 11 in this series, which focuses on the ethics of QL primary research, guides no. 16 and no. 17 on the Timescapes archive and data management planning, and guide no. 19 on secondary analysis.

KEY POINTS

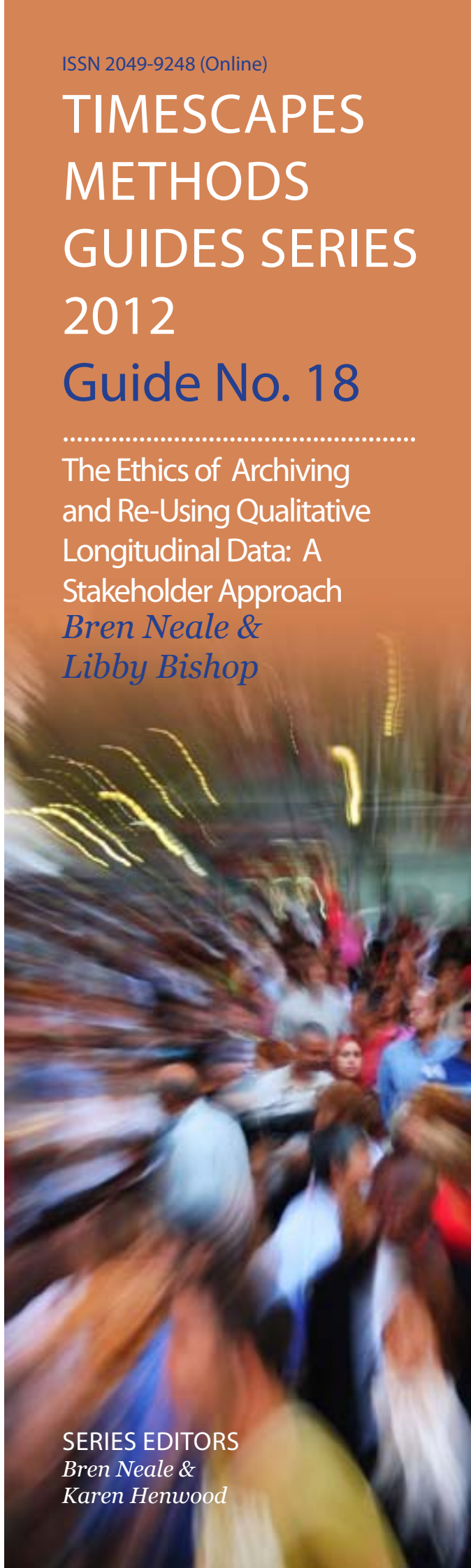
- A stakeholder approach to ethical practice is useful, since it allows flexibility to work out an appropriate balance between the needs and requirements of research participants, primary researchers, secondary users, funders, and the wider public. The approach was effective in Timescapes, enabling us to minimise risk while still enabling access.
- The complex nature of QL research and the extended time frames give rise to distinctive ethical challenges in the archiving and re-use of the data. This is particularly so where primary and secondary teams are engaged in analysis at the same time.
- Consent in QL research is an ongoing process rather than a one off event. Consent needs to distinguish between two dimensions of re-use – the infrastructure and arrangements for re-use which can be specified and agreed in advance, and the modes of analyses that the data may be subjected to, which cannot be specified or agreed in advance.
- QL data often present higher risks of disclosing identities than cross sectional data. However, in preparing data for archiving, a balance needs to be struck between the drive to preserve confidentiality and the drive to produce an authentic and robust dataset for posterity.
- There is never a perfect time to 'let go' of data: primary QL researchers value flexible time frames for archiving that mesh with the time frames of enquiry,
- The work of preparing QL data and metadata for archiving is complex, skilled, and time consuming. It requires adequate funding and acknowledgement for the primary researchers.
- Registration procedures for archive users offer ethical safeguards, but for QL research tighter access controls may benefit the primary teams and facilitate data deposit.
- The re-use of QL data requires attendance to the needs and ethical claims of both primary and secondary researchers, which are equally valid; bringing these researchers into conversation with each other may help to foster an ethic of mutual care and respect.

ISSN 2049-9248 (Online)

TIMESCAPES METHODS GUIDES SERIES 2012 Guide No. 18

The Ethics of Archiving
and Re-Using Qualitative
Longitudinal Data: A
Stakeholder Approach
*Bren Neale &
Libby Bishop*

SERIES EDITORS
*Bren Neale &
Karen Henwood*



BACKGROUND

From its inception, Timescapes was committed to archiving and re-using the data generated from our network of QL projects. This commitment was made with full knowledge that the nature of the data would make ethical data sharing especially challenging. At the same time, it is important to note that the features that make QL data challenging to archive and share are precisely the ones that make such data exceptionally valuable for long-term re-use. Large-scale QL studies are rare and create large volumes of data. They represent major investments of intellectual, monetary and personal resources, making them deserving of long-term curation. And, most significantly, longitudinal data becomes more valuable over time, as information about the sample becomes deeper and broader.

This belief in the potential wider benefits of sharing the data was a key motivation for addressing the ethical challenges of archiving the Timescapes datasets. In each area of practice we sought to meet the sometimes conflicting needs and interests of three key stakeholders in the resource: research participants, primary researchers and secondary users. Building relations of trust and respect between participants and researchers and between primary researchers and secondary users is of central importance. The primary/secondary researcher relationship is particularly salient for QL research. Longitudinal data, when gathered through repeated encounters over many years, blurs the primary/secondary divide. Over what length of time does primary analysis become secondary? If one revisits one's own data and brings it into conversation with other datasets, is this primary or secondary research? There is no fixed answer and no clear point at which primary analysis (by the original team) ends, and secondary analysis (by others) begins. What is clear is that where QL researchers revisit their data for cumulative analysis, they may be doing so alongside secondary analysts, who might be asking different questions of the data and producing different interpretations at the same time. This was the case within Timescapes, where it was a condition of funding that we demonstrate the value of sharing and re-using data during the life time of the project (Neale *et al*, 2012). Building an ethics of re-use that takes this overlap of research endeavours into account is, therefore, particularly important for QL research.

Most researchers in Timescapes subscribed to an ethical stance that emphasizes the particularities of any ethical question, and typically calls for negotiation in resolving ethical questions, rather than relying on the application of general rules (Mauthner and Edwards, 2010). Nevertheless, having a broad framework of considerations within which specific ethical issues may be worked out in practice is very helpful (Bishop, 2009). In this guide we do not give specific guidance on ethical protocols or problem solving, although we do refer readers to templates that we have developed for consent, transcribing, and anonymising. Our intention

is to provide a broad overview of issues around the ethical archiving and re-use of QL data, drawing on our experience within Timescapes.

RESEARCH DESIGN AND PRACTICE

The Stakeholder Approach

Timescapes has developed an innovative stakeholder model of archiving data that enables archiving to be seen as an integral part of the research process. Primary researchers who deposit their data with Timescapes are stakeholders in the resource and, given the long time frames for their research, are encouraged to re-use data as well as deposit. (Neale and Bishop, 2012). We found that this approach was vital in addressing the ethical challenges of archiving QL data. Only by this means could we take into account the perspectives of all those involved, including research participants, primary and secondary researchers, archivists, funders and the wider public. Their varied needs are set out below:

- Social researchers—Timescapes included—are committed to the welfare of research participants and to ensuring that they come to no harm from engaging with research (Hadfield, 2010). This duty of care was a foundational element of the archiving project, covering issues of consent and confidentiality, the ethical representation of accounts and interpretations, and assurances that secondary use would not render participants vulnerable to exploitation.
- The duty of care extends to the researchers themselves, who also appear in the data and whose professional work in producing and preparing a dataset is open to wider scrutiny (Neale and Bishop, 2012). Such scrutiny has different implications for early career and for established researchers. The timing of re-use also has an effect on primary teams.
- The wider professional community, including archivists, funders and secondary users, has a stake in matters of transparency – and in broadening evidence and advancing knowledge through the re-use of valuable data. This is consistent with scholarly values of being willing to expose one's work to scrutiny.
- Finally, the public has an interest, in that public funds contribute to the ability of researchers to generate such data. There seems to be a reasonable claim that publicly funded research should be available to the maximum extent possible.

Acknowledging multiple ethical claims does not nullify duties of care, or rights of participants, but rather recognises that ethics involve situations that arise when equally compelling duties may come into conflict (Bishop, 2009).

Below we set out a range of considerations that informed our approach to ethical archiving and re-use within Timescapes. These include: gaining permission for archiving; the preparation and representation of data by the primary team; the protection of data in the archive; and the ethical re-use of QL data.

Gaining Permission for Archiving: Informed Consent

The issue of participant consent for archiving and re-use is of central importance. It hinges on a number of key considerations: what, exactly, is informed consent in this context? What are participants consenting to, and to what extent can consent for re-use be deemed “informed” when future uses of data are unknown?

Our approach to the issue of informed consent was helped by the adoption of a broad interpretation of “informed”. While precise future uses are unknowable, it is possible to provide examples of the ways in which similar materials have been used, typical re-users, and so on. There is a related strategy - called generic or enduring consent in medical research - in which there is no time limit on consent, and consent is granted for unspecified hypotheses and procedures. This strategy has also been identified as appropriate for longitudinal social research (ESRC, 2010). For Timescapes, the extended time frames that enabled relations of trust and respect to develop between researchers and participants, proved to be an advantage. Where possible, written consent for archiving was obtained as part of the general consent process at the start of the study (see templates at: <http://www.timescapes.leeds.ac.uk/the-archive/researcher-documents.php>). Being transparent about consent for archiving at the outset is important where funders increasingly expect data to be archived. But the researchers were able to work with a model of process consent that entailed revisiting the topic of archiving over the course of the study, and at times when trust had developed and the majority of data had been collected. Participants were thereby able to reflect on their consent status as an ongoing process, (Cutcliffe and Ramcharan, 2002), and with increased understanding of the research and their part in it as the study progressed.

In further teasing out the question of what participants are consenting to, it is worth distinguishing between two dimensions of re-use. The first relates to the infrastructure and arrangements for re-use – by whom, under what conditions, and with what safeguards to maximise confidentiality (these include, for example, legal obligations such as copyright and institutional controls through licence agreements and Research Ethics Committees). As we show below, these dimensions are relatively straightforward to specify in advance, making them a viable basis for negotiating consent. The second dimension of re-use, on the other hand, relates to the kind of analysis that secondary researchers engage in and the interpretations that they produce. As we show below, it is not possible to specify these dimensions in advance. It could be argued

that participants should have ‘veto power’ over the use of their data, specifically the right to reject interpretations or conclusions over which they disagree. However for Timescapes this was not seen as a viable basis upon which to negotiate consent.

It is reassuring to know, in this context, that participants do not assume a monopoly on how data about their lives should be interpreted. Participants may have some input into processes of interpretation, particularly in participatory or action research, which are founded on principles of social justice and empowerment. But the production of new knowledge is more likely to be seen as the task of the researcher, who is in a position to place individual data in the context of the overall study and the wider research evidence. Whether this task is carried out by the primary team or other bona fide researchers may also make little material difference, as long as there are safeguards for maintaining the integrity of participants’ accounts. Where participants’ views on archiving have been sought, they have been generally positive about their accounts forming part of a social historical resource and contributing to the common good (Weller, 2010; Bishop and Neale, 2010). This was certainly our experience within Timescapes where 90% of participants gave full consent for archiving and a further 10% agreed to archive with restricted access or embargoes in place.



Preparing and Representing Data for Archiving

A number of ethical considerations arise for primary researchers in the preparation of data for archiving. For QL researchers, the investment in existing samples is deep – personally and intellectually. All researchers feel responsible for participants, but inevitably these feelings strengthen in long standing relationships. The pressures can create mixed effects, both a great commitment to handle people's accounts of their lives with utmost care, combined with a great sense of responsibility to get voices heard as widely as is appropriate (for further discussion of these themes, see the contributions to Henwood et al 2012).

A prime duty of care is to ensure confidentiality for research participants through the process of anonymising data (for an example of anonymisation guidelines see: <http://www.timescapes.leeds.ac.uk/the-archive/researcher-documents.php>). Yet, in comparison with most cross sectional interviews, the accumulation of personal, biographical, geographic, and other details over time increases the risk of disclosure. Within Timescapes the issue of confidentiality was heightened where data on personal lives was of a sensitive nature, where study samples were deemed vulnerable, or where samples were drawn from identifiable localities. An accumulation of references to local landmarks, groups or institutions in multiple waves of data may pose higher risks of disclosing identities. A further complexity is that QL designs may involve bringing research participants together at intervals, or creating ongoing participant networks in ways that could undermine confidentiality. The same issue arises where the unit of analysis is a particular family, community or organisation; here the risks to internal confidentiality and, therefore, the challenges of anonymising data – and re-using it ethically - are all the greater.

The pressure to protect data, however, is not the only consideration and it may run counter to the drive towards authenticity and robustness in the representation of a dataset. The production of QL datasets for archiving can be seen as a key output from a study, representing a major investment of time and skills by the primary team. Researchers are under an obligation to produce high quality datasets that represent data accurately and enable participants' accounts to be heard and understood in context. This may be best achieved through minimal alteration of the data; stripping out data may produce a sanitised version of participants' lives. Care is needed, for example, with the production of transcripts in order to retain the authenticity of the aural interview (for transcription guidelines see: <http://www.timescapes.leeds.ac.uk/the-archive/researcher-documents.php>). In this context, over-anonymising can strip a dataset of its integrity and diminish its intellectual meaning and scientific value, to the detriment of the primary team as well as the secondary user. Some forms of data, such as still

and moving images can be technically altered (e.g. blurring faces) but at high cost to the integrity and quality of the data.

As part of this drive to authenticity, primary researchers need to create adequate metadata (data about data) – e.g. fieldnotes and other contextualising material. This can enhance the status of a dataset for the originating team and its value for re-users, while also decreasing the risk of distortion by secondary researchers. However, the burden of this time consuming work may fall disproportionately on early career researchers whose work may go unrecognised. While data would not have been generated without funding and institutional support, it would also not exist without researcher labour (Maunther in Miller *et al*, 2012). The production of QL datasets for archiving, then, deserves recognition by the wider research community, and the process needs to be costed fully in research proposals and adequately resourced by funders.

For all these reasons, the preparation of Timescapes data, while informed by central guidelines from the archiving team, were implemented according to the needs of individual projects. Some researchers, following the norms of oral history, used a relatively "light touch" in anonymisation. Others, following more conservative social scientific norms, felt obliged to do more, and provided specific guides to future users that key information (e.g., geographical information) had been removed or changed (Hadfield, 2010). The production of metadata also varied across the teams, from 'gold standard' through to more modest documentation. Finally the timing of archiving also varied, with some teams depositing data in the early stages of the research, while others delayed until later in the process. There is no right time to archive data; indeed, there remains a deep ambiguity about the right moment, when QL data are, unavoidably and inherently, always provisional, contingent and in process. Whatever the timing, it represents a balancing act between the needs of primary and secondary researchers and indeed between protecting and openly representing the lives of the participants. Across Timescapes, then, the varied ways in which researchers prepared their datasets reflected different ways of striking an ethical balance between these competing considerations.

Protecting Data in the Archive: Registration Conditions and Access controls

The ability to control access to data in the archive, and specify rules for its use, is a cornerstone of ethical archiving practice, and provides a necessary complement to the strategies of seeking informed consent and altering data to hide identities. Within Timescapes, such controls on secondary use were particularly attractive in overcoming some of the conflicting interests surrounding anonymising and altering data.

Access to the Timescapes Archive – a digital resource – is via a password controlled registration process. This is reserved for bona fide researchers from authorised academic, policy related or public institutions; frivolous and commercial uses are excluded. Registered users sign a legally binding licence, which lays down strict terms and conditions. For example, they can only access the archive for the purposes of learning, teaching or research.



They sign up to strict protocols for the confidentiality of the research participants, including undertaking not to share data with third parties. The copyright in the material and the intellectual property rights of the primary researchers are also protected: secondary users are requested to acknowledge the original data creators in any publications arising from the secondary research. Data centres cannot offer cast iron guarantees that these agreements will not be breached, nor can they be held responsible if they are (although they will impose harsh sanctions on the breachers). There are cases, for example, 'Climate-gate' at UEA and tree ring data at Queens Belfast, where FOI (Freedom of information) was used to obtain data, (although, crucially, in neither case were these data protected by a licence with a data centre). Nevertheless, despite these risks, registration and licence agreements provide an important framework for the regulation and control of data use.

In Timescapes, our stakeholder approach necessitated further safeguards. We created a restricted level of access that depositors could use for certain data – e.g. data that are sensitive or difficult to anonymise. To gain access to such data, registered users must submit a further

application, which is screened and considered for approval by the primary team. While this places a burden of long-term stewardship on the primary researcher, it has some tangible advantages. It enables the primary researcher, who might still wish to revisit these data for longitudinal analysis, to retain some control over who uses the data and for what purpose. It also brings primary and secondary researchers into contact with each other, and can foster collaborative modes of working. Feedback from Timescapes researchers suggests that having these diverse options for sharing their data was crucial in their willingness to deposit data to the Archive. In this way, we sought to balance the needs of primary and secondary users as well as providing safeguards for participants.

The Ethical Re-Use of QL Data

Qualitative data re-use raises a number of ethical issues that have been the subject of much discussion and debate (Bornat *et al*, 2012). Working with QL data creates additional complexities because of the potential overlap between primary and secondary research. As indicated above controls can be placed on secondary users through the archive registration process. The protection of participants is a central concern – and this includes the need to maintain the integrity of participants' accounts and ensure that their views are not misrepresented or exploited. The duty of care also extends to the primary researchers, and not simply through acknowledgements in publications. Over time, primary QL researchers can become highly committed to the datasets and samples that they research. While this does not give them prior ownership of the data or a monopoly on interpretation, the relationships that they develop with participants through ongoing engagement with the research, and the contextual knowledge that they build up is extremely valuable; it provides a solid base upon which to build re-use. Secondary research may well be aided where secondary users can work with this dynamic and build relationships of trust with the primary team.

The literature in this area tends to focus on the ethic of care that secondary users owe to primary researchers - whose work is assumed to have been completed. Re-users are enjoined to avoid criticising what they may discern as outdated findings or outmoded ways of researching, and to be aware of the different cultural and intellectual environments in which classic datasets were produced (Johnson *et al*, 2010). The rationale that is increasingly adopted is to be 'better with/because of' – rather than better than - those who came before us' (Bornat *et al*, 2012). The ethical responsibilities of secondary users, however, should not detract from their own equally valid ethical claims. Secondary analysts have their own integrity, which is founded on intellectual rigour. They need well produced and fully contextualised datasets to work with; they need to be free to ask their own research questions of the data, engage in distinctive modes of analysis, and produce their own interpretations, even where these run counter to

primary understandings. The work of secondary analysts is best seen as an enhancement of primary research rather than a challenge or threat to it, and the endeavours of data re-use deserve respect and support from the primary researcher. In other words, for QL research, where both primary and secondary use may be ongoing, the ethic of care needs to run in both directions.

CONCLUSION

The ethos of qualitative data sharing is perhaps more advanced in the UK than anywhere else, yet challenges remain. The broader cultural shift requires further effort, encouragement, time and diplomacy - with ongoing commitments to training and capacity building, work with professional associations, and direct engagement with ethical debates over the re-use of data.

The domain of applied ethics is about making difficult choices in situations where no unambiguous options exist. Archives expose materials created for one purpose to the possibility of other uses; in these circumstances some risk is unavoidable (Danielson, 2010). The perspective from medical research may be useful here. The benefits of medical research are potentially great but they carry higher risks, sometimes to the point where matters of life and death are at stake. But avoiding harm to participants is not the only ethical consideration; if it were, human trials would not be possible. For data archives, where the risks are real but the stakes lower, the same principles apply. The challenge in this context is to minimise risk, while still enabling access.

The broader 'stakeholder' view of ethics that Timescapes developed for QL research may offer a useful way forward, for it acknowledges and seeks to balance the varied and ever shifting standpoints and priorities of primary and secondary researchers, archivists, funders, and research participants.

- Bishop, L. (2009). 'Ethical sharing and re-use of qualitative data'. *Australian Journal of Social Issues*. 44(3). [Online]. Available at: <http://www.data-archive.ac.uk/media/249157/ajsi44bishop.pdf>. [Accessed 24th February 2011]
- Bishop, L. and Neale, B. (2010). 'Sharing Qualitative and Qualitative Longitudinal Data in the UK: Archiving Strategies and Development'. *IASSIST Quarterly* (IQ). 34(3-4) and 35(1-2). pp 23-29
- Bornat, J., Raghuram, P. And Henry, L. (2012 in press) Revisiting the Archives – Opportunities and Challenges: a case study from the history of Geriatric Medicine. *Sociological Research Online*
- Cutcliffe, J and Ramcharan, P. (2002). 'Leveling the Playing Field? Exploring the Merits of the Ethics-as-Process Approach for Judging Qualitative Research Proposals'. *Qualitative Health Research*. 12. pp 1000-1010
- Danielson, E. S. (2010). *The Ethical Archivist*. Chicago: Society of American Archivists.
- ESRC. (2010). *Framework for Research Ethics* (FRE). [Online]. Available at: <http://www.esrc.ac.uk/about-esrc/information/research-ethics.aspx> [Accessed 28th March 2012].
- Hadfield, L. (2010). 'Balancing on the edge of the archive: the researcher's role in collecting and preparing data for deposit'. In: Shirani, F and Weller, S. (eds). *Conducting Qualitative Longitudinal Research: Fieldwork Experiences*. Timescapes Working Papers Series 2, 60–73. [Online]. Available at: http://www.timescapes.leeds.ac.uk/assets/files/WP2_final_Jan%202010.pdf
- Henwood, K., Neale, B. and Holland, J. (eds.). (2012). 'Advancing Methods and Resources for Qualitative Longitudinal Research: The Timescapes Initiative'. *Qualitative Research*. Special Issue Vol 12 .(1)
- Johnson J., Rolph, S. and Smith, R. (2010). *Residential Care Transformed: Revisiting 'The Last Refuge'*. Basingstoke: Palgrave.
- Mauthner, N. S. and Edwards, R. (2010). 'Feminist Research Management in Higher Education in Britain: Possibilities and Practices'. *Gender, Work & Organization*. 17. pp 481–502
- Neale, B. and Bishop, L. (2012). 'The Timescapes Archive: a stakeholder approach to archiving qualitative longitudinal data'. *Qualitative Research*. 12(1). pp 53-65.

RESEARCH TEAM CONTACT DETAILS

Bren Neale: b.neale@leeds.ac.uk
Libby Bishop: ebishop@essex.ac.uk