National Centre for Research Methods Working Paper

2/19

Analysing large volumes of complex qualitative data

Reflections from a group of international experts

Susie Weller, Rosalind Edwards, Lynn Jamieson and Emma Davidson

NCRM
National Centre for Research Methods

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Contents

Tracing changes in notions and practices of child feeding: a trajectory approach to qualitative longitudinal research
Irmak Karademir Hazır, Oxford Brookes University, UK.

Analysing young people's experiences of coping with problems, difficult situations and feelings: An evolving approach to analysing qualitative longitudinal evaluation research
Emily Stapley, Anna Freud: National Centre for Children and Families, UK.

5. The potential of computer assistance in handling Big Qual data

Can a computer do qualitative analysis?
Daniel Turner, Quirkos software, UK.

Computer-assisted text analysis beyond words
Gregor Wiedemann, University of Hamburg, Germany.

Navigating the landscape of qualitative data in surveys with automated semantic analysis
Elena Zaitseva, Liverpool John Moores University, UK.

Computational text analysis using R in Big Qual data: lessons from a feasibility study looking at care and intimacy
Emma Davidson, Justin Chun-ting Ho and Lynn Jamieson, University of Edinburgh, UK.

6. Collaboration in Big Qual analysis

Collaborating with original research teams: Some reflections on good secondary analytic practice
Susie Weller, University of Southampton, UK.

Facebook timelines: Young people's growing up narratives online
Sian Lincoln and Brady Robards, Independent researcher and Monash University, Australia.

The challenges of computer assisted data analysis for distributed research teams working on large qualitative projects
Rebecca Taylor, University of Southampton, UK.

Time, technology and documentation
Rachel Thomson, Sara Bragg and Liam Berriman, University of Sussex, UK.

Working in collaboration to develop the teaching of Big Qual analysis
Sarah Lewthwaite, University of Southampton, UK.

7. Ethical considerations

The ethics of secondary data analysis
Ginny Morrow, University College London, UK.

Following families
Jane Millar and Tess Ridge, University of Bath, UK.

Visual approaches in qualitative longitudinal research
Fiona Shirani, Cardiff University, UK.

Contributors

Åsa Audulv is a lecturer in the Department of Nursing Science, Mid Sweden University, Sweden and School of Occupational Therapy, Dalhousie University, Canada. She has conducted qualitative longitudinal research (QLR) into self-management among people with long-term health conditions.

Sue Bellass is a Programme Manager and Research Fellow in the Mental Health and Addiction Research Group, Department of Health Sciences at the University of York with research interests in severe mental illness and comorbid physical health conditions. Her doctoral thesis explored how intergenerational families are affected by young onset dementia over time.

Liam Berriman is a Lecturer in Childhood and Youth Studies, based in the Department of Social Work & Social Care, University of Sussex, UK. He has conducted research on children's relationships with digital media and technology as part of their everyday lives, the technologisation of toys in childhood, and the co-production of research using digital methods.

Libby Bishop is the Coordinator for International Data Infrastructures in the Data Archive at GESIS-Leibniz Institute for Social Sciences. She enables and enhances connections between GESIS and international data infrastructures, including the Consortium of European Social Science Data Archives (CESSDA) and the European Open Science Cloud (EOSC). Her research interests include data ethics, privacy, data sharing, and commodification of data and she has published on the methodological and ethical issues of reusing and sharing data.

Sara Bragg is an Associate Researcher at the Centre for Innovation and Research in Childhood and Youth, University of Sussex, UK. She also teaches at UCL Institute of Education, London. She has research interests in young people, participation, gender, sexuality, media, education, and creative research methods.

Justin Chun-ting Ho is a PhD Candidate in Sociology at the University of Edinburgh, UK. He has expertise in quantitative text analysis and an interdisciplinary interest in nationalism studies and computational social science. He is currently researching the construction of Hong Kong nationalism on social media.

Emma Davidson is a Leverhulme Early Career Research Fellow and co-director of the interdisciplinary consortium, the Centre for Research on Families and Relationships. Her research interests include the role of 'third spaces' in community identity and civil society, and she is currently working on an ethnographic study of public libraries in austerity.

Rosalind Edwards is Professor of Sociology, and co-director of the ESRC National Centre of Research Methods. A Fellow of the Academy of Social Sciences, Rosalind has researched and published widely on qualitative longitudinal research methods. She is a founding and co-editor of the *International Journal of Social Research Methodology*.

Nick Emmel is Professor of Social Research Methodology in the School of Sociology and Social Policy at the University of Leeds, UK. He continues to refine a model of vulnerability through on-going qualitative longitudinal research and revisiting yesterday's data. He has recently co-edited Doing Realist Research (2018) published by SAGE. Nick is currently writing about transdisciplinary approaches to research to address complex puzzles.

Joanna Fadyl is a Senior Lecturer and Deputy Director of the Centre for Person Centred Research at Auckland University of Technology in New Zealand. Her expertise is in rehabilitation and disability. The

study she discusses involved a large research team led by Professor Kathryn McPherson and Associate Professor Alice Theadom.

Jane Gray is Professor of Sociology at Maynooth University, Ireland. She has expertise in families, households and social change, as well as qualitative data management and sharing. She has completed studies such as 'Family Rhythms', has contributed to the development of the Digital Repository of Ireland, and is programme leader for the Irish Qualitative Data Archive.

Lynn Jamieson is Professor of Families and Relationships at the University of Edinburgh. She is founding director of the interdisciplinary consortium, the Centre for Research on Families and Relationships. Lynn is co-editor of the Palgrave Macmillan series, Families and Intimate Life and Associate Editor of the Policy Press journal, Families Relationships and Societies. Her publications include 'Intimacy: Personal Relationships in Modern Societies'.

Irmak Karademir Hazır is a Senior Lecturer in Sociology in the Department of Social Sciences, Oxford Brookes University, UK. Her research focuses on class cultures, embodiment, cultural tastes and hierarchies. She is also interested in using and teaching quantitative and qualitative research methods, and has used mixed methods designs in her work.

Sarah Lewthwaite is a Research Fellow in the ESRC National Centre for Research Methods and Southampton Education School, University of Southampton, UK. She has expertise in the learning and teaching of advanced research methods, as well as, the intersections between critical theory, accessibility, new technologies and student experience in higher education.

Sian Lincoln has expertise in contemporary youth, youth culture, social network sites and identity, and ethnography. Her work has been published widely and has appeared in a number of journals including *New Media & Society*, *Journal of Youth Studies and Social Media + Society* as well as in a number of anthologies.

Jane Millar is a Professor of Social Policy at the University of Bath, a Fellow of the British Academy and of the Academy of Social Sciences. Jane is well known for her work on lone parents and welfare to work, social security and tax credits, poverty and family policy.

Ginny Morrow Ginny Morrow is Visiting Professor at University College London, Institute of Education, in the Department of Social Science, Social Science Research Unit, and Thomas Coram Research Unit, and Research Associate, Young Lives, Department of International Development, University of Oxford. Her research focuses on children's work, violence affecting children, sociological approaches to the study of childhood and children's rights, the ethics of social research with children, children's understandings of family, and children and 'social capital'.

Bren Neale is Emeritus Professor of Life course and Family Research at the University of Leeds, UK and a fellow of the Academy of Social Sciences. She was director of the ESRC Timescapes initiative from 2007 to 2012. Bren is a leading expert in Qualitative Longitudinal (QL) research methodology and continues to provide training and support in this methodology for new and established researchers throughout the UK and internationally.

Ruth Patrick is a Lecturer in Social Policy and Social Work at the University of York, UK. Her research illustrates the ways in which qualitative longitudinal research can help us to understand popular and political narratives around poverty, welfare reform and austerity and lived experiences for those directly affected by recent and ongoing changes to the social security system.

Georgia Philip is a Research Fellow in the School of Social Work, at the University of East Anglia. She has research expertise in the areas of: fathers, father engagement, gender and parenting, qualitative (and longitudinal) research and the feminist ethics of care.

Tess Ridge is an Honorary Professor in the Department of Social and Policy Studies at the University of Bath, UK, and Fellow of the Academy of Social Sciences. Tess is well known for her children-focused work on childhood poverty, social exclusion, welfare and economic support.

Brady Robards is a Senior Lecturer in Sociology in the School of Social Sciences at Monash University. He studies digital culture, with a focus on how people use and produce social media and how social media platforms operate as memory archives made up of digital traces of life over years of use. Brady has studied social media use among particular groups, such as LGBTIQ+ people, young people, tourists, and in the context of alcohol consumption.

Fiona Shirani is a Research Associate at Cardiff University. She leads the fieldwork on a qualitative longitudinal work package as part of the interdisciplinary FLEXIS programme, which seeks to investigate how flexible energy systems can meet modern-day energy challenges.

Emily Stapley is a Qualitative Research Fellow in the Evidence Based Practice Unit (EBPU) at the Anna Freud National Centre for Children and Families and University College London. EBPU is a child and youth mental health research and innovation unit.

Anna Tarrant is a Leverhulme Early Career Fellow and Senior Lecturer in Sociology in the School of Social and Political Science at the University of Lincoln, UK. Her research interests include men's care responsibilities and support needs in low-income families, over time and across households.

Rebecca Taylor is a Lecturer in Sociology at the University of Southampton, UK. Her research focuses on conceptualising work, particularly unpaid forms of work, understanding individuals' working lives and careers, and work in different organisations and sectors.

Rachel Thomson is Professor of Childhood and Youth Studies at the University of Sussex, UK. She is responsible for three QL data sets: the 15 year project Inventing Adulthoods, the intergenerational study Making Modern Mothers and the born-digital Everyday Childhoods project. She has co-edited two special issues on QLR for the International Journal of Social Research Methodology in 2003 (6:3) and 2015 (18:3) and is working on a mixed methods longitudinal study of 'open ended' youth transitions in Denmark.

Daniel Turner is an experienced qualitative researcher based in the UK and founder of Quirkos. After completing his doctorate at Sheffield University, Daniel continued using qualitative research in health, long-term conditions and policy. He left academia to develop Quirkos, a simple and visual qualitative analysis tool initially designed for participatory analysis and now used in more than 250 universities across the world. He continues to deliver training on Quirkos, and writes a popular blog on qualitative methodologies.

Susie Weller is a Senior Research Fellow in the ESRC National Centre for Research Methods and the MRC Lifecourse Epidemiology Unit, University of Southampton, UK. She has expertise in qualitative longitudinal and creative research with children and families, and in the context of lived experiences of health care interventions. She is co-founder of the Big Qual Analysis Resource Hub (http://bigqlr.ncrm.ac.uk/).

Gregor Wiedemann works as a researcher in the Language Technology Group in the computer science department at Hamburg University, Germany. He has expertise in developing methods and workflows of text mining for applications in social sciences.

Sarah Wilson is a Senior Lecturer in Sociology in the Faculty of Social Sciences at the University of Stirling, UK. Sarah's research interests are in the sociology of families, relationships and personal life, with a

methodological focus on developing visual, audial and artistic qualitative research. Her recent work explores the representation of sociological work and overlaps with the visual arts.

Elena Zaitseva is Academic Research and Development Officer at the Teaching and Learning Academy, Liverpool John Moores University, UK. Her research interests are in the higher education student experience, learner identity and learner journeys.

Introduction: Reflections on analysing large volumes of qualitative data

Susie Weller and Rosalind Edwards, University of Southampton, UK.
Lynn Jamieson and Emma Davidson, University of Edinburgh, UK.

This working paper brings together the reflections of a wide range of international researchers to explore, showcase and reflect critically on the potentials and challenges of analysing large volumes of complex qualitative, and qualitative longitudinal (QLR) data, including archived material. Big Qual analysis is a new area for qualitative work and there is little guidance on how best to work with masses of qualitative material. The working paper comprises a set of blogs housed in the 'Big Qual Analysis Resource Hub' (http://bigqlr.ncrm.ac.uk/). We created this website to map the progress of our ESRC National Centre for Research Methods research project '*Working across qualitative longitudinal studies: a feasibility study looking at care and intimacy*' (2015-2019). As part of the project we developed procedures for working with multiple sets of in-depth temporal qualitative data (see Davidson et al. 2019; Edwards et al. 2019 for discussion of our methodological findings).

We have gathered together and made available the 27 blog post reflections from 32 authors in this working paper form because accounts of data management and analysis in qualitative research are often sanitised by the time they reach academic journals. Here, our contributors document and share publicly the trials and tribulations, intellectual commitments, contingencies and decision-making processes underlying such analysis, contributing to debates around good practice. We hope that this collection of reflections will promote further conversations about analysis/secondary analysis across large scale and/or multiple qualitative data sets. With guest posts from international scholars, from early career through to established researchers, on topics as varied as the ethics of using Big Qual data, using secondary qualitative material and computer-assisted qualitative data analysis software, this collection of reflections profiles the diversity of work taking place internationally.

The blog posts reflect on the experiences and perspectives of those conducting analysis across large-scale, multiple and/or QLR data sets, particularly the re-use of archived material. The possibilities comprise multiple permutations from bringing together two or more archived data sets through to combining archived material with a researcher's own primary data. Some of the blogs outlined below focus on secondary analysis and the re-use of archived data sets, including QLR material, whilst others are concerned with handling large volumes of qualitative data. In both cases the approach undertaken involves, to some degree, engaging with an amount of data that would be too challenging for a single researcher or small team to handle with qualitative integrity. The working paper is organised thematically and comprises contributions from guest bloggers accrued throughout the project.

The first section – Asking new questions of existing data - comprises three thought-provoking contributions written by leading experts in the field. Each blog focuses on the task of re-using pre-existing data. Bren Neale's post sets the scene for what follows. Drawing on her extensive expertise as Director of the Timescapes Initiative, she considers the diverse forms of archival data that may be re-used or re-purposed in qualitative (longitudinal) work. In so doing, Bren outlines the possibilities for, and progress made in developing ways of working with/across assemblages of archived material to capture social and temporal processes. This is followed by Nick Emmel's reflections on the new insights gleaned from revisiting his own existing data and, in the case of his QLR work, how new empirical accounts can re-shape how existing data is viewed. Libby Bishop then outlines some of the collections available for re-use at the UK Data Service, and offers an explanation about how a data archive is the perfect starting point for those new to QLR.

Comprising four blogs, section two is concerned with 'data sets in conversation'. In our own work, we have focused on developing ways of bringing together and analysing across multiple archived qualitative data sets. We have contributed the first blog in this section sharing our experiences of identifying and selecting from and/or within multiple data sets to create a new data assemblage to analyse. Anna Tarrant

then reflects on her work exploring men, poverty and lifetimes of care in which she brings two qualitative longitudinal data sets into conversation with her own empirical data. The third blog from Sarah Wilson focuses on using qualitative secondary analysis as a tool of critical reflexivity. In so doing, she draws on her work on vulnerability and the key insights gleaned from bringing data from a specific group of young people into conversation with archived data from a broader sample. In each case, the authors highlight how bringing different data sets into conversation has allowed them to test emerging theories and view their own data in new ways.

The emphasis of section four is on slicing data in different ways, with a particular focus on the decision-making processes surrounding the organisation of data particularly in complex QLR studies. In her blog Rachel Thomson highlights the parallels between debates about scale in Big Qual and QLR. She encourages us to re-think what we consider to constitute a case and the fluidity of cases and practices of casing. Similarly, Sue Bellas shares in detail her approach to analysing data from her intergenerational longitudinal study of early onset dementia, from multiple perspectives. Drawing on Rachel's work she describes the creation of family cases and her cross-case analysis across the four generations represented in the sample families. Jane Gray's blog focuses on using different temporal gazes to reconcile different temporalities when bringing together a data set comprising retrospective life story narratives with a set of qualitative longitudinal interviews from a prospective panel study.

Section five includes six blogs each focusing on different approaches to analysing large volumes of qualitative (longitudinal) data. The first post by Åsa Audulv draws on her preliminary findings to highlight the lack of transparency around approaches to QLR analysis in health research publications. She encourages researchers to be both transparent and proud! Georgia Philip's blog focuses on the practicalities, process and challenges of managing the volume and depth of data generated in a QLR analysis of men's experiences of the UK child protection system. In her post, Ruth Patrick draws on the 'Lived experiences of welfare reform study' to demonstrate the value of conducting both diachronic and synchronic analyses and in working iteratively across and within cases in QLR work. In so doing, she highlights some implications for re-users of such material. Joanna Fadyl then details her experiences of conducting the 'Traumatic Brain Injury (TBI) experiences study'; a QLR study about recovery and adaptation. She outlines her teams' approach to capturing change over time using a thematic approach with evolving coding and a data visualisation tool. In Irmak Karademir-Hazir's blog she outlines the trajectory approach she is currently using in her ethnographic and longitudinal research looking at the practices of foodwork (eating, cooking and feeding) in families with small children across different social classes. Emily Stapley's contribution completes this section. Emily discusses her evolving approach to analysing qualitative longitudinal evaluation data drawing on the analysis of interview data from a five-year QLR evaluation of HeadStart; a programme designed to improve the mental health and wellbeing of young people.

Whilst some of the papers in section five outline approaches using qualitative analysis software, section six focuses specifically on text mining tools designed to handle large volumes of qualitative material. Drawing on his collaborative work (with Kathrin Cresswell, Claire Grover, Claire Lewellyn, Aziz Sheikh and the late Jon Oberlander, University of Edinburgh) Daniel Turner asks a fundamentally important question 'can a computer do qualitative analysis?' In so doing, he considers the advantages and challenges of using Machine Learning to assist with coding and help researchers handle large volumes of data in a time and cost effective manner. Following on from this, Gregor Wiedemann's blog focuses on computer-assisted text analysis beyond words and concerns computational textual analysis and the opportunities it presents for qualitative research and researchers. Elena Zaitseva's focus is on navigating the landscape of qualitative data in surveys, material that often goes unanalysed, with automated semantic analysis. Elena outlines examples of her work using the text mining software, Leximancer. Emma Davidson, Justin Chun-ting Ho and Lynn Jamieson's blog completes this section. They consider the potentials and pitfalls of using R, a tool for computational text analysis, to get an overview of a large volume of qualitative data and to identify areas of salience to explore further.

Part of our aim has been to contribute to emergent good practice in a new field, concerning the relationship between secondary analysts and the original teams who created the data sets. Collaboration has been key to this. The penultimate section of this working paper explores collaboration in Big Qual analysis. We open this section with a blog on collaborating with original research teams. Susie Weller considers some of the possibilities and challenges of developing collaborative relationships between secondary analysts and members of the original teams who created the data sets. In so doing, she shows how attachments to data and notions of ownership – for both original researchers and re-users of the data – shift over time. Sian Lincoln and Brady Robards draw on their study 'Facebook Timelines' which explores the role of social media in mediating and archiving 'growing up' narratives. Using Facebook they provide the fascinating example of analyzing longitudinal digital traces by working with participants as co-analysts encouraging them to 'scroll back' and interpret their own personal archives. This is followed by Rebecca Taylor's contribution on the challenges of collaborating with nationally distributed colleagues. In her blog, Rebecca considers three possible ways of overcoming the challenges of conducting large-scale QLR analysis in geographically-distributed research teams and the possibilities, and indeed limitations, offered by computer assisted data analysis software. This is followed by Rachel Thomson, Sara Bragg and Liam Berriman's contribution. They encourage us to reconsider the idea of archiving data as the end point of a study. Working with research participants, the team have co-produced the publicly accessible archive Everyday Childhoods; a process that has enabled them to explore what it means to become data. Sarah Lewthwaite completes the section on collaboration. She draws on a recent NCRM collaborative project - Big Qual Analysis: Innovation in Teaching and Method - and reflects on the steps she took to collaborate over pedagogic development of approaches to and resources for, the teaching and learning of Big Qual analysis.

The final section focuses on ethical considerations. The first contribution is from Ginny Morrow, who draws upon her wealth of experience, to reflect on the ethical responsibilities of researchers sharing secondary data. Jane Millar and Tess Ridge then draw on some of the insights gleaned from their QLR study. They focus on two factors that affected the sharing and re-use of data; the construction of their sample, and their family-based theoretical approach. The final contribution from Fiona Shirani considers visual approaches in QLR. Fiona reflects on some of the ethical issues involved in using/re-using visual material and the reasons why visuals (and their meaning) can feel more 'distant' in secondary analysis than the written word.

It has been a privilege to put together this collection of blogs. We believe the contributions highlight the changing data and analysis landscape. We hope that it will be a useful resource (please see http://bigqlr.ncrm.ac.uk/ for a range of multimedia resources).

References

Edwards, R., Weller, S., Jamieson, L. and Davidson, E., (2019) Search strategies: Analytic searching across multiple data sets and with combined sources, in K. Hughes and A. Tarrant (eds.) *Advances in Qualitative Secondary Analysis*, Sage, Ch.6.

Davidson, E., Edwards, R. Jamieson, L. and Weller, S. (2019) Big data, qualitative style: a breadth-and-depth method for working with large amounts of secondary qualitative data, *Quality & Quantity*, 53(1): 363-376. [Open access]

1. Asking new questions of existing data

Research data as documents of life
Bren Neale, University of Leeds, UK.

Among the varied sources of data that underpin Qualitative Longitudinal (QL) studies, documentary and archival sources have been relatively neglected. This is despite their potential to shed valuable light on temporal processes. These data sources form part of a larger corpus of materials that Plummer (2001) engagingly describes as 'documents of life':

> "The world is crammed full of human personal documents. People keep diaries, send letters, make quilts, take photos, dash off memos, compose auto/biographies, construct websites, scrawl graffiti, publish their memoirs, write letters, compose CVs, leave suicide notes, film video diaries, inscribe memorials on tombstones, shoot films, paint pictures, make tapes and try to record their personal dreams. All of these expressions of personal life are hurled out into the world by the millions, and can be of interest to anyone who cares to seek them out" (p. 17).

To take one example, letters have long provided a rich source of insight into unfolding lives. In their classic study of Polish migration, conducted in the first decades of the twentieth century, Thomas and Znaniecki (1958 [1918-20]) analysed the letters of Polish migrants to the US (an opportunistic source, for a rich collection of such letters was thrown out of a Chicago window and landed at Znaniecki's feet). Similarly, Stanley's (2013) study of the history of race and apartheid was based on an analysis of three collections of letters written by white South Africans spanning a 200 year period (1770s to 1970s). The documentary treasure trove outlined by Plummer also includes articles in popular books, magazines and newsprint; text messages, emails and interactive websites; the rich holdings of public record offices; and confidential and often revealing documents held in organisations and institutions. Social biographers and oral historians are adept at teasing out a variety of such evidence to piece together a composite picture of lives and times; they are 'jackdaws' rather than methodological purists (Thompson 1981: 290).

Among the many forms of documentary data that may be repurposed by researchers, social science and humanities datasets have significant value. The growth in the use of such legacy data over recent decades has been fuelled by the enthusiasm and commitment of researchers who wish to preserve their datasets for historical use. Further impetus has come from the development of data infrastructures and funding initiatives to support this process, and a fledgling corpus of literature that is documenting and refining methodologies for re-use (e.g. Corti, Witzel and Bishop 2005, Crow and Edwards 2012, Irwin 2013). Alongside the potential to draw on individual datasets, there is a growing interest in working across datasets, bringing together data that can build new insights across varied social or historical contexts (e.g. Irwin, Bornat and Winterton 2012, Davidson et al. 2018).

Many qualitative datasets remain in the stewardship of the original researchers where they are at risk of being lost to posterity (although they may be fortuitously rediscovered, O'Connor and Goodwin 2012). However, the culture of archiving and preserving legacy data through institutional, specialist or national repositories is fast becoming established (Bishop and Kuula-Luumi 2017). These facilities are scattered across the UK (for example, the Kirklees Sound Archive in West Yorkshire, which houses oral history interviews on the wool textile industry (Bornat 2013)). The principal collections in the UK are held at the UK Data Archive (which includes the classic 'Qualidata' collection); the British Library Sound Archive, NIQA (the Northern Ireland Qualitative Archive, including the ARK resource); the recently developed Timescapes Archive (an institutional repository at the University of Leeds, which specialises in Qualitative Longitudinal datasets); and the Mass Observation Archive, a resource which, for many decades, has commissioned and curated contemporary accounts from a panel of volunteer recorders. International resources include the Irish Qualitative Data Archive, the Murray Research Center Archive (Harvard), and a

range of data facilities at varying levels of development across mainland Europe (Neale and Bishop 2010-11).

In recent years some vigorous debates have ensued about the ethical and epistemological foundations for reusing qualitative datasets. In the main, the issues have revolved around data ownership and researcher reputations; the ethics of confidentiality and consent for longer-term use; the nature of disciplinary boundaries; and the tension between realist understandings of data (as something that is simply 'out there'), and a narrowly constructivist view that data are non-transferable because they are jointly produced and their meaning tied to the context of their production.

These debates are becoming less polarised over time. In part this is due to a growing awareness that most of these issues are not unique to the secondary use of datasets (or documentary sources more generally) but impact also on their primary use, and indeed how they are generated in the first place. In particular, epistemological debates about the status and veracity of qualitative research data are beginning to shift ground (see, for example, Mauthner et al. 1998, Mauthner and Parry 2013). Research data are by no means simply 'out there' for they are inevitably constructed and re-constructed in different social, spatial and historical contexts; indeed, they are transformed historically simply through the passage of time (Moore 2007). But this does not mean that the narratives they contain are 'made up' or that they have no integrity or value across different contexts (Hammersley 2010; Bornat 2013). It does suggest, however, that data sources are capable of more than one interpretation, and that their meaning and salience emerge in the moment of their use:

> "There is no a-priori privileged moment in time in which we can gain a deeper, more profound, truer insight, than in any other moment. … There is never a single authorised reading … It is the multiple viewpoints, taken together, which are the most illuminating" (Brockmeier and Reissman cited in Andrews 2008: 89, Andrews 2008: 90).

Moreover, whether revisiting data involves stepping into the shoes of an earlier self, or of someone else entirely, this seems to have little bearing on the interpretive process. From this point of view, the distinctions between using and re-using data, or between primary and secondary analysis begin to break down (Moore 2007, Neale 2013).

This is nowhere more apparent than in Qualitative Longitudinal enquiry, where the transformative potential of data is part and parcel of the enterprise. Since data are used and re-used over the longitudinal frame of a study, their re-generation is a continual process. The production of new data as a study progresses inevitably reconfigures and re-contextualises the dataset as a whole, creating new assemblages of data and opening up new insights from a different temporal standpoint. Indeed, since longitudinal datasets may well outlive their original research questions, it is inevitable that researchers will need to ask new questions of old data (Elder and Taylor 2009).

The status and veracity of research data, then, is not a black and white, either/or issue, but one of recognising the limitations and partial vision of all data sources, and the need to appraise the degree of 'fit' and contextual understanding that can be achieved and maintained (Hammersley 2010; Duncan 2012; Irwin 2013). This, in turn, has implications for how a dataset is crafted and contextualised for future use (Neale 2013).

A decade ago, debates about the use of qualitative datasets were in danger of becoming polarised (Moore 2007). However, the pre-occupations of researchers are beginning to move on. The concern with whether or not qualitative datasets should be used is giving way to a more productive concern with how they should be used, not least, how best to work with their inherent temporality. Overall, the 'jackdaw' approach to re-purposing documentary and archival sources of data is the very stuff of historical sociology and of social history more generally (Kynaston 2005, Bornat 2008, McLeod and Thomson 2009), and it has huge and perhaps untapped potential in Qualitative Longitudinal research.

References

Andrews, M. (2008) Never the last word: Revisiting data, in M. Andrews, C. Squire and M. Tamboukou (eds.) *Doing narrative research*, London, Sage, 86-101

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

Bornat, J. (2008) Crossing boundaries with secondary analysis: Implications for archived oral history data, Paper given at the *ESRC National Council for Research Methods Network for Methodological Innovation: Theory, Methods and Ethics across Disciplines*, 19 September 2008, University of Essex,

Bornat, J. (2013) Secondary analysis in reflection: Some experiences of re-use from an oral history perspective, *Families, Relationships and Societies*, 2(2): 309-17

Corti, L., Witzel, A. and Bishop, L. (2005) (eds.) Secondary analysis of qualitative data: Special issue, *Forum: Qualitative Social Research*, 6(1).

Crow, G. and Edwards, R. (2012) (eds.) Editorial Introduction: Perspectives on working with archived textual and visual material in social research, *International Journal of Social Research Methodology*, 15(4): 259-262.

Duncan, S. (2012) Using elderly data theoretically: Personal life in 1949/50 and individualisation theory, *International Journal of Social Research Methodology*, 15(4): 311-319.

Elder, G. and Taylor, M. (2009) Linking research questions to data archives, in J. Giele and G. Elder (eds.) *The Craft of Life course research*, New York, Guilford Press, 93-116.

Hammersley, M. (2010) Can we re-use qualitative data via secondary analysis? Notes on some terminological and substantive issues, *Sociological Research Online*, 15, 1,5.

Irwin, S. (2013) Qualitative secondary analysis in practice' Introduction in S. Irwin and J. Bornat (eds.) Qualitative secondary analysis (Open Space), *Families, Relationships and Societies*, 2(2): 285-8.

Irwin, S., Bornat, J. and Winterton, M. (2012) Timescapes secondary analysis: Comparison, context and working across datasets, *Qualitative Research*, 12(1): 66-80.

Kynaston, D. (2005) The uses of sociology for real-time history, *Forum: Qualitative Social Research*, 6,1.

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

Mauthner, N., Parry, O. and Backett-Milburn, K. (1998) The data are out there, or are they? Implications for archiving and revisiting qualitative data, *Sociology*, 32(4): 733-745.

Mauthner, N. and Parry, O. (2013). Open Access Digital Data Sharing: Principles, policies and practices, *Social Epistemology*, 27(1): 47-67.

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., and Bishop, L. (2010 -11) Qualitative and qualitative longitudinal resources in Europe: Mapping the field, *IASSIST Quarterly: Special double issue*, 34 (3-4); 35(1-2)

O'Connor, H. and Goodwin, J. (2012) Revisiting Norbert Elias's sociology of community: Learning from the Leicester re-studies, *The Sociological Review*, 60: 476-497.

Plummer, K. (2001) *Documents of Life 2: An invitation to a critical humanism*, London, Sage.

Stanley, L. (2013) Whites writing: Letters and documents of life in a QLR project, in L. Stanley (ed.) *Documents of Life Revisited*, London, Routledge, 59-76.

Thomas, W. I. and Znaniecki, F. (1958) [1918-20] *The Polish Peasant in Europe and America Volumes I and II*, New York, Dover Publications.

Thompson, P. (1981) Life histories and the analysis of social change, in. D. Bertaux (ed.) *Biography and Society: The life history approach in the social sciences*, London, Sage, 289-306.

Revisiting yesterday's data today

Nick Emmel, University of Leeds, UK.

I have recently finished writing a paper about vulnerability. This is the third in an ongoing series of published papers; the first published in 2010 and the second in 2014 (Emmel and Hughes 2010, 2014, Emmel 2017). Each elaborates and extends a model of vulnerability. All three are based on the same data collected in a qualitative longitudinal research project, Intergenerational Exchange, a part of Timescapes and its archive. The second and third paper draw on newly collected data from subsequent research projects as well. In this blog I want to explore how interpretation and explanation are reconstituted and reconceived through engagement with these new data and theory, considering some methodological lessons in the context of qualitative longitudinal research.

At first sight the narratives told us about poverty, social exclusion, and the experiences of grand parenting by Bob and Diane, Ruth, Sheila, Geoff and Margaret, and Lynn, which populate these three papers seem fixed, even immutable. After all, I am still using the same printed transcripts from interviews conducted between 2007 and 2011, marked up with a marginalia of memos and codes in my micrographia handwriting, text emphasised with single and double underlines in black ink. But each time I get these transcripts out of the locked filing cabinet in my office I learn something new.

To start with there are the misremembered memories of what is actually in the transcripts. Many of the stories our participants tell, Geoff and Margaret's account of the midnight drop, Sheila bathing her kids in the washing machine, or Lynn walking into the family court for the first time, I have retold over and over again. In their retelling details have been elaborated, twisted, and reworked to make better stories so my students, service deliverers, and policy makers will think a little harder, I hope, about powerlessness, constrained powerfulness, and ways in which excluded people depend on undependable service delivery. In this way they are no different to the original stories, neither truth nor untruth, but narrated for a purpose, to describe experience in qualitative research. Getting the detail and emphasis right is important. The participants know their lived experience far better than I do. Re-reading the transcripts, these stories are reattached to their empirical moorings once again. But this is only the start of their reanalysis.

Rereading may confirm empirical description but past interpretations are unsettled by new empirical accounts. New knowledge has the effect, as Barbara Adam (1990:143) observes, of making the 'past as revocable and hypothetical as the future'. In the most recent of the three papers the apparently foundational role of poverty elaborated in our first paper is reinterpreted. New data from relatively affluent grandparents describe the barriers they face in accessing services and the ways in which these experiences make them vulnerable. This knowledge has the effect of reconstituting the original transcripts, shifting attention away from the determining role of poverty to relationships with service providers in which poverty may play a generative part. These data evoke new interpretations. But it is not only new empirical accounts that reshape this longitudinal engagement, new ideas are at play.

In this blog I have suggested that new empirical accounts change how we understand and interpret existing data. To ascribe reinterpretation only to these insights is not enough however. Explanations rely on more than reconstructing empirical accounts in the light of new insight. For a realist like me theories guide the reading of the original transcripts and the collection of new data. Theories are practical things, bundles of hypotheses to be judged and refined empirically. We started with a theory about time as a chronological progression of events, as is explained in the first paper. For our participants, they noticed little difference as recession merged with recession all the way back to the closure of the estate's main employer in 1984. This theory was found wanting when we came to looking at young grandparenthood and engagement with service provision in the second paper. A refined theoretical account of the social conscience of generational and institutional time supported explanation. These theories, like the empirical accounts of the social world they are brought into a relation with, are revocable and only ever relatively enduring.

To paraphrase the Greek philosopher Heraclitus, no researcher ever steps into the same river twice, for it is not the same river and it is not the same researcher. Revisiting yesterday's data today reminds us of these methodological lessons in qualitative longitudinal research.

References

Adam, B (1990) *Time and Social Theory,* Polity Press: Cambridge.

Emmel, N. (2017) Empowerment in the relational longitudinal space of vulnerability. *Social Policy and Society.* July.

Emmel, N. and Hughes, K. (2010) 'Recession, it's all the same to us son': the longitudinal experience (1999-2010) of deprivation, *21st Century Society*, 5(2): 171-182.

Emmel, N. and Hughes, K. (2014) Vulnerability, inter-generational exchange, and the conscience of generations, in J. Holland and R. Edwards (eds) *Understanding Families over Time: Research and Policy*, Palgrave, Basingstoke.

Data from the past and for the future: Qualitative longitudinal data available at the UK Data Service

Libby Bishop, GESIS-Leibniz Institute for Social Sciences, Germany.

You may already be a member of the tribe of qualitative longitudinal (QL) researchers if you are reading this. But what if you are just starting out? You might be curious about how others have done QL projects. Of course, there are published articles to look at, and there are many to choose from now. But wouldn't it be helpful to actually look at the data other researchers have used? To read in some detail what strategies were used to maintain contact between interviews? To read transcripts to discover, for example, exactly how the interviewer gently guided the respondent back to topics from the previous contact, without losing the thread of more recent events? All this, and more, is possible by looking at qualitative longitudinal data collections available for research at the UK Data Service (UKDS).

Below I provide a brief introduction to just a few of these collections, of which the UKDS has dozens. These are available to be downloaded and used by researchers (after having registered with the Service). Two of these studies are about older age, and another is on a timely issue: elections.

SN 851919 Maintaining dignity in later life: A longitudinal qualitative study of older people's experiences of supportive care

The aim of this study was to examine preparations for the end of life made by older people with supportive care needs and the factors that support or undermine a sense of dignity. Thirty-four participants in Bristol and Nottingham were recruited via GPs and day centres. All had health problems that required support and care to varying degrees, including family care and support, medical treatment, community nursing, home care services and moves to care homes. They were interviewed face-to-face on four occasions (on average) between June 2008 and January 2011 and contacted by telephone between interviews. Face-to-face interviews were recorded and transcribed verbatim.

SN 5237 Adding quality to quantity: Quality of life in older age, 2000-2002

The broad aim of the study was to define the constituents of quality of life in older age. The research questions were twofold: how do older people define and prioritise quality of life, and how do they feel it can be improved? This study represented a unique multidisciplinary and mixed methods collaboration between investigators with backgrounds in sociology, psychology, social gerontology, transport planning and clinical epidemiology. Following the fielding of the questionnaire, 80 respondents were selected for an in-depth interview to probe factors further affecting quality of life.

SN 6861 Qualitative election study of Britain, 2010

This research project recorded the views and concerns of Britons before and after the 2010 General Election. By conducting 14 focus groups with people in England, Scotland and Wales the project investigated, qualitatively, pre- and post-election views. The aim was to generate data that: 1) provided insights into the views and perceptions of citizens on politicians, party leaders, and political issues (e.g. civic duty, political alienation, political activism) before and after the general election; 2) allowed for analysis of the meaning that underlies their assessments, uncover sources of normative values, and make explicit the tacit assumptions participants use to reach their judgements. Three additional focus groups were conducted on the night of the first ever Leaders' Debates and the transcripts record people's expectations in advance of the debates and their reactions afterwards. As well as the focus group transcripts, the collection includes a quantitative file of results from the pre-focus group questionnaire given to participants. And watch this space – comparable data from the 2015 UK elections will be arriving shortly.

What could be better than QL data? Getting funded to do research with QL data! The ESRC has a programme, the Secondary Data Analysis Initiative, which does just that (https://esrc.ukri.org/research/our-research/secondary-data-analysis-initiative/). It offers funding for up to 18 months and £200,000 for research that collects no new data, but uses data from selected existing resources. One of the designated

resources is Timescapes, a rich lode of QL data. Another, the Qualitative Archives on Ageism and Conflict, is held at the Northern Ireland Qualitative Archive.

As always, if you want any help getting starting or looking for data, just get in touch http://data-archive.ac.uk/contact

2. Data sets in conversation

Selecting data sets to create new assemblages

Susie Weller and Rosalind Edwards, University of Southampton, UK.
Lynn Jamieson and Emma Davidson, University of Edinburgh, UK.

This blog focuses on the process of identifying qualitative material from multiple archived data sets to bring together to conduct secondary analysis. This process is the first stage in a four-step breath-and-depth method we developed for analysing large volumes of qualitative data (Davidson et al. 2019). We draw on our experiences of conducting the ESRC National Centre for Research Methods project, of which the Big Qual Analysis Resource Hub (http://bigqlr.ncrm.ac.uk/) is an outcome. Utilising different qualitative longitudinal research (QLR) data sets housed in the Timescapes Archive, our project aimed to explore the possibilities for developing new procedures for working across multiple sets of archived qualitative data.

The availability of volumes of complex qualitative data for secondary analysis is growing. Indeed, major research funding bodies in the UK regard the sharing of data as vital to accountability and transparency and, for some, it is a contractual requirement. Furthermore, the increasing influence of big data, which has until now, generally concerned large-scale quantitative data sets, highlights the potential for researchers to enhance further the value of existing qualitative investments. Yet, the full potential of archived qualitative has yet to be realised.

The development of central and local digital repositories opens up exciting possibilities for doing new research using existing data sets. With that comes the opportunity to bring together one or more data sets into a new assemblage in order ask new questions of the data, make comparisons, explore how processes work in different contexts, and provide new insights.

Major contemporary online qualitative archival sources established internationally for data preservation and sharing include (see also the Registry of Research Data Repositories http://www.re3data.org/):

The UK Data Archive	https://data-archive.ac.uk/
Finnish Social Science Data Archive	https://www.fsd.uta.fi/en/
Henry A. Murray Research Archive (Harvard)	https://murray.harvard.edu/
Qualitative Data Repository	https://qdr.syr.edu/
Irish Qualitative Data Archive	https://www.maynoothuniversity.ie/social-sciences-institute/research/iqda
Northern Ireland Qualitative Archive	http://www.ark.ac.uk/qual/ageism/ http://www.ark.ac.uk/qual/conflict/
Swiss Data and Research Information Services + Swiss Foundation for Research in Social Sciences	http://forscenter.ch/en/data-and-research-information-services/ http://www.css.ethz.ch/en/services/datasets.html
Timescapes Qualitative Longitudinal Data Archive	http://timescapes-archive.leeds.ac.uk/

Many of these data repositories have been designed with re-use in mind and material is accompanied by documentation about the original project such as: aims and objectives, the methodology, sample and methods, and units of analysis, as well as file types and formats; in other words, descriptive, structural and administrative 'meta data' about the data set. Registration, including signing an 'end user' agreement or

licence, is usually a requirement prior to gaining access and downloading data sets. Such agreements often contain clauses around the use, storage and sharing of data.

Identifying appropriate qualitative material for a given project involves exploring the data that is available in an archive or across several archives. You could bring together data from many different projects housed in one archive, as we have done. Alternatively, data sets from different repositories could be synthesised, or you could search for archived material to bring into conversation with your own data.

The aim of this initial search is to gain a precursory understanding of the nature, quality and 'fit' with the research topic of the available small-scale data sets. We saw parallels between this process and that of an archaeologist's aerial survey. We felt we needed to fly systematically across a data landscape to get a good overview. This part of the process is likely to be time-consuming. It can be wide-ranging, for example, locating data sets on a broad topic area, or it could be quite narrow, focused on searching for data to fit a specific substantive issue or set of research questions. As part of this initial identification of data sets we found it useful to explore some of the outputs produced by the original researchers.

The process of searching within a given archive varies. The UK Data Service (UKDS, https://www.ukdataservice.ac.uk/), for instance, features the 'Discover' search function for reviewing their data catalogue, which includes the option to filter for qualitative data sources. The search function in the Timescapes Archive (http://timescapes-archive.leeds.ac.uk/) allows browsing by project, concepts or descriptive word, enabling searches by criteria such as gender, employment status etc. This approach does rely on the keywords assigned to each data item by the original research team, so there may be data that is of interest that does not come up on a descriptive word search. New forms of searching are currently in development. In archives such as 'Qualibank' (https://discover.ukdataservice.ac.uk/QualiBank), accessed via the UKDS, detailed searches can be conducted across the content of the entire collection, although at present this comprises only a small collection of classic studies. Using international archives can raise further challenges of searching for terms in different/multiple languages or making appropriate translations.

Searches within an archive(s) are guided by the researchers' own questions, research topic, and the geographic or linguistic context and these help in the process of deciding which data sets or which parts of multiple small-scale data sets, to include or exclude from the larger, combined data set to be constructed, that we have referred to as our data assemblage. This unique assemblage can be viewed as a new data set, with its own methodological history and the potential to be curated and used by other researchers.

In our study, we surveyed the parameters of six of the core data sets deposited in the Timescapes Archive. We initially kept the six projects separate in order to get a sense of the scope and nature of each the data sets. We mapped the studies, explored the state and volume of the data, viewed any contextual material and metadata available, logged the research tools used, and gained an overview of the substantive emphasis of each project. We then used the qualitative analysis software, NVivo, to help us manage the volume of data and decided, as part of this process, to harmonise file names to aid retrieval and the reorganise the files from their original data sets into new groupings – gender and cohort generation - based on our substantive focus and chosen unit of analysis for cases. It was at this point that the individual data sets were merged into our new data assemblage.

The blog is based on our forthcoming chapter in Kahryn Hughes and Anna Tarrant's book 'Advances in Qualitative Secondary Analysis' (Sage).

References
Davidson, E., Edwards, R. Jamieson, L. and Weller, S. (2019) Big data, qualitative style: a breadth-and-depth method for working with large amounts of secondary qualitative data, *Quality & Quantity*, 53(1): 363-376. [Open access]

Reflections from the Men, Poverty and Lifetimes of Care study

Anna Tarrant, University of Lincoln, UK.

Like the 'Working across qualitative longitudinal studies' project, the Leverhulme Trust funded 'Men, Poverty and Lifetimes of Care' (MPLC) study has facilitated important opportunities for reflection on key methodological questions about the feasibility of working with multiple qualitative longitudinal datasets. Qualitative secondary analysis (or the re-use of qualitative data in its simplest form) is a relatively novel approach in the context of the much wider spread re-use of quantitative data, yet it has already provoked a great deal of debate, particularly in relation to issues of epistemology, ethics and context (see Irwin, 2013). An emerging area of concern within these debates focuses on the possibilities and pitfalls associated with bringing multiple datasets into analytic conversation and whether or not this is possible or even desirable.

In the first year of the MPLC study, I conducted a qualitative secondary analysis on two datasets from the Timescapes Archive, allowing me to reflect on some of these questions. At the outset of the proposed study, which aimed to explore men's care responsibilities in low-income localities, I identified the Following Young Fathers (FYF) and Intergenerational Exchange (IGE) studies (see the Timescapes website for more information about both studies http://www.timescapes.leeds.ac.uk/) as possible resources for exploring key themes in relation to this substantive area of focus. For the purposes of rigor, I employed a three stage methodological strategy that was attentive to the principles of the 'stakeholder ethics' model (Neale 2013) including:

1. Familiarisation with the datasets (by having individual conversations with available members of the original research teams and reading project outputs),
2. Holding a data sharing workshop to:
    a. Consolidate the familiarisation process and,
    b. To facilitate a more collaborative mode of working with the original project teams by bringing the datasets into analytic conversation with a focus on the broad themes of men and care,
3. The Qualitative Secondary Analysis itself.

These processes are discussed in much greater depth in a Timescapes Working Paper (https://timescapes-archive.leeds.ac.uk/publications-and-outputs/).

Qualitative secondary analysis across multiple datasets has been really insightful and productive. While a time-consuming and difficult process (particularly in becoming familiar with data generated by others), it has fed directly into the design and conduct of the second empirical phase of the MPLC study, for which, I am interviewing men living in low-income localities. In combination, the FYF and IGE datasets have also provided a sampling framework for the MPLC study. Since September 2015, I have focused on recruiting men living in low-income circumstances of different age groups, in order to explore men's trajectories and their care responsibilities over time. I have also been able to recognise the importance of men's wider interdependencies in low-income localities and this has prompted me to ask the participants in the MPLC study, specific questions about the significance of their wider support networks. In terms of substantive outcomes, I have gained greater insight into men's experiences of living on a low-income over time and how gendered inequalities mediate these processes. While the datasets are not directly comparable, the participants in both studies live in contemporaneous times and there are remarkable similarities across the datasets with regards to how men experience low-income life. In bringing the datasets into conversation, it has been possible to test my emerging theories with empirical data from both datasets.

For more information about the MPLC study please follow the study on Twitter @menpovcare and the study website http://menandcare.org.uk

Using qualitative secondary analysis as a tool of critical reflexivity

Sarah Wilson, University of Stirling, UK.

Maintaining a critical, reflexive approach to research when engaging in specialised work is not easy. Partly because of the need to convince funders of their expertise, researchers often focus on relatively circumscribed areas of inquiry, with samples drawn from particular social groups.

My own research has focused on samples characterised as 'vulnerable'; notably young people affected by parental substance misuse or living 'in care'. Often this work has been located within more 'applied' approaches to social research, and influenced by funders' concerns. Such work is valuable. However, the segregation often maintained between research with young people from more 'vulnerable' and more 'ordinary' backgrounds may reinforce perceptions that the experiences, values and aspirations of members of each 'category' are distinct. As Law (2009) argues, research is 'performative', helping to re-produce and reinforce perceptions of social groups. In the current political context, such distinctions may even implicitly reinforce the stigmatisation of 'troubled families'. As such, there is a need to find ways to subject one's own research practice to scrutiny.

To better situate my previous research, I engaged in qualitative secondary analysis of the longitudinal Timescapes 'Siblings and Friends' (SAF) study to prepare for a new project with 'looked after' young people: Young people creating belonging: spaces, sounds and sights (ESRC RES-061-25-0501). The idea was to reflect on my own approaches, and previous framings of interview questions in the light of the very rich SAF project data which involved predominantly 'ordinary' young people from across the UK. This proved to be an illuminating, if demanding, process that prompted further thought about both projects.

Importantly, this analysis highlighted significant commonalities between the experience of those included in 'ordinary' and 'vulnerable' samples. Notably, the SAF data included several accounts of strained family relationships, of parental mental ill-health and of undesirable housing conditions that suggested family circumstances comparable to those in my previous work on parental substance misuse. However, the SAF interview questions situated violence outside of the home. As Gillies (2000) argues, even where 'difficult' accounts within 'ordinary' samples are identified, they are often not written up. As such, the complexity and pain within 'ordinary' families may be under-estimated in research, and potentially more easily obscured within political discourse. Similarly, the everyday ambiguity and minor conflicts associated with 'ordinary' siblings and parents sharing limited space may be downplayed. Such ambiguities and tensions led several SAF respondents to seek out friends' homes, or private corners of their own, to escape from family life at least for a time. I had previously associated such strategies with young people affected by parental substance use, many of whom often spent time at friends' houses. However, this analysis suggested a more nuanced understanding of the importance to the latter group of employing strategies that could be presented as 'ordinary' teenage practices. The process of secondary analysis also highlighted uncomfortable omissions from my previous research in which, for various reasons, greater emphasis was placed on the respondents' own potential substance use, than on their school work and employment aspirations. The predominance of such concerns in the SAF accounts led me to worry that my own research had reflected and performed perceptions of education as less important to 'vulnerable' than to 'ordinary' young people.

In conclusion, qualitative secondary analysis is a 'labour-intensive, time-consuming process' that Gillies and Edwards (2005: para24) compare to primary data collection. However, it presents a useful tool to subject assumptions built up over a specialised research career to scrutiny.

This post draws on Sarah's 2014 article in Sociological Research Online, 'Using secondary analysis to maintain a critically reflexive approach to qualitative research' which you can read here:
http://www.socresonline.org.uk/19/3/21.html

References

Gillies, V. (2000) Young people and family life: analysing and comparing disciplinary discourses, *Journal of Youth Studies*, 3(2): 211-228

Gillies, V. and Edwards, R. (2005) Secondary analysis in exploring family and social change: addressing the issue of context, *Forum: Qualitative Social Research*, 6(1): art 44.

Law, J. (2009) Assembling the World by Survey: Performativity and Politics, *Cultural Sociology*, (3)2: 239-256.

Wilson, S. (2014) Using secondary analysis to maintain a critically reflexive approach to qualitative research, *Sociological Research Online*, 19(3), 21 http://www.socresonline.org.uk/19/3/21.html

Thomson, R. (2010). Creating family case histories: subjects, selves and family dynamics. In Thomson, R. (Ed.) *Intensity and insight: qualitative longitudinal methods as a route to the psycho-social.* Timescapes Working Paper Series No.3.

3. Slicing data in different ways

Case histories in qualitative longitudinal research – some thoughts

Rachel Thomson, University of Sussex, UK.

Questions of scale

Qualitative longitudinal research can play around with our ideas of scale. A study can seem to be 'small', following 6 cases for example, yet at the same time can be 'big', or perhaps a better word is 'deep' in collecting many instances of data for that case over an extended or just intensive period of time. Discussing this point Lyne Yates (2003) makes a case for QLR having a different kind of 'warrant' – or relationship with validity –moving us away from ideas of cases as being 'representative' in an abstract way –be that they are typical or that they may provide insight into a wider population through the operations of probability sampling. By following just cases over a period of 10 years (as we have in an ongoing study) we are able to understand relationships, sequences, consequences and antecedents in a concrete way- exploring the relationship between what we say and do, and between what we want and what we get – as researchers and as participants. More recently Liz Stanley has challenged the qualitative quantitative distinction on her work using collections of letters showing that in an era of digital data qualitative material and quantity and quantification are not mutually exclusive (Stanley 2015). Rather we might think of scale in terms of a zooming in and zooming out of perspective, and the potential to combine the affordances of the microscope and macroscope. Debates about scale within QLR parallel debates about scale within 'big data' and the kinds of digital tools that can be used to explore patterns, to zoom in for the close-up and to zoom out for the landscape or the map.

From cases to casing

QLR can be designed in different ways in order to reveal different kinds of cases. At the most basic level we might think of the case as a unit of analysis that we follow over time. So for example in our project Making Modern Mothers (http://modernmothers.org/), the unit of analysis was women about to have their first baby. Yet cases are not stable, especially when pregnant and in this study we expanded our case to include significant others (especially grandmothers) and children when they were born. These children are now the focus of a follow-on project that explores digital childhoods, yet the backstory of the family is a vital part of the case and family members play a key part in narrating the case of the child who is the focus and who as we watch moves from being a 'case' of a child into a teenager and an adult. Analytically we can also think of the case in other ways, for example thinking about all of the urban families together and considering their affinities and their difference from the rural families. We might also think of the case of social class, or cutting the data set in the opposite direction, from the diachronic to the synchronic, considering how the families responded to a key external event such as the 'credit crunch' that turned into the extended period of austerity through which we continue to live. Rather than thinking of cases as stable and defined simply through existence we might follow Charles Ragin (1992) to think about practices of 'casing' in social enquiry, a flexible analytic practice that pays due respect to the complexity of the social realm and which in linking ideas and evidence had the potential for the testing and emergence of theory.

The case history and the archive

The 'case' itself is an object and genre with a history linked to practices of natural history, collecting, sorting and narrating and reflections. Butterflies were collected and displayed in a case long ago in a way that has parallels with the ways that doctors and lawyers began to conceptualise case histories and case law. A special issue 'On the case' of the journal *Critical Inquiry* helps us as social scientists understand our practices in historical and cultural context as well as helping us see the kinds of spillages that echoes that may travel between medical, legal, scientific and literary uses (Berlant 2007). There is no definitive way of constructing or telling a case, yet we may find ourselves being drawn into particular tropes taking up associated forms of authority. When telling the story of an individual over time it may be hard to escape the perspective of the doctor or the social worker who is able to see and describe underlying causes or pathologies. Perhaps we need to deliberately disrupt these well-worn narrative tendencies by reading

materials against the grain, changing the direction of our analysis, or moving between individual and collective or conceptual cases self-consciously in order to find new perspectives.

In earlier work I suggested that we might make use of the notion of the archive more fully in our work learning from the critical work that has been done of reading the archive (Thomson 2007, McLeod and Thomson 2009, Thomson 2011). If we think of our data sets as archives, which can be organised into all sort of cases (individuals, institutional, geographical, temporal), we can also think about the kinds of stories that can be told from the archive, putting material together in a particular way will enable a particular history. Yet this is not definitive or exclusive. That material could be told in different ways by different analysts without taking away from the 'validity' of the material itself. Digital information systems allow individuals to build their own archives, copying and linking data from public collections and potentially making their own archives available to others. Sociological data sets are also made available to and interrogated by secondary analysts and there is a compelling case for social scientists to build on the lessons of historical and literary scholars about archival methodologies and epistemologies as well as understanding the new methodologies of the digital humanities. Having my data used by secondary analysts encourages me to believe that the potentials of this area are just beginning to be explored by sociologists – see for example http://www.whiteswritingwhiteness.ed.ac.uk/blog/archive-project-sendoff/

References
Berlant, L. (2007) On the case, *Critical Inquiry*: special issue, 33(4)
McLeod, J. and Thomson, R. (2009) *Researching Social Change: qualitative approaches*, Sage.
Ragin, C. (1992) '"Casing" and the process of social inquiry' in Ragin & Becker (eds) *What is a case: exploring the foundations of social inquiry*, Cambridge University Press.
Stanley, L. (2015) Operationalizing a QLR project on social change and whiteness in South Africa, 1770s – 1970s, *International Journal of Social Research Methodology,* 18(3): 251-65.
Thomson, R. (2007) The qualitative longitudinal case history: practical, methodological and ethical reflections, *Social Policy and Society*, 6(4): 571-582.
Thomson, R. (2011*) Unfolding lives. Youth, gender and change*. Policy Press
Thomson, R., Hadfield, L., Kehily, M.J. and Sharpe, S. (2010) Family fortunes: an intergenerational perspective on recession, *21st Century Society*, 5(2): 149-157.
Thomson, R. (2014) Generational research: Between historical and sociological imaginations, *International Journal of Social Research Methods*, 17(2): 147-156.
Yates, L. (2003) Interpretative claims and methodological warrant in small number qualitative longitudinal research, *International Journal of Social Research Methodology,* 6(3): 223-32.

The challenges of multiple perspectival qualitative longitudinal analysis: a strategy created for an intergenerational study of young onset dementia

Sue Bellass, University of York, UK.

Although dementia is often perceived to be a condition that occurs in later life, around 1 in 20 people with dementia are below the age of 65 (Alzheimer's Societ, 2015). Over the last two decades there has been increasing interest in developing qualitative understandings of the experience of the condition in younger people; however, almost without exception existing studies have used cross-sectional designs, providing only a snapshot of life with an unpredictable, dynamic condition. For my PhD I decided to use a qualitative longitudinal (QL) methodology to explore relationality over a twelve-month period by following five intergenerational families where one person had received a diagnosis of young onset dementia.

Since people with dementia are a marginalised, negatively positioned group (Sabat et al. 2011), I felt it was appropriate to democratise the research process to enable my participants to choose their preferred means of engaging with the study. This choice included the method of data collection (ethical approval was gained for interviews, audio/ video diaries, blogs and tweets) and, if participants opted for interviews, which family members would participate and where the interviews would take place. Ultimately, 18 participants chose to be interviewed, 16 of whom were interviewed in pairs or larger family groups, with two preferring individual interviews. Interviews were conducted in three waves at months 0, 6 and 12.

Analysing the data set has been a challenging process. As Henderson et al. (2012) note, despite increasing interest in QL methods, methods of analysing and representing complex QL data sets have rarely been explicated. I experienced this as a mixed blessing; on the one hand, there is space for creativity, flexibility and freedom, on the other, there is room for doubt to flourish! I have attempted to slice the data in different ways in order to interrogate the data set to best effect. Inspired by Thomson (2010, 2014), I treated each family as a unique case and also aimed to create a cross-case analysis across the four generations represented in the families.

Initially I attempted to analyse the group interviews at the 'family' level, however it quickly became apparent that divergent accounts were being obscured. Subsequently I took a multiple perspectival approach (Ribbens McCarthy et al, 2003), teasing apart individual experiences within the families, viewing them as cases within a case. For each person, I induced categories of experience then, to permit holistic re-engagement, organised the raw data in a time-ordered matrix across the three waves.

Then, again for each person, I created a longitudinal matrix adapted from Saldana (2003) to look for transitions and continuities, using motif coding, a form of coding which draws attention to recurring elements in experiences, and describing through-lines, a crystallisation of a participant's change over time. Although it could be argued that such an approach may disguise intersubjective creation of meaning, I consciously retained a focus on relationality, creating spaces within the matrix to capture data on meaning-making processes over time. Finally, I created an intergenerational matrix, organising the data by generation to look for patterns and themes, setting the data against the backdrop of the recent increasing public, policy and research interest in dementia to try and interweave biographical, generational and historical timescapes.

Example QL matrices

Qualitative research has faced criticism for lack of clarity regarding the relationship between theory and data, and this, I argue, is an important area to address as we continue to develop the contours of QL research. My own perspective has been influenced by Mills (1959), who describes a 'shuttle back and forth' between theory and data. I have utilised such an iterative approach, and have drawn on theory from the sociology of chronic illness and family and relationship sociology to develop understandings of the intergenerational experience of young onset dementia.

References
Alzheimer's Society (2015) *Dementia 2015: Aiming higher to transform lives*. London: Alzheimer's Society.
Henderson, S., Holland, J., McGrellis, S., Sharpe, S., and Thomson, R. (2012) Storying qualitative longitudinal research: sequence, voice and motif. *Qualitative Research*, 12(1): 16-34.
Mills, C.W. (1959) *The Sociological Imagination*. London: Penguin.
Ribbens McCarthy, J., Holland, J., and Gillies, V. (2003) Multiple perspectives on the 'family' lives of young people: methodological and theoretical issues in case study research. *International Journal of Social Research Methodology*, 6(1): 1-23.
Sabat, S.R., Johnson, A., Swarbrick, C., and Keady, J. (2011) The 'demented other' or simply 'a person'? Extending the philosophical discourse of Naue and Kroll through the situated self. *Nursing Philosophy*, 12(4): 282-292.
Saldaña, J. (2003). *Longitudinal Qualitative Research: analyzing change through time*. California: Alta Mira Press.

Working backwards and forwards across the data: Bringing together qualitative longitudinal datasets with different temporal gazes

Jane Gray, Maynooth University, Ireland.

Bren Neale (2019, p. 20) has contrasted qualitative longitudinal (QLR) methods that *prospectively* trace lives through time with approaches to the study of lives that reconstruct them through a *retrospective* gaze. Joanne Bornat and Bill Bytheway (2012) showed how different methods of data collection construct different temporalities within QLR. In this blog post I describe how the *Family Rhythms* project (https://www.maynoothuniversity.ie/iqda/family-rhythms) created an interesting opportunity to bring different temporal gazes and temporalities together, in a study funded by the Irish Research Council as a demonstrator project for secondary qualitative data analysis in Ireland. Inspired by new sociological approaches to understanding family life as configurations, practices and displays, Ruth Geraghty, David Ralph and I aimed to develop a fresh understanding of long-term patterns of family change by bringing retrospective life story narratives from the *'Life Histories and Social Change'* (LHSC) project together with qualitative interviews collected as part of the first wave of the prospective panel study *'Growing Up in Ireland'* (GUI).

Because the LHSC study was carried out with three birth cohorts of Irish people (born before 1935, 1945-54 and 1965-74), we initially aimed to treat the GUI interviews (carried out with nine-year old children and their parents) as a fourth cohort (born around 1998). However, we soon found that the different study designs created challenges for this simple 'additive' approach. First, while the LHSC study included a life history calendar instrument to collect retrospective data about the timing of events in participants' lives, the life story interviews were relatively unstructured, loosely guided by topic and life stage. By contrast, the GUI interviews were semi-structured with questions designed to map on to the broad themes covered within the quantitative panel study. More significantly, however, it soon became apparent that the different temporal gazes adopted within the studies affected the substantive content of the data. With their focus on remembering past events, the LHSC interviews (including the formal calendar) look 'backwards', whereas the GUI interviews have a pronounced 'forward-looking' focus on the children's anticipated futures. This was reinforced, in the case of GUI, by additional instruments including, for example, an essay writing exercise that invited the children to imagine what their lives would be like at age 13. These divergent temporal perspectives affected how people talked about their family lives. While the LHSC interviewees 'made sense' of their family lives by reconstructing them within a biography, the GUI interviewees situated them within everyday practices and contemporary relationships. Their temporal orientation is anticipatory and aspirational, rather than reconstructive and explanatory.

Of course, these differences were not absolute: many LHSC interviews include narrative segments about hopes for the future, while some GUI parent interviews include reflections on past family lives. Nevertheless, the divergent temporal perspectives of the studies meant that it was not possible to make straightforward thematic or life stage comparisons across cohorts. We addressed this challenge by adopting a 'temporal gaze' within our analysis, in a process that we have described as 'working backwards and forwards across the data' (see Gray, Geraghty and Ralph 2013, Geraghty and Gray 2017). In effect, this meant that we read both with and against the temporal 'grain' of the data (Savage 2005), often incorporating different generational standpoints. This can be seen most clearly in our analysis of the changing relationship between grandchildren and their grandparents. A reading that begins with children in the GUI study and works *backwards* across LHSC participants' childhood memories, reveals an exceptional degree of continuity in the quality of the relationship from the perspective of grandchildren, going right back to the earliest decades of the 20th century. However, a reading that begins with the childhood memories of the oldest LHSC participants and works *forwards* through memories and contemporary experiences from the perspectives of parents and grandparents, reveals significant change in the family, household and community contexts within which the grandchild-grandparent relationship was experienced across historical time. This analytical approach thus yielded new substantive and theoretical insights on the character of long-term patterns of family change.

Our strategy of 'reading backwards and forwards' emerged as a way of addressing the challenges presented by our efforts to work across qualitative longitudinal datasets with different temporal gazes and temporalities. However, what started out as a problem turned into an opportunity to develop higher level understandings of long-term patterns of family change by reading with and against the temporal grain of the datasets, illustrating the potential for including divergent temporal gazes within the corpus of QLR.

References

Bornat, J. and Bytheway, B. (2012) Working with different temporalities: Archived life history interviews and diaries. *International Journal of Social Research Methodology*, 15(4): 291-299.

Geraghty, R. and Gray, J. (2017) Family Rhythms: Re-visioning family change in Ireland using qualitative archived data from Growing Up in Ireland and Life Histories and Social Change. *Irish Journal of Sociology,* 25(2): 207-213.

Gray, J., Geraghty, R. and Ralph, D. (2013) Young grandchildren and their grandparents: continuity and change across four birth cohorts. *Families, Relationships and Societies,* (2)2: 289-298.

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

Savage, M. (2005). Revisiting classic qualitative studies. *Historical Social Research/Historische Sozialforschung,* 118-139.

4. Different analytic strategies for developing new understandings

Be transparent (and proud) - How can we better describe the practice of qualitative longitudinal analysis?

Åsa Audulv, Mid Sweden University, Sweden and Dalhousie University, Canada.

About 12 years ago I started a QLR project as part of my PhD work. At that time, I knew little about the traps, tricks and rewards of longitudinal analysis. I basically had the idea that our phenomenon of interest - the self-management of long-term health conditions - changed through an individual's illness trajectory and, because existing research heavily relied on one-time interviews, I thought a longitudinal design would provide more insight. In short, data collection with four interviews per participant, spanning over two years seemed like a design that could contribute new knowledge. I understood that this design could result in around 70 interview transcripts to analyse and, at that time, I only had vague ideas about how that analysis might be conducted.

Over the past year I have been working with colleagues on a literature review concerning different methodological approaches to QLR analysis. Our inclusion criteria have been articles within the field of health research, collecting qualitative data at several time-points. After reading 52 articles one thing that surprised us was how little was conveyed about the longitudinal aspects of the analysis. In total, 57.6 per cent (30 articles) did not mention how they had managed the longitudinal aspects. For example, they did not say anything about time point, change, or comparison over/through time-points in their analysis section. Since, the body of QLR work is small in comparison to qualitative studies it is possible that many authors were more used to describing approaches to analysing one-time data-collection studies and therefore did not really know how to outline the longitudinal aspect in their analysis. The limited amount of methodological literature might also add to this uncertainty. Further, it is possible that most peer-reviewers of QLR papers are experts in the substantive focus of the work, rather than on QLR methodology, so they might not spot this aspect during the peer-review process. There might also be pragmatic reasons, like limited space available. However, the fact that QLR studies are complicated and relatively unusual would add to the importance of explicit analysis descriptions regarding how such analysis was conducted.

In our review, 22 articles (42.3%) had some description of how they analysed the longitudinal aspect. However, the clarity and depth of the descriptions varied. Some described the longitudinal aspect as an integrated part of their whole analysis. These projects were often centred around investigating change. They typically described their analysis in several steps where the longitudinal aspects were included in almost every step. For instance, Johansen and colleagues (2013) conducted a study about addicted individuals' social motivations and non-professional support. In their description of the analysis the longitudinal aspect was well integrated (the bold indicates the longitudinal aspects):

> "…, we first conducted an open coding of the data **from phase 1 and phase 2**. Next, we used the framework analysis method **to track changes over time** [33], and facilitate axial coding and constant comparison. Relationships between the codes were explored **throughout all three phases** of the study and **individual changes** were covaried with dyadic events and events involving relationships with other people representing network support for either using or non-using. As such, narrative analyses were conducted for all dyads **to capture details about the support process and its consequence for recovery**. In this way, we were able to describe the support dynamics of each dyad, explore how the support was influenced by characteristics of the individual members and support arrangement, and theorize about the ways this affected recovery. In addition **to the tracking of thematic changes**, we also **utilized proportions as indicators of change** [34]" (Johansen, 2013, p.233)

Other studies described the longitudinal aspects as one isolated step, often at the end of the analysis description. This suggests that the first part of the analysis had been conducted with a focus on the phenomena with the longitudinal aspects brought in at a later stage to deepen the understanding and/or add another perspective. For example, do Mackintosh-Franklin et al (2014) describe:

> "*Findings from each interview stage were analysed separately, and only after separate analysis had taken place **were both data sets combined for final analysis**. Findings reported below are from the two final stages of this analytical process, using separate and combined interview sets.*" (Mackintosh-Franklin, 2014, p.202)

Some articles mentioned a longitudinal dimension to the analysis, but were not specific about how that analysis was conducted. For example, Salter et al (2014, p.2) describe: "Iteration between both data sets and the research literature helped inform the analysis at the explanatory level." Several studies described the use of tools and/or analysis strategies that are often employed for analysing longitudinal aspects. For example, matrices, flow charts, and/or comparing across parts or interviews. Some described these tools and analysis strategies clearly. However, it is more common that they are mentioned in passing and the reason and outcomes of using these practices remain unclear. For example, one article mentioned the use of matrices but did not describe if those matrices were compared to time-points, cases or both.

In conclusion, as the other blogs in this collection have shown, there are different ways to analyse QLR data, and thus different ways of describing the qualitative longitudinal aspects of analysis. First, we need to be clear about what aspects of a project are longitudinal and how we are going to analyse them. Secondly, by being transparent in our description of how we conduct the analysis we can make our approach and our justification for that approach clearer. In turn, that will make it easier for our readers to evaluate the quality of our work. In our review, 57.6% of the articles lacked a description of how they analysed time in their QLR. I would argue that would be 30 articles too many. A third reason to clearly describe the longitudinal aspects of an analysis is to raise awareness of our work. We should be proud of the approach we use. QLR opens up a wide range of possibilities. It can help us better describe our phenomena of interest, and collect richer data. By writing a succinct analysis section we are giving an example of how it can be done, teaching others about QLR, and showing the merits of such approaches. My longitudinal data collection lasted for two and a half years and included 81 interviews that generated 726 single-spaced transcribed pages. Eventually, it was presented in two research papers (Audulv, Asplund and Norbergh 2012, Audulv 2013) and I still think it was a rather cool project.

References:
Audulv, Å., Asplund, K. and Norbergh, K-G. (2012) The process of self-management integration. *Qualitative Health Research*. 22(3), 332-345
Audulv, Å. (2013). The over time development of chronic illness self-management patterns: a longitudinal qualitative study. *BMC Public Health*, 13:452
Johansen, A.B., Brendryen, H., Darnell F.J. and Wennesland, D.K. (2013). Practical support aids addiction recovery: the positive identity model of change. *BMC Psychiatry* 13:201
Mackintosh-Franklin C. (2014). The Impact of Experience on Undergraduate Preregistration Student Nurses' Responses to Patients in Pain: A 2-Year Qualitative Longitudinal Study. *Pain Management Nursing*, 15, (1): 199-207
Salter C, McDaid L, Bhattacharya D, Holland R, Marshall T, et al. (2014) Abandoned Acid? Understanding Adherence to Bisphosphonate Medications for the Prevention of Osteoporosis among Older Women: A Qualitative Longitudinal Study. *PLoS ONE* 9(1): e83552.

Working with qualitative longitudinal data
Georgia Philip, University of East Anglia, UK.

For the past two years I have worked with colleagues John Clifton and Marian Brandon on a qualitative longitudinal (QL) study of men's experiences of the UK child protection system. Alongside the twists and turns of the research relationships developed with our participants and the conceptual work involved in presenting their accounts, we have also encountered practical challenges of managing the volume and depth of data generated. This post briefly identifies some of these challenges, and our responses to them.

Our QL study involved 35 men who were fathers or father figures to a child with a newly made child protection plan, recruited between April and August 2015, and taking part for a period of 12 months. The study consisted of two in-depth interviews, at the start and end of the study period, and (approximately) monthly phone contacts with each man. Twenty-eight men participated for the full 12 months. We took a holistic approach, looking back at men's histories, relationships, fathering experiences and any past encounters with welfare agencies, and then accompanying them forward, into the current encounter with child protection and its impact on their lives.

Fatherhood
Our overall approach to the analysis was inductive and iterative, drawing on existing QL methodological literature (Neale, Henwood and Holland 2012). It also engaged us in thinking about 'time' in theoretical and methodological terms: as a concept, that shapes how lives are lived, narrated and imagined, and as a resource for examining a significant local authority process. Our practical approach to the management of the high volume of data was a combination of pre-emptive and responsive strategies. Three challenges we encountered were, how to analyse across and within our sample; how to facilitate data sharing across the research team; how to combine analysis of men's lives, and of the child protection system, in coherent way.

Early on, we decided to use NVivo Frameworks as a mechanism for managing the data (NatCen 2014, Ritchie et al. 2014), and we constructed a matrix to record aspects of men's lives, and of the unfolding child protection process. This enabled us to collate and analyse data from the outset rather than separating (and delaying) analysis from data collection. It also established a process for organising the data using the 'case and wave' approach adopted in other QL studies (Hughes and Emmel 2012, Thomson 2007) to look across the sample by time wave (we divided our 12 months into four three-month periods), and within it, at each man's individual 'case' However, whilst NVivo allowed us to develop a way of structuring our analysis, it did not, in practice, facilitate a reliable way of collaborating across the research team.

As the researchers, John and I had a group of men and an accumulating data set that we 'knew' better. This meant we needed to develop ways of sharing cases and checking our developing analysis, to build an integrated and credible understanding of the sample as a whole. We found that working independently on, and then trying to merge, copies of our NVivo project just wasn't viable, and the project files were unstable. Therefore we had to devise, or revert back, to other strategies for managing this. We continued using our original matrix, to summarise data over the four time waves, and to help compile the individual case studies, but did this using Word and sharing via a secure drive on the University network. We met monthly as a full team to discuss and compare our analysis, understand the developing cumulative picture, and review the ongoing process of data gathering. We also came to make extensive use of memo writing as a particularly useful means of condensing data, exploring pertinent issues within it, and discussing these with each other. We then took the decision that John and I each take the lead in analysing one of the two main domains of the data: men's encounter with the child protection process and their wider lives as fathers. This ensured that we both had to fully consider all participants' data and actively collaborate on integrating our work as part of the later, conceptual stages of the analysis.

This project has been intensely demanding and satisfying, at every stage. Finding ways of coping with rich, accumulating data, generated with increasing momentum as research relationships develop, has been just

one of these demands. Being committed to an inductive approach, which does justice to the men's own accounts, whilst also generating a coherent conceptual explanation and meaningful practice messages for social workers, is another. What we have offered here is a tiny glimpse into some of the practical strategies for meeting such multiple demands, which we hope may be useful for other researchers new to QL research.

For further details about the project please see: http://bigqlr.ncrm.ac.uk/wp-content/blogs.dir/sites/58/2017/06/Fathers-Research-Summary.pdf

References:
NatCen (2014) *Frameworks in NVIVO manual- Step by step guide to setting up Framework matrices in NVIVO.* London: NatCen Social Research.
Neale, B, Henwood, K. and Holland, J (2012) Researching Lives Through Time: an introduction to the Timescapes approach, *Qualitative Research*, 12(1): 4-15
Ritchie, J., Lewis, J., McNaughton Nicholls, C., and Ormston, R. (2014) *Qualitative research practice: A guide for social science students and researchers*, London: Sage.
Thomson, R. (2007). The qualitative longitudinal case history: practical, methodological and ethical reflections. *Social Policy and Society*, 6(4): 571–582.

Analytic strategies for working within and across cases in qualitative longitudinal research

Ruth Patrick, University of York, UK.

When I think about the – many – reasons why I am a fan of qualitative longitudinal research (QLR), I often remember Tess Ridge's reflection on her own journey moving from researching at one point of time to researching through and across time. Tess described her experience as equivalent to going from watching television in black and white to technicolor, such are the greater depths, richness and detail that qualitative longitudinal research enables.

This richness is a wonderful advantage of QLR, but it does create challenges for the research process, especially when it comes to data management and analysis. In 'The Lived Experiences of Welfare Reform Study', I have followed a small number of single parents, disabled people and young jobseekers as they navigate the changing social security context, and experience an increasingly punitive regime of welfare conditionality and benefit sanctions. This research (which remains ongoing) has generated rich data, which I have sought to analyse by working to develop both diachronic (tracking a case over time) and synchronic (looking across cases at one point in time) analyses, as well as exploring the iteration between the two (Corden and Nice 2006). To aid my data management, I use the qualitative analysis software package NVivo, with thematic codes emerging from a close reading and engagement with the data generated. Data analysis strategies include developing pen pictures for each case, which provide a short account of each individual's journey through welfare reform. The synchronic analysis is supported by the coding process and then efforts to climb the analytical conceptual scaffold working upwards from data management and coding to descriptive analyses, and finally to explanatory accounts (Spencer et al. 2003). Repeated immersion with the data is critical, as is taking the time to return to the data generated after each wave, as each re-analysis brings fresh insight. In looking to the iteration between the diachronic and synchronic, I find it helpful to explore patterns and anomalies in the data generated, and to identify common themes emerging through time between the cases.

One theme to emerge very strongly from this analysis is the extent of 'othering' that exists (Lister, 2004), whereby the people I interviewed seek to assert their own deservingness to social security by dismissing and being critical of the asserted 'undeservingness' of some 'other'. This 'othering' was widespread in my interviews, and there was some evidence of this increasing over time, as welfare reforms continued to take effect.

For example, Chloe, a single parent, talked negatively about immigrants in each of our three interviews between 2011 and 2013. However, her anger grew at each interview encounter and – by the third interview – in 2013 – she was employing threats of violence, and using racist language, in her articulation of how she felt towards this 'other' that she judged undeserving.

My initial diachronic analysis of Chloe's case found that her sense of anger and even hatred towards immigrants grew over time, and that this could have arisen because she was herself being increasingly adversely affected by welfare reform. As her own situation deteriorated, she hit out more stridently at the perceived 'other', with her anger borne out of alienation, poverty and disenfranchisement.

However, another analysis is possible. As qualitative longitudinal researchers remind us (see, for example, Neale, 2018), one of the advantages of QLR is that repeat encounters develop the relationship between researcher and researched, and then create the possibilities for disclosures in later interviews because of a strengthened relationship and improved trust. Could it be, therefore, that Chloe was only speaking more stridently because she felt more secure in our relationship, and comfortable to speak to me more openly?

There are no easy answers here, and to assume otherwise would be simplistic but analytical strategies can help. Further diachronic analysis of Chloe's case reveals that there were some significant disclosures made

in the third interview wave, which were not mentioned (although relevant) in the first and second interviews and so this might suggest a change in relationship as posited. At the same time, though, synchronic analysis of 'othering' in each of the interview waves shows that the increased presence and ferocity of 'othering' observed in Chloe over time was also observable in several of the participants.

What is important to recognise – above all – is that as a qualitative longitudinal researcher returns each time to a participant, their relationship is inevitably evolving and changing, and that this may alter what participants say and why they say it. Working within and between cases in an iterative manner could help a secondary analyst to understand more of the context of the interview, and to consider how the changing relationship between researcher and participant may have impacted on what is disclosed over time. To further support this, it is beneficial if secondary analysts can have access to any field notes or research diaries completed by the primary researcher(s), as these may help clarify how research relationships evolved over time, and any reflections from the researcher on how these affected the data generation process.

QLR is a wonderful method within a researcher's tool bag, but it – like any of the most powerful tools – needs to come with careful instructions and health warnings.

References
Corden, A. and Nice, K. (2006) *Pathways to Work: Findings from the final cohort in a qualitative longitudinal panel of Incapacity Benefit recipients*. Research Report No 398. London: Department for Work and Pensions.
Lister R. (2004) *Poverty*. Bristol: Policy Press.
Neale, B (2018) *What is qualitative longitudinal research?* London, Bloomsbury.
Neale, B. and Hanna, E. (2012) *The ethics of researching lives qualitatively through time*,
Spencer, L., Ritchie, J. and O'Connor, W. (2003) Analysis: Practices, Principles and Processes. In: J. Ritchie and J. Lewis (Eds.) *Qualitative Research Practice: A Guide for Social Science Research Students and Researchers*. London: SAGE Publications, pp.199-218.

Timescapes Method Guides Series, University of Leeds, Available at:
http://www.timescapes.leeds.ac.uk/assets/files/methods-guides/timescapes-series-2.pdf

Seeing the changes that matter: QLR focused on recovery and adaptation
Joanna Fadyl, Auckland University of Technology, New Zealand.

For QLR, our data collection period (48 months in total) was relatively short. Our focus was on understanding what helped or hindered recovery and adaptation for people with Traumatic Brain Injury (TBI) and significant others in their lives (family and close community). However, with 52 participants (and their significant others), the volume of data was significant. We interviewed our participants at 6, 12, 24 and 48 months after a TBI. At 48 months this was a subset of participants with diverse experiences.

The focus for our analytical approach was a type of thematic analysis based on Kathy Charmaz's writing on grounded theory. The purpose of our research was to build a picture of what recovery and adaptation looks like for a cohort of people over time. While we did do some analysis of 'case sets' (the series of interviews relating to a particular person) to understand and contextualise aspects of their stories, the focus of analysis was not as much on individuals as it was on looking at patterns across the participant group.

Of course, making sense of a large amount of rich data is always challenging, but the added dimension of change over time was something we spent a lot of time pondering. Because we were interested in exploring recovery and adaptation – and we were particularly interested in how this presented across a cohort – one of the biggest challenges was to find strategies to make the changes we were interested in visible in our coding structure so we could easily see what was happening in our data over time. We chose to set up an extensive code structure during analysis at the first time-point, and work with this set of codes throughout, adapting and adding to them at further time-points. We reasoned that this would enable us to track both similarities and differences in the ways people were talking about their experiences over the various time-points. Indeed, it has made it possible to map the set of codes themselves as a way of seeing the changes over time. To make this work well, we used detailed titles for the codes and comprehensive code descriptions that included examples from the data. At each time-point the code descriptions were added to, reflecting changes and new aspects, and at each time-point consideration was given to which particular codes were out-dated and/or had shifted enough to be inconsistent with previous titles and descriptions. We also considered the new codes that were needed.

I will illustrate with an example. A code we labelled 'allowing me to change what I normally do to manage symptoms and recover' at 6-months, needed extensions to the code description at 12 months to reflect subtle changes. Beyond that although data still fitted with the essence of the code that had been developing over time, we began to question the ongoing appropriateness of the code title. The later data related to the same idea, but it was no longer about managing symptoms so much as it was about navigating the need to do things differently than before the injury in order to cope with changes. This way of working with the code enabled us to reflect on the experience and processes for participants relating to 'allowing me to change what I normally do' over time. At the 24-month point it was 'in transition' – not quite a new code yet, but different enough to be an uncomfortable fit with the original title and description. The description now included this query and ideas that might help us re-consider it in light of new data in the future.

It was apparent that when analysing interviews with participants at 48-months, the data related to this idea had changed, and it was clear that it no longer fitted the existing code title or description. We needed to consider introducing a new code, one that had a key relationship with the existing one but captured the essence of our findings more clearly. Essentially, the idea of 'changing what I normally do' had expired because there was less of a tendency to refer to pre-injury activities as 'what I normally do'. However, negotiating having to do things differently than other people in order to manage life was still an issue for the participants who were experiencing ongoing effects. The change in the codes over time and the relationship between the 'old' and 'new' code were very visible using this system. The extensive code descriptions helped orientate us to the interview extracts that were most influential in shaping the code,

and the database we set up for recording our coding allowed us to create reports of every extract coded here so we could review and debate the changes with reference to the key data and the general 'feel' of what was coded there.

Another key strategy we used to help us explore the data over time was the use of data visualisation software. The software we used (QlikSense) is designed for exploring patterns in data and then directly drilling down into the relevant detail to look at what is going on (as opposed to seeing an overview – we did our overviews on paper). One example is where codes and groups of codes varied in their prominence (e.g. coding density or number of participants who contributed to the code) across different time-points. Seeing these differences prompted us to look at the code descriptions and the data coded there to consider if this pattern added to our understanding of how people's experiences were changing over time. We provide some more detailed examples of different patterns we explored in the paper that was published in *Nursing Inquiry* in 2017. The paper also gives some more detail and a slightly different perspective on some of the other discussion in this post. We invite you to read the paper and contribute to the conversation!

References
Fadyl, J. K., Theadom, A., Channon, A. and McPherson, K.M. (2017) Recovery and Adaptation after Traumatic Brain Injury in New Zealand: Longitudinal qualitative findings over the first two years. *Neuropsychological Rehabilitation* (open access)
Fadyl, J. K., Channon, A., Theadom, A. and McPherson, K.M. (2017) Optimising Qualitative Longitudinal Analysis: Insights from a Study of Traumatic Brain Injury Recovery and Adaptation. *Nursing Inquiry*, 24(2).

Tracing changes in notions and practices of child feeding: a trajectory approach to qualitative longitudinal research

Irmak Karademir Hazır, Oxford Brookes University, UK.

I am using qualitative longitudinal research (QLR) to explore how families with young children (1.5 to 4 years old) organise and negotiate eating/feeding practices at home and beyond. The families I work with have different levels of economic, cultural, and temporal resources at their disposal and they all try to manage them to maintain an emotionally and nutritionally rewarding food routine. My interviews have generated data that could be very interesting for a cross-sectional analysis, demonstrating different notions of healthy eating/feeding and class-cultural distinctions in food socialisation. However, I am more interested in the element of change in this particular study. In what ways do a variety of factors (e.g. parents' return to work; arrival of a new sibling; information received from professionals, social media, or the baby food industry) shape the period in which young children embody new food habits? What happens to adults' eating practices when they have new family members (e.g. changes in the gender division of labour; experiences of commensality; acquisition of new cooking practices)? How do parents negotiate their feeding principles as children grow? To understand how these processes unfold in time, I use a trajectory approach in my analysis.

Since I visit my families every six months over two years, the data collection period of this study can be considered short for QLR. However, given that the topic is concerned with a very dynamic moment in couples' lives, the distance between the time points works well. Inspired by Grossoehme and Lipstein's approach (2016) to data analysis in medical QLR, I chose trajectory analysis as an analytical approach, as opposed to recurrent cross-sectional analysis. Trajectory analysis prioritises unpacking how an experience changes over time as well as the factors surrounding the case, rather than solely identifying the differences between two time points. It is advised that researchers use time-ordered displays (sequential matrices), which would permit an understanding of 'what led to what'. To be able to employ trajectory analysis, the data collected from each stage should be coded individually first. After each stage, the themes are put into a matrix to show stability and change with time. As the example below shows, changes such as children starting school or a family's decision to become vegetarian between two stages will influence their feeding principles, routines, and emotional responses. When the coding of three stages is completed, the matrix will show the trajectory of food parenting experiences (from introduction of solids to school age) around the key themes identified.

What makes my analysis different from other QLR that I have read so far is that each stage of data collection in my fieldwork focuses on a different aspect of food practice (provision, preparation, management), and this, I think, complicates the analysis. The first stage of data collection took place in the homes of families, where we prepared food and in most cases ate together. In the analysis of this stage, I looked for themes explaining families' notions of good feeding/eating and how they organised their resources to enact and transfer these routines to their children. In the second stage, we went out food shopping together and talked about their preferences as I observed their food provisioning routines. Thus, each disposition that I identified in the analysis corresponded to a different set of practices in each stage, related to the provision, preparation, or emotional management of food work. As the example below shows, the practice extracted from the first time point to demonstrate a theme is usually related to preparation and cooking, whereas for the second time point the examples come from the shopping experience (i.e. provisioning). However, all examples are linked to the theme identified and show the trajectory of the dispositions/practices.

Thomson (2007) suggests that there are two aspects of longitudinal qualitative data analysis: the first is the longitudinal aspect of individual cases and the second is cross-sectional differences of structural context. She argues that in the analysis, researchers should develop case histories and then bring them into conversation with each other by comparing their various themes. Since I am interested in how social class shapes foodwork/feeding work in families, I decided to adjust the matrix to help me see the second aspect

that Thompson refers to: cross-sectional variations. To achieve this, I decided to colour-code each entry to indicate the level of economic and cultural resources of the family interviewed (e.g. green indicates that the family has high cultural capital/moderate economic capital). At the end of the three stages, the matrix will not only demonstrate how events unfolded for each individual family but also how similar processes are lived by families from different social classes.

Like in the case of all QLR, the amount of data that requires appropriate structuring is challenging, but I believe that a systematic trajectory analysis, supported by cross-sectional comparisons and a reflexive approach, will generate rich and insightful analysis.

Themes	Stage 1	Stage 2	Stage 3
Theme 1: Socialising the child to an organic and environmentally friendly diet	Family X grows veggies in the garden and teaches their children where food comes from and how it is recycled.	Family X seemed stressed during shopping due to the economic cost of an environmentally friendly diet. They have become vegetarian since the last stage. They choose their children's food from the organic range and buy regular products for themselves to balance the budget.	
Theme 2: Expanding a cultivated taste palette	Family Y uses temporal and economic resources intensely to introduce new dishes and snacks. In Family Z, the mother compares herself with 'perfect' mums in Facebook feeding groups and aims to imitate their style to deal with her 'lack of creativity'.	Family Y's child starts nursery. Lots of stress about the new food habits the child learns at school. Food practices at home begin to change: they buy meat only once a week to balance the high salt/high protein in school meals. Mother in Family Z returned to work in the meantime. Works long hours. Shares cooking duties with her husband and her in-laws now. Unfollowed the Facebook groups to manage stress. Baby is now old enough to eat whatever the family eats, which seems to reduce the stress involved in the acquisition and preparation of food.	

Table 1 Sample family matrix

References

Thomson R. (2007) The qualitative longitudinal case history: practical, methodological and ethical reflections. *Social Policy and Society* 6(4): 571-582.

Grossoehme D. and Lipstein E. (2016) Analyzing longitudinal qualitative data: the application of trajectory and recurrent cross-sectional approaches. *BMC Research Notes* 9(1): 136.

Analysing young people's experiences of coping with problems, difficult situations and feelings: An evolving approach to analysing qualitative longitudinal evaluation data

Emily Stapley, Anna Freud: National Centre for Children and Families, UK.

This blog draws on a five-year qualitative longitudinal (QLR) study. The work is part of the evaluation of HeadStart; a five-year, £56 million National Lottery funded programme set up by The National Lottery Community Fund to explore and test new ways to improve the mental health and wellbeing of young people aged 10 to 16 and prevent serious mental health issues from developing. Six local-authority-led partnerships in Blackpool, Cornwall, Hull, Kent, Newham and Wolverhampton are working with local young people, schools, families, charities, and community and public services to make young people's mental health and wellbeing everybody's business.

The aim of our study is to explore young people's experiences of coping with and receiving support for problems and difficult feelings or situations over a five-year period. The young people invited to take part in our study were those who were already receiving support from HeadStart or those who might do so in the future. Participants were in Years 5 or 7 at school (age 9 to 12) at the start of the study and (we hope!) will continue to be involved until they are in Years 9 or 11 (age 14 to 16). Working with two colleagues in the Evidence Based Practice Unit (EBPU) at the Anna Freud National Centre for Children and Families and UCL, and at the University of Manchester (both of whom are PhD students), we are conducting semi-structured interviews once a year with approximately 80 young people (10 to 15 at each HeadStart partnership).

I decided to conduct a cross-sectional thematic analysis of the interviews in the first year, drawing on Braun and Clarke's (2006) methodology. This decision was made in the context of the fact that:

1. We were working with such a large dataset (82 interviews);
2. We had always intended to present the themes arising across the dataset in the first year of the project, as a baseline report for the study as a whole (see Stapley and Deighton 2018).

We took a team approach, using the qualitative data analysis software package NVivo (v11) to facilitate our analysis of the wave 1 dataset. As part of this process, I initially developed a thematic framework relating to our research questions by coding 80 per cent of the interview transcripts. This involved giving brief labels to the extracts of the interview transcripts that related to our research questions, which described the content of the extracts, and then grouping all extracts with similar labels or codes together to form themes. The other two members of our team then each coded the remaining 20 per cent of the transcripts using my thematic framework. This resulted in refinements and additions being made where necessary to the thematic framework.

At the outset of the study, we made a pragmatic decision to analyse the data drawing on the interviews across the HeadStart partnerships, rather than to conduct individual pieces of partnership-specific analysis. This speaks to our remit as the HeadStart Learning Team responsible for the national evaluation of the programme, whereas site-specific qualitative data collection and analysis is being conducted locally by the individual partnerships. However, we did explore which themes from our analysis described above could be seen specifically in the interviews from each partnership (i.e. across all of the interviews in a given partnership, which themes from our thematic framework were present and which were not?). There was relatively little variation between the partnerships, in terms of the themes from our thematic framework that could be seen specifically in their interviews. Ultimately, any decision to bring together the national and locally-collected qualitative datasets will be influenced by the degree of heterogeneity in our aims/research questions, our capacity, and the instigation of appropriate data sharing agreements. Following our initial analysis of the wave 1 dataset, we had a decision to make in the second year about how to conduct diachronic analysis across waves 1 and 2. Sources such as Grossoehme and Lipstein (2016)

have been helpful in thinking about this. We are currently planning to use typology methods, such as ideal-type analysis, to explore the patterns or 'types' evident in the young people's experiences and perspectives, and the potential shift in this across the two years. For instance, do the young people (individually and in general across the sample) exhibit different patterns of coping behaviour and support use in the second year of the study, as compared to the first year, and why? What are the mechanisms or factors behind changes in the young people's wellbeing across the first and second years of the study? The ideal-type analysis process typically begins by the researcher(s) writing a 'case reconstruction' of each interview, in our case a summary of the content of each transcript. These case reconstructions are then systematically compared with each other by the researcher(s), which leads to the formation of a number of broadly similar groups of case reconstructions or, in other words, interviews representing similar types of experience (e.g. Stapley et al. 2017).

We are now about to go into wave 3, our third year of data collection, and are really looking forward to seeing our participants again, as they grow older and have new experiences, opinions and perspectives. The growing size of the dataset as we accumulate more interviews each year means that establishing clear baselines in our analysis at the outset of the study will be important to direct our focus over the course of the study. At this early stage, I would envisage our analytic approach evolving over time, depending on the findings from our analysis at each wave and the topics raised by the young people during data collection.

References
Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77-101.
Grossoehme, D. and Lipstein, E. (2016) Analyzing longitudinal qualitative data: The application of trajectory and recurrent cross-sectional approaches. *BMC Research Notes, 9*, 1-5.
Stapley, E. and Deighton, J. (2018) *HeadStart Year 1: National Qualitative Evaluation Findings - Young People's Perspectives.* London: CAMHS Press.
Stapley, E., Target, M., and Midgley, N. (2017) The journey through and beyond mental health services in the United Kingdom: A typology of parents' ways of managing the crisis of their teenage child's depression. *Journal of Clinical Psychology, 73*, 1429-1441.

5. The potential of computer assistance in handling Big Qual data

Can a computer do qualitative analysis?
Daniel Turner, Quirkos software, UK.

It seems that everywhere we look researchers are applying machine learning (ML) and artificial intelligence (AI) to new fields. But what about qualitative analysis? Is there a potential for software to help a researcher in coding qualitative data and understanding emerging themes and trends from complex datasets?

Firstly, why would we want to do this? The power of qualitative research comes from uncovering the unexpected and unanticipated in complex issues that defy easy questions and answers. Quantitative research methods typically struggle with these kind of topics, and machine learning approaches are essentially quantitative methods of analysing qualitative data.

However, while machines may not be ready to take the place of a researcher in setting research questions and evaluating complex answers, there are areas that could benefit from a more automated approach. Qualitative analysis is time consuming and hence costly, and this greatly limits the situations in which it is utilised. If we could train a computer system to act as a guide or assistant for a qualitative researcher wading through very large, long or longitudinal qualitative data sets, it could open many doors.

Few qualitative research projects have the luxury of a secondary coder who can independently read, analyse and check interpretations of the data, but an automated tool could perform this function, giving some level of assurance and suggesting quotes or topics that might have been overlooked.

Qualitative research could use larger data sources if a tool could at least speed up the work of a human researcher. While we aim in qualitative research to focus on the small, often this means focusing on a very small population group or geographical area. With faster coding tools we could design qualitative research with the same resources that samples more diverse populations to see how universal or variable trends are.

It also could allow for secondary analysis: qualitative research generates huge amounts of deep detailed data that is typically only used to answer a small set of research questions. Using ML tools to explore existing qualitative data sets with new research questions could help to get increased value from archived and/or multiple sets of data.

I'm also very excited about the potential for including wider sources of qualitative data in research projects. While most researchers go straight to interviews or focus groups with respondents, analysing policy or media on the subject would help gain a better understanding of the culture and context around a research issue. Usually this work is too extensive to systematically include in academic projects, but could increase the applicability of research findings to setting policy and understanding media coverage on contentious issues.

With an interdisciplinary team from the University of Edinburgh, we performed experiments with current ML tools to see how feasible these approaches currently are. We tried three different types of qualitative data sets with conventional 'off-the-shelf' Natural Language Processing tools to try and do 'categorisation' tasks where researchers had already given the 'topics' or categories we wanted extracts on from the data. The software was tasked with assessing which sentences were relevant to each of the topics we defined. Even in the best performing approach there was only an agreement rate of ~20% compared to how the researchers had coded the data. However this was not far off the agreement rate of a second human

coder, who was not involved with the research project, did not know the research question, just the categories to code into. In this respect the researcher was put in the same situation as the computer.
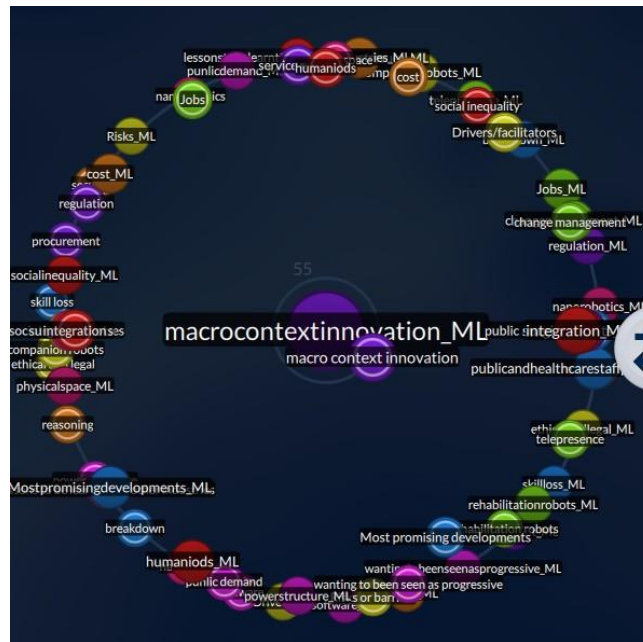

Figure 1: Visualisations in Quirkos allow the user to quickly see how well automated coding correlates with their own interpretations

The core challenge comes from the relatively small size of qualitative data sets. ML algorithms work best when they have thousands, or even millions of sources to identify patterns in. Typical qualitative research projects may only have a dozen or less sources, and so the approaches give generally weak results. However, the accuracy of the process could be improved, especially by pre-training the model with other related datasets.

There are also limitations to the way the ML approaches themselves work – for example there is no way at the moment to input the research questions into the software. While you can provide a coding framework of topics you are interested in (or get it to try and guess what the categories should be) you can't explain to the algorithm what your research questions are, and so what aspects of the data is interesting to you. ML might highlight how often your respondents talked about different flavours of ice cream, but if your interest is in healthy eating this may not be very helpful.

Finally, even when the ML is working well, it's very difficult to know why: ML typically doesn't create a human readable decision tree that would explain why it made each choice. In deep learning approaches, where the algorithm is self-training, the designers of the system can't see how works, creating a 'black box'. And this is problematic because we can't see their decision making process, and tell if a few unusual pieces of data are skewing the process, or if it is making basic mistakes like confusing the two different meanings of a word like 'mine'.

There is a potential here for a new field: one which meets the quantitative worlds of big data with the insight from qualitative questions. It's unlikely that these tools will remove the researcher and their primary role in analysis, and there will always be problems and questions that are best met with a purely manual qualitative approach. However, for the right research questions and data sets, it could open the door to new approaches and even more nuanced answers.

For further details please see: https://www.quirkos.com/index.html

Computer-assisted text analysis beyond words
Gregor Wiedemann, University of Hamburg, Germany.

In our digital era, amounts of textual data are growing rapidly. Unlike traditional data acquisition in qualitative analysis, such as conducting interviews, texts from (online) news articles, user commentaries or social network posts are usually not generated directly for the purpose of research. This huge pool of new data provides interesting material for analysis, but it also poses qualitative research with the challenge to open up to new methods.

Despite the long history of computer-assisted text analysis, it has stayed a parallel development with only little interaction with qualitative analysis. Methods of lexicometric analysis such as extraction of key words, collocations or frequency analysis usually operate on the level of single words. Unfortunately, as Benjamin Schmidt phrased it, "words are frustrating entities to study. Although higher order entities like concepts are all ultimately constituted through words, no word or group can easily stand in for any of them" (2012). Since qualitative studies are interested in the production of meaning, of what is said and how, there certainly are overlaps with lexicometric measures, but nonetheless their research subjects appear somewhat incompatible. Observation of words alone without respect to their local context appears as rough simplification compared to a hermeneutic close reading and interpretation of a text passage.

The field of natural language processing (NLP) from the discipline of computer science provides a huge variety of (semi-) automatic approaches for large scale text analysis, and has only slowly been discovered by social scientists and other qualitative researchers. Many of these text mining methods operate on semantics beyond the level of isolated words, and are therefore much more compatible with established methods of qualitative text analysis. Topic models, for instance, allow for automatic extraction of word and document clusters in large document collections (Blei 2012). Since topics represent measures of latent semantic meaning, they can be interpreted qualitatively and utilised for quantitative thematic analysis of document collections at the same time. Text classification as a method of supervised machine learning provides techniques even closer to established manual analysis approaches. It allows for automatic coding of documents, or parts of documents such as paragraphs, sentences or phrases on the basis of manually labelled training sets. The classifier learns features from hand coded text, where coding is realised analogously to conventional content analysis. The classifier model can be seen as a 'naïve coder' who has learned characteristics of language expressions representative for a specific interpretation of meaning of a text passage. This 'naïve coder' then is able to process and code thousands of new texts, which explicitly opens the qualitative analysis of categories up to quantification.

In my dissertation study on the discourse of democratic demarcation in Germany (Wiedemann 2016), I utilised methods of text mining in an integrated, systematic analysis on more than 600,000 newspaper documents covering a time period of more than six decades. Among others, I tracked categories of left-wing and right-wing demarcation in the public discourse over time. Categories were operationalised as sentences expressing demarcation against, or a demand for, exclusion of left-/right-wing political actors or ideologies from the legitimate political spectrum (e.g. "The fascist National Democratic Party needs to be banned" or "The communist protests in Berlin pose a serious threat to our democracy"). Using automatic text classification, I was able to measure the distribution of such qualitatively defined categories in different newspapers between 1950 and 2011. As an example, the following figure shows relative frequencies of documents containing demarcation statements in the German newspaper, the Frankfurter Allgemeine Zeitung (FAZ).

Democratric demarcation (FAZ)
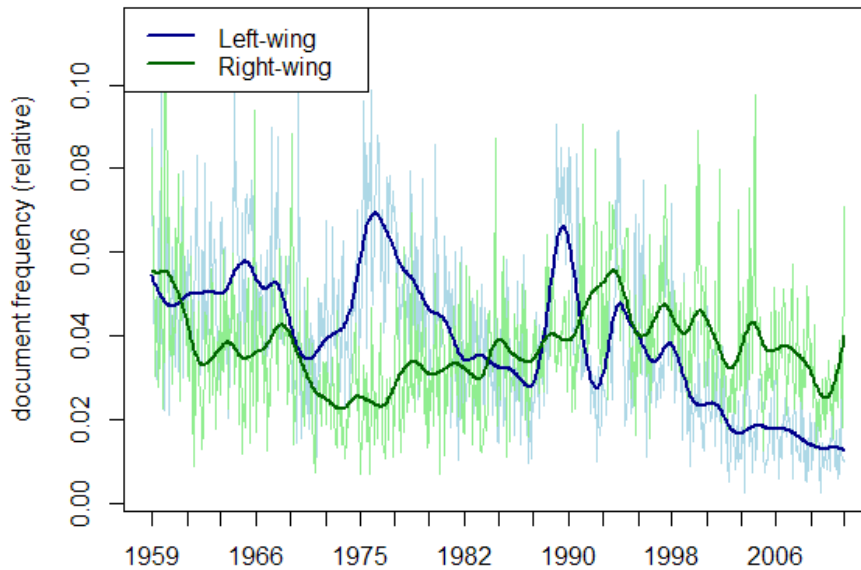
Distribution indicates that demarcation towards left-wing actors and ideology long-time superseded right-wing demarcation. Soon after 1990, the latter became the primary discourse subject of threats of German democracy. The enormous benefit of automatic classification is that it allows for easy comparison of publications (e.g. other newspapers) or relations with any other category. For instance, the distribution of "reassurance of democratic identity", a third category I measured, strongly correlates with right-wing demarcation, but not with left-wing demarcation. Such a finding can be realised only by a combination of the qualitative and the quantitative paradigm.

While computer-assisted methods support qualitative researchers clearly in their task of retrieving "what" is being said in large data sets, they certainly have limitations on the more interpretive task of reconstructing "how" something is said, i.e. the characterisation of how meaning is produced. It is an exciting future task of qualitative research to determine how nowadays state-of-the-art NLP methods may contribute to this requirement. In this respect, computational analysis extends the toolbox for qualitative researchers by complementing their well-established methods. They offer conventional approaches new chances for reproducible research designs and opportunities to open up to "big data" (Wiedemann 2013). Currently, actors in the emerging field of "data science" are a major driving force in computational textual analysis for social science related questions. Since I repeatedly observe lack of basic methodological and theoretical knowledge with respect to qualitative research in this field, I look forward to a closer interdisciplinary integration of them both.

Further reading

Blei, D.M. (2012) Probabilistic topic models: Surveying a suite of algorithms that offer a solution to managing large document archives. *Communications of the ACM*, 55(4): 77–84.
Schmidt, B.M. (2012) Words alone: dismantling topic models in the humanities. *Journal of Digital Humanities* 2(1). http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt
Wiedemann, G. (2013) Opening up to Big Data. Computer-Assisted Analysis of Textual Data in Social Sciences. *Historical Social Research,* 38(4): 332-357.
Wiedemann, G. (2016). Text Mining for Qualitative Data Analysis in the Social Sciences: A Study on Democratic Discourse in Germany. Wiesbaden: Springer VS, http://link.springer.com/book/10.1007%2F978-3-658-07224-7

Navigating the landscape of qualitative data in surveys with automated semantic analysis

Elena Zaitseva, Liverpool John Moores University, UK.

Reflecting on the quantitative-qualitative divide in large scale survey data almost twenty years ago, Bolden and Moscarola (2000) concluded that free text comments (e.g. responses to open questions in questionnaires) are 'poorly utilised, either being totally ignored, analysed non-systematically, or treated as an aside' (Bolden and Moscarola, 2000, p. 450). Two decades later and not much has changed. Examining thousands of fragmented open question responses, varying from a short phrase or sentence to mini-narratives or lengthy reflective accounts, remains a complex, time and resource consuming exercise for researchers. However, timely analysis of free text comments could help not only enhance understanding of quantitative results, but also reveal new discourses not necessarily anticipated by the survey's creators.

As part of a Higher Education Funding Council for England (HEFCE) funded project on the 'Sophomore Slump' that investigated disengagement and underperformance of second year university students, we undertook a comparative analysis of comments provided in a student survey deployed at each level of study (comparing themes from year one, two and final year students' feedback) (Zaitseva et al. 2013). Each data set comprised, on average, 250 pages of text - single spaced Times New Roman 12 point font.

My search for a user-friendly instrument that would allow us to instantly see the whole institutional landscape of student feedback for each level of study, and be able to detect differences and drill down into the particular areas or topics, led me to Leximancer – a tool for visualising the conceptual and thematic structure of a text, developed at the University of Queensland (Smith and Humphreys 2006).

The software automatically identifies concepts, themes (clusters of concepts) and connections between them by data mining the text, and visually representing the findings in the form of a concept map – a process called *unsupervised semantic mapping of natural language*. Based on an assumption that a concept is characterised by words that tend to appear in conjunction with it, the software measures how relevant one word is to a set of other words. Only words that pass a certain relevance weight threshold, established by the software, form concepts, although this parameter can be manually adjusted (Fig 1).
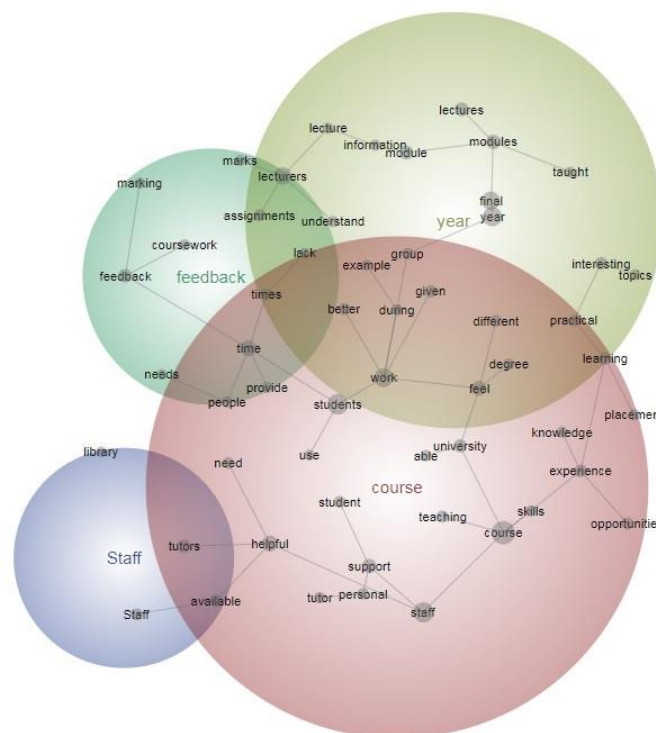

Figure 1. Example of a concept map generated by Leximancer

The tool not only determines the key concepts, themes and associated sentiments, but also provides useful information about the proximity of the concepts and their location. This is particularly beneficial for longitudinal and comparative analysis where underlying differences can be identified from the positioning of concepts on the map.

Although the 'mapping' process is completed automatically, making sense of the map and establishing meaning behind each concept is the researcher's task. The researcher has to 'dissect' the concepts and associated themes by exploring all instances (direct quotes) that contributed to the concept's creation, and undertake a more traditional interpretive/thematic analysis.

Using Leximancer in the 'Sophomore Slump' research helped uncover change in student attitudes and priorities as they progressed with their studies, showing how they moved from affectively oriented goals in the first year to a second year's learning and goal reaffirmation stage, and achievement and outcome-oriented learning in the final year.

Another research project where the capabilities of Leximancer were tested, was analysis of free text comments of postgraduate taught students at the sector level to identify the dominant themes within their feedback (Zaitseva and Milsom 2015). The Postgraduate Taught Experience Survey (PTES) dataset included responses of 67,580 students from 100 higher education institutions. The survey provided the opportunity to comment after each section (seven in total), and invited responses on the most enjoyable aspects as well as how the course experience could be improved. The overall data set comprised around 2,670,000 words which was the equivalent of 5933 pages (single spaced Times New Roman, 12 point font). An attempt to generate a concept map from a combined data set resulted in a densely populated map and thousands of quotes attached to each concept, so it was deemed unsuitable for analysis. The data had to be disaggregated by analysing responses from each section separately, and augmented by insights from the demographic data breakdown (e.g. looking at trends in responses of young and mature, part-time and full-time students) to be able to achieve at least some saturation in thematic exploration.

The analysis identified a number of new themes, including the heavy workload of part-time students which was often underrepresented in course-related marketing information, and its impact on student mental health and ability to achieve (Fig 2.); issues around 'levelness' of Masters programme delivery which, in some cases, was aimed at doctoral level and, in other cases, at final year undergraduate degree, and some other.
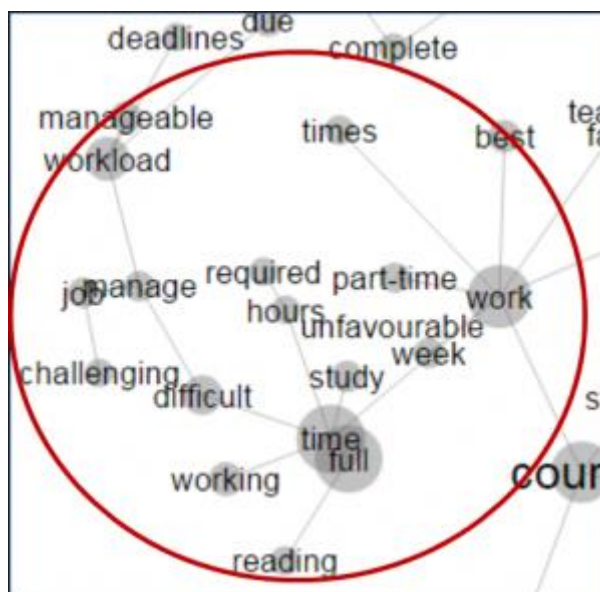

Figure 2. A fragment of part-time student experience concept map

Instruments such as Leximancer allow researchers to conduct analysis of large qualitative data sets in a time-efficient and consistent manner, as data pre-processing is done by computer. The concept map that emerges from this analysis captures 'the wisdom of crowds' (Dodgson et al. 2008) and is a text-driven, not a researcher-driven representation. But the researcher is able to interrogate the concept map and perform a more focused/tailored analysis by mining the text for 'deeper contextual associations' (Stewart and Chakraborty, 2010). The vaster the data source, the more nuanced the concept map will be.

Use of computer aided analysis increases reliability (as the top level of analysis is independent of the researcher's effect), and facilitates reproducibility of the findings as it is possible to retrace your thinking that may have contributed to the emergence of new ideas and research findings.

There are limitations to this type of analysis. Some concepts emerge strongly where they are represented by a narrow vocabulary. In the context of student surveys, words such as *lecture*, *library*, *feedback* or *exams* will have a strong presence on the concept maps. In contrast, other elements of student experience, such as *personal development* or *extracurricular activities*, will be identified from a broader pool of terms and have a greater likelihood of being diluted as a concept in the map. This can be mitigated by undertaking a tailored analysis, for example, through concept seeding, by adding concepts that have not passed publication threshold, but are of interest to the researcher.

Some concepts are relatively fixed in their meaning, while others are very broad. For instance, the concept *tutorial* is most likely to represent a single meaning in student feedback. At the same time, the concept *work*, being noun as well as a verb, might have multiple meanings. To fine-tune the analysis, more specific queries should be run to better understand all connotations related to the concept (e.g. group + work, part-time+ work).

Sentiment analysis needs to be occasionally verified by checking contextual understanding, but Leximancer also mitigates this by providing both indicators (favourable and unfavourable probability).

Without any doubt there are limits to what the software analysis can achieve. Complexity of language implies that automated semantic analysis methods will never replace careful and close reading of the text, but 'computer assisted methods are best thought of as amplifying and augmenting careful reading and thoughtful analysis' (Grimmer and Stewart, 2013, p. 2). These methods are vital to handling large volumes of qualitative data that might otherwise go un-analysed.

References
Bolden, R. and Moscarola, J. (2000) Bridging the Quantitative-Qualitative Divide: The Lexical Approach to Textual Data Analysis, *Social Science Computer Review,* 18(4): 450-460.
Grimmer, J. and Stewart, B. (2013) Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts, *Political Analysis* Advance Access, 1-31, available online: https://web.stanford.edu/~jgrimmer/tad2.pdf
Smith, A. and Humpreys, M. (2006) Evaluation of Unsupervised Semantic Mapping of Natural Language with Leximancer Concept Mapping, *Behavioural Research Methods,* (38): 262–79
Stewart, G. and Chakraborty, A. (2010) Strategy Content Analysis for Service Identification: A Case Study on Government Agencies. *5th Conference on Qualitative Research in IT,* Brisbane, available online: https://researchonline.jcu.edu.au/25633/1/QUALIT10.pdf
Zaitseva, E., Milsom, C. and Stewart, M. (2013) Connecting the Dots: Using Concept Maps for Interpreting Student Satisfaction. *Quality in Higher Education*, 19(2): 225–47.
Zaitseva, E. and Milsom, C. (2015) *In their own words: Analysing students' comments from the Postgraduate Taught Experience Survey*, York: Higher Education Academy, available online: https://www.heacademy.ac.uk/knowledge-hub/postgraduate-taught-experience-survey-2015

Computational text analysis using R in Big Qual data: lessons from a feasibility study looking at care and intimacy

Emma Davidson, Justin Chun-ting Ho and Lynn Jamieson, University of Edinburgh, UK.

The use of computational text analysis has increased rapidly across the humanities and social sciences. Much of this growth has centred on taking advantage of the breadth of new digital sources of data and the rich qualitative material they provide. Despite this progress, the application of these approaches to qualitative methods in the social sciences remains in its infancy. Our project is one such endeavour (http://bigqlr.ncrm.ac.uk/). Together with colleagues - Professor Rosalind Edwards, Dr Susie Weller and Professor Lynn Jamieson - it involved secondary analysis of six of the core studies stored in the Timescapes Qualitative Longitudinal Data Archive. Using a substantive focus on practices of care and intimacy over time, and across the life course, we wanted to explore the methodological possibilities of working with large volumes of qualitative data. We also wanted to address the scepticism that 'scaling up' could damage the integrity of the qualitative research process.

The breadth-and-depth method

From the outset, our intention was to develop an approach, which integrated computer-assisted methods for analysing the *breadth* of large volumes of qualitative data, with more conventional methods of qualitative analysis that emphasise *depth*. This would – we hoped – take us away from the linearity implied by 'scaling up', towards an iterative and interpretative approach more akin to the epistemological position of the qualitative researcher. A discussion of our breath-and-depth method is detailed in Davidson et al. (2019). One of our first analytical steps was to 'pool' the data into a new assemblage classified by gender and generation-cohort. Too large to read or analyse using conventional qualitative research methods, we looked to computer-assisted methods to support our analysis. What we were seeking was a way of 'thematically' mapping the landscape of the data. Rather like an archaeologist undertaking geophysical surveying, we anticipated using this surface survey to detect 'themes' for further exploration. Once identified, these themes would be analysed using shallow test pit sampling, the aim of which is to ascertain if they are worthy of deeper investigation. We expected a recursive movement between the thematic mapping and the preliminary analysis. So, where a possible theme proved to be too ambiguous or tangential, it would be eliminated, followed by a return to the thematic mapping to try again. If a theme(s) relevance is confirmed, the move to in-depth interpretive analysis can be made.

Thematic mapping and computer-assisted text analysis

There are, of course, various ways of undertaking computer-assisted approach to thematic mapping. And as part of the project we experimented - more and less successfully - with various text analytics tools, including Leximancer, Wordsmith, AntConc and R. In each, we were broadly interested in text analysis, exploring for instance word frequencies, word proximity and co-location, conducting searching for words within pre-defined thematic clusters (for example, relating to performing practical acts of care and intimacy), as well as keyword analysis.

We wanted to explore R since it provided the ability to write the programming language ourselves and change the form of analysis according to any emergent results. This not only meant that we were in control of the programming steps, but also that these steps were transparent and understood. The limitation - of course - is that we were a team of researchers whose skills were primarily in qualitative data analysis! And while we were capable of undertaking statistical analysis, we had no prior experience of statistical programming languages, nor of natural language processing. It became clear that in order to proceed we didn't just need a skilled R user to produce the analysis for us, but a collaborator who could go on this journey with us. This proved a difficult task since the majority of those we approached were skilled in computational methods, yet were not familiar or sufficiently interested to collaborate in a project where the focus was on qualitative research methods. This reluctance perhaps reflects the tendency for qualitative methods to use small-scale and intensive approaches which focus on the micro-level of social interactions. Computational scientists, conversely, have focused on big data to understand social phenomena at an

aggregate level. By seeking to bring these skills together, our study presented possible collaborators not only with an unfamiliar form of data, but also an unfamiliar approach.

Using R to analyse Big Qual data

We were - eventually - lucky enough to recruit Justin Chun-ting Ho to the project, a doctoral candidate from the University of Edinburgh and in collaboration we developed a plan for the proposed work. A priority was to conduct a comparative keyword analysis to identify 'keyness' by gender and generation cohort. We were also keen to 'seed' our own concepts by creating pre-defined thematic word clusters, and examining their relative frequency across the different categories of data. How does the frequency of positive emotion words, for example, compare between the youngest and oldest men?

Using keyword analysis, we were able to gain both general insights, as well as potential areas for further exploration. We found, for example, that relationship and emotion words occurred with a greater frequency amongst women, as did words related to everyday or practical acts of care, such as 'feed' and 'dress'. Conversely, we found that words relating to work and leisure activities were most common amongst men. Changes across the life-course were also noted, with family - predictably – becoming more salient feature of life for older generations. As an example, the figure below shows a comparison of the oldest and youngest women, and the shifting focus from friends to family.
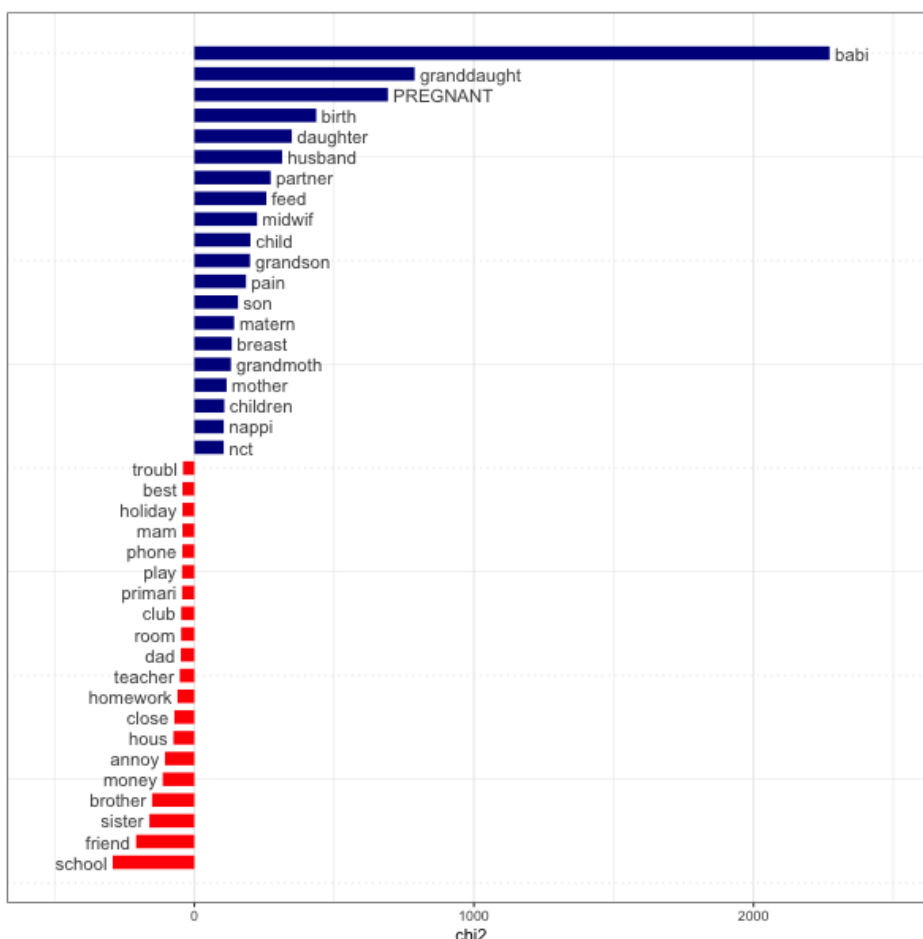

Figure 1: Comparative keyword analysis: pre-1950 (oldest) versus post 1990 (youngest) women

We were also, however, aware that the results reflected the complexity of speech itself (for example, the meaning of 'care'), while some concepts were structured strongly by individual projects (for example, the frequent use of siblings was to large extent a function this terms prevalence in one of the core Timescapes projects, rather than coming from naturally occurring speech). It also raised the question of the extent to which examples of care and intimacy were neglected due to the parameters used to define keyness - that is, we were looking at the keyness of *all* words, not just those words related to care and intimacy.

These reflections were themselves useful since it provides us an opportunity to critically evaluate the tacit theory underpinning our understandings of what constitutes practices of care and intimacy. Where we benefited from R was in its flexibility, since we were able to explore a range of alternate forms of analysis to integrate further. For example, we went on to manually identify care and intimacy keywords, and to combine them into thematic 'clusters' that share some characteristic (for example, conflict words, relationship work, words describing practical acts of care and words describing formal childcare). We then counted the frequency of words from each cluster using R to show the thematic differences between interview transcripts of different genders and generations. In this way, we were able to augment human effort with the power of machine; qualitative analysis allowed us to identify the themes while computational techniques could show the prevalence of such themes within the corpus, which would otherwise be too big for qualitative analysis. This thematic analysis, in turn, provided further outputs, which identified specific themes (including 'love' and 'arguments' for exploration through shallow test pit analysis, see Davidson et al. 2019).

Reflections and moving forward

The project, overall, has shown that text analytics can provide a unique opportunity for qualitative researchers seeking to interrogate large volumes of qualitative data. We concur with Gregor Wiedemann's contribution in this collection that these methods can extend and complement the qualitative researchers' toolkit. Our work with R has provided tangible benefits, and crucially supports the breadth-and-depth approach developed by the project. However, unlike pre-programmed and commercially available software such as Leximancer, R requires a certain level of competency in statistical programming language – and crucially the time and resources to invest in developing these skills. It is perhaps for this reason that our analysis ultimately relied on Leximancer, and its accessible, user-friendly interface.

Qualitative researchers are not unique - many social scientists, regardless of methodological orientation, do not have these skills. Yet given the rise of big data and possibilities it offers the social sciences, the value of text analytics is likely to grow – as will demand for these skills. To bridge this chasm, investment is needed in training, capacity building and opportunities for multi-disciplinary working. The method developed through our project aims to form such a bridge across disciplines. However, it also revealed the importance of developing text analytic skills directly into bids for funding – either through a named collaborator equally invested in the project outcomes, or sufficient resources for the training and development of the team. Looking forward, we anticipate with excitement the collaborative opportunities that the 'big data' era presents to qualitative researchers.

References
Davidson, E., Edwards, R., Jamieson, L. and Weller, S. (2019) Big data, Qualitative style: a breadth-and-depth method for working with large amounts of secondary qualitative data. *Quality and Quantity*. 53(363): 363-376.

6. Collaboration in Big Qual analysis

Collaborating with original research teams: Some reflections on good secondary analytic practice

Susie Weller, University of Southampton, UK.

With colleagues, I have been conducting secondary analysis across six of the core studies housed in the Timescapes Qualitative Longitudinal Data Archive (https://timescapes-archive.leeds.ac.uk/). The Timescapes project sought to scale up qualitative longitudinal (QLR) work. It was a five-year study comprising a set of empirical projects documenting change and continuity in identities and relationships over the lifecourse. The initiative also pioneered new approaches to archiving and re-using QLR data. Seven teams from five Higher Education Institutions in the UK conducted the original studies. As a secondary analysis team, we came to these data sets not just as secondary analysts, but also primary researchers. I conducted one of the Timescapes studies – Siblings and Friends (http://www.timescapes.leeds.ac.uk/research/siblings-friends/index.html) – with Rosalind Edwards, and Lynn Jamieson was part of the Work and Family Lives (http://www.timescapes.leeds.ac.uk/research/work-family-lives.html) project. Not only did this connection help us understand better the origins of the data, but it also facilitated relationships with the original researchers.

Having been heavily invested in our own QLR studies, we were mindful of the very particular nature of the long-term connection between researchers and participants. Our perception was that even though the original teams had archived their data for the purpose of re-use, we ought, in our negotiations about the secondary analysis of their material, to be sensitive about such long-term connections and the emotional investment made by the researchers. For us, our initial ideas about good secondary analytic practice involved developing approaches to sustained collaboration with the original researchers. Of course, some secondary analysts might regard the engagement of primary researchers as an interference, instead viewing the data as embodying new knowledge or alternative insights, which do not require the explicit involvement of the original researchers. Our approach was guided by a duty of care, and was shaped by our own understandings of the temporal and emotional investment involved in QLR.

With these concerns in mind, we contacted former Timescapes colleagues at the outset to inform them of the purpose of our study and our plans to use their archived material. In the early stages, we liaised with individuals via email, asking project-specific questions about, for instance, the research context, data set structure and their own analysis. Whilst our intention was to be inclusive, in practice we liaised with only one or two members of the original team; those with whom we had strong professional relationships. Later, we took a more formalised approach inviting members of the original teams to complete an online consultation with questions asking them about their changing connection to the data, feelings and concerns about data sharing and re-use, and the forms of consultation or connection (if any) they would consider appropriate/valuable. We received responses from all teams over varying timescales, some of whom have contributed to this collection.

Most of the responses were from the researchers who had produced the data. Their willingness to contribute to our work on good practice in qualitative secondary analysis may be regarded as acts of cooperation and we have relied heavily on the goodwill of these colleagues, some of whom we have known for many years. In 2017, with NCRM colleagues Melanie Nind and Sarah Lewthwaite, we were awarded additional funding to build capacity and develop resources for the teaching of our new breath-and-depth approach to Big Qual analysis. This opportunity enabled us to work more closely with some of our former colleagues through action-oriented training events. We have since shared details of the resources produced via our final correspondence with the original teams.

We soon came to realise that, whilst our initial ideal was to foster sustained collaboration, this was not something that the original researchers necessarily wanted, expected or could accommodate. Some had

left academia for new ventures, or were not available. Others had developed different interests and had moved on from their Timescapes work. Few were still using their own project material. Our consultation revealed that of the 19 who responded to a question about their connection to the data, seven explicitly stated that their attachment had declined over time (one person reported having never felt any connection). Furthermore, of the 14 who replied to a question asking their opinion on appropriate levels of contact between original researchers and secondary analysts, three did not want any contact at all. Conversely, our engagement with material from studies other than our own gave us a greater (and growing) sense of connection to the broader Timescapes collection.

Whilst original team members may wish to collaborate they may not have the time or funds to do so. Yet, it may well be junior/field researchers who are best placed to enlighten secondary analysts on the minutiae of a project. We were, however, concerned that sustained collaboration, which relies largely on the goodwill of colleagues, could result in exploitation. It is important to acknowledge the hidden labour involved in such collaborations and to think through the possibilities for formalising the process to some degree. This could involve a variety of options from acknowledging the investments of data generators in project outputs through to developing joint ventures, or incorporating willing original researchers in grant design and budgets. This might be particularly appealing for fixed-term contract researchers.

That said our consultation showed that some of our Timescapes colleagues felt increasingly detached from 'their' archived data over time, whereas we became more attached to it. We merged data from the six projects into one assemblage organising the material by gender and cohort-generation. This was a time-consuming process and we engaged with the data over the course of four years, albeit on a part-time basis. The labour we invested in this process meant that we became attached to it as our production, thereby shifting our perception of ownership. Indeed, we are currently in the process of preparing our data assemblage for deposit in the Timescapes Archive as a teaching data set. Archiving and data re-use implies that the knowledge production has not ended. Secondary analysis disrupts usual understandings of collaboration introducing it as emergent, iterative, unexpected.

Facebook timelines: Young people's growing up narratives online

Sian Lincoln, Independent researcher.
Brady Robards, Monash University, Australia.

In 2014 Facebook celebrated its tenth birthday. To mark this first decade, we edited a special issue of *New Media & Society* that reflected on the extent to which the site had become embedded into the everyday lives of its users. It was also evident at this point that there was now a generation of young people who had literally 'grown up' using the site. This prompted us to design a new research project, and a new research method in the process. Facebook Timelines is a qualitative study with young people in their twenties who joined the site in their early teens. We were particularly interested in this age group because they had used Facebook throughout their teens and many found themselves at a 'crossroads' moment in their life when they are beginning to think seriously about post-education working life and 'professional identity'. Using a combination of qualitative interviewing, time-lining and the 'scroll back method', we worked with 40 young people to find out how they (and their friends) had disclosed their 'growing up' experiences on the site. In this respect, the Facebook Timeline (also known as the profile) was used as a 'prompt' and the years upon years of disclosures on the site acted as 'cues' for what often became elaborate and in-depth stories of teenage life.

One of our core interests here was how 'growing up' stories are recorded and made visible on social media. Given Facebook's longevity, it has become a digital archive of life for many – a longitudinal digital trace. We wanted to interrogate this further by working with our participants as co-analysts of their own digital traces. How do young people make sense of these longitudinal digital traces? How do these traces persist and re-surface, years later, as people grow up and enter into new stages of their lives?

Time-lining: going back to pencil and paper

What key or critical moments have you experienced in your teenage years, since joining Facebook? As a period of turbulence and change, we were keen to ask this question and explore what our participants perceived to be those important, life defining events or rites of passage that have come to define them. A simple print out of a timeline enabled our participants to consider this question and to map out those moments as they remember them. These included: going to high school, leaving school, getting a part time job, going to a first gig, family weddings, births and deaths, going into full time employment, going to university, the beginning and end of relationships and all manner of important moments. Our participants were then invited to log into their Facebook profile using a laptop, tablet or phone, depending on the participants' preference, to consider how the moments they recalled 'mapped onto' their Facebook Timeline.

The 'scroll back' method

At this point, our participants were asked to 'scroll back' to their very first post on the site. It is common for them to have an emotional response to early disclosures on the site; embarrassment being the most typical. For us, this was interesting because their response acted as the first 'marker of growing up' they encounter in the 'scroll back' and represented a form of self-reflexivity and self-realisation. In addition, their responses were physical: the covering of the eyes, a slight wince, even turning away from the screen when confronted with a younger self and evidence of their digital trace dating back some years. Consider a 24-year-old confronting their 16-year-old self, as mediated on Facebook. Once the 'scroll back' begins, participants click chronologically through their years of disclosures, opening up year after year of their Facebook archive provoking narration and description of the content. This method serves to be empowering for the participant as it places them in control of which moments they wish to talk about and which they do not; which to discuss and which to pass over. However, because of the sheer amount of content – much of which is forgotten (particularly the earlier stuff) – there is a danger that participants will be confronted with challenging, difficult moments from their past at which point the participant is asked whether they wish to continue. Often they do, seeing this as a 'therapeutic moment' to reflect with hindsight on the event. Some saw it as an important life moment and thus it remained in their Timeline.

Importantly, we recruited our participants not just as interviewees or 'subjects' of observation. We worked with our participants as co-analysts of their own digital traces. Having our participants sign-in to their own Facebook accounts and scroll back through their Timeline profiles in front of us allowed us to see their Facebook histories 'from their perspective'. If we were to analyse these digital traces without the involvement of the participants themselves, we'd be limited in multiple ways: first, in terms of what we could actually see, but second – and for us, more importantly – in terms of the stories that certain disclosures prompted. Often, our participants would be able to 'fill in the blanks' or provide crucial context and explanation for in-jokes, vague status updates, or obscure images that we alone would have had little capacity to fully understand. Thus, our analysis here really hinged on the involvement and insight of our participants themselves.

'Scroll back' and narratives of growing up

The Facebook Timelines project has clearly under-lined the significance of Facebook in the lives of young people in their twenties as a key platform for sharing their everyday life experiences. While some participants claim to be 'partial' Facebook users today amidst broader claims of 'Facebook fatigue' and a more complicated 'polymedia' environment including Instagram, Snapchat, dating and hook-up apps, and so on, scrolling back through participants' Timelines has affirmed just how embedded and central Facebook is in their lives. Further, their changes in use from 'intense' to more silent (but still present) 'disuse' tells us much about their growing up and claims to being 'more mature' equating to disclosing less. Additionally, the amount of 'memory work' the site is doing on their behalf (so many forgotten moments were unveiled through 'scroll back') makes getting rid of Facebook for good almost an impossibility.

Facebook Timelines offer immense opportunities for longitudinal researchers, however the depth of many profiles certainly presents analytical challenges as essentially these are not profiles that have been created for a research project. For us, and as we mention above, 'analysis' of the Timelines was embedded into the scroll back method from the start with participants analyzing their own digital traces as a core part of the research process. Drawing on Thomson and Holland (2003) we then considered the data 'cross-sectionally in order to identify discourses through which identity is constructed, and longitudinally at the development of a particular narrative over time' (2003: 236). We did this with the participants as they scrolled back, then cross-referenced these discourses with other participants by analyzing the interview transcripts using the themes defined by our participants (for example, relationships, travel and education). Overall, we felt this approach gave our participants a genuine feeling that they had witnessed, unfolded and given voice to, a self-narrative of their growing up on Facebook.

Related publications

In this article, we expand on the 'scroll back' method in much more detail:
- Robards, B and Lincoln, S. (2017) Uncovering Longitudinal life narratives: Scrolling back on Facebook, *Qualitative Research*. http://journals.sagepub.com/doi/abs/10.1177/1468794117700707

These publications report on findings from our study:
- Lincoln, S. and Robards, B. (2016) Editing the project of the self: Sustained Facebook use and growing up online in *Journal of Youth Studies*. http://www.tandfonline.com/doi/full/10.1080/13676261.2016.1241869
- Robards, R. and Lincoln, S. (2016) Making it "Facebook Official": Reflecting on Romantic Relationships Through Sustained Facebook Use in Social Media + Society (special issue: Making Digital Cultures). http://journals.sagepub.com/doi/full/10.1177/2056305116672890

Further background to our research:
- Lincoln, S. and Robards, B. (2015) Being strategic and taking control: Bedrooms, social network sites and the narratives of growing up, *New Media & Society*, 18(6): 927-943: http://journals.sagepub.com/doi/abs/10.1177/1461444814554065

- Lincoln, S. and Robards, B. (2014) Editorial: Ten Years of Facebook, Special Issue: Ten Years of Facebook in *New Media & Society*, 16(7): 1047-1067.
http://journals.sagepub.com/doi/full/10.1177/1461444814543994

References

Thomson, R. and Holland, R. (2003) Hindsight, foresight and insight: the challenges of longitudinal qualitative research. *International Journal of Social Research Methodology*, 6(3): 233-244.

The challenges of computer assisted data analysis for distributed research teams working on large qualitative projects

Rebecca Taylor, University of Southampton, UK.

Academics, like many other groups of workers in the digital economy, often find themselves working in geographically distributed teams spanning multiple locations connected by increasingly sophisticated digital technologies. Teleconferencing tools like Skype; cloud based file storage/hosting services such as Google docs and Dropbox; and project planning tools such as Trello, enable groups of researchers to meet, talk, write, share and edit documents, plan, manage and conduct research and even analyse data despite their separate locations.

If you are a researcher involved in large scale qualitative studies, such as qualitative longitudinal research (QLR), where projects can potentially span decades and short-term contracts mean that researchers move between institutions, it is highly likely that you will, at some point, be operating in a distributed research team working across institutions, geographical locations and maybe even time zones. QLR in particular tends to amplify the challenges and opportunities of other qualitative methodologies (see e.g. Thomson and Holland 2003); the difficulties of managing multiple cases over multiple waves in terms of storage, labelling and retrieval are even more demanding when carried out remotely. In fact any large data set creates challenges for a distributed team. Providing access to data across institutions necessitates organising access rights and often the use of a VPN (Virtual Personal Network). Cloud based collaboration solutions may lack institutional technical support and the required level of data security raising legal and ethical problems for the storage of non-anonymised transcripts, observation notes and other documents.

These issues are all in play when it comes to analysing a geographically-distributed team's data. The overwhelming array of CAQDAS (Computer Assisted Qualitative Data Analysis Software) packages offer multiple functionality for managing and manipulating qualitative data but are less helpful when it comes to facilitating distributed team working. Our recent experiences as a research team spread across two institutions with members also working mainly from home, provides a useful case study of the issues. As we looked at the CAQDAS packages currently available it became apparent that our options were dependent on where the software was situated – locally, institutionally, or in the cloud:

Option A: Working locally

This traditional model involved packages (such as NVivo, MAX Q) uploaded onto individual computers so that all team members worked on their own local version of the project. For the team to work together on the data and see everyone's coding and new transcripts, required that researchers all send their projects to a team member who would merge them together and redistribute a new master copy of the project. In a distributed team, this meant finding a way to regularly transfer large project files safely, securely and easily between team members with all the attendant hazards of version control and file management. The size of project files and the security issues around cloud based storage ruled out the more straightforward options like email or Dropbox and the remote desktop route made any sort of data transfer brain numbingly complicated because there was no way to move documents between the home computer and the remote desktop. We had one option for data transfer – a University of Southampton download service for large files which used high levels of encryption.

Option B: Working institutionally

This model made use of server-based packages which stored the data centrally such NVivo Server ('NVivo for Teams' with V11) enabling team members to work on the project simultaneously using an institutional local area network (LAN). In the case of NVivo Server this mitigated the need for a regular time consuming merge process. However, for those members of the team at other institutions or not working on campus it required using remote desktop solutions which were slow and unwieldy and made file transfers (for example when importing a new transcript into the software) difficult. We worried about this process given the software's reputation for stability issues when used with a potentially intermittent

network connection. More importantly, it required a different type of Institutional software licence which was an expense we had not budgeted for and implied considerable delay as we negotiated with the university about purchase and technical support.

Option C: Working in the cloud

Thinking more creatively about the problem we looked at online (and thus not institutionally located) packages such as US-based Dedoose (try saying that with an American accent – it makes more sense) designed to facilitate team-based qualitative and mixed methods data analysis. We could, it seemed, all work online on the same project from any PC or laptop in any location without the need to merge or transfer projects and documents – Were all our problems solved? Sadly not. Consultation with IT services in our own institutions revealed that such sites used cloud storage in the US and were therefore deemed insecure – we would be compromising our data security and thus our ethical contract. So we were back to square one or in our case Option A – the old school model; a laborious and time-consuming (but ultimately secure) way of working; individual projects on our individual desktops with regular or not so regular transfers and merges.

It's worked Ok – we are now writing our third journal article. Yet as the funding ended and we lost our brilliant Research Fellow to another short term contract we have tended towards more individualised analysis, the merge process has largely fizzled out as no one has time to manage it and the software serves primarily as a data management tool. It is clear that in the contemporary HE landscape of intensification, and metricisation of research, the tools for distributed team working need to be super-effective and easy to use; they need to make collaborative qualitative analysis straightforward and rewarding irrespective of the geographical location of individual team members. Distributed working arrangements are certainly not going away.

References

Thomson, R. and Holland, J. (2003) Hindsight, foresight and insight: The challenges of qualitative longitudinal research, *International Journal of Social Research Methodology*, 6(3): 233-244.

Time, technology and documentation

Rachel Thomson, Sara Bragg and Liam Berriman, University of Sussex, UK.

There is a tradition within qualitative longitudinal research of returning to earlier studies building on the places, people or data sets of earlier research. In some disciplines this kind of iterative practice is well established, for example the long term ethnography in anthropology where generations of scholars 'pass the mantle' of responsibility for tracking the historical evolution of a community. Within sociology we talk of 'revisits' that can take the form of re-engaging with the methods/data or sites of earlier studies and earlier research selves if revisiting our own work. These kinds of reflexive contemplations have the potential to historicise social research practice, helping us to see how our research questions, methods and technologies are part and parcel of the knowledge economies that we as researchers are part of, and how these change over time. In general terms, designing time into a research process has enormous potential for making things visible in new ways, including the contingent modes of production of social research.

So paradoxically, by holding certain things constant, temporal methods have the capacity to help us notice change. For example following the same participant over time reveals all kinds of transformations but also a consolidation of something that in retrospect we understand as always having been there. Repeating a method over time has a similar analytic dividend providing a bridge to consider relations of sameness and mutability, difference and repetition. Generations within a family, institution or a society can also be thought of through the same prism – enabling us to tease apart biographical and historical time, life stages (such as early career, or young adulthood) and contexts (post- Brexit austerity). Designing generations into social research increases the power and the complexity of any investigation.

Our new book *Researching Everyday Childhoods* is a culmination of several threads of methodological development in the field of qualitative longitudinal research. The project focuses on children and young people and what it is like to live and grow in a culture that is saturated by digital technology. It is also a book about what it means for researchers to operate in the same environment, recognising how our practice is transformed by new tools and changing relationships of expertise and authority. The book is a mediation on a shift from analog to digital knowledge that encompasses all of the actors involved: the researchers, the participants, the funders, the audiences, the publishers, the data. This is achieved by anchoring the empirical project to our own pasts – the seven year old children in the study are the yet to be born babies in our earlier intergenerational study of new motherhood. The researchers following them have known their families for almost a decade and this 'back-story' forms part of the relationship and data shadow for their cases. We have also adapted methods first trialled in the motherhood study: a day in a life, object based conversations and 'recursive interviews' where fragments of data and analysis from the research relationship are represented and responded to in the present.

Yet the study also brings in the new in a deliberate way. New participants in the form of a panel of teenagers, and new researchers bringing fresh perspectives, research questions and skills into the team. Importantly the project has sought to address the limits of our earlier research.

This includes the idea of starting rather than ending with the archive. Where previously we had promised confidentiality and anonymity as a condition of the research, in this project we invited participants to work collaboratively with us to co-produce a publically accessible archive. The practice of 'curation' is as important to us as 'data generation' and we are aware that professional social researchers no longer have a monopoly over such knowledge practices and the resulting knowledge relations. Working in collaboration with the Mass Observation Archive and our participant families we have created a new multi-media collection as well as an open access online interface – something that has involved us entering the archive itself, exploring what it means to become data, to be available for unknown audiences and unforeseen modes of secondary analysis. Thinking through what is the same and what might be different, we move more deeply into an era of digital data in which notions of indelibility, anonymity and trust change their character. We cannot confidently make promises about a future that we are yet to apprehend. We can

however engage in the analytic and ethical labour necessary to ensure that we are thinking together in a way that is transparent, reflexive and accountable. Our book Researching Everyday Childhoods: Time, Technology and Documentation in a Digital Age does just that. We are pleased that it is also open access, meaning that along with the public archive it may be used as a resource for teaching and collaboration.

Working in collaboration to develop the teaching of Big Qual analysis

Sarah Lewthwaite, University of Southampton, UK.

Research teams increasingly collaborate across complex divides. Working in geographically distributed, interdisciplinary and cross-functional teams can be challenging – particularly in areas of methodological innovation, such as Big Qual. Added to this, the impetus to build research capacity in cutting-edge methods can mean research teams become teaching teams. Collaborating as a teaching team adds complexity, in several key areas.

Traditionally, research methods teaching has lacked 'pedagogical culture', with an absence of resources, research and discursive material that methods teachers can draw upon to develop teaching. This matters because methods are pedagogically distinctive in the social sciences. Learners require theoretical understanding, procedural knowledge and technical skill (Kilburn, Nind and Wiles 2014), as well as an ability to put forward a method whilst simultaneously subjecting that method to sustained scrutiny (Bourdieu 1992). Methods education can also be characterised by a focus on teaching with and through data (Lewthwaite and Nind 2016). Such requirements demand certain pedagogic responses – fostering reflexivity, learning by doing, and so forth. Experiential learning has been cited as the 'signature' pedagogy of qualitative research; however, when conducting research with archives, 'experience' and notions of the 'field' are redefined. This gestures to particular 'pedagogic content knowledge' (or PCK), (Shulman 1986) - the pedagogic specificity - of working with archives and Big Qual analysis, amongst qualitative methods. Collaborating to develop PCK for Big Qual analysis from scratch is a challenge. Whilst acknowledging that context, learners, and different modes of teaching all impact on PCK, to begin to answer this challenge, we found the following steps useful in facilitating joint working.

1. Develop shared pedagogic language

Advanced research methods are frequently taught by content experts; researchers who may not have a background in education. As a result, talk about pedagogy may not come easily. To facilitate conversations, we worked with the Big Qual team (http://bigqlr.ncrm.ac.uk/) to develop a 2-page glossary of pedagogic terms (Lewthwaite and Nind 2018), offering definitions of salient pedagogies with which to work. Beginning these conversations, teams may find that they have already invested methodological language with pedagogy, in methodological writing, conference presentations and seminars. With tools for dialogue, such implicit pedagogic knowledge can be more readily made explicit. These are verdant starting points for teaching teams.

2. Sequence content

The Big Qual team employed a metaphor for a 'breadth-and-depth' method for Big Qual analysis (Davidson, et al. 2019) dividing the method into four steps. This sequenced approach provided a useful framework both for the ordering and chunking of content in class, and the division of labour for the teaching team, in planning and delivery. Importantly, in practice this raised three key issues. First, the necessity of stressing the whole of the method – and maintaining a logical, iterative thread that connects across the steps (e.g. the use of a worked example across the piece) – so a method isn't reduced to its constitute parts. Second, orientating students within this framework, so they can understand at any given point where they are in the relation to the overview. Third, the importance of step-by-step annotated lesson plans. These detailed who was responsible for what, the timing and the nature of delivery at every stage. In a distributed team, where physical planning meetings are difficult, annotated lesson plans, and the sharing of presentation slides and notes, handouts and materials (linked below), were crucial to the team as a whole for grasping what would happen and when. As a shared teacher-resource, the lesson plan could then be developed after teaching, on the basis of team reflection and student feedback, to see where improvements could be made.

3. Pedagogic dialogue and reflection

The sharing of materials gestures to how teaching might be done, but does not address potential pedagogic conflict amongst individuals within a team. Pedagogy evokes values and approaches, as well as discrete

actions. To this end, it is useful to discuss as a team underlying assumptions concerning what the teaching will convey to learners. How and why are the team invested in these methods, or particular ways of teaching it? What is the team trying to articulate when they articulate the method? Will teaching be student-centred and dialogic? Will it call upon learner expertise? To this end, dialogue and reflection on teaching is essential to the development of coherent team-teaching. Innovative methods frequently rely upon incremental advance rather than revolution (Wiles, Bengry-Howell, Crow and Nind 2013), so drawing upon prior experience and teaching resources can offer a useful way into teaching, but these must be (re)purposed effectively to the task at hand. Using cycles of planning, action and reflection helps to develop teaching. Learning from each other (building a local pedagogical culture), through discussion, is essential. In our work, we proposed a typology of pedagogy for methodological learning, to facilitate discussion and draw out implicit and unreflected knowledge. This encourages teachers to reflect upon their teaching approaches, strategies, tactics and tasks (moving from an *approach* – how a teacher goes about their pedagogic work in a way that coheres around a theory, principles or a set of values to the operational, *task* level – what it is learners are required to do. See Nind & Lewthwaite, f/c). By attending to values in both pedagogy and method, teams are better equipped to address sticky questions. For example, when teaching with secondary data, particularly teaching and learning challenges are raised. Archives can be challenging for learners, being built predominantly for archiving – rather than teaching or learning. When getting learners 'hands-on' with an archive, should learners be able to generate or apply their own (authentic) search terms to the archive? Or should teachers supply a tried-and-tested route through Search? A learner-generated search approach may be more engaging, being authentically connected to a learners research interests. However, the search may not return any data. This is an authentic lesson in the potential frustrations of archival research, but it may disengage students from the method at an early stage. Alternatively, teacher-guided search can ensure students can access and navigate data, but without offering a 'teachable moment' regarding the difficulty of archival search. By considering team values – such sticky issues can be evaluated for more informed pedagogic decision-making. Is authentic and experimental learning foremost? Or is modelling, exposition and demonstration paramount at an early stage? Do the team want to prioritise student-centred, or teacher-led approaches? How and when should these change?

This is one of the pedagogic issues specific to Big Qual analysis that will arise in teaching (another example gravitates around learner diversity – how to bridge the divergent qualitative and quantitative understandings). However, by using active and reflexive approaches to pedagogic development in dialogue, as a team, team-teaching can be hugely beneficial. Come together to debrief after teaching. Collect meaningful student feedback for team reflection. Look for ways to smooth transitions between teachers, and broker more communal pedagogic content knowledge. Feeding into and out of this process is an impetus to share your approaches, strategies, tactics and tasks with peers and wider teaching networks. We have sought to do this with the *Teaching Big Qual Analysis: Innovation in method and pedagogy* project. In this way, pedagogical culture can be built, to sustain methodological developments and build a resource base that wider publics can benefit from.

For other teaching resources stemming from the Big Qual Analysis – Innovation in Method and Pedagogy project, please visit: https://www.ncrm.ac.uk/resources/online/teaching_big_qual

References

Bourdieu, P. (1992) "The Practice of Reflexive Sociology (The Paris Workshop)", in *An Invitation to Reflexive Sociology*, edited by P. Bourdieu and L. Wacquant, 217–260. Chicago: University of Chicago Press.
Davidson, E., Edwards, R., Jamieson, L. and Weller, S. (2019) Big data, Qualitative style: a breadth-and-depth method for working with large amounts of secondary qualitative data. *Quality and Quantity*. 53(1): 363-376.
Kilburn, D., Nind, M., and Wiles, R.A. (2014) "Learning as Researchers and Teachers: The Development of a Pedagogical Culture for Social Science Research Methods?." *British Journal of Educational Studies*. 62(2): 191–207.
Lewthwaite S. and Nind, M. (2016) Teaching Research Methods in the Social Sciences: Expert Perspectives on Pedagogy and Practice, *British Journal of Educational Studies*. 64(4): 413-430.

Lewthwaite, S. and Nind, M. (2018) <u>A glossary for methods teaching</u> – *NCRM quick start guide.* Manual. NCRM.

Nind, M. and Lewthwaite, S. (f/c) A conceptual-empirical typology of social science research methods pedagogy. *Research Papers in Education.*

Shulman, L. (1986) <u>Those who understand: knowledge growth in teaching</u>, *Educational Researcher.* 15(2): 4–14.

Wiles, R., Bengry-Howell, A., Crow, G. and Nind, M. (2013) <u>But is it innovation? The development of novel methodological approaches in qualitative research.</u> *Methodological Innovations Online.* 8(1): 18-33.

Other resources

Lewthwaite, S., Jamieson, L., Weller, S., Edwards, R. and Nind, M.(2019) *Teaching how to analyse large volumes of secondary qualitative data.* National Centre for Research Methods online learning resource. Available at https://www.ncrm.ac.uk/resources/online/teaching_big_qual/

7. Ethical considerations

The ethics of secondary data analysis – respecting communities in research

Ginny Morrow, University College London, UK.

For the past 10 years, I have been involved with Young Lives (https://www.younglives.org.uk/), a longitudinal study of children growing up in Ethiopia, India, Peru and Vietnam, which has been an amazing experience and a great privilege. As well as being Deputy Director since 2011, I have been 'embedded' in Young Lives as the ethics lead – though it is vital that ethics are not the responsibility of one person, but shared across the whole team.

Young Lives encounters all kinds of ethics questions and dilemmas, and for this guest blog, I have been asked to explore the ethics of secondary data analysis. Arguments about the promises and pitfalls of archiving (qualitative) data are well-rehearsed, as outlined in discussions by Natasha Mauthner and others.

A few years ago, as an ESRC-funded National Centre for Research Methods node (2011-14), Young Lives qualitative research team had a very productive and enjoyable collaboration with colleagues at TCRU in London and Sussex, Family Lives and Environments, as part of Novella (Narratives of Varied Everyday Lives and Linked Approaches), in which Young Lives qualitative data formed the basis for narrative and thematic analysis of children's and their families relationships to the environment in India (Andhra Pradesh) and England (see Catharine Walker's thesis, http://www.novella.ac.uk/about/1056.html). Based on our experiences, we produced a working paper exploring the ethics of sharing qualitative data– and we identified a number of challenges, which we hope have helped other researchers as they grapple with the demands of sharing data.

We argued that sharing data and undertaking secondary analysis can take many forms, and bring many benefits. But it can be ethically complex. One of the considerations that we discussed was responsibilities to participants and to the original researchers, and the need to achieve a contextual understanding of the data by identifying and countering risks of misinterpretation. We highlighted the importance of developing and maintaining trusting relationships between research participants, primary and secondary researchers.

Novella involved a team of qualitative researchers, and we did not fully discuss the ethics of secondary analysis of survey data, bar touching on questions of informed consent. But one of the questions that I've long been concerned about, based on experiences at Young Lives of seeing research based on our publically-archived survey data being used in ways very far from the intentions of our study (which is to explore childhood poverty over time), is the following: how do the people we study and write about, feel about the interpretation and use we make of their data? Might they object to how their data are used, and how they are represented in research findings and other media dissemination?

So I was fascinated to learn about the EU-funded project, entitled TRUST, that has led to the generation of the San Code of Research Ethics, launched by the South African San Institute a couple of weeks ago (this video - https://www.youtube.com/watch?v=HOdw3mv7JSo - gives a great insight to the project).

The San Code of Ethics calls for respect, honesty, justice and fairness, and care – and asks that the San Council, which represents the San Community, is involved in research from inception, design, through to approval of the project, and subsequent publications. The San are not the only indigenous people to create codes of ethics demanding they are fairly respected in research, and the impetus for this initiative has come from genomics research, but the points about respect are relevant for all research. Two points are worthy of much more attention in research ethics:

> 1. Failure by researchers to meet their promises to provide feedback, which the San Council say they have encountered frequently, and which they see as an example of disrespect; and

2. 'A lack of honesty in many instances in the past. Researchers have deviated from the stated purpose of research, failed to honour a promise to show the San the research prior to publication, and published a biased paper based upon leading questions given to young San trainees'

The technicalities of all of this may be challenging, but demand our attention, so that open, honest, and continuous communication can take place, and the hurt caused by lack of justice, fairness and respect can be avoided in the future.

References
Mauthner, N.S. (2016) Should data sharing be regulated? In A. Hamilton and W.C. van den Hoonaard (eds.) *The Ethics Rupture: Exploring alternatives to formal research-ethics review*. University of Toronto Press, pp. 206-229.

Following families
Jane Millar and Tess Ridge, University of Bath, UK.

Our longitudinal qualitative research started about 15 years ago, with a project to explore the impact of moving into paid work on the lives of lone mothers and their children. This was a very hot policy topic at the time, with a major policy drive to increase lone-parent employment.

Our sample consisted of 50 lone mothers who left Income Support for paid work in 2001/2002. We interviewed the mothers and their children first in 2004, and then again in 2005 and 2007. We have published a number of reports and articles, looking at various aspects of the move into work and employment sustainability, see our project webpage - The family work project: earning and caring in low-income households (http://www.bath.ac.uk/casp/projects/current/workwelfare/FamilyWorkProject.html).

In 2016 we returned to 15 families, chosen to reflect the range of family and employment experiences and circumstances[1]. The long-term nature of the study has provided a unique insight into how these families managed work and care through some very challenging economic times.

Every longitudinal study starts at a particular point in time, and from the conceptual and methodological decisions and priorities at that time. Such decisions have implications throughout the project, and beyond. Here we discuss two factors that affected the question of data archiving and re-use: how we found the sample, and our family-based theoretical approach.

For the sample, we were interested in exploring the transition into work, and what helped and what hindered. So we wanted to interview lone mothers who recently started working. We found our sample through the (as it was then) Inland Revenue and the Department for Work and Pensions, who agreed to draw a sample to our specifications. These were that the women should have at least one child aged 8 to 14, been receiving Income Support, started work and received tax credits between October 2002 and October 2003, and lived in specified areas of the country (see our DWP-published report on the first three rounds - https://lx.iriss.org.uk/sites/default/files/resources/rrep536.pdf). This gave us a very well specified and robust sample. But one of the conditions was that we should not share the data, even in anonymised form, due to concerns about confidentiality and privacy. So, we agreed not to place the transcripts in a data archive, or make available in other ways. Data archiving had not been a condition of the funding for the project, so there were no issues there.

Times have changed, and the general view now is in support of open access to all sorts of research data, with a growing literature on the issues and challenges of this in respect of qualitative research. We do agree this is important and are very interested in the way that the 'Working across qualitative longitudinal studies' (http://bigqlr.ncrm.ac.uk/) project is taking this forward. Understanding and practice have developed much beyond where we were fifteen years ago. As Bren Neale discusses in her blog (see section 1), the debate has moved on: 'a concern with whether or not qualitative datasets should be used is giving way to a more productive concern with how they should be used, not least, how best to work with their inherent temporality'.

Still, in some ways we are relieved not to have had to address the issues of how to anonymise our material in ways that would enable further analysis that could be properly grounded in the actual interview content. This difficulty would have been compounded for us by a key feature of our research design, which was that we interviewed both the mothers and the children.

Our initial starting point was that the move into paid work, and then trying to sustain work over time, was something that involved the children as well as the mothers. The children's lives would change, and they would have their own perspectives and experiences. In order to explore this 'family-work-project' as we called it, we needed to interview both the mothers and the children.

We did find that the mothers and children shared a commitment to the family-work-project and that this was a key factor in enabling the women to sustain work. But in analysing the interviews, and presenting the management of work and care as a family issue, we were also very aware of the importance of maintaining within-family privacy and confidentiality. The family-work-project sometimes involved painful adjustments and compromises and, as time passed, some of the ambivalence came more to the fore.

For some participants, the challenges of managing family life with low and insecure incomes over many years did come at a heavy cost to family relationships, at certain points in time. From the first round, the interviews had been carried out separately with the mothers and the children. And we took the decision to analyse and present these separately as well, in order to maintain within-family privacy. Thus in our articles and reports, particularly those using all waves of the data, we have focused on the mothers and the children separately. Where, for example, we wanted to discuss how the mother responded to her child's situation and decisions we did so without directly identifying the link to the child's account. But we did, of course, know that link ourselves. And we are not sure it would have been possible to anonymise the transcripts to ensure such protection and keep that separation between the accounts of family members. We struggled with this ourselves, and so are very aware of the challenges. In making the data anonymous, there would, we think, inevitably have to be some loss of the overall family perspective.

Developing approaches to informed consent that can recognise the family perspective in the analysis of the data would therefore be useful. However, it is not always the case that a longitudinal study is funded over several waves with one funder, and in our case we sought funding as we progressed and the study developed. This was demanding, but at the time no funding would have been available for three or more waves of research. Getting and maintaining informed consent over time is particularly challenging and requires considerable ethical rigour to ensure that participants – families in this case – do not have an 'obligation' to continue in the study and are aware that their data may be used elsewhere.

Using, and re-using, longitudinal qualitative data from interviews is an ongoing process that is far from straightforward. It is important to be aware of potential issues in the design of the research, insofar as possible. But issues and tensions also emerge during the course of the research, and these cannot always be anticipated.

Visual approaches in QLR
Fiona Shirani, Cardiff University, UK.

Both qualitative longitudinal research (QLR) and visual methods have seen a surge in interest in recent years, yet relatively little attention has been given to the intersection of the two. This is perhaps surprising given many of the arguments in support of visual or multimodal methods take on a particular resonance in the context of QLR. In this short blog, I draw on my experience of designing and undertaking visual activities in two QLR projects - Energy Biographies (http://energybiographies.org/) and Timescapes (http://www.timescapes.leeds.ac.uk/) - to highlight some relevant issues.

Methodological innovation is an important element of research, and visual approaches have been key to enhancing creativity in qualitative work. QLR provides greater scope for methodological innovation and experimentation due to the extended timescales and flexible nature of the approach. For example, an activity that would take up too much time for a one-off study represents a smaller proportion of a QLR project. There is also time and space for reflection between interviews, giving the researcher an opportunity to hone and adapt activities for later waves of data collection.

Beyond methodological innovation, an advantage of incorporating a range of activities is the potential to make the research experience more engaging for participants. QLR requires a significant commitment from participants and maintaining the sample over time is an important concern. Whilst some participants enjoy the format of a qualitative interview, others may relish the opportunity to direct conversation through a photo-elicitation exercise, for example. Activities can also be conducted between interviews, serving as opportunities to both maintain contact and collect further data.

In both Timescapes and Energy Biographies we used visual activities to encourage participants to talk across extended time frames, thinking about their past memories and anticipated futures. Thinking temporally is a particular concern of QLR research yet can prove challenging for participants. Having a tangible reference point in the form of a visual representation can help anchor discussions.

Alongside these benefits, consideration must be given to some of the challenges arising from using multimodal approaches in QLR. Most notably, whilst accumulation of information about the individual is a strength of QLR, it also raises issues around anonymity and confidentiality. Adding visual data further complicates this issue and requires careful thought about how visual artefacts produced during the research should be analysed and presented, as well as challenges for archiving and data re-use.

In our Timescapes and Energy Biographies projects we have primarily used images as a means of eliciting talk; therefore analysis has focused on pictures and their accompanying text. We have found that narratives often go beyond what is represented in the image and therefore have argued the importance of attending to both talk and text. However, there are many possibilities for the analysis and creative presentation of multimodal data, and participant-generated images also inspired some elements of our public engagement exhibition. We also explored the possibility of working with images alone during a multimodal workshop with academic colleagues, where feedback indicated that the everyday nature of the images made them accessible for people to imbue with their own interpretations. However, there are clearly important ethical issues to consider in presenting images without the contextual information of their production and asking others to engage creatively with them.

Visual or multimodal approaches have much to offer QLR and combining the two could provide multiple benefits to future research.

For further details see: Shirani, F. and Henwood, K. (2010) Continuity and change in a qualitative longitudinal study of fatherhood: relevance without responsibility, *International Journal of Social Research Methodology*, 14(1): 17-29.