Moving data into and out of an institutional repository: Off the map and into the territory

Libby Bishop*

Abstract
Given the recent proliferation of institutional repositories, a key strategic question is how multiple institutions—repositories, archives, universities and others—can best work together to manage and preserve research data. In 2007, Green and Gutmann proposed how partnerships among social science researchers, institutional repositories and domain repositories should best work. This paper uses the Timescapes Archive—a new collection of qualitative longitudinal data— to examine the challenges of working across institutions in order to move data into and out of institutional repositories. The Timescapes Archive both tests and extends their framework by focusing on the specific case of qualitative longitudinal research and by highlighting researchers' roles across all phases of data preservation and sharing. Topics of metadata, ethical data sharing, and preservation are discussed in detail. What emerged from the work to date is the extremely complex nature of the coordination required among the agents; getting the timing right is both critical and difficult. Coordination among three agents is likely to be challenging under any circumstances and becomes more so when the trajectories of different life cycles, for research projects and for data sharing, are considered. Timescapes exposed some structural tensions that, although they can not be removed or eliminated, can be effectively managed.

Introduction
Institutional digital repositories are growing at a rapid rate driven by factors such as institutions promoting their intellectual capital, researchers' seeking greater control over output dissemination, research councils and other funders requiring data to be offered for deposit, and the growing support for open access initiatives (ESRC Data Policy, 2000). A key strategic question is how multiple institutions—repositories, universities and others—can best work together to manage and preserve research data. A number of UK reports have begun to recognise the urgency of this issue (Gibbs, 2007; Lyon, 2007). This paper uses the Timescapes Archive to examine the challenges of working across institutions in order to move data into and out of institutional repositories.

This paper will first describe key features of a framework Green and Gutmann (2007) propose for how partnerships among social science researchers, institutional repositories and domain repositories should best work. Next, the Timescapes Project and Archive will be described with a focus on its distinctive features. In a number of areas, the Timescapes Archive both tests and extends their framework. Each topic of metadata, ethical data sharing, and preservation will be discussed in detail. Lessons learned and next steps make up the conclusion.

Green and Gutmann framework
The Green and Gutmann framework lays out roles and relations among researchers, institutional repositories and domain repositories in relation to the life cycle for social science research. They argue that institutional repositories can play a key role by mediating between researchers (as depositors and users) and domain repositories. As definitions are not fixed in this area, I am following their typology. The distinguishing features of institutional repositories are that they: cover diverse disciplines, have tended to focus on research outputs rather than data, are committed to simplifying deposit procedures (at times employing less elaborate metadata and documentation), and believe in sustainability but often lack resources for long-term preservation. Domain repositories, by contrast, often have a disciplinary theme (e.g., social science), are focused primarily on data, not outputs, and are committed to preservation while also assuring usability of their collections (Green and Gutmann, 2007; 4).

Green and Gutmann present a comprehensive model describing roles for all three agents across all phases of research. This paper will highlight the parts of their model most relevant for the Timescapes Archive. The institutional repository is especially active early in the research life cycle. Early on, it can raise issues of data sharing and archiving to researchers, well before most projects will have considered such matters. It is also well positioned to mediate between researchers and the domain repository. The domain repository can provide information and support in several ways areas: advice and forms regarding the gaining of consent for data sharing, technical advice on collecting metadata, and assisting the transfer or sharing of data and documentation to a domain repository for

preservation. The Green and Gutmann model is based on "cooperation and specialisation" among the agents. "The next step in the evolution of digital repository strategies should be an explicit development of partnerships between researchers, institutional repositories, and domain-specific repositories" (Green and Gutmann, 2007, 16).

The Timescapes Project and Archive
The Timescapes project is a £4.5 million, five year ESRC-funded study designed to shed light on the dynamics of personal relationships over the life course and the identities that flow from those relationships. Timescapes entails a consortium of five universities conducting seven empirical projects that investigate the life course. Over 400 participants will contribute data and the archive will hold at half a terabyte of data. A key objective of this initiative is to establish a working archive of data as a resource for sharing among researchers, other authorised users, and for future historical purposes.

The Archive is developing a partnership between institutional and domain repositories by implementing a structure of "disaggregated preservation" (Knight and Hedges, 2007). An institutional repository at Leeds will receive incoming content and support data preparation, metadata collection and enhancement, and data sharing. This repository will extend the existing MIDESS system at Leeds, using DigiTool software, and is designed to accommodate multi-media file formats. This system is now called the Leeds University Digital Objects (LUDOS). The repository will send appropriately prepared (e.g., compliant with OAI-PMH and DDI standards) to the UK Data Archive (UKDA) for preservation. Dissemination versions of files (whether produced at Leeds or UKDA) will be available from both locations. Thus the repository will have primary responsibility for ingest and dissemination with preservation "disaggregated" to the UKDA.

A key element in this design choice was the fact that several JISC projects had successfully used similar designs including Sherpa DP2 (Knight, 2005; Knight and Hedges, 2007), the Preserv Project (Hitchcock, et al., 2007) and the PLEDGE Project (MIT Libraries, 2008). Knowledge of these projects and consultation with their project staff were critical in providing reassurance that the disaggregated preservation model was reasonable for Timescapes. Among other reasons for adopting disaggregated preservation was the obvious one of not recreating preservation services if existing ones fit for purpose are available (Knight and Hedges, 2007). LUDOS is shifting out of its pilot phase and will establish this preservation service, along with associated digitisation, collection managements and preservation policies. In sum, the Timescapes Archive will provide extended functionality by integrating with existing technical and administration infrastructures at

The University of Leeds Library, University of Leeds Information and Systems Support, and at the UKDA.

Distinctive features of the Timescapes Archive
Although the differences between qualitative and quantitative data are often overdrawn, in regards to archiving, there are some distinctions that affect each stage of data processing from ingest through to preservation. When considering Timescapes, it is useful to consider its distinctive products and processes. Its product is an archive of qualitative longitudinal data. Several features of this data pose particular challenges. The Timescapes collection will be predominantly qualitative data rather than numeric data. Much of it will be in traditional forms, such as interviews, but the collection with include other file formats such as images, audio and video. The data are also longitudinal and will be dynamically incrementing over the five years of the project. The most important implications are the capture of appropriate metadata for objects in the collection, and the personal and sensitive nature of the data that will require special treatment to comply with requirements of the Data Protection Act and to meet ethical, as well as legal, confidentiality commitments to research participants.

With respect to its process, the Timescapes project is also distinctive in its ambition to simultaneously conduct and synchronise primary research, preservation and data reuse. By doing so, the project promotes a central role for researchers not only in the primary research project, but in the data archiving process as well. Attention to the repository and researcher interaction has been highlighted by the JISC Digital Repository Review's objective of better integrating repository and researcher work practices (JISC, 2005, p. 8).

Not only the data, but the methodology and structure of the programme are qualitative. The definition of qualitative is highly contested, but typically, research begins with questions or aims, not usually formal hypotheses. The data gathering process is often emergent and flexibly adapted during the course of the research (Mason, 2002; Silverman, 1985). Although the Green and Gutmann model can encompass both, there is a feature of qualitative research that matters, and that is its iterative nature or what Berg calls "spiralling" (2004). For example, in a typical survey, a sample would be established early and not change unless drop outs required replacements. In qualitative research, after an initial phase of data collection, new themes could emerge, calling for new participants, questions, or data types. The general point is that the timing and phasing the relationships among institutional and domain repositories and researchers are complex and made more so by the non-linear nature of the qualitative research process. The specific forms of this complexity will be elaborated in discussions below on metadata, ethical use of data, and preservation.

Metadata – challenges for qualitative data
While all digital materials require good metadata, data (as distinct from research outputs) pose challenges for adequate metadata in part because of the more complex file formats involved (Heery and Powell, 2007). Because it is data, it needs more extensive metadata and contextual material to render it "independently understandable" (to meet OAIS standards) than textual research outputs. Unlike research outputs with relatively standardised formats, qualitative research data and documentation are highly diverse. Firstly, they are diverse in technical file formats (txt, doc, tif, jpeg, wav, etc.). Even within a format, a text document can be, to name just a few: interview, focus group, diary, field note, analytical note, and memo. It is also generally accepted that qualitative data needs extensive contextual information to enable effective reuse (Fielding, 2004). Much of this may fall into familiar metadata categories such as "interview" for type of data, but ideally context should also include information about the project background and even the social and institutional conditions in the wider environment that might have shaped project design (Bishop, 2006).

Not only is there a great deal of metadata to capture, but the knowledge of that metadata is widely distributed. Each agent—institutional repository, domain repository and researcher—brings specialised knowledge to metadata production. Detailed descriptive metadata intended to support resource discovery is usually known best by the producer of the data. "The metadata required to access, understand, and manipulate scientific datasets will continue to be largely the preserve of domain-experts" (Heery and Powell, 2007). But there is evidence of problems of getting depositors to provide adequate metadata. A disproportionate share of processing staff time is devoted to collecting and correcting adequate metadata (Beagrie, et al., 2008). Similar problems have been encountered in repositories of learning objects (Ryan and Walmsley, 2003).

In contrast, the domain repository expertise lies in domain knowledge and technical expertise in resource discovery and preservation:

> "…the domain-specific repository has specialized knowledge of data management approaches to data in a specific scientific field, for example, domain-specific metadata standards (the DDI in the case of the social sciences), as well as the ability to expose the research products to the field in a way that will have the greatest impact" (Green and Gutmann, 2007; 16).

Green and Gutmann (2007) suggest that institutional repositories broker key relationships and thus enable the creation of more and higher quality metadata.

Defining a metadata schema for Timescapes
The starting point for defining a metadata schema for

the Timescapes Archive was a commitment to openness and using (or building upon) existing standards. One requirement was conformity with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). In building a schema for Timescapes, we used UKDA standards as a starting point. The study description and catalogue record created at the UKDA for each dataset follow the international standard for social science data, the Data Documentation Initiative (DDI). The study metadata is also mapped to the Dublin Core standard, and is compliant with the Open Archives Initiative (OAI) and Z39.50 for metadata harvesting and sharing. The UKDA is generally compliant with the OAIS Reference Model with some "additions and alterations" based on the specific type of material processed (UKDA Preservation Policy, 2008). Timescapes is attempting to both meet existing requirements and to anticipate changes (e.g., expanded use of XML, METS and multi-media data) at the UKDA.

The Timescapes metadata schema has been produced by the project's technical officer in consultation with the Leeds Library staff members. The metadata standards used were chosen because they are being actively used and they are supported by organisations that are responsible for creating standards for digital archives and preservation systems. Image metadata is captured using the "NISO Metadata for Images in XML" (MIX) standard developed by the Library of Congress . Preservation metadata is captured using the "Preservation Metadata: Implementation Strategies" standard developed by the Library of Congress . As with the MIX standard, DigiTool can automatically generate PREMIS-compliant metadata recorded during the deposit and ingest process.

Descriptive metadata is used to support several methods for searching the Timescapes archive. The first is the use of logical collections which are a feature of DigiTool that allow the setting up of "slices" (e.g., women) through the metadata in the form of predefined searches on metadata fields. We have also created a set of baseline metadata elements that will be used to support the creation of logical collections and other searching. This will be mapped to the Metadata Object Description Schema (MODS) standard developed by the Library of Congress. We have also created a specification for detailed metadata providing very detailed information about the subject (personal characteristics, employment, education, living arrangements and so on). An XML schema is in development that will be used to capture this metadata. We are using Microsoft InfoPath software to create a user-friendly form and interface to capture this metadata.

Metadata foregrounds researchers' roles in archiving process
A central focus of work has been diverse efforts to engage researchers in collecting and providing metadata. Activities

included providing researchers with transcription and spreadsheet templates and involving researchers in defining the descriptive metadata and the initial filters to appear on a resource discovery page.

What emerged from the experience with metadata in the Timescapes project is the extremely complex nature of the coordination required among the agents. The first point of timing coordination was between the institutional repository and the domain repository. We began the metadata schema intending to follow and, as much as possible, fully replicate standards in use at UKDA. This was largely successful, but illustrative differences appeared. First, the UKDA had not settled on metadata specifications for audio and video. And though it is in the process of doing that work, the scheduled completion date is later than what Timescapes required. The situation is similar regarding item-level metadata. This is metadata that applies to a unit smaller than the full dataset or collection, such as a single interview for a qualitative collection. Currently, the UKDA provides some item-level metadata in the form of a spreadsheet that is part of the documentation for a study. This metadata does not enable the user to search for key categories (e.g., gender) across all collections. (Qualidata Online does have this capability, but currently holds only a small share of total holdings.) Work is proceeding in these areas, but it is not expected to be completed in time to meet Timescapes need for item-level metadata for resource discovery and, even more urgently, for access control.

Finally, there is the role of the researchers and the research life cycle to add to the mix of institutional repository and domain repository data life cycles. Because researchers are using a flexible, emergent model for data collection, they can not be certain about the scope of data to be collected. This posed a challenge when the metadata specification called for some points to be nailed down earlier than was comfortable for some researchers. For example, in choosing filters for resource discovery, researchers complained that they had not yet started collecting data, how could they be expected to know how to search it or to define key subject categories? Ultimately, these requirements did not cause lasting problems. Although initial choices had to be made to develop the search, few of these choices are permanent. It is clear that a key point of communication involves making clear to researchers that decisions must be made, but also making clear which choices create "lock-in" and which are more flexible. Generally, this has been achieved by good communication among the research archivist, technical officer and researchers.

Ethical data sharing
The collection, use, publication and dissemination of data is subject to an extensive array of guidelines for its legal and ethical use, ranging from requirements of the Data Protection Act, to Review Ethics Committees' requirements, to guidelines issued by various professional bodies such as the British Sociological Association and the Medical Research Council. Although Timescapes data pose particular ethical challenges, even relative to qualitative data generally, the project is explicitly focused on developing strategies to enable sharing and preservation of even this most challenging data.

Domain repositories and the UKDA in particular, have long history and great expertise in finding ways to ethically share and reuse data. There are three inter-related strategies available to make data shareable: gaining informed consent for sharing and archiving, altering data to protect identities (e.g., anonymisation) and controlling access. Timescapes is following the UKDA model of integrating all three of these strategies and extending UKDA procedures in selected areas to address the particular requirements of qualitative longitudinal data. This paper looks in detail the area of informed consent with a brief note on anonymisation.

Informed consent is an ethical requirement for most research. It must be considered and implemented throughout the research life cycle, from the inception of planning to publication and including making provision for data sharing. Researchers and archives share strong commitments to ethical use of data, from point of collection through to preservation. But as we have seen with metadata, agents occupy different locations in research and data life cycles which influence the particulars of how these commitments are perceived and acted upon.

If research data to be archived at the UKDA contains personal or sensitive data about informants, explicit consent is needed, ideally in writing, for such data to be processed by UKDA. If a person can not be identified—if the data are fully anonymised—then the DPA no longer applies. The challenge for much qualitative data is that it can be difficult to assure absolute anonymisation, and thus the safest stance from a legal perspective is to have written consent in place.

For researchers, the legal framework is far from clear (Thomas and Walport, 2007). The vast majority of researchers are deeply committed to ethical use of data and see these standards as more binding than any legal formalities. However, it is not surprising that researchers emphasise their areas of experience: recruiting, contact with participants, and publication. In most cases, their attention to data sharing is a lower priority. As more funders recommend data sharing, this issue is growing in prominence, however, most funder mandates remain voluntary recommendations with infrequent formal enforcement.

The process of producing a model consent form acceptable to all the projects took a great deal of time, including drafting the form, holding consultations, incorporating feedback and keeping the wording of the Timescapes form aligned with changes and updates in UKDA policies. This process of engaging researchers has yielded a standardised consent form that covers areas of consent for participation, research outputs and data sharing and archiving. The outcome has been positive, though not ideal: most but not all the teams have agreed to use the form, although some will use recorded verbal consent or defer seeking consent for archiving.

A similar issue arose concerning the second strategy for data sharing: anonymisation. A set of guidelines were drafted with instructions about what content to anonymise and formats for doing so. The system had to be easy to teach to transcribers (some projects works with pools of transcribers and have little or no control over their quality), and to make it possible to easily convert files to XML.

As with the metadata case, informed consent showed that understanding agents' locations in their respective life cycles is essential for successful collaboration. The UKDA's duty is to meet legal requirements regarding data sharing, and it recommends written informed consent. Researchers want to minimising burdens on participants and thus they prefer fewer formal procedures and more discretion. As in the metadata example, the role of researchers as participants in design processes featured in dealing with consent.

Preservation
It is probably the area of preservation where the Timescapes archive is most clearly attempting to follow the strategy of "cooperation and specialisation" advocated by Green and Gutmann. The logic behind this strategy is to maximise effective use of resources, and in theory, it is a compelling argument. At the experiential level in Timescapes, we are convinced is it the right approach, however, it has proved important not to underestimate the challenges of coordination.

The platform for the Timescapes archive, LUDOS, is a development involving the University Library and Information Systems Services. The time invested has been well spent, enabling the creation of a partnership for the parallel development of the platform and archive.

Nonetheless, there have been hurdles to overcome. Timescapes is committed to compliance open source, non-proprietary tools where available. However, the MIDESS project at Leeds was in place and already running DigiTool, a proprietary package. Choosing a new software platform such as Fedora would have forced Timescapes outside of the existing Library-managed repository system at Leeds. The Library stood to benefit from a large project

committed to depositing in its repository and Timescapes needed an institutional base at Leeds. We deemed that a vital key to long-term sustainability for the Timescapes archive would be its embeddedness in the wider Library and IS infrastructures at Leeds, and we agreed to accept the limitations of the proprietary software application.

Regarding the UKDA relationship, the agreement to collaborate with the UKDA for preservation was obvious; as an ESRC funded project, Timescapes is obliged to offer its data for deposit. Irrespective of any mandate, the UKDA goals of preservation, balancing authenticity with usability are shared by Timescapes, making the UKDA an ideal long-term home for the collection.

Timescapes is currently actively engaging with the UKDA, taking the lead in some areas and following UKDA policy in others. For now, we are following a "pull" model, that is, letting the service provider offer guidance on preservation, then attempting to implement the necessary procedures as possible within the Timescapes project. In addition to the metadata cooperation already discussed, we have recently sent data already deposited at UKDA (from a "feeder" project for Timescapes) to Leeds. The data and metadata are being used to test the draft metadata schema and InfoPath ingest and deposit procedures. The objective is that Timescapes data will be processed to a very high level of quality, standard compliant, and more "deposit-ready" than the typical dataset received at UKDA.

Researcher-centred archiving
Green and Gutmann's (2007) central insight is their understanding of critical roles for all three agents, researchers and institutional and domain repositories in preservation and situation those agents in their own life cycles. They focus in particular on the ways institutional and domain repositories can collaborate. Timescapes extends this model by highlighting researchers' roles across all phases of data preservation and sharing, including selecting which data are to be preserved, enriching that data with contextual and other metadata, and identifying and promoting opportunities for reworking the archived data. In doing so, Timescapes is (implicitly if not always explicitly) following some principles of cooperative, or participatory design, at least to the extent that users' (i.e., researchers') needs are met and that high standards of usability are achieved.

Typically, participatory design projects have a dimension of user-empowerment. This is a consideration in Timescapes in the following way. Repositories and archives may be used in a managerialist fashion to reduce researcher autonomy. This can come about through greater centralised control of research and teaching resources by requiring sharing on conditions not discussed or negotiated with researchers. One objective of Timescapes is to promote alternatives to this top-down model. This is not to say that

researchers' interests should dominate the world of data sharing, but they should be equal partners—along with repository and archive experts, university administrators, and others, including some segments of the public with interests in data sharing—in the development of such systems.

In addition to engaging users in metadata collection, interface design, and consent and anonymisation guidelines, Timescapes is also seeking researcher involvement very early in promoting the archive for reuse. One of the sychronicities of Timescapes is that fact that we are planning and designing for secondary analysis while primary data are being collected. This is unconventional and uncomfortable for researchers. Their response is (though usually voiced more diplomatically): "why hassle me about reuse now; I have not even recruited my sample for the primary research yet?" It is hoped that this early planning will help to assure an active community of researchers committed to reusing data as soon as materials become available.

Several factors about Timescapes may enable this vision to be achieved. The programme will produce the largest dataset of its kind with a substantive focus on the life course and a methodological focus in qualitative longitudinal methodology. These thematic foci allow Timescapes to target potential re-users from an existing community of researchers, many of whom already have a history of sharing and collaboration.

Our strategies for building a community of users include encouraging affiliated projects (where affiliates will be required to deposit and share their new data and to re-use Timescapes data), securing secondary analysis studentships, and providing mobile and in-house training workshops and a help desk. We also aim to showcase data sharing and re-use among the seven Timescapes projects. Data sharing sessions have already taken place among projects with common interests (e.g. parenthood, childhood, older lives).

Active promotion of the archive as a specialist data resource has begun, starting with an extensive consultation with potential end users. However, Timescapes is already acting as a magnet for researchers interested in developing affiliated projects. The key to this interest lies in the thematic focus of the archive, enabling researchers who share broad substantive interests (e.g., family, relationships, life course) to also share data management protocols, methodologies, research data, and outputs.

Conclusion

Green and Gutmann have been excellent tour guides and their map has proved itself an accurate one as Timescapes has navigated the difficult terrain of archiving qualitative longitudinal data. They clearly grasp the need for all three agents, researchers, institutional repositories and domain

repositories, to engage as equal partners in producing and sharing data. Particularly useful insights are gained by situating agents in the contexts of their local practices: research and data life cycles.

The Timescapes archive has deepened the usefulness of this framework by providing details of how cooperation and specialisation can work in a live project. And Timescapes has extended their model by using the case of qualitative longitudinal data to demonstrate the necessity for finely-tuned timing if coordination is to work. The metadata example showed how this coordination can happen. In its role as domain expert, the UKDA has demonstrated the value of relying on international social science standards such as DDI. Timescapes is advancing development in the areas of audio and video, specific to its needs. The institutional repository at Leeds, by working closely with researchers, has greatly expanded and regularised the metadata that will be collected for Timescapes. This can only enhance resource discovery for the Timescapes data and the potential for comparative, mixed methods research with other, large-sample quantitative, data. Informed consent demonstrated the power of specialisation with the UKDA focused on compliance with complex legal requirements and researchers on their ethical responsibilities. Finally, the model of disaggregated preservation adopted clearly showed that institutions can specialise, yet still work together to use valuable data sharing resources in the most efficient way possible.

Even the best maps and guides can not remove every obstacle from a journey. Coordination among three agents is likely to be challenging under any circumstances, and becomes more so when the trajectories of different life cycles, for research projects and for data sharing, are considered. Timescapes exposed some structural differences that, although they can be managed, can not be removed or eliminated. Repositories, both institutional and domain, tend toward needing fixity and formality in areas such as standard and guidelines. Timescapes methodology tends toward less fixity and formalisation, especially in early phases of work. So researchers, by and large, are pushing to keep things loose while repositories need to nail things down. The tension is, of course, compounded in Timescapes because of its commitment to engage researchers throughout the data sharing process.

What have we learned so far? First, that the tensions of coordination are not "solvable"; no amount of planning or anticipation will remove them. The tensions are inherent in the different roles and perspectives of the various agents. If all agents need to participate (and all evidence suggests the benefits are worthwhile), then effort has to be put toward managing the tensions constructively. This management takes resources, and that much at least can be planned for. Cross-institutional teams need to be used with regular, substantive meetings, not mere semi-annual formal

sessions.

In the face of day to day frustrations, much remains positive. If we are to create archives that researchers will not merely use, but actively support and fight for, then these archives have to be built, from the beginning, with researcher input. Equally important, if those archives are to be sustainable and obtain long-term funding, we have to embed archives in on-going institutions and demonstrate efficient use of all-too-scarce (and given prospective economic conditions, likely to become more scarce) resources for this valuable endeavour.

*Contact: Libby Bishop, Research Liaison Office-UKDA at University of Essex and Research Archivist at University of Leeds. E-mail: e.l.bishop@leeds.ac.uk

References
Beagrie, N., Chruszcz, J., and Lavoie, B. "Keeping research data safe." JISC Report, May 2008. http://www.jisc.ac.uk/publications/publications/keepingresearchdatasafe.aspx

Berg, B. Qualitative Research Methods. Boston: Pearson Education, 2004.

Bishop, L. (forthcoming 2008) 'Archiving for the future: the archivist as researcher' In: M. Brugidou, et al. (eds.) Secondary Analysis in Qualitative Research: Challenges for Human and Social Sciences. Paris: Lavoisier.

Bishop, L. (2006) "A proposal for archiving context for secondary analysis". Methodological Innovations Online 1(2). http://sirius.soc.plymouth.ac.uk/~andyp/viewarticle.php?id=26.

ESRC Data Policy, April 2000

http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/DataPolicy2000_tcm6-12051.pdf).

Fielding, N. (2004). 'Getting the most from archived qualitative data: epistemological, practical and professional obstacles', International Journal of Social Research Methodology, 7(1), pp. 97-104

Green, A. and Gutmann, M. (2007) "Building partnerships among social science researchers, institution-based repositories and domain specific data archives". OCLC Systems & Services: International digital library perspectives 23(1): 35-53. http://deepblue.lib.umich.edu/handle/2027.42/41214 [Open Access version]

Gibbs, H. (2007) DISC-UK DataShare: State-of-the-Art Review. DISC-UK, August 2007. http://www.disc-uk.org/docs/state-of-the-art-review.pdf

Heery, R. and Powell, A. (2006). Digital repositories roadmap: looking forward. Bath: UKOLN/Eduserv. http://www.ukoln.ac.uk/repositories/publications/roadmap-200604/

Hitchcock, S., Brody, T., Hay, J.M.N., and Carr, L. (2007) "Digital preservation service provider models for institutional repositories: towards distributed services," D-Lib Magazine 13(5/6), January 2007. http://www.dlib.org/dlib/may07/hitchcock/05hitchcock.html

Joint Information Systems Committee (JISC), Digital Repositories Programme. (2005) Digital Repositories Review.

http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/digitalrepositoriesreview2005.pdf

Knight, G. (2005) "Sherpa-DP OAIS report: An OAIS compliant model for disaggregated services". Arts and Humanities Data Service Report. http://ahds.ac.uk/about/projects/sherpa-dp/sherpa-dp-oais-report.pdf

Knight, G. and M. Hedges. (2007) "Modelling OAIS Compliance for Disaggregated Preservation Services." The International Journal of Digital Curation 2(1) June 2007. http://www.ijdc.net/ijdc/article/viewArticle/25/0

Lavoie, B. (2004) The Open Archival information System Reference Model: Introductory Guide. Digital Preservation Coalition (DPC) and OCLC. http://www.dpconline.org/docs/lavoie_OAIS.pdf

Lyon L. (2007) Dealing with data: roles, responsibilities and relationships, Consultancy Report. Bath: UKOLN. http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf

Mason, J. Qualitative Researching. London: Sage, 2002.

Thomas, R. and M. Walport. Data Sharing Review. Ministry of Justice. http://www.justice.gov.uk/reviews/datasharing-intro.htm

MIT Libraries. PLEDGE. [accessed 16 Sept 2008]. Available from http://pledge.mit.edu

Ryan, B. & Walmsley, S. (2003) Implementing metadata collection: a project's problems and solutions. Learning technology, Vol. 5, no. 1, Jan. 2003. http://lttf.ieee.org/learn_tech/issues/january2003/index.html#3

SherpaDP2. Project Overview. [accessed 25 Nov 2008]. http://www.sherpadp.org.uk/sherpadp2.html

Silverman, D. Qualitative Methodology and Sociology. Aldershot: Gower, 1985.

UK Data Archive. Preservation Policy, 2008. http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0308.pdf

Footnotes

1 http://ludos.leeds.ac.uk/ludos/

2 http://www.loc.gov/standards/mix/

3 http://www.loc.gov/standards/premis/

4 http://www.loc.gov/standards/mods/