| **Final report cover sheet** **IE Programme: Repositories: start-up** | JISC |
|---|---|

| **Name of Proposed Project:** | **Lifespan Initiative for the Research and Data Archive Repository (Lifespan RADAR)** |
|---|---|
| **Name of Lead Institution:** | Royal Holloway, University of London |
| **Name(s) of Project Partner(s):** | Bifulco, Smith, Aguilar, Hannan (Lifespan, H&SC, RHUL), Pope (IT, RHUL) |

**Full Contact Details for Primary Contact:**

Name: Professor Antonia Bifulco            Position: Director of the Lifespan Research Group

Email: a.bifulco@rhul.ac.uk

Tel: 0207 307 8615                         Fax: 020 7636 2268

Address: Lifespan Research Group, Royal Holloway, University of London, 11 Bedford Square, London WC1B 3RF

| **Length of Project:** | **1 year** | |
|---|---|---|
| **Project Start Date:** 1 April 2009 | **Project End Date:** 31 March 2010 | |
| **Total Funding Requested from JISC:** £30,000 | | |

**Funding Broken Down over Financial Years (April - March):**

| **April 09 – March 10** | **April 10 – March 11** | **April 11 – March 12** |
|---|---|---|
| £30,000 | none | none |

**Document information:**

**Final report**

**Authors: Ananay Aguilar and Antonia Bifulco**

**Date: Date 31 March 2010**

**File name: 100331 Final report**

**URL: http://lifespancollection.org.uk/blog/**

**Project abstract**

The project explored the application of an information environment for research data collections in the frame of the Digital Object Repository (DOR) held by Royal Holloway, University of London. The research data for the project was provided by the Lifespan Research Group, which, in line with the prevailing research councils' policies on data sharing and preservation, has engaged in the project to offer its vast audio-tape and interview schedule collection to a wider research community. The project involved the design of a digitisation strategy that embraces and defines best-practice solutions for long-term digital data preservation and the creation of processes of submission of, and accessibility to, current and future critical datasets that ensure compliance with data security, copyright legislation, licensing, and associated audit functions.

# Lifespan RADAR
# Lifespan Initiative for the Research and Data Archive Repository

# JISC Final Report

**Ananay Aguilar and Antonia Bifulco**
**Contact: a.bifulco@rhul.ac.uk**
**Date 31 March 2010**
**URL: http://lifespancollection.org.uk/blog/**

# Table of Contents

## Acknowledgements

## Executive Summary

Turning the Lifespan Collection (LC) into a digitally searchable and accessible collection for secondary analysis is a project that has long been overdue. The LC can be seen as the last and arguably the most ambitious initiative that arose from a longer tradition of using life history approaches and socio-historical contexts to understand psychosocial risks for psychiatric disorder in the community. It includes an intergenerational sample of almost 600 family members selected through members living in North London in the 1990s, who gave their life stories from childhood to the present day in audio-taped interviews. In addition to the specific research interests of the original projects, the collection now represents a highly interdisciplinary resource, with potential for use across the humanities and the social and health sciences.

This project explored the application of an information environment for research data collections in the frame of the Digital Object Repository (DOR) held by Royal Holloway, University of London. The research data for the project was provided by the Lifespan Research Group, which, in line with the prevailing research councils' policies on data sharing and preservation, has engaged in the project to offer its vast audio-tape and interview schedule collection to a wider research community. The project involved the design of a digitisation strategy that embraces and defines best-practice solutions for long-term digital data preservation and the creation of processes of submission of, and accessibility to, current and future critical datasets that ensure compliance with data security, copyright legislation, licensing, and associated audit functions.

Our first task was to finalise the website with detailed information on the LC and to create a blog that would inform the public of the progress on the JISC-funded Lifespan RADAR project[1]. The creation of the website was fundamental to creating an overview of the large collection and associated research findings and projects. The blog allowed us to trace all events related to the Lifespan RADAR project, while offering direct research outputs to the public.

The project then went on to focus on its main task: to explore all aspects of digitising the LC and to make it available to the broader research community. For the development of the digitisation workflow, including an encompassing metadata schema and an appropriate ethical and legal framework, we consulted available best-practice solutions and shaped them into tools suited to the needs of the LC with the help of the IT Department. For the ethical and legal considerations the available literature was less effective. This is because guidelines generally refer to born-digital archives or research projects at planning stage, and less to collections where returning to copyright owners may not be cost-effective. In order to reach conclusions for future action, we engaged in scrutinising relevant legislation, organised a study day and sought professional supervision, as well as testing in ethics applications the feasibility of some re-use projects.

This exercise and the collaboration with the IT Department translates into a precedent for the DOR in creating space for research data collections.

Our aims and objectives led to the following achievements:

- Through discussions and the disseminating work, the collection can now be seen as a viable highly interdisciplinary resource, with potential for use across the humanities and the sciences, broader than the specific research interests of the original projects.

- Lifespan has increased academic awareness outside its immediate environment, of the unique and irreplaceable nature of its data. In the climate generated by the ESRC's Dataset Policy of preserving and disseminating research data for secondary analysis, it is fundamental to raise awareness of individual collections and create a culture of re-use.

- Lifespan has successfully engaged in the debate of data re-use and secondary analysis now active in the social sciences. The consideration of important legal and ethical aspects renders a unique perspective to the discussion, to provide pathways for other such projects.

- Lifespan RADAR has provided an initiative for similar collections to be archived within RHUL and beyond. Reports of the start-up experience both illustrate, and contribute to, the exemplar implementation of the tools and solutions already delivered by JISC and other institutions. In

---

[1] http://lifespancollection.org.uk/blog/

particular, Lifespan RADAR dealt with the contentious issues surrounding sensitive research data (for example involving mental health data and early life abuse) for which seeking renewed consent or copyright may not be cost-effective. Other projects at RHUL wishing to submit to the Digital Object Repository, as well as similar projects within the community will be able to learn from this experience.

- These activities raised the profile of the LC, and the Lifespan Research Group and its distinctive methodology more generally, within the wider research community.

- While samples taken from the LC have been used for training in the measures developed by the research group, the project has allowed us to explore the possibility of future postgraduate educational use of the materials. Interactive teaching tools would be one of the resources potentially available in the near future.

- To test the ethical rigour of the approach, three test projects were devised, two with outside agencies/universities for re-using parts of the LC data under the supervision of the original PI (Bifulco). The third examined its use within RHUL for educational purposes. Initial response from the Royal Holloway Ethics Committee is favourable, and we await the final confirmation of permission. This sets a precedence for ethical agreement for the data re-use which can be built on for future uses.

- Two applications to further the survival of the collection have been, or are, underway. The first involved an unsuccessful application to the Wellcome Trust Research Resources in Medical History. This was to undertake the digitisation and preservation of a significant section of the LC. It was to involve the extraction of quantitative data for use as reference tool for the digitised audio files and an interactive web portal for the exploration of the potential for research by external researchers. The second application is to the MRC Methodology Research Programme - research on methods development for analysis of secondary studies. This is to utilise the LC as a base for further developing the lifespan interviews on social adversity, to new modes involving on-line interview applications. This will involve the re-contacting of the LC sample for validating the new modes of interview and the use of LC data as comparison to the other psychiatric patient groups investigated using the new modes. The project application will be submitted, in collaboration with KCL in June 2010.

## Background

The Lifespan Research Group (Lifespan) is the descendant of the research group directed by Professor George Brown, one of the pioneers of social science, whose distinctive contextualised life history interviews utilised in the study of psychosocial risks for psychological disorder in the community is of international standing. The data collected from 1990-1999 and held by Lifespan is a unique data set for its provenance, size and content. It was funded by two Medical Research Council programme grants (worth a total £3.5 million) and includes around 3400 hours worth of audio-taped interviews with individuals and families living in the community. This includes rated schedules in the form of paper records (both handwritten and electronically word processed) and an electronic quantitative data set. This collection is now held by Prof. Bifulco at Lifespan where her group has established a strong reputation in the clinical psychology field as well as in applied research with health and social care services, and knowledge exchange through training programmes, evaluation of public and social services, and practice partnerships. It is therefore timely to join the current climate for archiving and data re-use generated by the ESRC's Dataset Policy by preserving and disseminating this valuable collection for secondary analysis. Related efforts include the establishment of The Life History, Health and Identity Research Cluster, composed of researchers primarily at RHUL in collaboration with colleagues at Universities of Manchester, Leeds and Essex, who have been drawn to the LC's resource for secondary analysis.

At the time of our application, the IT department had purchased and began to implement a blended repository to support the storage, curation, discovery and delivery of e-library, e-learning and e-research digital assets at Royal Holloway. The implementation project has most recently delivered open access research publications and past exam papers but has not yet included research data sets. Therefore, our project was designed to offer an opportunity to develop a scaleable storage infrastructure to cope with these higher data volumes and investigate how well the repository platform would fit with the preservation strategies and data policies of this type of data. In this way, the project was designed to help IT meet the college's strategic aim to provide an integrated information environment that would fully support the complete research lifecycle from project inception to publications. Finally, we aspired to offer both the LC resource and the experience of scoping its archiving and re-use experience to the wider research community.

## Aims and Objectives

Whilst the aims and objectives of the project remained unchanged throughout the year, the order and focus changed as it progressed. In particular, the decision was made to first concentrate on the dissemination exercise, since without evidence of outside interest in the collection by the academic community, the project would be redundant. The second priority was to concentrate on the ethical and legal issues surrounding the collection due to the potential obstacles envisaged, since without this the project could not progress. The third priority was to develop the metadata schema in order to test the feasibility of archiving such a large collection with suitable access. The objectives that achieved least input due to foreseen risks related to the direct input of our host institution (RHUL) to developing the digital storage, as outlined in our application and project plan. However, we continue pursuing this objective.

More details on the individual objectives, are as follows:

### 1. Explore options for use in new research projects and create appropriate networks

Lifespan RADAR focussed on the exploration of options for use in new research projects and the creation of appropriate networks. At the start of the project, members of our team attended various events discussing the implications of digitising and archiving research data, including re-use and secondary analysis, and the economies and politics of memory circulation. Highlights included two conferences: 'Digital economies and the politics of circulation', at Columbia University, USA and 'Archiving and Reusing Qualitative Data: Theory, Method and Ethics across Disciplines, organised by CRESC[2], at Manchester University. We were invited to provide a seminar at Manchester University, hosted by CRESC, where we had the opportunity to explore the interests and needs of the academic

---

[2] ESRC Centre for Research on Socio-Cultural Change: http://www.cresc.ac.uk/

community for the implementation of a framework with interdisciplinary considerations, illustrated by the LC. In November 2009, we organised a study day on 'Research, access and consent', specifically dealing with exploring the ethical and legal issues surrounding research data collected in the past. All these events led to collaborations with established institutions dedicated to archiving and sharing research material (Timescapes, University of Leeds, Qualidata – University of Essex and CRESC, University of Manchester), and selected pilot-studies in collaboration with external research agencies.

## 2. Research best-practice preservation solutions with appropriate legal and ethical considerations

In order to offer the LC for secondary analysis, the sensitive and retrospective nature of the LC (for example covering childhood abuse, and lifetime mental health issues) demanded consideration of an appropriate ethical and legal framework that was greater than expected. The participation in, and organisation of, events centred on the implications of digitising and archiving research data, highlighted the scant literature and expertise in the ethical and legal implications of these activities. This provided an opportunity to lead relevant discussions within the wider research community and generate interest amongst researchers in various related fields. Our research findings have led to an authoritative statement of our ethical and legal position. We have tested this in a preliminary Ethics Application to Royal Holloway Ethics Committee[3] with three examples of re-use based within RHUL but with outside collaboration (from Kings College London and British Association for Adoption and Fostering). This will be accompanied by guidelines for similar ventures engaged in the analysis of retrospective consent and risk management strategies, and finally, to a peer-reviewed publication discussing these issues in further detail.

## 3. Implement an information environment suitable to the research group's needs and epistemological underpinnings

The DOR was launched at the end of the last calendar year with a collection of RHUL's research publications and past exam papers. The need to host collections to store primary research data presented a new challenge for the repository. The first step was to draw up an ontology of the Lifespan data collection to understand the entities included and attributes involved. This allowed the IT Department to create a proof of concept collection within the repository to provide us with a suitable information environment. Once we reached a common understanding of the LC and the tools available to the DOR we could proceed to create a schema that was extensible and re-usable beyond the initial Lifespan RADAR use case. The DOR has a second proposed research data collection that can be used to test how re-usable the schema is. Additionally the college Open Access Publications Policy (OAPP) will require us from 1 September 2010 to store datasets alongside publications if required by research funders, and the experience of developing the LC will assist with this.

## 4. Lay the groundwork for and explore the potential use in e-learning and teaching

While anonymised training material using the LC has been developed and is in frequent use, the ethical and legal requirements in relation to greater use of the collection in e-learning and teaching need further investigation. This is being explored, in particular in relation to postgraduate research and assessment training for social work and clinical psychology.

## 5. Develop a framework for other similar start-up ventures within RHUL complementing the Digital Object Repository's (DOR) e-learning and teaching infrastructure

The conclusions reached through the above activities allowed the creation of useful guidelines in ethical and legal issues, as well as other recommendations in relation to the creation of an encompassing digitisation workflow and appropriate metadata schema. The repository schema can be used as a starting point for other collections containing research items, speeding up the analysis and design phases of future research data collections. RHUL has previously shared its Research Publication schema with other institutions using 'Equella' and will also be willing to share the Research Item schema, providing a useful framework for similar ventures across the Equella community, not just at RHUL.

---

[3] 'Re-using the Lifespan Collection' – 20/1/2010 – Application to RHUL Ethics Committee (A Bifulco).

**6.     Secure funding for population of the Digital Object Repository and further storage space**

RHUL has committed to enhance and implement new information environments as part of their corporate and information strategy. However, investment in appropriate storage space for audio-visual data such as ours could not in fact be accommodated in 2009-2010. This was a risk envisaged in our application. However, the experience of the LC project has ensured that analysis of storage requirements for research data sets will now be included into the repository capacity planning models to try and anticipate requirements. Such a model may help secure specific funding for large datasets that cannot be accommodated through planned capacity growth.

**7.     Create a communication layer that incorporates future interoperability with similar projects in the social sciences (specifically in the interest of Lifespan), with RHUL's strategic alliances in the South West London Academic Network (St. George's University of London and Kingston University), and other research links within the wider community**

During a period of planned merger of RHUL with St George's, University of London during this project period, the college's attention was diverted from the development of interoperable networks. With the abortion of the merger and with funding cuts to universities, this has forestalled development of interoperability with the further obstacle of reduced independent funding. We are committed to search for alternatives to fund the development of the LC and have put forward one application to the Wellcome Trust Research Resources in Medical History in June 2009 for digitising the collection, which was unsuccessful, and are planning one to the MRC Methodology Research Programme in June 2010 for re-using the collection with the aim of also developing electronic transcripts of the audio-taped material.


## Methodology

For the development of the digitisation workflow, including an encompassing metadata schema and an appropriate ethical and legal framework, we consulted best-practice solutions where available. In particular, for the digitisation workflow we corresponded with the British Library Sound Archive and independent engineers. Relevant literature accessed included the IASA guidelines. For the ethical and legal considerations the available literature was less pertinent. This is because available guidelines generally refer to born-digital archives or research projects at the planning stage, and less to collections where returning to copyright owners may not be cost-effective. In order to reach conclusions, we engaged in scrutinising relevant legislation, organised a study day and sought professional supervision.

In relation to the development of the DOR and the implementation of our strategy, the project counted with the help of Alison Pope, the Business Analyst at the IT Department, who has been in charge of launching the DOR. Royal Holloway currently has two live collections: Royal Holloway Research Online containing research publications and a digitised exam paper collection. Royal Holloway Research Online (RHRO) uses a custom scheme that was strongly informed by the Scholarly Works Application Profile (SWAP), which itself builds upon the Functional Requirements for Bibliographic Records (FRBR). When presenting metadata for harvesting via OAI-PMH this custom schema is transformed into the oai_dc schema based on unqualified Dublin Core. It has not yet been considered transforming this schema into other standard formats. The exam paper collection uses a simple Dublin Core based schema and is not exposed outside of the institution.

The need to host collections to store primary research data presented a new requirement for the repository. Our approach was to reach a common understanding of the ontology of the LC and of the tools available to the DOR. The initial draft schema was still too oriented towards the specific Lifespan ontology to be re-usable as a base schema for other collections. We therefore began to think of the object as a generic research item with curation and preservation needs, so we used existing knowledge, along with the research done as part of the Lifespan project to think about how a new schema for a research item might work. In developing the schema it was necessary to balance conforming to standards and adopting best practice with the implementation constraints posed by our specific repository tool and user requirements. The resulting draft schema will therefore not be a pure implementation of any one approach but will adopt aspects drawn from research information

management, research publication and preservation models to create a schema for a repository collection that is both interactive and accessible, and aware of the ethical and preservation issues around such a collection. We take an iterative approach to schema development so this initial draft schema will be refined throughout the implementation cycles, starting with a small proof of concept implementation.

## Implementation

### 1.    Dissemination

Lifespan RADAR focussed on exploring the options for use in new research projects and the creation of appropriate networks. This involved participating in and organising events to generate general discussions about archiving and re-use of research data and specific discussion about the LC.

**1.1**    The creation of a **website** describing the Lifespan Collection provided a first portal to our project. Our entire progress was then recorded in our **blog** (http://lifespancollection.org.uk/blog/). It includes our various reports, including that of the digitisation strategy, the developments in the creation of a metadata schema and the final report. Our ethical and legal position will be available in due course.

**1.2**    **Participation in events** included:

- Conference on 'Archiving and Reusing Qualitative Data: Theory, Method and Ethics across Disciplines, organised by CRESC, at Manchester University, 19-20 March 2009

- Conference on 'Digital economies and the politics of circulation', at Columbia University, New York, 3-4 April 2009 (participation was co-funded by the Music Department at RHUL)

- 'Research in the open: How mandates work in practice', Repositories Support Project, RIBA, 29 May 2009

- Information Environment and VRE programmes start-up meeting, JISC, Leicester University, 7-8 July 2009

These events offered a wide perspective on the economies and politics of memory circulation, the practicalities of technical implementations and the ethical and legal considerations. With a better picture of the current debates we were in a position to focus our own debates.

**1.3**    We provided a **seminar at Manchester University, hosted by CRESC**. CRESC has participated in the debate of re-use and secondary analysis with highly regarded publications and has engaged in secondary analysis of research material held in archives such as the Mass Observation Archive. With this in mind, Prof Bifulco presented the background of the LC, research findings, current implementations and ideas for future use. Dr Smith focussed on a social and oral history approach to the LC, one of many interdisciplinary approaches. Ananay Aguilar and Leonie Hannan presented the JISC-funded project, Lifespan RADAR, and its aim of creating a framework for offering the LC to the wider research community for re-use. The seminar gave us the opportunity to explore the interests and needs of scholars in a range of fields for the implementation of a framework with interdisciplinary considerations for re-use. It was also essential in generating interest in the LC and Lifespan's activities more generally.

**1.4**    Preparation of the **study day 'Research, access and consent'**. See below for details.

**1.5**    As part of our efforts in disseminating our findings, we plan to publish a case study in dealing with ethical and legal issues in a **peer-reviewed publication**. The following journals have been investigated as relevant to this issue in terms of their remit and we will submit our paper to one or more of the following academic journals.

- Sociological Theory
- Social Psychology Quarterly
- Theory, Culture & Society: Explorations in Critical Social Science
- Theory and Society
- Qualitative Inquiry
- International Journal of Social Research Methodology
- Quality and Quantity: International Journal of Methodology

- International Journal of Qualitative Methods
- History of Psychology
- Forum Qualitative Sozialforschung / Forum: Qualitative Social
- Qualitative Research Journal

The cost of this activities reaching beyond the formal lifetime of this project has been accommodated in the final budget report and entails Research Assistant time until the completion of the paper and its first submission (see financial report).

## 2.     Research, access and consent

In order to offer the LC as a resource for secondary research analysis, the sensitive and retrospective nature of the LC demanded careful consideration of an appropriate ethical and legal framework. In order to share our ideas, capture expert knowledge and gain feedback, we organised a study day on 'Research, Access and Consent' in Central London. The preparation of the study day, both our own presentation and the invitation of keynote speakers, highlighted the relatively narrow literature and expertise in this field. Thus, the day provided an opportunity to lead relevant discussions within the wider research community and generate interest amongst researchers in various fields. The day was well attended with participants from diverse disciplines (social history archivists; social scientists; psychiatric epidemiologists; digital archiving specialists; legal and ethical experts). The study day proved to be successful in generating interest in the LC for a diverse strand of research purposes.

In preparation for our presentation, we investigated the ethical permissions and consent forms utilised 20 years ago when the LC originally collected, studied relevant legislation (in particular, the 1988 Copyright, Designs and Patents Act and the 1998 Data Protection Act), and corresponded with the legal IP and Contracts Manager at RHUL.

Professor Bifulco chaired the morning session and after outlining the features of the LC with Aguilar and Hannan, speakers included Dr Philip Johnson (RHUL), Dr Libby Bishop (Timescapes, University of Leeds & Qualidata, UKDA, University of Essex) and Dr Niamh Moore (CRESC, University of Manchester). They presented their own views on the ethics of reuse and discussed the various legal and ethical implications of re-using research data. (This including the ethical implications of destroying unique data collections funded with public money). Discussions centred on the agency of research subjects versus the authority and independence of researchers, risk management issues, including the implementation of levels of access and practical issues in re-contacting the research participants for permissions to archive and re-use their interviews,

Dr Graham Smith (History Department, RHUL and co-investigator of the project), chaired the afternoon session, which featured proposed pilot studies using the LC. While Smith's study focussed on the qualitative aspects of the oral interviews, particularly on the history of food, Dr Helen Fisher (Institute of Psychiatry, King's College, London) used the quantitative elements of the collection on for comparison with her own psychiatric data. Julia Feast and Margaret Grant, Social Workers from the British Association for Adoption and Fostering (BAAF) proposed using the methods developed by Lifespan for a study of trans-national adoption to collect adult adversity experience and to utilise LC as a matched comparison group. These studies proved both interest generated by the LC, and the feasibility of re-using the data in a wide range of topics and disciplines. This underlined the need for preserving and developing the material for new research. One outcome proposed was to put forward an ethics application to RHUL Ethics Committee to test out the ethical feasibility of re-using the data on this relatively small scale.

A report on the study day was published on our blog. The conclusions have provided an authoritative statement of our ethical and legal position for data re-use. This is to be accompanied by guidelines for similar ventures engaged in the analysis of retrospective consent and risk management strategies, and, finally, to a peer-reviewed publication discussing these issues in further detail.

## 3.     Digitisation strategy and workflow

With the help of experts at the British Library Sound Archive, and sound engineers Andrew Hallifax and the partners at 'All you need is ears'[4], along with relevant literature, we gained a thorough understanding of the possibilities at hand for digitising the audio collection. We explored the market for appropriate partners for the enormous task of digitising the 2200 tape recordings and reached

---

[4] Andrew Halifax, Stephen Greaves and Jenny Grigg, Bishopston, Bristol.

important decisions in relation to the quality of the new format and the best storage solutions. We explored a workflow that would include shipping and handling of the tapes, its transfer into the new format and the submission to the DOR at RHUL. A report on this workflow design was published on our blog and serves as an example for similar ventures. However, further examinations of ethical and legal considerations, together with funding limitations, imposed restrictions on the applicability of this preliminary design. We have now gained a wide perspective that will allow us to swiftly explore alternatives, should the opportunity arise  (for example digitising on site).

### 4. Metadata schema

A metadata schema has been designed and will be implemented in due course. We explored guidelines by the UKDA, by Richard Gartner for JISC and by the University of Queensland proved exceptionally useful. In addition, we followed metadata schemas as used by the British Library and smaller, similar, archives like Timescapes. With the guidance of the Business Analyst at the IT Department at RHUL, Alison Pope (co-investigator on project), we engaged in exploring the various possibilities of conceptualising the collection within RHUL's platform, Equella. While creating schemas that would reflect the research programme's structure or the selection of participants seemed seductive, we settled for a conservative item-based approach. The record attached to each item will thus include information in relation to the research programme, research topic, research subject, object type and manifestation. The implementation of the schema will consider access issues based on the ethical and legal position.

We have several reference models and metadata standards to draw upon, but none of which fully encompass our objectives with this collection. As such the aim of our proposed schema is to describe, curate and preserve records about a research item which can be understood in CERIF terms as a sub-class of Research Result and in PREMIS terms as an intellectual entity. The research item is a primary research data asset, created as a result of research activity and that is manifested in one of more digital or physical objects. We are beginning to form the ontological basis for such a research item, and the metadata schema, taxonomies and related repository functions necessary to implement this as a collection in Equella.

## Outputs and Results

### 1. Research, access and consent

Our most important findings relate to issues of access and consent. The characteristics of the LC created the need of an approach to the ethical and legal considerations for which there are no conclusive guidelines. Thus, Lifespan RADAR provided the opportunity to lead relevant discussions within the wider research community. Our analysis allowed us to reach conclusions as to our own collection, but most importantly, it provided the foundations to create authoritative guidelines and recommendations for similar ventures. A peer-reviewed publication will offer further summary and discussion on the implications of our approach.

Our situation can be summarised thus:

- The consent forms in existence from the original MRC data collection, complied with ethical considerations recommended at the time by the relevant research councils. They were reviewed and accepted by the MRC and the relevant Health Authorities. They are fully compliant with current guidelines, but did not include any instruction to destroy the data and asked for agreement for future follow-up to continue the research in the future.

- Under the Data Protection Act, the research material falls under the category of sensitive data, for which consent for research purposes has been sought. Under these circumstances, the data must be used exclusively for research and may be kept indefinitely for this purpose.

- Copyright in the LC is held by RHUL, apart from the interviewees' responses to the interview questions in the sound recordings.

Seeking the assignment of the copyright in the responses may not be cost-effective. This is because of the great number of individuals, the significant amount of time since the last contact and their high

mobility. Under these circumstances, working under the provision of fair dealing is a more cost-effective strategy. While the provision of fair dealing allows for great flexibility it has not clear boundaries. Therefore, a risk management approach is deemed necessary.

In a preliminary statement, which awaits final inspection by our legal advisor, we have complied with six principles which we consider deal with legal and ethical re-use issues:

1. The research activities and the research results do not at any point involve the identification of the research respondents. To reflect this, the person identification is fully coded numerically. All research records refer only to the id number. A list of names and addresses (correct in 2000) are kept entirely separately and under electronic password.

2. The Lifespan research group will continue to analyse and publish on the LC under the provision of 'fair dealing' for the purposes of private study and non-commercial research.

3. The LC will also be available for re-use provided that new research is supervised by the Director Prof Bifulco (original PI to the LC project), under the same considerations as above.

4. The Lifespan will continue the efforts to preserve the LC. The tapes are reaching their life end and the transfer do a lasting format is a matter of urgency.

5. The tapes shall not leave the premises of the Lifespan for the purpose of research, by members of the team or otherwise.

6. Once the LC becomes accessible online, the metadata schema will include several password-protected levels of access to provide security and anonymity to reflect our commitment to the points above. This includes ensuring that no copies of the interviews are available for download or in any other format.


## 2.    Design and workflow

An important aspect of our project was to explore all aspects of digitising the LC and make it available to the broader research community. We thus engaged in understanding the details involved in **digitising the collection** and creating a **metadata schema** that reflects our commitment to engage interdisciplinary interest.

This understanding and the collaboration with the IT Department translates into a **precedent for the DOR** in creating space for research data collections.


## 3.    Tangible deliverables

Our first task was to finalise the **website** with detailed information on the LC and to create a **blog** that would inform the public of the progress on the JISC-funded Lifespan RADAR project. The creation of the website was fundamental to creating an overview of the large collection and associated research findings and projects. The blog allowed us to trace all events related to the Lifespan RADAR project, while offering direct research outputs to the public. These included a report on our digitisation strategy, on our first effort on generating a metadata schema, and on our study day. Our full reports and publications will also be accessible from the blog.

The metadata schema will be implemented shortly. Considerable progress has been made on its conceptualisation, on clarifying available options and on creating a workflow in collaboration with Alison Pope. Once implemented, the schema will provide a **catalogue** with encompassing discovery tools including our various specifications. By using semantic units rather than defined metadata nodes we expect to be able to make no presumption as to the research structures, classification taxonomies, identifier types, format types, etc., that may be used and so create a schema that is extensible and re-usable beyond the initial Lifespan RADAR use case. This will then be tested with a similar collection to see how re-usable the schema is.

## Outcomes

The aims and objectives of the project remained unchanged throughout the year. However, the order and focus changed considerably. In particular, we decided to concentrate on dissemination and the ethical and legal issues surrounding the collections for the reasons outlined above. These led to important achievements that reach well beyond the development of our own work and that of the college:

- Through the disseminating work, the collection now represents a highly interdisciplinary resource, with potential for use across the humanities and the social and health sciences, broadening the specific research interests of the original projects. The interdisciplinary potential has been demonstrated through the participation of researchers in various fields within the project and the events organised. As well as scholars in psychology and genetics, these have included scholars in the fields of history, sociology and social work.

- Lifespan has increased awareness of the unique and irreplaceable nature of its data. In the climate generated by the ESRC's Dataset Policy of preserving and disseminating research data for secondary analysis, it is fundamental to raise awareness of individual collections and create a culture of re-use.

- Lifespan has successfully engaged in the debate of re-use and secondary analysis now active in the social sciences. The consideration of important legal and ethical aspects renders a unique perspective to the discussion.

- Lifespan RADAR has provided an initiative for similar collections within RHUL and beyond. Reports of the start-up experience illustrate the exemplar implementation of the tools and solutions already delivered by JISC and other institutions, but also add to it. In particular, Lifespan RADAR dealt with the thorny issues surrounding sensitive research data for which seeking renewed consent or copyright may not be cost-effective. Other projects at RHUL wishing to submit to the Digital Object Repository, as well as similar projects within the community will have this experience readily available.

- These activities raised the profile of the LC, and Lifespan more generally, within the wider research community.

- While samples taken from the LC have been used for training in the measures developed by the research group, the project has allowed us to explore the possibility of postgraduate teaching. Interactive teaching tools would be one of the resources available.

- To test the ethical rigour of the approach, three test projects were devised, two with outside agencies/universities (British Association for Adoption and Fostering and King's College London) for re-using parts of the LC data under the supervision of the original PI (Bifulco). The third examined its use within RHUL for educational purposes. Initial response from the Royal Holloway Ethics Committee is favourable, and we await the final confirmation of permission. This sets a precedence for ethical agreement for the data re-use which can be built on for future uses.

- Two applications to further the survival of the collection have been, or are, underway. The first involved an unsuccessful application to the Wellcome Trust Research Resources in Medical History[5]. This was to undertake the digitisation and preservation of a significant section of the LC with extraction of quantitative data for use as reference tool for the digitised audio files and an interactive web portal for the exploration of the potential for research by external researchers. The second project is to the MRC Methodology Research Programme (research

---

[5] Preserving the Lifespan Collection: Life history interviews of psychiatric disorder and related experiences in three generations of London families. Prof Antonia Bifulco & Dr Graham Smith (June 2009) - Wellcome Trust Research Resources in Medical History Application - Unsuccessful

on methods development for analysis of secondary studies). This is to utilise the LC as a base for further developing the lifespan interviews on social adversity in relation to psychological disorder, to new modes involving on-line interview applications. This is an issue currently needed in gene-environment studies on large numbers, where the measures of the social environment require investigator-interview approaches, but scale and respondent locations preclude the usual face-to-face interview methods. The aim is to develop new methods to be tested for validity against the original tools, by recontacting over 100 individuals from the original LC sample. Once developed these tools will be tested on two ongoing genetic studies at the Institute of Psychiatry, King's College London for both feasibility and predictive validity. This will involve the use of LC data as comparison to the other psychiatric patient groups investigated. The project application will be submitted, together with KCL in June 2010.

## Conclusions

Lifespan RADAR is an example of how a valuable research data set can be developed into an accessible collection for re-use and secondary analysis. It illustrates the different stages and aspects involved in the task and the importance of disseminating the collection from an early stage. It further shows how clearing pending ethical and legal issues is fundamental to the progression of all other aspects: the creation of an appropriate digitisation, the design of a metadata schema, and finally, the use for educational/teaching purposes. This was not clear at the beginning of the project, nor was it clear what position to take from the information available. This project not only exemplifies the importance of dealing with these issues, but showcases how to incorporate them into the project. The project thus points towards the importance of committing to a position that considers and manages the potential risk involved in offering access to sensitive research data collected in the past and will offer a report of this experience to the wider community.

The conclusions reached through the above activities allowed the creation of useful guidelines in ethical and legal issues, as well as other recommendations in relation to the creation of an encompassing digitisation workflow and appropriate metadata schema. The repository schema can be used as a starting point for other collections containing research items, speeding up the analysis and design phases of future research data collections. The DOR has a second proposed research data collection to test this. Additionally the college Open Access Publications Policy (OAPP) will require from 1 September 2010 to store datasets alongside publications if required by research funders, and the experience of developing the LC will assist with this. Finally, RHUL has previously shared its Research Publication schema with other institutions using Equella and will also be willing to share the Research Item schema, providing a useful framework for similar ventures across the Equella community, beyond RHUL.

## Implications

Lifespan RADAR offered the opportunity to launch relevant discussions on the ethical and legal considerations of making sensitive research data collection available to the wider research community. In particular, Lifespan RADAR dealt with the thorny issues surrounding sensitive research data for which seeking renewed consent or copyright may not be cost-effective. Our analysis allowed us to reach conclusions as to our own collection, but most importantly, it provided the foundations to create authoritative guidelines and recommendations for similar projects. This discussion will lead to a peer-reviewed publication to offer further discussion on the implications of our approach.

Lifespan has alerted the research community of the unique and irreplaceable nature of the LC. In the climate generated by the ESRC's Dataset Policy of preserving and disseminating research data for secondary analysis, it is fundamental to raise awareness of individual collections and create a culture of re-use. The detailed consideration of important legal and ethical aspects renders a unique perspective to the discussion of re-use and secondary analysis now active in the social sciences.

Lifespan RADAR has provided an initiative for similar collections within RHUL and beyond. Reports of the start-up experience illustrate the exemplar implementation of the tools and solutions already delivered by JISC and other institutions, but also add to it. Other projects at RHUL wishing to submit

to the Digital Object Repository, as well as similar projects in the wider research community will have this experience readily available.

The next step is to finalise the MRC methodology application to preserve and re-use parts of the collection to develop new methods to aid in genetic research. Other sources of funding will also be sought for preservation of the full collection.

## Recommendations

Resolving ethical and legal issues is fundamental to the progression of all other aspects associated with developing a fully accessible collection. This includes an appropriate digitisation workflow that doesn't interfere with the ethical and legal position, a metadata schema that incorporates different levels of access reflecting these issues, and finally, the use for future secondary research, and for teaching purposes without infringement. A number of these issues have been resolved in the scoping project and further efforts will be made to realise the full potential of the Lifespan Collection by full archiving and testing the possibility of wider re-use of the series. Further funding will be sought to realise these aims.

## References

**Guidelines**

British Academy and the Publishers Association: *Joint guidelines on copyright and academic research: guidelines for researchers and publishers in the humanities and social sciences*, April 2008, http://www.britac.ac.uk/policy/copyright-guideines/index.cfm

British Library: Archival sound recordings, http://sounds.bl.uk/TextPage.aspx?page=ethicalusage

CIBER, UCL (2009): Copycats? Digital consumers in the online age (A CIBER Executive Summary for the Strategic Advisory Board for Intellectual Property Policy), www.sabip.org.uk/sabip-ciberreport.pdf

Economic and Social Data Service, ethical and legal considerations, http://www.esds.ac.uk/aandp/create/ethical.asp

International Association of Sound and Audiovisual Archives (IASA): The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy, http://www.iasa-web.org/downloads/publications/TC03_English.pdf

International Association of Sound and Audiovisual Archives (IASA): Copyright & Other Intellectual Property Rights, http://www.iasa-web.org/copyright.asp

JISC: IPR Toolkit: overview of this toolkit and key issues, http://www.jisc.ac.uk/media/documents/publications/scaiprtoolkitoverview.pdf

JISC Legal Information: Legal Guidance for ICT Use in Education, Research and External Engagement, http://www.jisclegal.ac.uk/legalareas/copyrightipr/iproverview.aspx

National Archives: Copyright guidance, http://www.nationalarchives.gov.uk/policy/?source=ddmenu_services5

National Preservation Office: Caring for CDs and DVDs, www.bl.uk/npo/pdf/cd.pdf

Oral History Society: Is your oral history legal and ethical?, http://www.oralhistory.org.uk/ethics/index.php

RHUL position on copyright: http://www.rhul.ac.uk/research-and-enterprise/wmi/CategoryTemplate.asp?LEVEL=SECONDARY&DID=1&CID=35&TID=0 as of 1 March 2010

UKDA: Manage and share data, http://www.data-archive.ac.uk/sharing/sharing.asp

UK Government "Intellectual Property Portal" http://www.intellectual-property.gov.uk/

Welsh Repository Network, Aberystwyth University: Multimedia Deposits: Complications and Considerations with Intellectual Property Rights, http://www.wrn.aber.ac.uk/objects/multimedia/index.htm

Preservation Metadata Maintenance Activity, http://www.loc.gov/standards/premis/

UKOLN: Scholarly Works Application Profile, http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile

International Federation of Library Associations and Institutions (IFLA): Functional Requirements for Bibliographic Records, http://www.ifla.org/publications/functional-requirements-for-bibliographic-records

European Current Research Information System: CERIF 2008 1.2 - Full Data Model, http://www.eurocris.org/cerif/cerif-releases/cerif-2008/

**Statutes**

CDPA 1988: *Unofficial consolidated text of the 1988 Copyright, Designs and Patent Act amended as of May 2007*, www.ipo.gov.uk/cdpact1988.pdf

Data Protection Act 1998 (c.29)

All Acts of Parliament can be found at http://www.opsi.gov.uk/acts.htm

Information Commissioner, http://www.dataprotection.gov.uk/dpr/foi.nsf

**Archives and research centres with which we corresponded**

CRESC, ESRC Centre for Research on Socio-Cultural Change: http://www.cresc.ac.uk/

Timescapes, University of Leeds: http://www.timescapes.leeds.ac.uk/

Qualidata, Economic and Social Data Service: http://www.esds.ac.uk/qualidata/about/introduction.asp

**Academic books and journal articles**

Ethical and legal considerations

Cook, Nicholas: 'Sound law and the scholar', *Popular Music and Society*, (forthcoming)

MacNeil, Heather: *Without consent: The ethics of disclosing personal information in public archives*. Lanham: Scarecross Press, 1992

MacQueen, Hector: 'My tongue is my ain': Copyright, the Spoken Word and Privacy' (2005) *Modern Law Review*, vol 68(3), 349-377

Re-use of qualitative data

Bishop, Libby: 'A reflexive account of reusing qualitative data: beyond primary/secondary dualism', Sociological Research Online, vol.12, no.3, May 2007 (downloaded at http://www.socresonline.org.uk/12/3/1.html on 20/10/08)

Bornat, Joanna: 'Recycling the evidence: different approaches to the reanalysis of gerontological data', Forum: Qualitative Social Research, vol.6, no.1, January 2005 (downloaded at http://www.qualitative-research.net/fqs/ on 25/10/08)

Corti, Louise: Progress and problems of preserving and providing access to qualitative data for social research – the international picture of an emerging culture, Forum: Qualitative Social Research, vol.1, no.3, December 2000 (downloaded at http://www.qualitative-research.net/fqs/ on 25/10/08

Corti, Louise & Ahmed, Nadeem: Digitising and providing access to socio-medial case records: the case of George Brown's work, Forum: Qualitative Social Research, vol.1, no.3, December 2000 (downloaded at http://www.qualitative-research.net/fqs/ on 25/10/08)

Fielding, Nigel: Getting the most from archived qualitative data: epistemological, practical and professional obstacles, International Journal of Social Research Methodology, vol.7, no.1, 97-104, 2004

Fink, Anne Sofia: 'The role of the researcher in the qualitative research process: a potential barrier to archiving qualitative data', *Forum: Qualitative Social Research*, vol.1, no.3, December 2000 (downloaded at http://www.qualitative-research.net/fqs/ on 25/10/08)

Hammersely, Martyn: 'Towards a usable past for qualitative research', International Journal of Social Research Methodology, vol.7, no.1, 2004, pp.19-27

Law, Margaret (2005): 'Reduce, reuse recycle: issues in the secondary use of research data', *IASSIST*, Spring 2005, www.iassistdata.org/publications/iq/iq29/iqvol291law.pdf

Mauthner, N., Parry, O. & Backett-Milburn, K.: 'The data are out there, or are they? Implications for archiving and revisiting qualitative data', Sociology, vol.37, no.3, 1998, pp.413-431

Moore, Niamh: '(Re)using qualitative data?', Sociological Research Online, vol.12, no.3, May 2007 (downloaded at http://www.socresonline.org.uk/12/3/1.html on 20/10/08)

Parry, O. & N. Mauthner: 'Back to basics: who re-uses qualitative data and why?' *Sociology* vol. 39, no. 2, 2005, pp. 337-342

Parry, O. & N. S. Mauthner: (2004) 'Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data', *Sociology* vol. 38, no. 1, 2004, pp. 139

Potter, Jonathan & Wethrell, Margaret: *Discourse and social psychology: beyond attitudes and behaviour*. London: Sage, 1987

Savage, Mike: 'Changing social class identities in post-war Britain: perspectives from Mass-Observation', Sociological Research Online, vol.12, no.3, May 2007 (downloaded at http://www.socresonline.org.uk/12/3/1.html on 20/10/08)

Silva, Elizabeth: 'What's [yet] to be seen? Re-using qualitative data', Sociological Research Online, vol.12, no.3, May 2007 (downloaded at http://www.socresonline.org.uk/12/3/1.html on 20/10/08)

Summerfield, Penny: *Reconstructing women's wartime lives: discourse and subjectivity in oral histories of the Second World War* (Manchester, Manchester University Press, 1998)

Thompson, Paul: Re-using qualitative research data: a personal account, Forum: Qualitative Social Research, vol.1, no.3, December 2000 (downloaded at http://www.qualitative-research.net/fqs/ on 25/10/08)