

Project Document Cover Sheet

Project Information			
Project Acronym	DataShare		
Project Title	DISC-UK DataShare		
Start Date	March 2007	End Date	March 2009
Lead Institution	EDINA, Edinburgh University		
Project Director	Peter Burnhill, Mark Brown		
Project Manager & contact details	Robin Rice, Edinburgh University Data Library, Main Library Bldg., George Square, Edinburgh EH8 9LJ R.Rice@ed.ac.uk 0131 651 1431		
Partner Institutions	Universities of Edinburgh, Oxford, Southampton and London School of Economics		
Project Web URL	http://www.disc-uk.org/datashare/		
Programme Name (and number)	JISC Repositories and Preservation Programme: Repositories Start-up and Enhancement projects strand		
Programme Manager	Andrew McGregor		

Document Name			
Document Title	Southampton Data Survey: Our Experience and Lessons Learned		
Author(s) & project role	Harry Gibbs, University of Southampton, Project Officer		
Date		Filename	
URL	http://www.disc-uk.org/publications.html		
Access	<input type="checkbox"/> Project and JISC internal	<input type="checkbox"/> General dissemination	

Document History		
Version	Date	Comments
1	17 March 2009	

Table of Contents

1. Introduction	2
2. Scope of the Survey	2
3. Data Survey Team	2
4. Methodological Development	2
4.1 Stage 1	2
4.2 Stage 2.....	3
4.3 Stage 3.....	3
5. Ethical Approval	3
6. Online Questionnaire	3
7. Administration of the Questionnaire	4
8. Follow-up Interviews	4
9. Reporting	5
10. Lessons Learned	5
10.1 Rigorous Approach	5
10.2 Ethics Approval.....	5
10.3 Questionnaire Design.....	5
10.4 QuestionMark Perception	6
10.5 Audio Formats.....	6
10.6 Time.....	6
11. Successes	6
11.1 Participation	6
11.2 Data Gathered.....	6
Appendix 1: Questionnaire	7
Appendix 2: Interview Consent Form	16
Appendix 3: Questionnaire Participant Information Sheet	17
Appendix 4: Generic Interview Schedule	19

1. Introduction

The Data Audit Framework Methodology Development (DAFD) project, led by the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow, has developed a framework methodology that, “provides organisations with the means to identify, locate, describe and assess how they are managing their research data assets.”¹ The Data Audit Framework Methodology² was piloted during the course of the DAFD project and was subsequently used by four Implementation projects at University College London, Imperial College London, University of Edinburgh and King’s College London.

As part of the JISC funded DISC-UK DataShare project, this work has been built upon at the University of Oxford and the University of Southampton. This document describes the Data Survey that took place at Southampton, with a view to sharing our experiences and lessons learned with others who might be considering undertaking a data audit.

2. Scope of the Survey

The survey took place within the School of Social Sciences at the University of Southampton. The School has a total of 124 researchers across its six Divisions: Economics, Gerontology, Politics, Social Statistics, Social Work, and Sociology and Social Policy. In line with the DAFD project, we restricted our scope to research data used for analysis, although this could be primary, secondary, qualitative, quantitative and in any format. Our sample group was researchers employed by the University of Southampton.

3. Data Survey Team

The Data Survey team consisted of myself, the DataShare Project Officer and Academic Liaison Librarian for Social Sciences, and Teresa McGowan, a research assistant in the School of Social Sciences. It was felt that Teresa’s contacts and good reputation within the School would encourage participation, and that the survey would benefit from her subject knowledge. Teresa was employed for 301 hours from August 2008 to February 2009, although her efforts were concentrated from December to February when data collection and analysis took place. My role was to plan and oversee the survey.

4. Methodological Development

Our methodology evolved considerably between August and December. We began with the DAF methodology but gradually moved away from the audit approach with feedback from School staff and as we learned more from the experiences of the other DAF projects. Essentially, our methodological plans were made and adjusted in three stages, as follows.

4.1 Stage 1

When we began in August the early DAF pilots were reporting their first ‘lessons learned’. A common finding was that interviewing researchers had proved time consuming and a number were in the process of developing online questionnaires, as an alternative. We took the decision to use a questionnaire but also to gather more in depth information from a small number of participants in follow-up interviews.

Our initial plan was to adapt the DAF audit forms³ (3A & 3B) to suit the types of data held within the School and to use the questionnaire to gather the information to complete the customised form. It was felt that to comprehensively complete both forms would place too

¹ The Data Audit Framework (DAF): <http://www.data-audit.eu/index.html>

² The Data Audit Framework (DAF) Methodology http://www.data-audit.eu/DAF_Methodology.pdf

³ The Data Audit Framework (DAF) Methodology Audit Forms http://www.data-audit.eu/DAF_Methodology.pdf (see pp. 33-36)

great a burden on researchers. The most pertinent metadata fields were selected in consultation with School staff and the adapted form was presented at a meeting of the Division of Social Statistics with an outline of our planned methodology.

It became apparent at an early stage that it wouldn't be possible to conduct a complete audit of all data held in the School. One issue of debate, therefore, was whether to ask about data used in the most recent project or over a given period of time. Some staff commented that they worked on multiple projects simultaneously, so found 'most recent project' unhelpful. Time periods are also problematic because some researchers will use/create large quantities of data in one year, whilst others may only work with a small number of datasets over a career.

The key benefit to the School was proposed as the compilation and dissemination of an internal list of datasets held within the School to facilitate the sharing of expertise and de-duplication of effort. However, staff seemed sceptical of the benefits of this.

4.2 Stage 2

At this time the DAF Implementation group at Imperial College circulated a copy of their questionnaire, which asked for a limited amount of metadata describing their most important data holdings and allowed the participant to detail more than one dataset, should they wish. It also asked participants more general questions about their data management practices.

This questionnaire was adapted to suit the Social Sciences and discussed with the Deputy Head of School for Research who was broadly supportive of the approach but advised strongly that we would not get responses to a questionnaire that took more than 8 minutes to complete.

4.3 Stage 3

The questionnaire was reviewed in light of this time restriction. We felt that providing metadata to describe datasets would be the most time consuming part of the process because it would likely require time to recall or check information. It also seemed that DAF Implementation groups were concluding that their findings in regard to data management processes and gaps in support for data management were most interesting.

With valuable input from Luis Martinez Uribe, Oxford's Digital Repository Research Co-ordinator, we developed a set of questions that asked broadly about the type of data held but focussed primarily on the management of that data. We made heavy use of tick box answers in an attempt to make the questionnaire as quick and simple to complete and, therefore, to maximise the response rate. The questionnaire used can be found in Appendix 1.

5. Ethical Approval

The School of Social Sciences requires that any study involving human participants must be approved by the School Ethics Committee. An application to the Committee must include a full description of the proposed research ideally with questionnaires, interview schedules and consent forms appended. We found it difficult to comply with these requirements at an early stage of planning. In addition, we had to complete a risk assessment and research governance form, all of which took an unanticipated 28 hours. The questionnaire consent form is included in Appendix 1 and the interview consent form is in Appendix 2.

6. Online Questionnaire

We chose to use software called QuestionMark Perception, primarily because it is supported by the University of Southampton. Our experience suggests that it is primarily designed for online quizzes and student assessment, with surveys being a secondary function. It proved difficult and time consuming to implement complex question filtering and we were forced to

simplify our question order. We also found that the methods for reporting results were inadequate for our purposes. Some key issues follow.

- It was not possible to produce a single report that presented all responses from all participants and we were forced to manually combine two reports.
- The main Excel report was not suitable for analysis or export, and had to be manually transferred to SPSS.
- The summary report was found to have significant inaccuracies meaning that we were unable to make use of it.

7. Administration of the Questionnaire

The link to the questionnaire was sent to all School staff in an email briefly explaining the purpose of the Data Survey and with an Information Sheet attached. The information sheet is in Appendix 3.

We had a slight concern that the long link generated by QuestionMark Perception may break in transit and considered using TinyURL.com⁴. This worked well in tests but we decided that the intermediary webpage looked too commercial and may discourage participation.

The questionnaire was available for two weeks and two reminders were sent during this period, including one from the Deputy Head of School for Research. Twenty-three full responses were received on the first day whereas only three/two/one were received on subsequent days, respectively. The reminders prompted small flurries of responses. This suggests that researchers tended to complete the questionnaire at the time of opening the email.

From the 124 researchers, we received 47 full responses (38%) and 8 partial responses. Of the full responses 12 told us that they hold no data, leaving 35 who told us about their data and data management practices. Of these 35, 17 volunteered to be contacted for interview (although one did not provide a contact name).

We realised too late that by asking participants to tell us which Division they belonged to at the end of the questionnaire, we were not able to identify the Divisional affiliation of those who told us they hold no data. Similarly, because we didn't ask for names, or some other means of identifying individuals, it was not possible to know if partial respondents had responded in full at another time. Therefore we took the decision to disregard partial responses in the analysis.

The 35 full responses from those holding data took between 2.5-13 minutes to complete, with only six exceeding the target 8 minutes and the majority taking around 4-6 minutes.

8. Follow-up Interviews

We aimed to interview a researcher from each Division, six in total, that between them worked with a variety of data types, including qualitative, quantitative, primary and secondary. We also selected researchers with a range of experience levels, from early career to Professor. Interviews were intended to take one hour each and in reality took between 40-55 minutes.

A generic interview schedule, see Appendix 4, was developed and then adapted for each interviewee, according to their questionnaire responses.

Contrary to the experiences of some of the DAF pilots, we found it remarkably simple to schedule the initial interviews. Initial responses to email invitations were received from all six

⁴ Tinyurl: <http://tinyurl.com/>

interviewees within one working day, although the firming up of details did prove more time consuming.

Unfortunately, our first week of interviewing fell at the same time as the first significant snow for 17 years(!) and three interviews were postponed as a result. A fourth was postponed owing to illness and two more due to researchers' exam commitments. One researcher commented that the beginning of February was bad timing.

Due to pressure on time, we were not able to reschedule one of the interviews but successfully completed five. All interviewees proved willing to talk openly and gave interesting interviews.

The interviews were recorded to mini-disc, and MP3 player as back-up. The mini-disc recorder produced far superior sound quality but was difficult and time consuming to transfer to other media. Despite significant effort, we had to concede that we lacked the expertise to convert the files to the format required by the transcriber. The MP3 player files were easy to transfer and did not require conversion but were of lower quality, making transcription more difficult.

The transcriber worked to a high standard although we were required to fill small gaps and to anonymise the transcripts. The transcription itself also took a considerable amount of time. Roughly speaking, transcription of a one hour interview takes seven hours. We received the last transcript after seven working days of completing the last interview.

9. Reporting

We agreed at an early stage that the main survey report should remain internal to the University and that it shouldn't be actively circulated beyond the School and the Library. It was felt that researchers may be discouraged from participating openly if they thought details of their working practices were to be more widely distributed.

10. Lessons Learned

10.1 Rigorous Approach

As an early career Social Scientist working among peers, Teresa felt it necessary to follow rigorous academic practice, such as transcribing interviews and using thorough qualitative analysis techniques. Such an approach is not addressed by the DAF Methodology and it is worth noting that it increases the time required considerably.

10.2 Ethics Approval

The requirement to seek ethical approval is inconsistent between institutions and probably between departments. Of the other DAF pilot institutions, only the University of Glasgow was required to apply for ethical approval. Furthermore, the application process at Glasgow⁵ was far less burdensome than that imposed by the School of Social Sciences (see section 5).

Before planning a data audit, auditors are advised to find out whether they will be required to apply for ethical approval and, if so, factor in the time to do so.

10.3 Questionnaire Design

The questionnaire (Appendix 1) is flawed in not asking for Divisional affiliation at the start, meaning that those who responded that they hold no data cannot be identified with their Division.

⁵ The Data Audit Framework (DAF) Methodology: http://www.data-audit.eu/DAF_Methodology.pdf (see Appendix 2)

Participants' names were not requested because we felt that staff would feel more comfortable responding to an anonymous questionnaire. However, the fact that individuals could not be identified meant that it was impossible to tell if a given respondent had responded more than once. It also made it more time consuming to manually combine reports (see section 6). With hindsight we may have asked for individual identifiers but would still have held concerns about the effect of this on the response rate.

10.4 QuestionMark Perception

To our cost, we found that QuestionMark Perception was not really fit for our purposes (see section 6). If we were to run a further survey, we would certainly look for alternative online questionnaire software and assess its functionality carefully before beginning. At least one DAF pilot ran a successful questionnaire using Bristol Online Surveys⁶.

10.5 Audio Formats

As with questionnaire software, interview recording methods and the conversion of audio file formats require full and careful testing (see section 8). Having discussed this with Social Scientists, we feel that investment in a digital recorder would have been justified.

10.6 Time

We have been struck by how time consuming the Data Survey has been. During the planning stage, we felt that that DAF Methodology⁷ underestimates the time required but we have exceeded even our own expectations! Much of this is as a result of unanticipated issues including having to gain ethical approval, the limitations of Perception, problems with audio file formats and inclement weather. The lesson here is that contingency time is a must.

11. **Successes**

11.1 Participation

We were very pleased with the questionnaire response rate (38%) and that almost half of respondents holding data offered to take part in a follow-up interview. It is not possible to explain this with any certainty, but we suspect that the simplicity of the questionnaire played a part. We also feel that the fact that we have well established links with School staff worked to our advantage. Having said that, a number of interview volunteers were unknown to both of us, suggesting that there is interest in the subject of data management. Another factor is that social scientists understand how difficult it can be to find research participants – our interviewees with experience of collecting qualitative data, in particular, expressed this as a motivation for participating.

11.2 Data Gathered

The interviews, in particular, revealed a very interesting picture of the diversity of data that are held in School, and of the research and data management processes that take place. It is hoped that the internal survey report will help the School to understand the breadth of its data holdings and the data management issues that staff face in the course of their research. It is also intended that, through the Library, the report will contribute towards the development of data services across disciplines at the University of Southampton.

⁶ Bristol Survey Tool: <http://www.survey.bris.ac.uk/>

⁷ The Data Audit Framework (DAF) Methodology: http://www.data-audit.eu/DAF_Methodology.pdf (see p. 60)

Appendix 1: Questionnaire

You may re-use or adapt this documentation for research or private study with acknowledgement to McGowan, T. & Gibbs, T. A. (2009) Southampton Data Survey: Our Experiences & Lessons Learned [unpublished]. University of Southampton: UK.

Welcome to the Research Data Survey questionnaire

Thank you for participating in this survey which aims to find out about research data held by staff in the School of Social Sciences and improve our understanding of the data management processes you employ.

For the purpose of this study 'research data' is data that you currently hold that has been collected and/or used in the course of your research at the University of Southampton. Research data can be primary data collected by you or your research group or secondary data provided by a third party. It may be quantitative or qualitative e.g. survey results, interview transcripts, databases compiled from documentary sources, images or audiovisual files.

Data that you 'currently hold' is all the research data that you currently store anywhere. For example, in your 'My Documents' folder, on the shared 'R' drive, a PC or laptop, on portable media such as CDs or memory sticks, or on paper.

It would help us greatly if you respond to this questionnaire even if you do not currently hold any research data (you will only be required to answer 2 questions).

The questionnaire is a maximum of 25 questions and should take no more than 10 minutes to complete.

Thank you for your time.

Participant consent form

Please read the following statements carefully before agreeing to take part in this study;

I have read and understood the participant information sheet (attached to the email in which you received this link).

I understand that;

- All results from this study will be anonymous. Information extracted from this questionnaire and any subsequent interview will not, under any circumstances, contain names or identifying characteristics of participants.
- I am free to withdraw from this study at any time without penalty.
- I am free to decline to answer particular questions.
- Whether I participate or not there will be no effect on my progress in employment in any way.

I consent to take part in this study on the terms described above; Yes No

[IF NO, GO TO END]

1. Do you currently hold any research data?

- Yes
- No **[GO TO END]**

2. Thinking about the primary data you hold, what type of data is it? *[Please select all that apply]*

- I don't hold any primary data **[GO TO QUESTION 4]**
- Cross sectional survey data
- Longitudinal survey data
- Interview/focus group transcripts
- Database compiled from documentary sources
- Image files
- Audio files
- Audio-visual files
- Other

If other, please specify

3. Who funded the collection of the primary data you hold? *[Please select all that apply]*

- ESRC
- EU-EDULINK
- Leverhulme Trust
- Nuffield Foundation
- UK Government department
- Wellcome Trust
- Other

If other, please specify

4. Thinking about the secondary data you hold, who collected this data? *[Please select all that apply]*

- I don't hold any secondary data [\[GO TO QUESTION 6\]](#)
- Datastream (Thomson Reuters)
- Eurostat
- International Labour Organization (ILO)
- Measure DHS
- Organisation for Economic Co-operation and Development (OECD)
- Office for National Statistics (ONS)
- US Census Bureau
- World Bank
- World Health Organization (WHO)
- Other

If other, please specify

5. What type of secondary data is it? *[Please select all that apply]*

- Cross sectional survey data
- Longitudinal survey data
- Interview/focus group transcripts
- Database compiled from documentary sources
- Image files
- Audio files
- Audio-visual files
- Macro-economic time series data
- Stock market data
- Company level data
- Other

If other, please specify

6. *The remaining questions relate to all the data you currently hold, both primary and secondary;*

When using or creating this data, did you collaborate with anyone else?

- Yes
 No [\[GO TO QUESTION 9\]](#)

7. How did you share data when you were collaborating? *[Please select all that apply]*

- By emailing files to colleagues
 Using a shared storage facility
 Using portable storage such as CDs, DVDs, memory sticks etc
 Other

If other, please specify

8. Did you encounter any practical problems when you were collaborating? *[Please select all that apply]*

- No
 Finding suitable shared storage space
 Lack of file naming conventions made it difficult to identify files
 Lack of version control caused confusion
 Legal issues arising from international transfer of data
 Problems establishing ownership of data
 Other

If other, please specify

9. Where do you store your data (excluding back up copies)? *[Please select all that apply]*

- On paper
- My Documents
- Shared drive (R-drive)
- Hard drive of office PC
- Hard drive of laptop PC
- Memory stick/USB/Flash drive
- CD/DVD
- External hard drive
- Other

If other, please specify

10. Have you ever experienced any problems storing your research data due to the size of the files?

- Yes
- No [\[GO TO QUESTION 12\]](#)

what problems

11. How did you overcome these storage problems? *[Please select all that apply]*

- Requested additional storage space from iSolutions
- Purchased an external hard drive
- Saved to portable media
- Other

If other, please specify

12. Is the data that you currently hold backed up anywhere?

- Yes, all of it is
- Yes, some of it is
- No, none of it is [\[GO TO QUESTION 14\]](#)

13. Where do you back up your data?

- On paper
- My Documents
- Shared drive (R drive)
- Hard drive of office PC
- Hard drive of laptop PC
- Memory stick/USB/Flash drive
- CD/DVD
- External hard drive
- Other

If other, please specify;

14. Do you deposit your data with a data service, such as the UK Data Archive?

- Yes, all of it [\[GO TO QUESTION 20\]](#)
- Yes, some of it
- No, none of it

If yes, please tell us which service/s you deposit with;

15. Do you think that any of your data needs to be preserved by the University for your own use or that of others?

- Yes
- No [\[GO TO QUESTION 17\]](#)

16. If you would like someone from the University Library to contact you about preserving your data please enter your name below;

17. Thinking about your data that is not deposited with a data service, could any of this data be re-used by others?

- Yes, all of it could be re-used [\[GO TO QUESTION 20\]](#)
 Yes, some of it could be re-used
 No, none of it could be re-used

18. Thinking about your data that can't be re-used or shared, please tell us why *[Please select all that apply]*;

- Confidentiality or data protection issues
 Licence agreements prohibit sharing
 The data is not fully documented
 The data is in a format that is no longer widely readable [\[IF SELECTED GO TO QUESTION 19, OTHERWISE GO TO QUESTION 20\]](#)
 Other

If other, please specify;

19. Please provide brief details of the data you have that is no longer widely readable (e.g. what software/hardware the data is on, its age etc);

20. Would you like to receive any additional support with managing your data? *[Please select all that apply]*

- Training
- Written guidance
- Help with writing data management plans for research bids
- Additional personal storage
- Additional shared storage
- None
- Other

If other, please specify

21. Which Division do you work in?

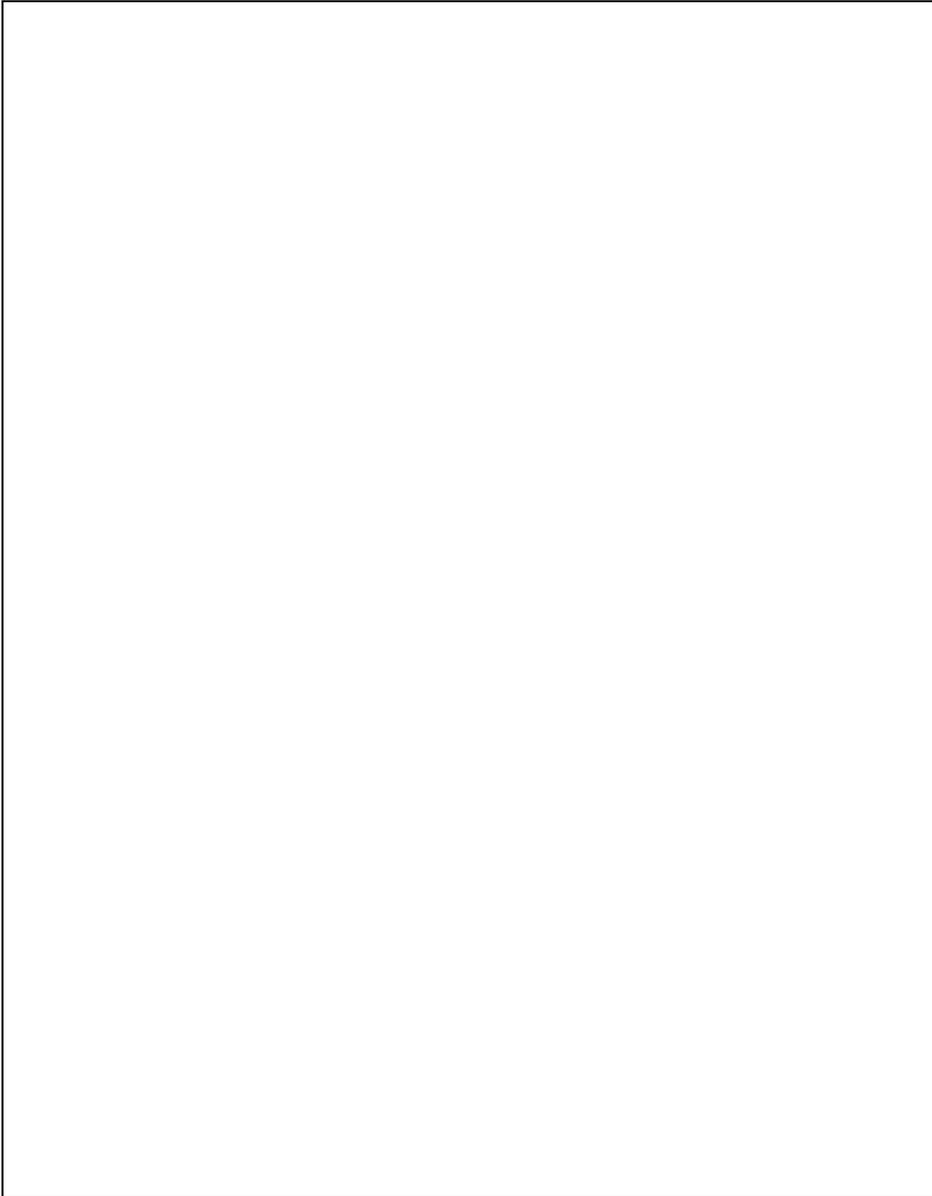
- Economics
- Gerontology
- Politics and International Relations
- Sociology and Social Policy
- Social Statistics
- Social Work Studies

22. Would you be prepared to participate in a follow up interview to explore data management issues in more depth (max. 1 hr)?

- Yes
- No

If yes, please provide your name and email address so that we can contact you;

- 23.** If you would like to expand on any of your above answers or make further comment, please do so here;



Appendix 2: Interview Consent Form

You may re-use or adapt this documentation for research or private study with acknowledgement to McGowan, T. & Gibbs, T. A. (2009) Southampton Data Survey: Our Experiences & Lessons Learned [unpublished]. University of Southampton: UK.

Interview Participant Consent Form

Title of Study: Research Data Survey in the School of Social Sciences

Iconsent to participate in the research project specified above.

1. I have read and understood the participant information sheet.
2. I agree to the interview being recorded.
3. I understand that:
 - I am free to withdraw from this study at any time without penalty.
 - I am free to ask for the recording to be stopped at any time without penalty.
 - I am free to decline to answer particular questions and to retract answers at a later date, if I so wish.
 - Whether I participate or not, or withdraw after participating, there will be no effect on my progress in employment in any way.
 - All information that the researcher extracts from this interview will not, under any circumstances, contain my name or any identifying characteristics.
 - The report that results from this study will remain internal to the School of Social Sciences.

Participant's signature.....Date.....

I certify that I have explained the study to the volunteer and consider that she/he understands what is involved and freely consents to participation.

Researcher's signature.....Date.....

NOTE: This consent form will remain with the researcher for their records.

Appendix 3: Questionnaire Participant Information Sheet

You may re-use or adapt this documentation for research or private study with acknowledgement to McGowan, T. & Gibbs, T. A. (2009) Southampton Data Survey: Our Experiences & Lessons Learned [unpublished]. University of Southampton: UK.

Questionnaire Participant Information Sheet

Title of Study

Research Data Survey in the School of Social Sciences

Names of Researchers

Teresa McGowan, School of Social Sciences

Harry Gibbs, Hartley Library

Purpose of Study

This study will pilot the 'Data Audit Framework Methodology' which is currently under development with Joint Information Systems Committee (JISC) funding. The methodology aims to enable institutions to find out what research data they hold, where it is located and who is responsible for it.

It is intended that the School of Social Sciences will benefit from increased knowledge of data holdings. The University's Data Preservation Group is also interested in the findings of the research, with a view to improving data management across the Institution. In addition, JISC will benefit from receiving feedback on their methodology.

Study Methods

We would like you to take part in a short online questionnaire asking you about research data you currently hold. Please fill in as much of the questionnaire as you can.

Following the administration of the questionnaire, if you have indicated your willingness to take part you may be approached to take part in an interview. The purpose of the interviews will be to provide an opportunity to discuss data management issues in some depth.

Participation in the Study

Your participation is completely voluntary. Whether you choose to participate or not, or withdraw after participating, there will be no effect on your progress in employment in any way. You may decline answering any questions you feel you do not wish to answer. There are no anticipated risks to your participation.

The online questionnaire should take no longer than 10 minutes to complete. Interviews will last no more than 1 hour and, as far as possible, will be scheduled at your convenience. Interviews will take place either in the interviewer's office or in your own office. Interviews will be recorded and you will be free to ask for the recording to be stopped at any time or for your comments to be off the record.

Anonymity and Confidentiality

Questionnaire submissions and interview recordings will remain confidential. During transcription, interviews will be fully anonymised, the resulting reports will not, under any circumstances, contain any identifying characteristics of participants.

Ethical Approval

This study has been approved by the School of Social Sciences Ethics Committee.

Access to the Study Reports

An internal report on the survey results will, in the first instance, be passed to the Deputy Head of School (Research & Enterprise). Copies of the report will also be made available to School staff. It is anticipated that a report to JISC detailing Southampton's use and interpretation of the 'Data Audit Framework' will be of little interest to participants but you will be welcome to a copy on request.

Further Information

If you have any questions about participation in this study, please feel free to discuss these with either Teresa McGowan or Harry Gibbs.

Teresa McGowan
Tel: 02380 592579
Email: T.McGowan@soton.ac.uk

Harry Gibbs
Tel: 02380 598584
Email: hgibbs@soton.ac.uk

Many thanks for your help with this study.

Appendix 4: Generic Interview Schedule

You may re-use or adapt this documentation for research or private study with acknowledgement to McGowan, T. & Gibbs, T. A. (2009) Southampton Data Survey: Our Experiences & Lessons Learned [unpublished]. University of Southampton: UK.

Introduction

INTRODUCE

My name is Teresa McGowan and this is **Harry Gibbs**. Harry is the School of Social Sciences **librarian** and I am a **research assistant** here in the School.

RESEARCH

We are working together on a **project funded by** the Joint Information Systems Committee (JISC). **JISC has developed a framework methodology** aimed at helping institutions find out what research data they hold, where it's located and who is responsible for it. **We are using** an adaptation of that framework today to test its usefulness and to help the School of Social Sciences find out more about data management and what can be done to aid staff in the use and management of their data.

THANKS

Thank you for agreeing to take part in this interview. Based on the information that we receive we will produce **two reports, one for JISC** simply discussing how we used and modified their framework, and a **second for the University** which we hope will be used to improve data management in the school.

RECORDING AND CONFIDENTIALITY

We would like to **record our discussion** as it is so difficult to write down everything that is said, and we don't want to miss anything. What you say in this interview will be **anonymous** – your names will not be recorded on the transcripts and only **me, Harry and one transcriber** will have access to the recording and notes. **No reports or publications** that are produced will identify you in any way.

WANT TO KNOW

Thank-you for taking part in the questionnaire, **the purpose** of this interview is to find out more about **the data you hold** that has been collected or used in the course of your research at this University and **your experience of managing** this data. **There are no right or wrong answers, we are just interested in what you have done and how you did it.**

We want this to be more like a **discussion** than a question and answer session. We have a list of **x** things we are interested in but it is important to us that you tell us about what is important to you. If there is anything I ask that **you don't understand** please tell us and we can explain further. If there is **anything you want to ask us** you can do that too. *(If they ask questions that anticipate later discussions, ask if it's OK to leave it until later)*

Do we have your permission to proceed?
(Record some thoughts about participants, body language etc)

Discussion Guide

1 Could you please tell us a bit about your area of research?

2 We can see from the questionnaire that you hold xxx data, please could you give us some more details about the **xxx data that you compiled from documents?**

Data Holdings		
Name of Interviewee	Primary Data	Secondary Data
[interviewee's name]	[list of primary data types held by individual, as reported in questionnaire]	[list of secondary data types held by individual, as reported in questionnaire]

Tick	Metadata Heading	Notes
	ID	A unique identification assigned by the research team
	Author	Person, group or organisation responsible for the intellectual content of the dataset
	Owner(s)	Current legal owner(s) of the dataset
	Source	The source(s) of the information found in the data asset
	Purpose	Reason why the asset was created, intended user communities or source of funding / original project title
	Title	Official name of the data asset, with additional or alternative titles or acronyms if they exist
	Description	A description of the information contained in the data asset and its spatial, temporal or subject coverage
	Subject	Data topics and keywords describing the subject matter of the data
	Geographical coverage	The countries, regions, cities etc covered in the data
	Time period covered	The date (or date range) covered by the data
	Date of collection	The date (or date range) on which the data was collected (for social surveys this will often be the same as the time period covered)
	Sample size & description	The number of individuals surveyed and characteristics
	Current location	Path or www address where the data can be found
	Format	Physical formats of dataset, including file format information
	Size	Size of the data in Mb/Gb
	Restrictions	Access restrictions placed on user of secondary data or restrictions owner would place on reuse of primary data
	Documentation available	Documentation that is available (e.g. user manuals, code books), including references to its location
	Retention period	Planned retention period for the data & ideal retention period

3 As I am sure you are aware, increasingly funding bodies want researchers to include a **data management or data**

sharing plan in the funding application. Have you ever experienced this?

OR

TO THOSE WHO ASKED FOR HELP WRITING DATA MANAGEMENT PLANS:

3 Can you tell us about your experience of data management or data sharing plans?

Prompts		Tick
YES	Which funder?	
	How did you find this experience?	
	How did it influence your actual data management/data sharing?	
NO	What do you think about in terms of data at the bid writing stage?	
	How far did planning influence your actual practice? (Did it go to plan?)	

4 TO COLLABORATORS:

In the questionnaire you mentioned you had some **problems collaborating**, could you tell us some more about this?

OR

Please can you tell us about one **experience of collaborating**?

Problems collaborating	
[name]	[list of problems reported in questionnaire]

	Prompts	Tick
Who, what where?	Who with?	
	Where were they geographically?	
	How many of you were sharing?	
Sharing methods;	What did you do?	
	What methods did you use to share?	
How did you deal with;	version control	
	file naming conventions	
	legal issues transferring data	
	How was ownership decided?	
Confidential data?	Have you ever shared confidential data?	
	How did you do it?	

OR

4 TO NON COLLABORATORS:

How do you deal with the **day-to-day management** of data?

Prompts	Tick
Version control	
File naming system	

5 BACK UP

You told us you use xxx methods to **store and back up data**, can you tell us why you chose these methods?

Storage location		
	Main	Back up
[name]	[list of main storage methods reported in questionnaire]	[list of back up methods reported in questionnaire]

	Prompts	Tick
What affects your choices?	Anticipated lifespan	
	Importance	
	Confidentiality	
	Physical space	
Have you ever experienced any problems with;	Data loss	
	Old formats	
	File size	

AND

TO THOSE WHO DO NOT BACK UP ALL DATA

You said **you don't back up all** your data, can you explain to us why you don't?

6 You mentioned that you had **data storage problems** and [list of problems reported in questionnaire], can you tell us a bit more about what happened?

	Tick
Was the unforeseen expense a problem for the project?	
Did the problem affect the project due to lost time?	

7 In the questionnaire you said you did not have anything that should be preserved by the university, do you have anything that you think should **be preserved by yourself or anyone else?**

	Prompts	Tick
<p style="text-align: center;">YES</p> <p>Could you tell us about it?</p>	How could it be best preserved?	
	Why should it be preserved?	
	Who should preserve it?	
	For how long?	
<p style="text-align: center;">NO</p> <p>Why doesn't it need preserving?</p>	Already preserved? (BY UKDA)	
	Data not reusable [why?]	
	Time	
	Money	

OR

7 You mentioned in the questionnaire that you have something that you think the **university should preserve** for the future, could you tell us about it now?

Prompts	Tick
How could it be best preserved?	
Why should it be preserved?	
Who should preserve it?	
For how long?	

- 8 **What support** is available to you to help you manage your data?

OR

- 8 In the questionnaire you said you would like some additional support in carrying out data management, **what support is available to you now?**

Prompt	Tick
How sufficient is it?	
How would you like data management support to look in an ideal world?	

- 9 *(Summarise what's been said, then;)* Is there anything you can think of I haven't asked or anything you wanted to say that has not been covered?

INVITE HARRY TO ASK ANY QUESTIONS ON ANYTHING SHE WOULD LIKE TO FOLLOW UP/CLARIFY

Closure

Thank-you for allowing us to talk to you today, it has been very interesting to listen to your views. We will email you to let you know when the results are available.

Thanks again for coming today, we are very grateful for your help.