JISC

Project Information			
Project Identifier	*To be completed by JISC*		
Project Title	Leeds RoaDMaP (Leeds Research Data Management Pilot)		
Project Hashtag	#leedsrdm		
Start Date	01/01/2012	**End Date**	30/06/2013
Lead Institution	University of Leeds		
Project Director	Brian Clifford, Deputy University Librarian and Head of Learning and Research Support, b.e.clifford@leeds.ac.uk		
Project Manager	Rachel Proudfoot,		
Contact email	r.e.proudfoot@leeds.ac.uk		
Partner Institutions	Digital Curation Centre, F5, National Instruments		
Project Webpage URL	http://library.leeds.ac.uk/roadmap-project		
Programme Name	JISC Managing Research Data Programme 2011-13		
Programme Manager	Simon Hodson		

Document Information			
Author(s)	Rachel Proudfoot		
Project Role(s)	Project Manager		
Title	RoaDMaP Project: Timescapes Case Study Report		
Reporting Period	*From 01-01-2012 to 30-06-2012*		
Date	12/06/2013	**Filename**	Timescapes_case_study_report_v1
URL	*If this report is on your project web site*		
Access	☐ Project and JISC internal	☒ General dissemination	

Document History		
Version	**Date**	**Comments**
0.1	22/03/2013	Draft for circulation to case study lead
0.2	15/04/2013	Revision circulated to other case studies for comment
0.3	21/05/2013	Revised with input from case study lead
1.0	12/06/2013	Final revision

CONTENTS

RoaDMaP Project: Timescapes Case Study Report

1. Project Background

1.1 Case study lead

Professor Bren Neale is the Director of the Timescapes ESRC Qualitative Longitudinal study and archive and Professor of Life course and Family Research at the University of Leeds. Professor Neale is the Timescapes case study lead for the RoaDMaP project and a member of the University's Research Data Working Group.

1.2 Timescapes

Timescapes was funded from 2007-12 under the ESRC Changing Lives and Times initiative. Nine projects investigated how personal and family relationships develop and change over time, focusing on relationships with significant others: parents, grandparents, siblings, children, partners, friends and lovers.[1] One of the projects was designed to showcase the production of case history material for archiving and another showcased secondary analysis of the empirical datasets. The Timescapes programme created a specialist archive of Qualitative Longitudinal (QL) social science research data for sharing and re-use. "QL data, which is gathered over time through in depth interviews and ethnographic methods, explores the lived experience of change and continuity in the social world and gives insights into how and why micro processes unfold."[2]

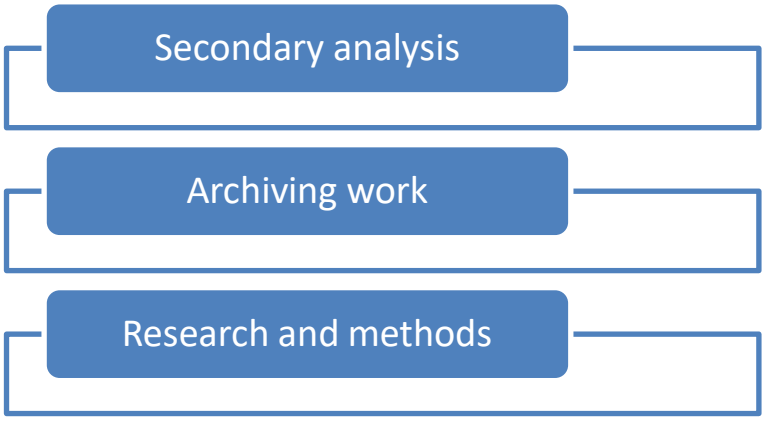


Figure 1: ***Timescapes was an innovative and complex programme, generating original research, building an archive to house it and demonstrating re-use of the data.***

Timescapes staff have extensive experience of managing research data, including developing plans, policies, training, metadata specifications and technical infrastructure. The projects also worked closely with the UK Data Archive[3] and were able to draw on UKDA's expertise in wide range of data management areas such as metadata standards, the ethics of data sharing and creation of data management plans. As a partner in the

[1] Timescapes web site http://www.timescapes.leeds.ac.uk/about/ [accessed 15 Mar 2013]

[2] Bren Neale (2013) *Timescapes: Archiving and sharing Qualitative Longitudinal data* [blog post] http://blog.library.leeds.ac.uk/blog/roadmap/post/132 [accessed 21 Mar 2013]

[3] UK Data Archive http://data-archive.ac.uk/

RoaDMaP project, Timescapes offered a rich vein of experience and an opportunity to capture reflections on data re-use in practice.

1.3 Timescapes data archive

The University Library created a digital repository (LUDOS – Leeds University Digital ObjectS) to house a variety of digital material not catered for by its open access publication repository (White Rose Research Online). LUDOS, which ran on the DigiTool platform from ExLibris, was launched in 2007. At the time, the Library was keen to work closely with Timescapes to demonstrate that LUDOS could support research data and thus help build the business case for the long term development of the service. From a Timescapes perspective, using institutionally supported infrastructure promised more sustainability within the institution than a specially commissioned repository for the data.

1.4 The case for the Timescapes data archive

The Timescapes Archive complements the UK Data Archive (which holds the largest collection of digital data in the social sciences and humanities in the United Kingdom). When Timescapes was funded in 2007 there were few QL datasets available for re-use and one of Timescapes' aims was to address "the discernible gap between the growing ethos of re-use and the actual practices of qualitative researchers" (Neale and Bishop 2012 p.55). The Timescapes Archive has evolved to meet the particular needs of QL data and manage it in ways which facilitate both data deposit and data re-use. Within the Timescapes Archive, "data files can be searched for thematic content, using conceptual and descriptive keyword searching at case level. Such refined search and retrieval can facilitate a more complex and nuanced discovery of temporal data than is usually the case in generic data facilities." (ibid. p.54)

The Timescapes Archive offers managed access to the data it holds and also facilitates contact between the researchers generating the data and potential secondary users, in this way "as well as the more usual indirect sharing of data via the archive, Timescapes enables a collaborative and personalized mode of data sharing " (ibid. p56).The Stakeholder Model is explored in Section 8.

2. Methodology

2.1 Interviews

A number of interviewees were identified in consultation with Professor Bren Neale to investigate the nature and challenges of Timescapes research data and reflect on lessons learnt from the creation and potential migration of the Timescapes data archive. Interviewees included researchers, administrators and Library staff. Interviewees are listed in the Acknowledgements section of this report. Unless indicated otherwise, the quotations in text boxes or speech bubbles in this report are taken directly from our RoaDMaP interviews.

2.2 Two elements of the case study

It became clear from preliminary discussions with the Timescapes researchers and Library staff that RoaDMaP might help progress an immediate data management issue – the need to establish a long term, sustainable platform for the Timescapes data. In 2012, the Library decided to migrate its digital library (LUDOS) from DigiTool to the open source EPrints repository platform; the Timescapes Archive was not included in the migration by mutual consent as, at the time, it was planned that the Timescapes data would move to a separate repository run on the Fedora digital repository system. Unfortunately, the migration to Fedora did not occur, largely due to cessation of project funding and changes in project personnel. Given the circumstances, we thought RoaDMaP could take the opportunity to capture reflections from Library staff

directly involved with the Timescapes project and explore whether the Library's recent experience of migrating the Leeds Digital Library from DigiTool to EPrints could help inform Timescapes' data migration options. TheTimescapes case study therefore expanded to consider both:

(i) Timescapes' research data management practice and lessons learnt

and

(ii) The longer term future for the Timescapes data archive.

2.3 RoaDMaP work packages

As planned, Timescapes interacted with RoaDMaP primarily through 5 work packages:

- **Requirements analysis (WP2)**: several interviews and discussions took place; we also utilised the research data methods literature and lessons learnt from the Timescapes project.
- **Institutional RDM policies (WP3):** the implications of the University's RDM policy, including who might be involved in implementation and how to promote the policy, was discussed with interviewees.
- **Data management planning (WP4):** Professor Neale completed a Data Management Plan for a Timescapes follow on project.
- **Repositories and metadata (WP5)**: the pre-existing Timescapes data archive provided an unusual and valuable opportunity to look at the design of a data repository and to consider requirements for migrating the data to a new repository platform; this has been explored through interviews and meetings between Timescapes researchers and Library staff.
- **Training (WP7)**: training requirements were gathered from the interviewees; some illustrative materials were sourced from Timescapes for our training sessions (e.g. data management plans).

3. Data sets

The Timescapes Archive holds qualitative longitudinal (QL) research, investigating people's subjective experience of their lives and following them over time.

Data in the archive includes:

- Textual data e.g. interview transcripts

Box 1: Photographs

"Photos taken by subjects as part of our project are very ordinary – but fascinating. Part of me thinks they should be kept and archived, another part that they've served their purpose – and they undermine the concept of anonymity.
"We might say researchers can have sight of but not publish the photos. It seems a shame to embargo them indefinitely."

- MaxQDA[4] coded transcriptsPhotographs
- Audio recordings
- Field diaries
- Self-portraits
- Relational maps (see *Figure 2*)
- Timelines
- Video diaries

[4] Proprietary software for organising, coding and visualising qualitative data

Box 2: Audio files of interviews

"Often researchers want to go back to the audio interview, hearing the sound of the voice, because you can pick up on such a lot through that route which you might lose in a transcript."

Figure 2: *Example Relational Map (from the Young Lives and Times project[5])*

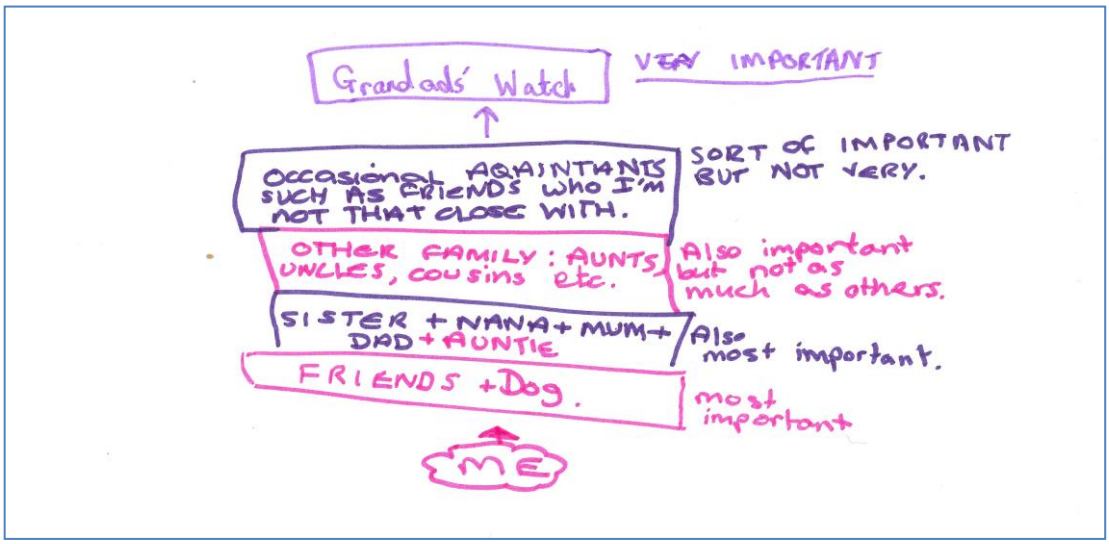


Table 1: ***Timescapes data archive file formats***

File extension	.rtf	.jpg	.xml	.pdf	.doc	.mp3	.jp2	.jpeg	.tif	.wma
% files	45%	24%	13%	10%	7%	<1%	<1%	<1%	<1%	<1%

Total size of the Timescapes Archive: 11 GB[6].

The archive stores eight social science datasets from the original Timescapes projects plus data from a network of affiliated projects, comprising around 3,000 files. To put this in perspective, the Timescapes End of Award Report (2012 p.31) suggests a "conservative estimate is that this represents a fifth of the number of qualitative data files currently held in the UK Data Archive. This is a substantial achievement given the complexities and challenges of archiving QL data, and the relatively short time frame for development to date."

A small collection of non-digital material has been archived with Special Collections in the University of Leeds Library.

[5] Young Lives and Times Exhibition Board http://www.timescapes.leeds.ac.uk/assets/files/FSS-Event/project-2.pdf

[6] Some data are still being processed for accession. Decisions as to where to deposit some audio, image and video files are still under review, due to their large size. (Timescapes ESRC End Of Award Report p.16)

4. Key players

A number of key players were identified during the RoaDMaP project – we gathered input from the first four categories below via interviews.

- Timescapes projects PIs and researchers
- Archiving staff
- Library staff providing technical support
- Non-Timescapes projects (affiliated projects) providing data for the archive
- Re-users of the data from across academia and policy organisations
- Interviews also revealed the importance of Faculty based research support staff – particularly those advising research at the pre-award stage.

5. Established Practice and Challenges

5.1 ESRC's data deposit requirements

"The ESRC requires all research data arising from ESRC-funded research to be made available to the scientific community in a timely and responsible manner unless there are exceptional reasons why this cannot happen. It is, therefore, the responsibility of the grant holder to formally offer any data created or repurposed during the lifetime of the grant to the UKDS [UK Data Service[7]] within three months of the end of the grant."

The ESRC recognises that not all data can be made openly available: "while we strongly advocate further re-use of data, we also recognise that some research data are more sensitive than others, and believe that it is the responsibility of the grant holder to consider all issues related to confidentiality, ethics, security and copyright before initiating the research."[8]

The Timescapes funding application to ESRC made the case that the Timescapes Archive would act as a working archive for the data and a "specialist satellite" (Neale 2007) of the UK Data Archive. The relationship between the Timescapes Archive and the UK Data Service is explored in paragraph 1.4 above and in Box: 8 *Institutional data repositories and subject repositories*.

5.2 Data Management Planning

The researchers we interviewed were aware of ESRC's requirements for research data management and archiving, for example, that a detailed data management plan is required, addressing the whole research data lifecycle. However, interviewees suggested that data management plans are not routinely created by other research colleagues. Indeed, one interviewee suggested there may still be an expectation amongst some researchers that after the final report and any publications have been written, data should be shredded.

[7] UK Data Service is managed by UK Data Archive

[8] http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx

For Timescapes, managing and preserving data is not an end in itself - it is data **re-use** which is the goal of the exercise. By the nature of the projects within the Timescapes programme, the Timescapes researchers had to consider a range of data management issues. This is not to say that all issues were anticipated or solved through initial data management planning, some issues, particularly around data sharing, had to be negotiated later in the project.

To support the Timescapes projects – and other QL researchers – a rich resource of Methods Guides has been created[9], including *Data Management Planning in QL Research* which emphasises that "since the QL research process is cumulative and cyclical, rather than linear, the tasks of QL data management are also cumulative and cyclical."[10] The report analyses data management planning activities, suggesting some lend themselves to a centralised approach whilst others may be more devolved (see Box 4). The Methods Guides also address use of the Timescapes Archive, secondary analysis of data and the ethics of data sharing and re-use.

Box 3: Support for data management planning

"Every time we do a data management plan, we're looking to see what we should be doing, should we be updating our knowledge? It's a continual process of updating. I've found it really useful with the last two bids I put in..to work with our IT Manager in our Faculty who has checked over that plan, made suggestions and so on. He has a format within which we can write the plan. It's good to feel you have that infrastructural support there - so you're not working in isolation and that what you're coming up with as a plan fits the institution. We need more of that."

Box 4: Centralised and devolved data management planning

"Some activities, such as security and file naming, benefit greatly from high levels of standardisation and thus a more centralised model; others, such as consent procedures, must be customised to specific projects and therefore are best decentralised (even if, for example, a standard consent form is used as an initial template)… However, finding the right balance is more art than science, and, like the doing of QL research itself, remains a continuous and always provisional undertaking." (Bishop and Neale 2012b).

5.3 Known challenges or issues

The QL data is challenging from a data archiving and re-use perspective. As one interviewee put it:

> *"What Timescapes did was take on every single complication you could throw at qualitative data and say 'can we archive it and can we do secondary analysis on it?'. So this material is personally sensitive, it's about families and personal relationships and it's longitudinal and so it doesn't 'close' naturally – when do you let go of a longitudinal project?"*

[9] Publications and Outputs, Timescapes web site http://www.timescapes.leeds.ac.uk/resources/publications

[10] Bishop, L. and Neale, B. (2012) *Data Management Planning in QL Research*. http://www.timescapes.leeds.ac.uk/resources/publications [accessed 1 March 2013]]

Challenges included:

Data management

- Managing multiple file formats – including deciding which data formats to keep and balancing this with anonymity requirements (for example see Box 1)
- File versioning (e.g. raw and anonymised versions)
- File naming conventions

Ethical considerations

- Balancing anonymisation with data integrity and meaning: the nature of the data means it can become 'over anonymised', losing important contextual elements (see Box x: Access to Research Data).
- Balancing anonymisation and retaining the authentic voice of the research participants. Some of the Timescapes researchers were from an oral history background and therefore particularly keen on "preserving some degree of 'fit' between the accounts given by participants, and the accounts produced by researchers" (Neale 2013b) . Participants "may value their voices going on record and contributing to documentary evidence about important social issues" (Neale 2013b).
- Ensuring participants cannot be indirectly identified – for example, by combining geographical and biographical data.
- Gaining participant consent, which is always provisional and subject to alteration over time. Issues include when to obtain consent for archiving (it may be better handled after the participants are more familiar with the researchers and the type of information they are gathering) and *how* to gain consent – for example, some researchers noted that a detailed, written consent form would not have been appropriate for use with their research subjects

Capturing metadata and context

- Agreeing appropriate metadata and contextual documentation
- Ensuring that tacit knowledge does not reside solely with researchers but is captured to enable re-use
- Deciding what material to share: for example, some interviewees noted difficulties in deciding whether to share their personal field notes as important contextualising documents.

When to share data

- QL studies can last for years, or even decades. Researchers can become heavily invested in the data, which may have many possibilities for primary analysis and 'reworking' over time, with no obvious closure or archiving point. A particularly interesting feature of the Timescapes approach is the 'stakeholder model' of data sharing and re-use which suggests a collaborative model to balance the requirements of research participants, primary and secondary researchers, data archivists and funders. As Neale suggests, archiving QL data is particularly complicated because "QL data are, unavoidably and inherently, always provisional, contingent and in process " (Neale 2013b).

The meaning of 'open access'

- The term 'open access' is a cause of concern for some of the researchers we interviewed. Whilst they are keen to support data sharing and re-use, fully open, registration-free access to data is not what is envisaged for the majority of content in the Timescapes Archive. Many of the researchers noted that consent had been obtained from participants on the understanding that future access would be restricted to 'bona fide' researchers. Similarly, the depositing researchers had agreed to make their data available on the understanding that access to the data would be controlled.
- The interviewees stressed very strongly that completely open access would make consent for archiving more difficult to obtain from participants and would result in a barrier to archiving data in their field.
- There was also some concern about research funder open access requirements and whether arrangements for controlled access to data fully met Funder expectations for data sharing.

Who prepares data for archiving?

- "The burden and responsibility for [managing data and producing data sets] may fall disproportionately on early career researchers whose contribution may go unrecognised" (Neale 2013b).

How data is presented for re-use

- One issue is how best to organise the data to facilitate analysis: for example, data is collected in successive 'waves' over time; within this there are individual 'cases' e.g. of individuals, organisations, groups; there will be underlying themes within and across the research data. To make sense of the data, it should be organisable by project, by wave, and by case to create a longitudinal case study, and by theme.

Resourcing and Sustainability

- Building in sufficient resource for data management and preparation tasks: some interviewees noted they had not built sufficient data management and preparation resource into previous bids. Estimating appropriate staff costs is difficult.
- A sustainable financial model for the Timescapes Archive which will provide support outside specific project timescales is required.
- As the DigiTool platform is no longer being supported by the Library, the Timescapes Archive needs to migrate to a new, supported repository platform.

Box 5: Controlled access to research data

The Timescapes Archive has a four tier access model:

- **Public** (openly available 'taster' data)
- **Registered access to anonymised data** (requestor must register for use and briefly state case for access; administrator grants access to 'bona fide' researchers for a specific period of time)
- **Restricted Access to sensitive or difficult to anonymise data** (access decisions are made on a case by case decision by the originating research team and, if granted, access is time limited)
- **Embargoed data** (the metadata and data are unavailable for a specified period of time)

One approach to making data more openly available is careful anonymisation. However, anonymisation may not always be an appropriate solution: for example, some research is based in a very specific geographic location and it would not be too difficult to identify individuals; one interviewee noted the risk that by removing names, places, and relationships "you have lost the very essence of what makes qualitative data qualitative data."

Several interviewees stated access to some data should be restricted to 'bona fide researchers': Professor Neale suggested this term covers "professional researchers employed at educational , academic or policy related institutions who are likely therefore to be bound by professional ethics when using data". Researchers are concerned about inappropriate exploitation of their data and may have gained consent from participants on the basis the results would only be available to other academic researchers. One researcher felt they might restrict the types of questions they would ask their subjects if the data was to be made more openly available, another that openness might lead to self-censorship in how data is presented.

One interviewee suggested a way to approach requests for restricted data: "if requestors are looking for mere descriptive accounts to illustrate a story they are telling without any commitment to further analysis, we might not grant the request. The nature of the research question is the starting point for negotiation."

A number of interviewees felt granular access control would be an essential feature of any institutional research data archive.

5.4 Training requirements

Most interviewees had no formal training in data management but had learned techniques as part of their general research practice and professional development. Some interviewees suggested areas where additional support could be useful:

- Consent requirements for archiving – including how to address this if you are already struggling to recruit a sample and/or researching sensitive issues
- Advice on costing appropriate research data management resource into bids
- Awareness raising around what restricted access and embargo options may be available

- Promotion of the University's research data policy to encourage researchers to think through data management from the start of the research process
- More support around data 'handover' – perhaps a particular issues for QL data. The notion of handover encompasses not just initial deposit but the idea that at some point in the future, responsibility for the data will move from the originating researchers to the repository administrators.

Interviewees suggested various mechanisms for training delivery including:

- Research data management in Masters level research methods training
- Working with Doctoral Training Centres
- Offering specific training to researchers at grant application and award
- More web content for researchers, possibly provided by the Library.

6. Benefits for the case study project and for RoaDMaP

6.1 Timescapes

- Involvement with RoaDMaP has been an opportunity to highlight the innovative work of the Timescapes programme, including data management expertise and resources to support data management planning and archiving practice.
- Involvement with RoaDMaP has also provided an opportunity to reflect on the successes of the Timescapes Archive as well as areas which might have been tackled differently or could be improved in future .
- The RoaDMaP interviews with stakeholders will help inform plans to migrate the Timescapes data.
- RoaDMaP also raised the possibility of assigning DOIs to the Timescapes datasets and whether this should be undertaken before or after migration. It is hoped this will be explored further towards the end of the RoaDMaP project.

6.2 RoaDMaP

- RoaDMaP has been able to draw on the research data management expertise of those involved with Timescapes – some of whom are members of the RoaDMaP Project Team: Professor Bren Neale, Brenda Phillips, who was the Timescapes Archiving Officer, and Bo Middleton from the Library who was (and is) the main liaison point between Timescapes and the University Library.
- The example of an existing, active archive with all the wealth of practical experience this brings has been invaluable to the project and has certainly informed our thinking about the functions of an institution-wide research data repository.
- The collaborative nature of the Timescapes Archive has been extremely useful to inform thinking on support roles and skills for a data repository (see Box x: Library perspectives)

6.3 Library (participation in Timescapes and RoaDMaP)

- Working with Timescapes has been invaluable to learn more about the way researchers work – for example, it was very useful to be exposed to the discussion of ethical concerns about making data available and how these might be addressed.
- Working closely with RoaDMaP has kept the Library up to date with the current institutional position on plans to introduce an institutional data repository and this has helped inform the decision about whether and how to migrate the Timescapes data. Some of these issues are yet to be resolved but RoaDMaP has acted as a catalyst to progress the conversation.

6.4 Mutual benefits

- It has been useful to explore the pros and cons of the current Timescapes data repository platform (DigiTool).
- Timescapes has provided a valuable 'test case' for the data repository functional requirements drawn up by the RoaDMaP project.

7. Library perspectives

7.1 Working together with Timescapes

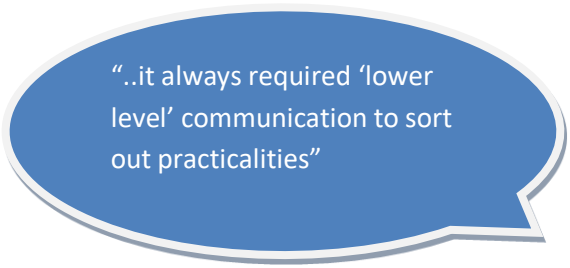
The Library became involved with Timescapes partly to help build a business case for the Library's emerging digital library service (LUDOS), including proof of concept that LUDOS could manage research data. This married well with Timescapes' requirement for a digital archive. The collaboration also offered the potential to include institutional infrastructure contribution in future research bids. The Library provided some technical support, drawing on in-house expertise in the DigiTool platform and was able to assist Timescapes in appointing to a project technical post. Helping set up the archive also provided the Library an opportunity to learn more about research culture. The relationship between researchers and the library has been mutually beneficial and complementary. One Library interviewee noted that the culture of the Library Service and researchers working on a specific project can be very different: researchers can be caught up in the excitement of their research; the Library brings a longer term, pragmatic perspective.

Interviewees noted the importance of having a clear, shared understanding of respective roles and responsibilities when a central service works closely with research project. Libraries are inclined to say 'yes', being geared towards providing help and support for customers. The library may need to manage boundaries and expectations, including working out sustainable cost-recovery models.

a

7.2 Communications

Regular meetings, about every two months, took place between Timescapes and Library staff at the set-up stage of the Timescapes Archive, as well as email communications (not all the key players were based at the University of Leeds). As the LUDOS service developed, the Library ingested a small number of datasets which were not related to Timescapes. Technical staff involved noted that there was a similar pattern of communication required in most cases (including Timescapes): typically, a high level steer at PI level, formal meetings, and a range of more informal communications with whoever was working most closely with the data e.g. a research assistant, a secretary, technical staff.



7.3 Hardware and software

The Timescapes programme employed a Timescapes Archive Technical Officer; the Library allocated a dedicated segment of DigiTool over which Timescapes staff had administrative control. This allowed a level of autonomy for Timescapes to develop their own customisations – for example, metadata schema – but also required some central configuration and troubleshooting at the Library end. Although this devolved model has worked well in some ways, technical understanding has not always been shared between the project and central service, and decisions and customisations have not always been documented in a readily accessible way. This is making the planning of data migration to a new repository more challenging.

There are pros and cons to using third party software compared with open source software but it could be argued that there are so many unknown requirements with data – and these may change as data is re-used over time – that the flexibility of open source becomes more attractive. LUDOS – now known as the Digital Library – has recently migrated from DigiTool to EPrints[11].

"It's convinced me open source is the way forward.

Though the Library provided the repository platform, the server was managed by the central IT department; a clear service level agreement with central IT would have been helpful to ensure a fully supported, secure environment for the Timescapes data (indeed, all content in the LUDOS service).

7.4 Timescapes migration issues

Most Leeds Digital Library content is openly available; it is Timescapes' restricted access layers which caused most concern for the Library interviewees who recognised the need to honour consent conditions from research participants and data depositors, and the need to fully understand the Timescapes access arrangements in DigiTool in order to re-create these using a different, as yet to be decided, repository platform.

The controlled access layers have underlying workflows associated with them: in particular, granting access to registered content and putting data requestors of the most restricted content in touch with researchers. Currently these are manual workflows and the Library interviewees questioned to what extent these processes are scalable and manageable in future. By migrating the data and continuing to 'host' the Timescapes Archive, to what extent would the responsibility for these access workflows fall to the Library in the absence of dedicated Timescapes staff (for example, at the end of specific project funding)? As well as scoping the technical requirements and accompanying costs of data migration, the Library and Timescapes staff have had to discuss options for longer term, post-project funding. This is a tricky but essential conversation - and an issue which is relevant for all data sets hosted by the institution, not specifically Timescapes data.

7.5 Potential RDM roles for the library?

There was no consensus across the interviewees about how active a role the Library should play in research data management (in general, not specifically Timescapes), but the main areas suggested for Library support included:

"You can only promote a service to researchers at the point at which you've got something which works in a good practical sense. Policy or 'coming soon' are of no interest – it's more bureaucracy and verbiage from the centre – whereas if there's something to offer you'll get engagement.."

- Advocating local – and other – data repository services.
- Feeding user requirements into the institutional research data management service.
- Providing metadata advice: opinions varied as to the exact form this might take. Timescapes was unusual in having significant metadata expertise within the project; there may be a role for the Library to support other projects without local expertise, though scalability would need to be considered carefully. One Library interviewee noted

[11] University of Leeds Digital Library. http://digital.library.leeds.ac.uk/

researchers tend to assume their data is unusual or unique and cannot be accommodated within existing metadata schemas; there may be a role for the Library in unpicking this issue.

- Advising on preservation and encouraging early preservation activity. Preservation was a strong, recurring theme in the Library interviews but also an area identified as a Library skills gap.
- Recommending file formats.
- Advising on finding and citing datasets.
- Support for data management planning – for example, supporting use of a data management planning tool.
- Supporting repository technical infrastructure and day to day service provision.
- Digitisation services.
- Making available more research resources from Special Collections.
- Keeping up to date with data discovery mechanisms to make sure any locally housed data remains findable.
- IPR and licensing advice.

Overall, involvement with Timescapes has offered a valuable learning opportunity for the Library and the Library /Timescapes ties have proved invaluable for the RoaDMaP project. The pressing need for data migration is forcing some challenging, but vital, conversations about how to cost long term data curation and what roles and responsibilities are part and parcel of any data repository service.

8. The Stakeholder Model of Data Sharing and Re-Use

Neale and Bishop (2012) note the importance of encouraging researchers to see data deposit as an integral part of the research process rather than "an administrative task that is 'tacked on' to the end of a project" (p56). Timescapes depositors were encouraged to re-use their own data and consider links with other related datasets, thus combining primary and secondary analysis. "The processes involved in a stakeholder model of archiving reveal that this is not simply a matter of bringing researchers to the archive but of bringing archiving more effectively into the realm of research" (ibid, p. 64). Cross-comparing data can enable researchers to see their own data with fresh eyes: Irwin, Bornat and Winterton (2012) suggest secondary analysis encourages researchers to 'hear silences' in their data and, "re-interrogate their own data within a revised conceptual framework"(p.66).

The Timescapes Archive model for controlled data access gives depositors a stake in deciding who has access to the data and how it is used; it also "facilitates consultation between primary and secondary users." (Timescapes Final Report 2012). The stakeholder approach aims to balance the interests of various interested parties , managing risks for the primary researchers and the research participants but still promoting access for secondary users.

Timescapes also provides a useful exploration of how researchers and archivists can work together. Developing a common language is essential (for example, the Timescapes team developed a shared understanding of the term 'metadata' by initially using the more widely understood term 'context'). Researcher and archivist priorities may differ, but it is possible to negotiate solutions to balance their sometimes competing needs (Neale and Bishop 2012).

Neale (2013b) explores 'Stakeholder Ethics', suggesting the elongated time frame of QL research requires particular ethical strategies. Ethical considerations include, "consent as an ongoing process; sustaining confidentiality when time magnifies the possibility of revealing identities; the ethical representation of lives in the cumulative construction, analysis and display of research outputs; and the delicate issue of the longer term stewardship and use of research data." As well as applying predefined ethics frameworks (pro-active ethics strategy), it is important to engage with ethical questions as they arise over time, refining ethical protocols and practices as necessary (re-active ethics strategy). For example, archiving consent may be revisited once data have been collected and participants are better able to reflect on the data and how it might be re-used.

Several interviewees were enthusiastic about the Stakeholder model: “A stakeholder model that tries to respect diverse players and diverse interests and find a way to bring them together productively and ethically is the way forward.”

9. What we wish we’d known before our research project / lessons learnt

9.1 Timescapes Archive

- Setting up the Timescapes Archive was a pioneering venture and it was difficult to get the balance of staffing right and in retrospect, the project could have included more resources to create a strong infrastructure team, perhaps including Library and IT staff. Through Timescapes and RoaDMaP, as an institution Leeds is gaining a more informed understanding of the resources required not only for short term data management and ingesting into the Timescapes Archive (or a wider institutional data repository), but longer term costs for core repository staffing and longer term data management.
- Two interviewees who had deposited data into the Timescapes Archive noted it would have been useful to have had a clearer idea of how the data would be presented and organised in the archive; because the data structure did not mirror the depositors’ structure, it made it more difficult to review and check the data.
- Some interviewees reflected on areas of research practice which might have been usefully documented and included with the research data e.g. difficulties in recruiting the sample, approach to interview / diary transcription.
- Any data repository platform should be able to support both searching and browsing options and should be able to display related records / documents together.
- Don’t embed key knowledge too narrowly in staff; manage the expertise risk.
- Full documentation of a data repository is essential for any subsequent migration.

9.2 Preparing data for archiving / when to deposit

- Describing some (non-Timescapes) research, one interviewee noted: “we didn’t bid for enough money for data archiving. We thought, ‘everything is digital, it will just go in’ but it turned out it was much more time consuming. We had to do things like clean up the tapes, the transcripts, some people wanted to be anonymised, some people wanted bits removed – there was quite a lot of running around.” There was a general consensus that it is easy to underestimate the staff resource required for appropriate data management and data preparation.
- Some researchers were keen on centralised support for data management, particularly preparing data for archiving, arguing time spent within the project on RDM was valuable time lost from doing the research itself. Others had a different view: “data management should be seen as an output of the research. It’s not in competition with the fieldwork, it’s a necessary part of the research process and should be funded”.
- The Timescapes End of Award Report (2012) suggests that attention to data and metadata preparation can save time further down the line: “when data copied to the UK Data Archive were processed for long term preservation, normal processing times were cut approximately by half” (p.31)

9.3 Controlled access to data

- Most interviewees were aware of funder requirements and other drivers to make research data openly available, but felt that the decision on level of openness should lie with the researchers and any data repository system should be able to support access control: “it is precisely and exactly those levels of control that makes many qualitative researchers feel OK about depositing their data.”

- One interviewee suggested the idea of data handover had not been fully explored within the research data management agenda. Perhaps after a fixed embargo period, the data could be available on request and then more openly.
- Although access control is vital, some interviewees stressed the importance of keeping the system as simple as possible – otherwise it could become unwieldy to administer. One suggestion was set embargo periods for particular purposes e.g. 'the data embargo period to allow depositors to publish papers is x months'.
- Consent agreements could address different scenarios including re-use of the data in the event the participant withdraws from the study or dies and whether the participant has any rights to see their own data.
- Timescapes found that if handled sensitively, many participants were willing to grant archiving consent; 95% agreed to data archiving (Timescapes End of Award Report 2012 p.31)

9.4 Data Ownership / legal aspects

- Researchers need clearer guidelines from the institution on data ownership – for example, what happens to data if the researcher moves institution; if they want to write papers based on the data etc. This issue is not generally included in standard inductions.
- A more fully worked through understanding of the implications of data archiving for Freedom of Information requests would be beneficial.
- One area we did not investigate in detail in RoaDMaP was the interplay between levels of openness and end user agreements / re-use licences. Timescapes users are asked to abide by an End User Licence.

9.5 Embedding research data management practice

- Some research groups designated an individual who had responsibility for data management and included data management as a standard agenda item for project meetings.
- School / Departmental pre-award administrators are a key group: they can walk researchers through processes using checklists and guidelines
- The support from the Faculty IT Manager with writing data management plans was particularly appreciated – though this would not be scalable across all research bids and Faculties at the institution.

Box 6: Senior researcher on data management planning..

"There's a craft element to all research and in learning a craft you learn on the job. You have to go and get your feet dirty. It's going to have to develop over time. It can't be parachuted in, I suggest."

9.6 Re-use in practice

- The Timescapes secondary analysts made some interesting observations about the realities of re-use and its potential frustrations. A number of strategies for engaging with data as a re-user are outlined in a Timescapes working paper (Irwin and Winterton 2011b); the paper notes the significant potential benefits of data re-use whilst acknowledging "secondary analysis is a challenging undertaking.. It can be frustrating for researchers to become 'users' when they may be more used to controlling the who, what and why of research design and data."(ibid p. 14).

- The process of depositing data for re-use presents intellectual and emotional complexities for primary researchers, as well as ethical and practical challenges.

9.7 Metadata

Box 7: Metadata

Timescapes reviewed the out-of-the-box metadata schemas in DigiTool but decided their requirements were very specific and they needed to develop their own. The Metadata Specification is included here as Appendix 1 and the full Timescapes metadata schema is available online: http://www.timescapes.leeds.ac.uk/assets/files/timescapes/mdspec.xsd

Metadata requires a lot of planning and effort; this work can only be done within the project but if unduly onerous, it presents a barrier and can cause researchers to disengage from data deposit. Ensuring researchers are not put off by 'archiving language' is also important: "I didn't used to call it metadata - we always called it contextual data.."

Being involved in Timescapes required the researchers (and Library staff) to think through what it means in practical terms to collect, manage, store and share data and metadata. This has had an impact on subsequent research bids because of greater appreciation for the processes involved and resourcing required.

A recognition of the value of high quality, or 'gold standard' metadata, was shared by many of the interviewees who were concerned about the potential for future re-users to 'misinterpret' the data:

> "We.. go out in the field and we create good field work notes. It might describe the interaction, our own role in the process, and where we are coming from as researchers. It would describe the environment in which we do an interview and as much as we can garner about the lives of the people.. Researchers are now thinking if we are going to archive this, we have to .. record this kind of data in a particular way for re-use purposes. So we are moving towards this idea of gold standard metadata which accompanies our data."

The Timescapes research programme included a secondary analysis strand looking within and across the Timescapes datasets. This work drew together data from different studies to achieve new insights and enhance understanding. A working paper from the Timescapes WP series (Irwin and Winterton 2011a) suggests a 'wish list' of contextual information to facilitate data re-use. The list is reproduced here as Appendix 2, Timescapes Desirable Documentation.

9.8 Building the service

- A number of interviewees advised it was important not to over promise what is on offer as the RDM service emerges; it would be better to introduce the service gradually and be clear about what support can be provided.
- Similarly, when designing the service and considering charging/cost recovery, we should have a clearly defined support offer. Anything outside the standard offering should be negotiated beforehand and include a clear exit strategy – particularly in relation to any special functionality or requirements.

10. Implications for an institutional data repository

10.1 Scoping the data repository

- It was challenging to build the Timescapes data archive before the data it would accommodate was fully understood. However, this may be the case for institutions building local data repositories – it is possible to improve knowledge of institutional data assets but it may be necessary to make some important decisions about the repository structure based on specific use cases and incrementally enhance the repository as the breadth of data it needs to house increases. One interviewee suggested the major challenge is trying to judge the likely evolution of whatever technology is implemented.
- Legacy data: though the interviews were primarily about the Timescapes data, interviewees raised the issue of other legacy data from projects – sometime in digital form, sometimes in analogue – which would need additional resource to manage and make available. Institutions may need to have clear policies on this – or advise on what level of resource might be needed to ingest older data into institutional data archives.

10.2 Searching and navigation

- Display of results has been an issue in DigiTool; ideally the repository should be able to display data and accompanying documentation together.
- Navigation through the archive – although keyword searching can be helpful, users expressed a desire to browse through contents. Ideally a repository will present a map or overview of what data is available so prospective re-users can orient themselves within the data collections and be offered routes into and through the data.
- It will be important to ensure data can be readily located by potential re-users via external search services, keeping up to date with new discovery mechanisms.

10.3 Access Control

- It is clear the repository will need to offer controlled access to at least some data sets. Any data request mechanism must address what happens if the contact is unavailable – nobody wants valuable data sets to become orphaned: alternative contacts should be specified and /or, after an agreed length of time, the responsibility for granting access could devolve to the archive administrators.

10.4 Repository management

- The repository manager needs to understand what is being ingested so good tools for inspecting the deposited data are vital.
- If data has been migrated from other system, it may have been cited using a system-specific identifier – we may need to retain these in addition to minting DOIs.
- The repository will need to offer appropriate end user agreements and/or standard licensing of data sets.
- Ideally a repository should not only link data and published outputs but flag what re-use of the data has taken place. The repository might require acknowledgement as part of re-use conditions, and/or ingest pertinent secondary analysis resources alongside the original data set.

10.5 Repository Ecosystem

- One, as yet, unresolved issue is whether the Timescapes Archive should be managed within the institutional data repository or whether it can be managed as a 'special case' because it manages data generated outside the University of Leeds. It can be argued that Timescapes should continue to be a separate repository in its own right, specialising in QL data – in which case, it would effectively be a national / subject repository hosted at the University of Leeds. The Timescapes Archive is set to expand: it is planned that several collections of

thematically related datasets will be added, for example data collections on third sector research will ingested as part of a new ESRC funded project, Changing Landscapes. There are other repositories of subject specific research data at the university; some further thinking about how these services inter-relate and are supported is needed. There is also the issue of how the institutional based Timescapes Archive relates to the UK Data Archive in the longer term (see Box 8: Institutional data repositories and subject repositories).

Box 8: Institutional data repositories and subject repositories

Timescapes' close relationship with the UKDA meant several of the interviewees had thought about the relationship between an institutional based data repository and a national data centre:

"The national archives now really don't have the resources to archive everything ..I think there is certainly scope for specialist collections that ..bring together collections of thematically related data sets. My understanding is that there are new national data policies emerging which will mean that we are going to have more pressure on us as institutions to look after our data at institutional level rather than shipping it off to the national resource."

"I can imagine a distribution of specialist services either around methods, or data, or format – a GIS specialist, these sorts of things – what I'm looking for is what's the logical relationship for the hierarchy between the generic data centre and the repository – they shouldn't be treading on each other's toes."

It is planned that, over time, the original Timescapes data will deposited into the UKDA . The Timescapes Archive acts as a platform to host 'live' data and facilitate data discovery and data re-use. UKDA is seen as providing a preservation function. This relationship holds true for the data generated by the original Timescapes Programme but discussions are ongoing about the role of UKDA in supporting new datasets ingested into Timescapes from affiliated projects or as a result of follow on funding for Timescapes.

One model for institutional and data archive relationship:

Local archive (e.g Timescapes) **Ingest and dissemination**	Data centre (e.g. UKDA) **Preservation**

As a thematic repository, the Timescapes Archive has been successful in attracting a network of affiliated datasets, offering the possibility of contributions to ongoing costs from affiliated projects.

Most interviewees were agnostic about whether data is deposited in a subject vs. an institutional repository, the most import factor is findability. One of our interviewees posed a scenario which will challenge institutional data repository and any national data catalogue service:

"..suppose I'm interested in investigating perceptions of vulnerability in low income communities or among service providers – where would I go [to find the relevant data] is the big question in my mind."

11. Next steps

Data management planning

- A key outcome from the interviews is the emphasis on pre-award staff; RoaDMaP has offered some research support staff training but more is needed. We have also run a specialist event with DCC targeted at pre-award staff. Ideally, to support professional and service development, we should build a bank of data management plans for projects which have successfully bid for funding, including how the elements of the data management plan were costed into the project bids.
- The data management planning support provided by the Faculty IT Manager was highly valued by the RoaDMaP researchers; however, this model is unlikely to scale across other Faculties where the volume of bids is greater. More work should be done on mapping data management planning functions against available staffing and to ensure the institution widens its skills and knowledge base in the area.

Timescapes repository migration

- RoaDMaP has produced a draft set of data repository functional requirements; the Timescapes data archive will be reviewed against these.
- A decision will be made in 2013 whether to migrate the Timescapes Archive and, if so, to what repository platform; regular meetings are already in place to progress this issue.
- RoaDMaP will investigate at what point to assign DOIs to the Timescapes data: should this be before or after migration?
- The feasibility of reproducing the Timescapes access control model is being discussed with the EPrints team in Southampton and the issue of access requirements has been raised with the wider repository community. Responses suggest other institutions also require access control though, of course, there is concern about undue restriction of content, about the scalability of manual request processes and the (perhaps unintended) consequences of applying inappropriate re-use licences.
- Timescapes may wish to seek further input from re-users of the data – particularly those who have accessed the data primarily via the online archive rather than directly from the originating researchers. Feedback on user experience could be used to inform the evolution of the Timescapes Archive, regardless of which software platform it sits on.

Funding model and sustainability

- We need a sustainable cost model to help researchers scope what type of storage and curation will be needed over time including what costs can be recovered against grants. Timescapes affiliated projects (which received methodological and data support) included £3000 in their bids to pay Timescapes for data archiving. This figure was a best guess and will be reviewed as more evidence of costs and staffing requirements is gathered. We will also look in more detail at the Archaeology Data Service charging model.

Institutional RDM Policy

- Some interviewees noted the danger that the policy is seen as ‘just a paper’ and not meaningful in their day to day role. To help engagement we will look at mapping the Policy against job areas and mapping the research data lifecycle against ‘who does what’; the ‘who’ may be different across research projects and will definitely vary across institutions but it should be possible to capture the ‘what’.

Research funders

- There was some feeling that more support should be forthcoming from research funders who require data management plans and data deposit; we are investigating how to utilise funder visits to explore/embed data management activity.
- One interviewee suggested it may be helpful to offer text or arguments about the potential added value of data archiving that researchers could include in grant proposals to funders who are less actively engaged with the data re-use agenda.

Future bids

- The relationship between the Timescapes Programme and the University Library continues to be strong, with active discussion of what level of support can be provided (in both directions) for data migration but also to handle data sets arising from new / future project funding. Timescapes recently bid successfully to ESRC for another project with a data sharing strand: '*Changing Landscapes for the Third Sector: Enhancing Knowledge and Informing Practice'*. The Library was involved in discussions about archiving costs for inclusion in the bid and an Archiving Officer post attached to this project will be based in the Library. Further collaborations around the Timescapes Archive are being explored. Thus the relationship continues to be mutually beneficial, offering the Library scope to expand and consolidate its role in research data management and Timescapes the opportunity to locate its data in an institutionally based infrastructure which, hopefully, will lead to long term stability and sustainability for the archive.

12. Summary and Key Points

- The RoaDMaP project benefitted from the expertise of the researchers, administrators and Library staff involved in the Timescapes programme. This case study aims to capture some of that expertise, including the realities of data management and re-use in practice.
- The Timescapes Archive illustrates some key characteristics of Qualitative Longitudinal data and demonstrates how it can be shared for re-use.
- Our interviews highlighted that controlled access may be necessary for some data. If data control is applied, repositories will need to decide how to manage this in the long term, particularly if manual processes are involved, and how to manage the handover of responsibility for the data from the depositor to the repository service.
- The interviews gave an insight into the process of creating a research data repository, including the benefits and challenges of working together across research teams and central services.
- The Timescapes Archive provides a useful test case for the data repository functional requirements drawn up by the RoaDMaP project which we hope to apply to an institution-wide data repository service.
- The Timescapes Archive provides a focus for QL data generated by researchers at a number of institutions. It throws into relief the need to understand how institutional data repositories housing research data produced at the host institution fit with institutionally-based repositories with a subject or thematic focus housing data generated outwith the institution. Some interviewees questioned whether every institution 're-inventing the wheel' was the most efficient data repository model.
- The Timescapes Archive illustrates one potential model for the relationship between institutionally based repositories and national data centres, in this case the UK Data Archive.

13. Recommendations (for Jisc)

- Further investigation at the national / international level of how the components of the repository landscape can most effectively fit together (institutional data repositories, institution-based thematic or subject data repositories, data centres) would be valuable. This could include consideration of how limited resource can be most effectively directed and the pros and cons of institutions or consortia specialising in particular types of data. This could be by subject areas and/or by characteristics of the data, such as longitudinal data.
- The term 'open access' means different things to different people; in the data context, 'controlled' or 'managed access' may be a more appropriate description for how data is handled in practice. Some standardisation of terminology could be useful. The issue of licensing goes hand in hand with access. There may be a danger that researchers will tend towards a conservative approach to data sharing and apply restrictions if these are offered, regardless of whether these are strictly necessary for the data or not. Spelling out the potential problems of multiple licences and conditions for future data use would be valuable, as would illustrating the benefits accruing with greater levels of openness.
- In our interviews and in the discussions between the Timescapes programme and the Library, resourcing and sustainability came up time and again. Institutions need more help in understanding the nature of data archiving roles, how to cost activities and infrastructure into bids and appropriate resourcing for a centrally managed data repository service.

Acknowledgements

Many thanks to the Timescapes and Library staff who generously gave their time to be interviewed for the RoaDMaP project.

Timescapes researchers, archive staff and affiliated projects:

- Professor Bren Neale, *Timescapes Director*
- Dr Nick Emmel, Senior Lecturer, Sociology and Social Policy (University of Leeds), *Timescapes Co-I*
- Dr Carmen Lau-Clayton, *Timescapes Research Fellow* and *Director of the ESRC Following Young Fathers Study* (Timescapes follow on project)
- Dr Sarah Irwin, Reader in Sociology, University of Leeds, *Timescapes Co-I*
- Professor Joanna Bornat, Emeritus Professor of Oral History, Open University, *Timescapes Co-I, Project Director of The Oldest Generation*
- Dr Bill Bytheway, Senior Research Fellow, Open University, *Co-I of The Oldest Generation*
- Professor Caroline Glendinning, Professor of Social Policy, University of York, *Timescapes affiliated project*
- Dr Libby Bishop, Universities of Leeds and Essex, *Senior Research Archivist for Timescapes* (two part time posts, one at each institution)
- Brenda Phillips, University of Leeds, *Timescapes Data Archiving Officer*

Library staff:

- Bo Middleton, Head of Digital Library Services, University of Leeds
- John Salter, Systems Team Officer, University of Leeds
- Dr Jonathan Ainsworth, Computer Officer: web development, University of Leeds

References

BISHOP, L. and NEALE, B. 2012a. Introduction to the Timescapes Archive. *Timescapes Methods Guide Series (16)*, [Accessed 6 March 2013]. Available from: http://www.timescapes.leeds.ac.uk/assets/files/methods-guides/timescapes-archive-introduction.pdf

BISHOP, L. and NEALE, B. 2012b. Data Management Planning in QL Research. *Timescapes Methods Guide Series (17)*, [Accessed 4 March 2013]. Available from: http://www.timescapes.leeds.ac.uk/assets/files/methods-guides/timescapes-data-management-planning.pdf

ESRC. 2013. *Research Funding Guide March 2013*, [Accessed 4 March 2013]. Available from: http://www.esrc.ac.uk/_images/Research-Funding-Guide_tcm8-2323.pdf

IRWIN, S., BORNAT, J. and WINTERTON, M. 2012. Timescapes secondary analysis: comparison, context and working across data sets. *Qualitative Research*, **12**(1), [Accessed 10 April 2013], pp.66-80. Available from: http://dx.doi.org/10.1177/1468794111426234

IRWIN, S. and WINTERTON, M. 2011a. Debates in Qualitative Secondary Analysis: Critical Reflections. *Timescapes Working Paper Series (4)*, [Accessed 9 March 2013]. Available from: http://www.timescapes.leeds.ac.uk/assets/files/WP4-March-2011.pdf

IRWIN, S. and WINTERTON, M. 2011b. Qualitative Secondary Analysis in Practice: An extended guide.. *Timescapes Working Paper Series (7),* [Accessed 9 March 2013]. Available from: http://www.timescapes.leeds.ac.uk/assets/files/secondary_analysis/working%20papers/WP7-Nov-2011.pdf

NEALE, B. 2007. Timescapes: study overview, [Accessed 15 March 2013]. Available from: http://www.timescapes.leeds.ac.uk/assets/files/TIMESCAPES_Blueprint.pdf

NEALE, B. 2012**.** Timescapes: Archiving and sharing Qualitative Longitudinal data**.** 26 Nov. *RoaDMaP Blog* [online]. [Accessed 15 March 2013]**.** Available from: http://blog.library.leeds.ac.uk/blog/roadmap/post/132

NEALE, B. 2013**.** Timescapes: Archiving and sharing Qualitative Longitudinal data – part2**.** 13 March. *RoaDMaP Blog* [online]. [Accessed 15 March 2013]**.** Available from: http://blog.library.leeds.ac.uk/blog/roadmap/post/155

NEALE, B. 2013b. Forthcoming November. Stakeholder Ethics in Qualitative Longitudinal Research: a temporal gaze on research relationships and the use and re-use of data. *Methodological Innovations Online. Special issue on Research Ethics in Challenging Contexts.*

NEALE, B and BISHOP, L. 2012. The Timescapes Archive: a stakeholder approach to archiving qualitative longitudinal data. *Qualitative Research*, **12**(1), [Accessed 5 April 2013], pp.53-56. Available from: http://dx.doi.org/10.1177/1468794111426233

RDMRose. 2013. Session 7.1: Case studies of researchers and research projects: Sociology . Transcription of excerpts from an interview with Professor Bren Neale. [Accessed 21 March 2013]Available from: http://rdmrose.group.shef.ac.uk/?page_id=10

TIMESCAPES. 2012. EPSRC End of Award Report.

TIMESCAPES. Timescapes web site. [Accessed 1 May 2013] Available from: http://www.timescapes.leeds.ac.uk/

UKDA. UK Data Archive web site. [Accessed 15 May 2013] Available from: http://data-archive.ac.uk/

UKDS. UK Data Service web site. [Accessed 15 May 2013] Available from: http://ukdataservice.ac.uk/

Appendix 1: Timescapes Metadata Specification

Technical

The metadata for file, image, audio and video also contains the following attributes:

1. embargo – should this resource be embargoed – true or false
2. embargoDate – a date and time that the embargo will apply till
3. dark – if true the resource will not be accessible from the archive

File

Name	Value	Type	JHOVE
Object identifier		URI	1
Filename		String	
Extension		String	1
Size	Kilobytes	Integer	1
Format	MIME + Originating software	String	1
Format version		String	1
Language	ISO 639-3	String	

Image

Name	Value	Type	JHOVE
Object identifier		URI	1
Filename		String	
Extension		String	1
X resolution	No of pixels per resolution unit	Integer	1
Y resolution	No of pixels per resolution unit	Integer	1
Resolution unit		Integer	1
Width	Pixels	Integer	1
Height	Pixels	Integer	1
Compression		String	1
Colour space	RGB, CMYK, YPbPr	String	1

Orientation	Portrait or Landscape	String	1

Audio

Name	Value	Type	JHOVE
Object identifier		URI	1
Filename		String	
Extension		String	1
Bit rate		Integer	1
Sample rate	Kilohertz	Integer	1
Sample size	Bit size	Integer	1
Duration	seconds	Integer	
Compression		String	1
Channels	1,2 or 5	Integer	1
Encapsulation		String	

Video

Name	Value	Type	JHOVE
Object identifier		URI	1
Filename		String	
Extension		String	1
Duration	Seconds	Integer	1
Frame dimensions		Integer	1
No of frames		Integer	1
Frame rate	Frames per second	Integer	1
Codec		String	?
Aspect ratio	4:3, 16:9	String	1
Scan type		String	

Audio		Boolean	1

Descriptive

Element	Value	Cardinality	Type	Comment
Project Name		1	String	
Gender	Male/Female	1	String	Controlled vocab
Age Group		1	String	Controlled vocab
Time Dimension	Year	1	String	
Data Type			String	
Social Class	MIME	1	String	Controlled vocab
Generation	Child/Parent/GrandParent	1	String	Controlled vocab
GUID		1	String	
Ethnicity		1	String	
Case ID		1	String	

Appendix 2: Timescapes Desirable Documentation

Text taken from Irwin and Winterton (2011a)

The list below headlines some key domains we believe would ease the task of data re-users and therefore facilitate effective secondary analysis. This is a 'wish list' from the SA project team, who have been working during Timescapes projects research period, and started before the archiving of much Timescapes data. We are therefore working on moving ground in terms of what is being archived (as we have typically been working with data provided directly to us as the SA team). As we see it this list is quite minimal, although in terms of project resources it may appear onerous. In particular, a profile of projects' participants is a task some, but not all, projects will be undertaking. Non-Timescapes readers need be aware of the very significant amount of work that goes into archiving data including many tasks not discussed here, such as anonymising transcripts and other data.

Desirable documentation which would facilitate the task of secondary analysts

An orientation to the project:

1. **Record of publications, presentations** etc which explain and /or draw on the archived data.
2. **Research design**

a) Brief context and logic of research design (in each data collection period if appropriate).

b) Is research exploratory or question driven? What are the questions?

c) What is the fit with Timescapes logic? Description of qualitative longitudinal design, including account of the kind of QL study undertaken. What are the key questions which benefit from the longitudinal design?

3. What were the **sampling decisions** and how do they relate to the research questions?

a) Was the desired sample achieved? How does the sample relate to wider empirical evidence across the population and/or theoretical issues?

b) Are there implicit as well as planned ways in which the sample is structured (for example, does opting into the study have implications for understanding the extent to which participants 'stand for' a wider population, or provide insights into wider experiences (e.g. even within their target populations, studies will often recruit particular kinds of people, in respect of background, education, identification with project aims etc). Would it be helpful to offer reflections /insights into this?

4. **An overview of what data is provided as part of the project.**

5. Projects would usefully supply **a descriptive profile of each participant**

What form should this take? A descriptive profile could relate to units of data (interview; diary; by wave); and / or could be an overview of the participant. Ideally it would be both.
Most useful would be a descriptive but broadly 'factual' account, describing a person's circumstances, biography, key life course changes, family/household context, etc and an indication of the points at which they were interviewed.

It is sometimes said that which facts to include will reflect the primary analysts' interests, and may encourage particular readings of the data. However, if it is broadly 'factual' and follows a similar format across the projects it would certainly facilitate the work of re-users.

6. **Relevant contextual information**?

We propose inclusion of information about the context of interviews (or other forms of data collection) if this is deemed relevant to secondary analysts. We do not document precisely what contextual information would be usefully supplied: this is being steered by project teams themselves and by the Timescapes Archive team. We suggest that it would be useful (with Bishop 2006[12]) to think of different levels of context. At the most detailed level this could pertain to annotations within transcripts (see endnote 2). At a meso level it could pertain to a note about the fieldwork context. At a macro level it could usefully include notes regarding the broader socio-demographic, economic and social policy context in which the research was conducted.

[12] Bishop, L. (2006) A proposal for archiving context for secondary analysis, *Methodological Innovations Online* 1 [2].

7. An opportunity to highlight areas within the data which might be rich for (further) analysis. Should projects take an opportunity to state areas they think might be ripe for analysis by secondary analysts? This might be construed as overly directional. Alternatively it could be construed as an additional resource for secondary users (including students who might come to the material with exploratory purpose).