MANAGING AND SHARING DATA

UK•DATA ARCHIVE

First edition, 2011

Authors: Louise Corti, Veerle Van den Eynden,
Libby Bishop, Bethany Morgan-Brett.

All online references available 31 August 2011.

Published by:
UK Data Archive
University of Essex
Wivenhoe Park
Colchester
Essex
CO4 3SQ

ISBN 1-904059-82-1

Designed and printed by
Print Essex at the University of Essex
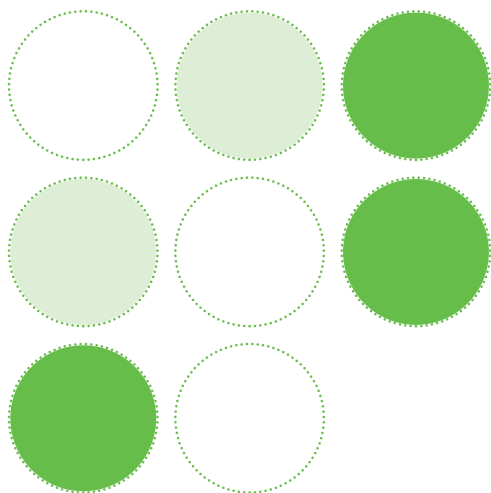
© 2011 University of Essex

50% recycled
material

CONTENTS

INTRODUCTION

THIS SET OF TRAINING RESOURCES PRESENTS A SUITE OF FLEXIBLE TRAINING MATERIALS FOR PEOPLE WHO ARE CHARGED WITH TRAINING RESEARCHERS AND RESEARCH SUPPORT STAFF IN HOW TO LOOK AFTER RESEARCH DATA.

The Training Resources sit alongside the UK Data Archive's popular Managing and Sharing Data: best practice for researchers, the most recent version published in May 2011. The materials have been designed and used as part of the Archive's daily work in supporting ESRC applicants and award holders. The UK Data Archive has over forty years experience in selecting, ingesting, curating and providing access to social science data. We are a designated Place of Deposit for The National Archives and are internationally acknowledged experts in this field.

Our recent ESRC Researcher Development Initiative project, 'Data Management and Sharing for Researchers Training', has enabled us to modularise, test, and refine the materials, creating off-the-shelf materials ready for instructors to use. Versions of these training materials have been used in our workshops aimed at diverse audiences – junior to senior researchers from a range of disciplines, junior and senior lecturers, postgraduate students, research support staff in universities, colleges and research organisations, and Research Ethics Committee (REC) members.

These materials are intended to be used flexibly and cover training in the seven key areas of data management identified by the UK Data Archive:

- sharing data - why and how
- data management planning for researchers and research centres
- documenting data
- formatting data
- storing data, including data security, data transfer, encryption, and file sharing
- ethics and consent
- data copyright

WHO SHOULD USE OUR TRAINING RESOURCES?

The intention is for them to be used at an introductory level by those who:

- provide formal advice on data management and sharing, including: University research grant administrators, REC members and IT support staff; research centre managers; and librarians supporting institutional data management requirements
- deliver training on research methods - these materials are highly complementary to introductions to research methods, covering areas that are fundamental (but often overlooked) in training students how to look after research data well

Our approach and materials have already been adopted by other organisations, such as the JISC's 2011 projects delivering postgraduate training in data management, and organisations belonging to the International Association for Social Science Information Services and Technology (IASSIST). Some modules may require more background knowledge for the presenter than others, for example consent and legal issues.

USING OUR MATERIALS

Our Training Resources are divided into seven sections, each presenting a similar suite of materials. Each module may contain resources of the following type:

- **topic overview PowerPoint presentation(s):** covering the key points and setting the scene for the topic; the presentations are, naturally, country specific, especially where legal and ethical issues relating to data are described; all presentations can be downloaded from our website
- **presenter's guide to exercises:** written as we administer the training exercises in our own training - not all exercises have a presenter's guide as they are self-explanatory
- **exercises – introduction:** the contextual materials that explain the exercise to participants
- **exercises and exemplars:** the exercises or discussion materials that participants receive
- **exercises – answers:** suggested answers to the exercises or discussion topics, which can be handed out and read after the exercise has been completed

We have provided electronic bundles of material on our website for download and use. They may be updated from time-to-time so please check our website: www.data-archive.ac.uk/create-manage/training-resources

REMEMBER...

- the materials can be used as they are or adapted to personal teaching styles. Enjoy them!
- to cite the UK Data Archive if you use our materials in your training - the materials are under a Creative Commons Attribution-NonCommercial-ShareAlike licence, and cannot be used for commercial gain or re-publishing
- to properly cite the source of any real data in any examples you use and
- we are very happy receive feedback on their use and suggestions for improvement or extension - email: datasharing@data-archive.ac.uk.

TRAINING RESOURCES OVERVIEW

DATA MANAGEMENT TOPIC	TOPIC OVERVIEW	TOPICS	EXERCISES	RESOURCES
SHARING YOUR DATA- WHY AND HOW	Data created from research are valuable resources that can be used and re-used for future scientific and educational purposes. Sharing data facilitates new scientific inquiry, avoids duplicate data collection and provides rich real-life resources for education and training.	• Why share research data? • Research funders and journals • How to share your data	• Barriers to sharing data	• Website content • Brochure content • PowerPoint presentation
DATA MANAGEMENT PLANNING	A data management and sharing plan helps researchers consider, when research is being designed and planned, how data will be managed during the research process and shared afterwards with the wider research community.	• Data management plans and the research lifecycle • Data Management Checklist • Roles and responsibilities • Costing data management • Strategies and resources for centres	• Data management plans • Data lifecycle planning	• Website content • Brochure content • PowerPoint presentation • Data Management Checklist • Lifecycle diagram • Data Management Costing Tool • Data Management Resources Library • Data Inventory
DOCUMENTING YOUR DATA	A crucial part of making data user-friendly, shareable and with long-lasting usability is to ensure they can be understood and interpreted by any user. This requires clear and detailed data description, annotation and contextual information.	• Documenting data • Study-level documentation • Data-level documentation • Metadata • Context debate	• Context for qualitative data	• Website content • Brochure content • PowerPoint presentation
FORMATTING YOUR DATA	Using standard and interchangeable or open lossless data formats ensures long-term usability of data. High quality data are well organised, structured, named and versioned and the authenticity of master files identified.	• File formats • Data conversions • Organising files and folders • Quality assurance • Version control and authenticity • Transcription	• File naming • Qualitative data transcription • Formats quiz	• Website content • Brochure content • PowerPoint presentations

TRAINING RESOURCES OVERVIEW

DATA MANAGEMENT TOPIC	TOPIC OVERVIEW	TOPICS	EXERCISES	RESOURCES
STORING YOUR DATA	Looking after research data for the longer-term and protecting them from unwanted loss requires having good strategies in place for securely storing, backing-up, transmitting, and disposing of data. Collaborative research brings challenges for the shared storage of, and access to, data.	• Making back-ups • Data storage • Data security • Data transmission and encryption • Data disposal • File sharing and collaborative environments	• Data security breaches • Data storage and security quiz • Encryption • Checksum	• Website content • Brochure content • PowerPoint presentations
ETHICS AND CONSENT	A combination of gaining consent for data sharing, anonymising and regulating access to data will increase the potential for making people-related research data more readily and widely available.	• Legal and ethical issues • Definitions: personal, confidential and sensitive personal data • Informed consent and data sharing • Anonymising data • Access control	• Consent forms • Anonymisation - qualitative data • Anonymisation - qualitative data	• Website content • Brochure content • PowerPoint presentations
COPYRIGHT	Copyright is an intellectual property right assigned automatically to the creator, that prevents unauthorised copying and publishing of an original work. Copyright applies to research data and plays a role when creating, sharing and re-using data.	• Definitions • Who owns copyright? • Re-use of data and copyright	• Copyright scenarios	• Website content • Brochure content • PowerPoint presentation

SHARING YOUR DATA – WHY AND HOW

DATA SHARING DRIVES SCIENCE FORWARD

DATA CREATED FROM RESEARCH ARE VALUABLE RESOURCES THAT CAN BE USED AND RE-USED FOR FUTURE SCIENTIFIC AND EDUCATIONAL PURPOSES. SHARING DATA FACILITATES NEW SCIENTIFIC INQUIRY, AVOIDS DUPLICATE DATA COLLECTION AND PROVIDES RICH REAL-LIFE RESOURCES FOR EDUCATION AND TRAINING.

RESOURCES OVERVIEW

AREAS OF COVERAGE

- Why share research data?

- Research funders and journals

- How to share your data

RESOURCES AVAILABLE

- PowerPoint presentation on 'Benefits of managing and sharing your data'

- Exercise 1: Barriers to sharing research data (no answers)

- Exercise 2: Reasons not to share data

- Exercise 2 Answers: Reasons not to share data

- www.data-archive.ac.uk/create-manage/ training-resources

EXERCISE ONE

BARRIERS TO SHARING RESEARCH DATA

Write down five reasons why researchers may be reluctant to share their research data.

	BARRIERS TO SHARING RESEARCH DATA
1	
2	
3	
4	
5	

EXERCISE TWO

REASONS NOT TO SHARE DATA

The following table gives examples of reasons not to share data. Discuss or write down possible solutions to these reasons.

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
1	My data is not of interest or use to anyone else.	
2	I want to publish my work before anyone else sees my data.	
3	I have not got the time or money to prepare data for sharing.	
4	If I ask my respondents for consent to share their data then they will not agree to participate in the study.	
5	I am doing highly sensitive research. I cannot possibly make my data available for others to see.	
6	I am doing quantitative research and the combination of my variables discloses my participant's identity.	
7	I have collected audiovisual data and I cannot anonymise them, therefore I cannot share these data.	
8	I have made promises to destroy my data once the project finishes.	
9	My data have been gathered under complete assurances of confidentiality.	
10	My data collection and resulting transcripts are in a foreign language.	
11	It is impossible to anonymise my transcripts as too much useful information is lost.	
12	My data collection contains data which I have purchased and it cannot be made public.	
13	Other researchers would not understand my data at all - or may use them for the wrong purpose.	
14	There is IPR in the data.	

REASONS NOT TO SHARE DATA

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
1	My data is not of interest or use to anyone else.	It is! Researchers want to access data from all kinds of studies, methodologies and disciplines. It is very difficult to predict which data may be important for future research. Who would have thought that amateur gardener's diaries would one day provide essential data for climate change research? Your data may also be essential for teaching purposes. Sharing is not just about archiving your data but about sharing them amongst colleagues.
2	I want to publish my work before anyone else sees my data.	Data sharing will not stand in the way of you first using your data for your publications. Most research funders allow you some period of sole use, but also want timely sharing. Also remember that you have already been working with your data for some time so you undoubtedly know the data better than anyone coming to use them afresh. If you are still concerned you can embargo your data for a specific period of time.
3	I have not got the time or money to prepare data for sharing	It is important to plan data management early in the research data lifecycle. Data management ideally becomes an integral part of your research practice, reduces time and financial costs and greatly enhancing the quality of the data for your use too.
4	If I ask my respondents for consent to share their data then they will not agree to participate in the study.	Don't assume that participants will not participate because data sharing is discussed. Talk to them – they may be less reluctant than you might think, or less concerned over data sharing! Make it clear that it is entirely their decision, whereby they can decide whether their data can be shared, independent of them participating in the research. Explain clearly what data sharing means, and why it may be important. But they are still free to consent or not. You can always explain what data archiving means in practice for their data. If you have not asked permission to share data during the research, then you can always return to gain retrospective permission from participants.
5	I am doing highly sensitive research. I cannot possibly make my data available for others to see.	The first thing is to ask respondents and see if you can get consent for sharing in the first instance. Anonymisation procedures can help to protect identifying information. If these first two strategies are not appropriate then consider controlling access to the data or embargoing for a period of time. Also data that is held in the UK Data Archive is not publically available. Only registered researchers can gain access to the data.
6	I am doing quantitative research and the combination of my variables discloses my participant's identity.	Quantitative data can be anonymised through processes of aggregation, top coding, removal of variables, or controlled access to certain variables (i.e. postcodes).
7	I have collected audiovisual data and I cannot anonymise them, therefore I cannot share these data.	Visual data can be anonymised through blurring faces or distorting voices, but this can be time consuming and costly to carry out. It can mean losing much of the value of the data. It is better to ask for consent to share data from participants in an unanonymised form, and/or control access to the data.
8	I have made promises to destroy my data once the project finishes.	Why were such promises made? Always avoid making unnecessary promises to destroy data. There is usually no legal or ethical need to do so, except in the case of personal data. But that certainly would not apply to research data in general. Also consider where you have received this advice from? You may need to negotiate with research ethics committee or ethics boards about this agreement.

EXERCISE 2 ANSWERS: REASONS NOT TO SHARE DATA

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
9	My data have been gathered under complete assurances of confidentiality.	Again why was such an assurance made? It is best to avoid unnecessary promises. Anonymisation procedures can be implemented to protect identities, but confidentially can never be completely guaranteed. You can also consider controlling access to the data.
10	My data collection and resulting transcripts are in a foreign language.	This should not be a problem. The UK Data Archive can accept foreign language transcripts although translations into English are preferred.
11	It is impossible to anonymise my transcripts as too much useful information is lost.	Get in touch with us at the UK Data Archive. We may be able to help and it might not be as difficult as it looks. Also, access controls on the data may be a better solution than anonymisation if too much useful information would be lost.
12	My data collection contains data which I have purchased and it cannot be made public.	It is important to know who holds the copyright to the data you are using and to obtain the relevant permissions. You need to be aware of the licence conditions of the data you are using and what you can and cannot do with the data.
13	Other researchers would not understand my data at all - or may use them for the wrong purpose.	Producing good documentation and providing contextual information for your research project should enable other researchers to correctly use and understand your data.
14	There is IPR in the data.	This should not be a problem if you seek copyright permission from the owner of the intellectual property rights. This is best done early on in the research project, but could be sought retrospectively.

DATA MANAGEMENT PLANNING

PLAN AHEAD TO CREATE HIGH-QUALITY AND SUSTAINABLE DATA THAT CAN BE SHARED

A DATA MANAGEMENT AND SHARING PLAN HELPS RESEARCHERS CONSIDER, WHEN RESEARCH IS BEING DESIGNED AND PLANNED, HOW DATA WILL BE MANAGED DURING THE RESEARCH PROCESS AND SHARED AFTERWARDS WITH THE WIDER RESEARCH COMMUNITY.

RESOURCES OVERVIEW

AREAS OF COVERAGE

- Data management planning tasks and the research lifecycle

- Data Management Checklist

- Roles and responsibilities for data management

- Costing data management

- Strategies and resources for research centres and hubs

RESOURCES AVAILABLE

- PowerPoint presentation on 'The data lifecycle and data management planning'

- Exercise 1: Using the data management lifecycle for your research planning

- Data lifecycle planning - real life example

- Exercise 2: Using the Data Management Checklist for your research planning

- Exercise 3: Using a data inventory for planning data management in Centres

- www.data-archive.ac.uk/create-manage/ training-resources

EXERCISE ONE

USING THE DATA MANAGEMENT LIFECYCLE FOR YOUR RESEARCH PLANNING

Data often have a longer lifespan than the research project that creates them. Researchers may continue to work on data after funding has ceased, follow-up projects may analyse or add to the data, and data may be re-used by other researchers.

Annotate the UK Data Archive's Data Lifecycle diagram overleaf to reflect your own research data lifecycle and the points at which you think data management procedures should be implemented.

THE UK DATA ARCHIVE'S DATA LIFECYCLE
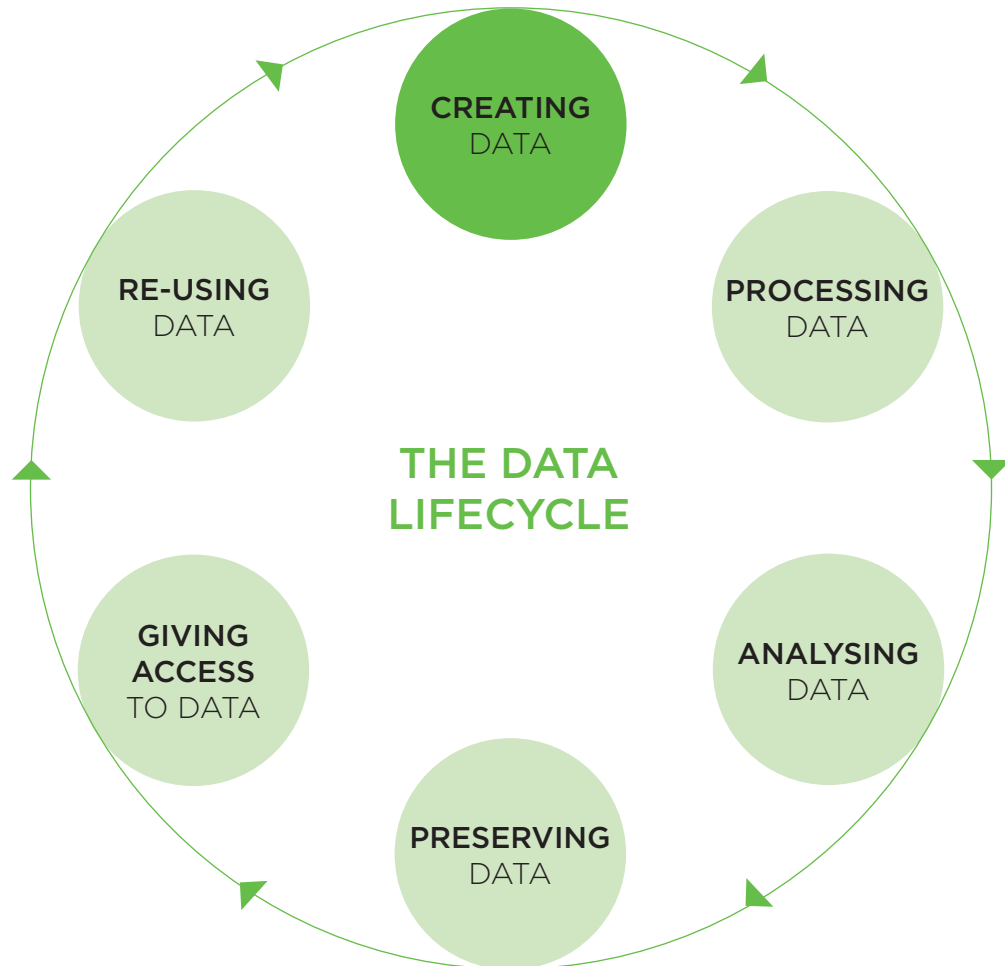
CREATING DATA
- design research
- plan data management (formats, storage etc.)
- plan consent for sharing
- locate existing data
- collect data (experiment, observe, measure, simulate)
- capture and create metadata

PROCESSING DATA
- enter data, digitise, transcribe, translate
- check, validate, clean data
- anonymise data where necessary
- describe data
- manage and store data

ANALYSING DATA
- interpret data
- derive data
- produce research outputs
- author publications
- prepare data for preservation

PRESERVING DATA
- migrate data to best format
- migrate data to suitable medium
- back-up and store data
- create metadata and documentation
- archive data

GIVING ACCESS TO DATA
- distribute data
- share data
- control access
- establish copyright
- promote data

RE-USING DATA
- follow-up research
- new research
- undertake research reviews
- scrutinise findings
- teach and learn

THE DATA
LIFECYCLE

CREATING
DATA

RE-USING
DATA

GIVING
ACCESS
TO DATA

PRESERVING
DATA

ANALYSING
DATA

PROCESSING
DATA

17

DATA MANAGEMENT PLANNING
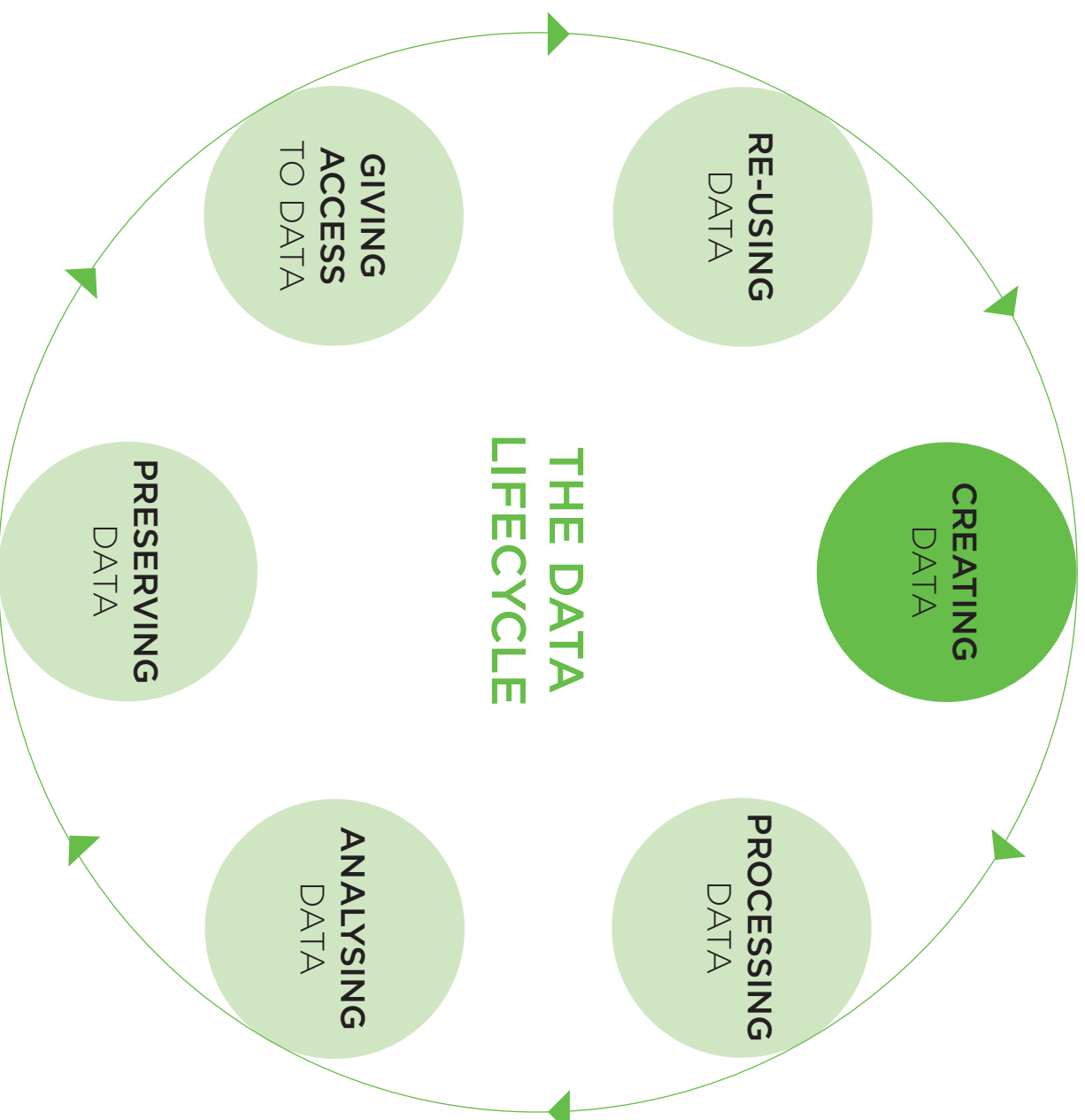
EXAMPLE: DATA LIFECYCLE PLANNING

Project: Negotiating Midlife - A Psycho-Social Investigation into the Subjective Experience of Ageing

Developed good file naming + folderstructure from the start

Starting to back up data on memorystick and laptop.

Developed interview schedule.

Designed consent forms (didn't include permissions to archive + share data)

Approached research participants.

. Pilot
. interviews - asked for written consent (Talkad) then through project
changed consent form to include consent to archive + share data.

Gathered detailed field notes
↳ good context to the interviews - interviews-...includes
consent to archive + share data.

Found transcribers

carried out transcription. Regular use five different transcribers. no strategy to maintain consistency.

Paper copies of transcripts stored in unlocked filing cabinets at home with consent forms.

Prepared for data collection

Fieldwork + collecting data

Processing data

Designed Research Project

Analysis of data

Chose PSEUDONYMS

Sharing data

Dissemination-conferences+publications

Writing up

Dissemination

Analysis conducted + saved in maxQDA

Reluctant to share fieldnotes as they are very disclosive (could apply access controls if archived)

Got retrospective consent to share + archive an interview which didn't originally sign per out.

Went back to explore the copyright of data used from ons.

Anonymisation

MAXQDA data exported

EXERCISE TWO

USING THE DATA MANAGEMENT CHECKLIST FOR YOUR RESEARCH PLANNING

Use the data management checklist to help point to relevant data management topics you need to consider when planning your research project.

DATA MANAGEMENT CHECKLIST	NOTES
DATA MANAGEMENT PLANNING	
Who is responsible for which part of data management?	
Do you need extra resources to manage data, such as people, time or hardware?	
DOCUMENTING YOUR DATA	
Are your structured data self-explanatory in terms of variable names, codes and abbreviations used?	
Which descriptions and contextual documentation can explain: what your data mean, how they were collected and the methods used to create them?	
How will you label and organise data, records and files?	
Will you apply consistency in how data are catalogued, transcribed and organised, e.g. standard templates or input forms?	
DATA FORMATTING	
Are you using standardised and consistent procedures to collect, process, check, validate and verify data?	
Which data formats will you use? Do formats and software enable sharing and long-term validity of data, such as non-proprietary software and software based on open standards?	
When converting data across formats, do you check that no data or internal metadata have been lost or changed?	

DATA MANAGEMENT CHECKLIST	NOTES
STORING YOUR DATA	
Are your digital and non-digital data, and any copies, held in a safe and secure location?	
Do you need to securely store personal or sensitive data?	
If data are collected with mobile devices, how will you transfer and store the data?	
If data are held in various places, how will you keep track of versions?	
Are your files backed up sufficiently and regularly and are back-ups stored safely?	
Do you know what the master version of your data files is?	
Who has access to which data during and after research? Are various access regulations needed?	
ETHICS AND CONSENT	
Do your data contain confidential or sensitive information? If so, have you discussed data sharing with the respondents from whom you collected the data?	
Are you gaining (written) consent from respondents to share data beyond your research?	
Do you need to anonymise data, e.g. to remove identifying information or personal data, during research or in preparation for sharing?	
Who has access to which data during and after research? Are various access regulations needed?	
COPYRIGHT	
If you are purchasing or re-using someone else's data sources have you considered how that data might be shareable?	
Have you established who owns the copyright of your data? Might there be joint copyright?	

COSTING TOOL: DATA MANAGEMENT PLANNING

ACTIVITY-BASED DATA MANAGEMENT COSTING TOOL FOR RESEARCHERS

If you were to give your research data to a researcher who did not know your research, what additional edits, changes or upgrades would you need to make to the data, for this researcher to understand and use them?

This tool can be used to estimate the additional costs – above standard research procedures and practices – that are needed to make research data shareable beyond the primary research team. It can be used to help prepare research grant applications.

We have provided a table indicating the activities you might list and cost to enable good data management. The likelihood is that such additional activities require extra researcher or administrative staff time input. Our model does not include data storage, preservation and dissemination activities. When data are deposited with a professional Data Centre, such as the UK Data Archive, these activities are often covered centrally.

STEP 1:
Check the activities in the table and tick those that may apply to your proposed research.

STEP 2:
For each selected activity, estimate the additional time and/or other resources needed and cost this e.g. people's time or physical resources needed such as hardware or software. Consider whether you need a dedicated data manager.

STEP 3:
Add these data management costs to your research application.

STEP 4:
Plan the data management activities in advance to avoid them competing with the need to focus on research excellence.

CAUTION:
Remember that when your research project nears it end you do not want these additional data management activities to compete with delivery of your planned outputs, authoring publications and the timely delivery of your project.

At this later stage the costs of preparing data for sharing may be significantly higher.

ADDITIONAL ACTIVITY	COMMENTS AND SUGGESTIONS	✓	COST
CONSENT FOR DATA SHARING Do you need to ask participants for their consent for data to be shared? Essential for qualitative interviews; possibly less so in quantitative surveys where data can be more easily anonymised.	• when consent for data sharing is considered as part of standard consent procedures – very low or no additional cost • when participants need to be re-contacted or re-visited after research to obtain retro-active consent for data sharing – could be high cost • does this require preparation of information sheets, consent forms or extra time for consent discussions/staff training?		
DIGITISATION Do analogue or paper-based research data need to be digitised to increase their potential for sharing?	• is additional equipment or software needed for conversion? • if simply image scanning text - may be low cost • if Optical Character Recognition required then checking for accuracy needed – may be high cost		
TRANSCRIPTION Will you transcribe qualitative data (e.g. interviews) as part of your research; or will you need to do this specifically so data can be shared and re-used?	• if part of research practice – very low or no additional cost • if not planned as part of research practice – potentially high additional cost • is full or partial transcription needed? Is translation needed? • if working with many trancribers or services, costs may be higher • consider cost of developing procedures, templates and guidance for transcribers		

RECOMMENDATIONS: STRATEGIES FOR RESEARCH CENTRES

Presented here are a set of strategies and tools the UK Data Archive recommends for research centres and departments for improving data management practices. These were developed from working closely with ESRC Centres and Programmes.

1. DATA MANAGEMENT STRATEGIES

Research centres and departments benefit from dedicated central research co-ordination, which can provide important input into planning and implementing data management and sharing activities.

Research hubs benefit from a centralised approach to data management providing both economies of scale and a lasting framework.

Benefits include:

- researchers can share good practice and data management experiences with each other, thereby building capacity, collective knowledge and resources for the centre
- a uniform approach to data management and creating standard policies on various data-related procedures and activities
- a record of who owns data and keeping track of projects and data over time, especially when researchers move
- data can be stored in a central location
- all researchers and staff are aware of duties, responsibilities and funder requirements regarding research data, with easy access to relevant information

Research hubs can provide shared guidance and tools:

- data management responsibilities assigned to named individuals
- standardised forms and templates e.g. for consent procedures, ethical review, data management plans
- data quality control standards, data transcription standard, confidentiality agreement for data handlers
- file sharing and storage procedures
- security policy for data storage and transmission
- data retention and destruction policy
- data copyright and ownership statement for the centres and for individual researchers
- standard data format recommendations
- version control and file naming guidelines
- information on funder requirements or policies regarding managing and sharing data that apply to projects or the centre
- research data sharing strategy, e.g. via institutional repository, data centres, website

2. DATA MANAGEMENT RESOURCES LIBRARY

This is a one-stop place where a research hub can hold all its relevant data management and data management planning guidance documents, for researchers to use.

Local resources can include:

- a Data Inventory for each project
- template consent forms and information sheets for use by projects
- ethical review forms
- guidelines on anonymisation
- version control guidance
- file naming convention guidance
- institutional IT management and back up procedures.
- transcription layout guidance and coversheets
- local statement on copyright
- local statement on retention and destruction of data
- local information on the Data Protection Act (1998), Freedom of Information (FoI) Acts in England (2000) and/or Scotland (2002), Environmental Information Regulations (2004)

External resources can include:

- ESRC Research Data Policy (2010)
- ESRC Framework for Research Ethics (2010)
- UK Data Archive Managing and Sharing Data: best practice for researchers (2011)
- UK Data Archive Managing and Sharing Data: training resources (2011)
- UK Data Archive advice on data formats
- Cabinet Office Data Handling Procedures in Government: Final Report (2008)
- JISC Freedom of Information and research data: Questions and answers (2010)
- UK research funders data policies compiled by the Digital Curation Centre

3. CHECKLIST OF RECOMMENDATIONS

TOPIC	DATA MANAGEMENT RECOMMENDATION
DATA MANAGEMENT STRATEGY Choosing the level of centre- or programme-level co-ordination and decision-making in planning and implementing management and sharing of research data	• level of centralisation or researcher autonomy for data management determined by type of research activities, methodological and discipline diversity, size of research hub, single or cross-institutional entity • allocate clear roles and responsibilities for all data management aspects; assign a data management co-ordinator for each project • develop a data management resources library • create a data inventory • recommended file structures for projects
ETHICS AND CONSENT Informed consent for people to participate in research and for use of information collected	• consent not prohibiting data sharing • gain consent for sharing, preservation and re-use of research data
COPYRIGHT Intellectual property rights of data	• identify source and ownership of third party data • establish conditions of use, copyright constraints and redistribution rights • statement on intellectual property rights of generated data
ANONYMISATION Procedures to protect the identity of individuals, organisations or businesses and avoid disclosure	• plan and anonymise data early in research • create anonymisation log of edits, replacements, removals or aggregations made
DOCUMENTATION Material explaining how data are created, what they mean, their content and structure, and alterations done	• create and sustain comprehensive documentation of data creation methodology • define specialist terms and acronyms used
RE-USE Plans for sharing and future use of data	• know relevant funders' data sharing policies • identify which data centres and institutional repositories can preserve your data and distribute them for re-use
DATA FORMATTING Layout, editing and structuring of data	• guidance for researchers and transcribers on transcription procedures and use of uniform editing and layout • use coversheet for transcripts of qualitative interviews • use non-disclosure agreement for administrative staff and transcribers accessing confidential data • list of recommended transcribing services
FILE FORMATS Formats and software in which research data are created, analysed, stored or preserved	• encourage use of standard or open lossless formats
VERSION CONTROL Checks and procedures to establish which version of a file is the most current	• establish procedures to control which version of a file is the agreed shareable version, or master version
STORAGE AND BACK-UP Steps to store data and back-up procedures to make copies to restore originals in case of data loss	• know your institutional arrangements and capacity for data storage • know your institutional procedure and regularity of data back-up, especially for remote and cross-institutional working
FILE SHARING AND STORAGE Strategy for sharing and storing research data and providing shared access	• for cross-institutional collaborative research, assess what kind of virtual research environment or file sharing and storage system you will use
TRANSMISSION Methods of moving data files from one location to another	• only send non-disclosive data in unencrypted form • encrypt confidential or personal data before transmission or password protected files • explore secure storage and sharing of files and folders with others across the internet using file synchronisation, virtual research environments or file sharing software
DESTRUCTION Procedures for safe and secure erasing of data at the conclusion of research project or when needed	• policy on retention and disposal of data

ADDITIONAL ACTIVITY	COMMENTS AND SUGGESTIONS	✓	COST
ANONYMISATION Do you need to remove identifying information or conceal the identity of participants before data can be shared? Anonymisation needs to be consistent throughout a data collection.	• for quantitative data (e.g. survey data) – may be low cost if pre-coded • for qualitative textual data (e.g. interview transcripts) – may be high cost as entire texts will need to be read and checked • for audiovisual data – may be very high cost, reduces the usefulness of data; access control is preferable where necessary • cost depends on how sensitive or complex data are and how much identifying information is recorded in the data - if only removal of names is required, cost is low • if anonymisation is planned before collection or transcription/digitisation – cost can be lowered		
DATA SECURITY AND ACCESS Protect data from unauthorised access, use, and disclosure	• for disclosive or sensitive data, determining conditions for controlling access to shared data may require extra time, team discussion and discussion with any host archive.		
FORMATTING AND ORGANISING Are your data files, spreadsheets, transcripts, etc. all in a uniform format or style and named with unique and clear file names?	• if planned by developing templates and structures for spreadsheets, databases, transcripts and file structures - low cost • if needed afterwards - higher cost		
DATA DESCRIPTION Are data in a spreadsheet or database clearly marked with variable and value labels, code descriptions, missing value descriptions etc.? Do textual data need description of context?	• if carried out as part of data creation or data input – lower cost • if needed afterwards – higher cost		
CLEANING Do quantitative data need to be cleaned, checked or verified? Do qualitative data need to be spell-checked?	• if carried out as part of data entry and preparation for analysis – low cost • if needed afterwards – higher cost		
DOCUMENTATION Do you have documentation for the data that describes the methodology of how data were gathered, processed, quality controlled, etc?	• if all data creation steps are well documented and documentation is kept well organised during research – low cost • if documentation needs to be written specifically afterwards – higher cost		
METADATA Do structured metadata need to be created when data are shared via a data centre or archive, e.g. completing a deposit form for the UK Data Archive?	• completing a UK Data Archive deposit form may take 1-2 hours		
FILE FORMAT Do data need to be converted to a standard or open format with long-term validity?	• for audiovisual data, converting to open digital formats can be time-consuming or require special equipment or software • is additional software or hardware needed for conversion?		
COPYRIGHT Do other parties hold copyright in the data; if so, do you need to seek copyright clearance?	• is time required to seek copyright clearance? • is legal input required?		
PLANNING ROLES AND RESPONSIBILITIES Do you need to plan data management for your research and allocate roles and responsibilities for various data management activities?	• if multiple partner institutions, researchers or funders are involved in research – higher cost • if staff training is required - higher cost		
OPERATIONALISING DATA MANAGEMENT What measures are needed to implement and operationalise data management throughout the research cycle?	• does putting your planned data management into practice throughout your research take time and resources e.g. regular team meetings, collaborative environments, additional staff?		

USING A DATA INVENTORY FOR PLANNING DATA MANAGEMENT IN CENTRES

Look at the basic Data Inventory tools provided by the UK Data Archive and evaluate how they might be used within your own organisation. Decide who is best placed to keep records about research projects and data and how information might be updated. Could this inventory be combined or integrated with records you already keep in your organisation, e.g. for publications, project outputs?

THE UK DATA ARCHIVE'S DATA INVENTORY FOR RESEARCH HUBS

An inventory of data collections created and acquired by researchers helps to keep track of data resources held in an organisation. A data inventory for a project or research centre can record:

• a definition of the data collection
• how they have been created
• from where they were obtained
• who owns them
• who has access, use and editing rights
• who is responsible for managing them
• how they will or can be shared

An inventory can be used as a tool to plan data management measures when research starts, and to keep track of management implementation during the research cycle via a regular update strategy. The inventory can:

• record storage and back-up strategies
• keep track of versions
• record data quality control procedures
• record any changes to planning

An inventory facilitates data sharing and complying with any requirements from funders or publishers regarding data sharing, as it contains much essential top-level information to enable future re-use and to make data archiving easy.

DATA INVENTORY TEMPLATES

The UK Data Archive developed an example Data Inventory - in line with the requirements for data management of the ESRC research data policy - which is currently structured as a MS Access database. Forms in the database can be used to add information on projects, related data collections (or datasets) and particular data and documentation files within data collections. Data management responsibilities can be allocated to named people. The database uses controlled vocabularies for types of data formats. This inventory can be used at research centre, department or project level and can be adapted for your own purposes. See www.data-archive.ac.uk/create-manage/strategies-for-centres/data-inventory

Alternatively, a simple MS Excel Data Inventory spreadsheet can be used. This could be derived from the ESDS data inventory template, developed in 2005 for research centres to co-ordinate data archive submissions. See www.esds.ac.uk/aandp/create/datatemplate.xls. It uses the following fields to keep track of each project and its data:

• Funding source (e.g. Centre budget, new ESRC award, Programme, other)
• Research Programme or Initiative
• Project title
• PI First Name
• PI Surname
• Other Researchers/Data Contacts
• Project Start Date dd-mm-yyyy
• Project End Date dd-mm-yyyy
• New datasets created? (Y/N)
• How many datasets created
• Dataset Type (qualitative, quantitative, mixed methods, model)
• Offering data for deposit at ESDS? (Note problematic issues surrounding data sharing)
• Other Plans to Share Data

DOCUMENTING YOUR DATA

MAKE DATA CLEAR TO UNDERSTAND AND EASY TO USE

A CRUCIAL PART OF MAKING DATA USER-FRIENDLY, SHAREABLE AND WITH LONG-LASTING USABILITY IS TO ENSURE THEY CAN BE UNDERSTOOD AND INTERPRETED BY ANY USER. THIS REQUIRES CLEAR AND DETAILED DATA DESCRIPTION, ANNOTATION AND CONTEXTUAL INFORMATION.

RESOURCES OVERVIEW

AREAS OF COVERAGE

- Documenting data

- Study-level documentation and context

- Data-level documentation

- Metadata

- Context debate

RESOURCES AVAILABLE

- PowerPoint presentation on 'Benefits of describing and contextualising research data'

- Exercise Presenter's Guide: Context for qualitative data

- Exercise: Context for qualitative data
  - Mothers and Daughters – interview extract
  - Mothers and Daughters – contextual information
  - Domestic Cooking – interview extract
  - Domestic Cooking – contextual information

- www.data-archive.ac.uk/create-manage/training-resources
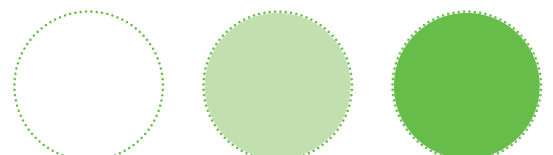
PRESENTER'S GUIDE

1. This exercise is designed to work with either one or two extracts. For shorter sessions, or if working in one small group, one exercise may be used. With longer, larger sessions, both can be used and groups can compare.

2. Distribute the data extract WITHOUT any accompanying context or biographical information. Normally, researchers would have additional information, but for this exercise, the idea is to have no context in order to mimic situations in which little or no context might accompany archived data.

3. For section three of the EXERCISE: CONTEXT FOR QUALITATIVE DATA consider levels of context from the transcript/micro level settings to interview settings to prevailing social conditions.

4. The Blaxter transcript has been transcribed in a Scots dialect. One topic area to explore with participants is whether this contextual information influences their reading. Would a vocabulary guide be useful as additional context?

EXERCISE MATERIALS

Blaxter, M., *Mothers and Daughters: Accounts of Health in the Grandmother Generation, 1945-1978* [computer file]. Colchester, Essex: UK Data Archive [distributor], July 2004. SN: 4943.

- Biographical information and interviewer notes for interview with G19
- Extract from interview G19 – *Mothers and Daughters*

Short, F., *Domestic Cooking and Cooking Skills in Late Twentieth Century England, 1996-1997* [computer file]. Colchester, Essex: UK Data Archive [distributor], August 2007. SN: 5663.

- Biographical information and interviewer notes for interview with 'L.A.'
- Extract from interview with 'L.A.' – *Domestic Cooking and Cooking Skills in Late Twentieth Century England*

CONTEXT FOR QUALITATIVE DATA

1. Read one (or both) interview extracts.

2. Identify one or two analytical points of interest, or research questions that this data might address – based on this short extract.

3. Consider your points or questions and think about what kinds of contextual information would help you.

4. Now read the contextual information provided for the extract(s) you reviewed. How does the information provided fit (or not) with what you thought would be useful?

5. Reflect now on the beginning of this exercise. Were you able – even without context – to come up with some kind of issue or question from the data?

6. Do you think there might be occasions when working with 'just the data' could be beneficial?

EXAMPLES OF LEVELS OF CONTEXT
UK Data Archive: www.data-archive.ac.uk/create-manage/document/study-level

Timescapes:
www.timescapes.leeds.ac.uk/assets/files/secondary_analysis/seconday_analysis_desirable_documentation.pdf

MOTHERS AND DAUGHTERS - INTERVIEW EXTRACT

G19:   An' a bottle o' Dettol wis aye... an' TCP.. that wis her. She niver gaed to the doctor. She wis a walkin' doctor. She didnae need a doctor...

we wis never nae weel. Ken. We hidnae much clathes, it wis a gym costume an' a white blouse an' that woman brushed it doon every night wi' Dettol an' water... it wis brushed, this wis to keep the germs aff ye. It must have worked, cos we wis niver nae weel, but we wis brought up on, fit, soup, tatties... butter, eggs.. fit else? Veggies. A real... a substantial meal. I mean, we niver got steak an' things like that.. I mean there wis meatless days, some days you couldnae get meat. I believe we wis healthier than what my kids were. They got a' this... a' the goodness has gone oot o' the grub, I think.

LP: Uh-huh... you think that food's got quite a lot to do with it [staying healthy]?

G19:   Well, my bringin' up made us healthier than the brinin' up I've gien them... comin'... wi' the things we ate. Wi' the things that we was forced to eat or do withoot. But nowadays... baked rice... now, you got baked rice. It wis made wi' eggs an' there wis currants in it, an' this wis a luxury, mind?

Husb: Nowadays, they'd sooner buy a tin.

G19:   Nowadays they get a tin an' there's nae eggs in it an' the goodness is oot o' it. Like... have you ever had frozen stuff an' you've cooked it an' you feel as though it didnae taste right... efter ha'in a fresh bit o' steak an' onion. Ken fit I mean? The juice... there's nae the juice in it. Well, that's whit we find wi' the things nowadays an' a, the richt good is out o' them... the body-buildin' material.. afore you eat it.

I mean, tinned soup, I would niver hae it in the hoose unless it wis maybe Karen [daughter] comin' in an' I wis gaun away in a hurry an' gettin' a tin o' soup... I wouldnae gie it to him [husband]... we were nae brought up like that, we wis brought up to get a' thing oot o' the groun' and intae a pot... My father grew a' thing. As I say we niver had the money that they've nowadays. My mother could have niver bought four tins, five tins of soup to pit in a pot. It would have cost her very little to put on a pot o' soup an' gettin' the full body o' that pot o' soup, ken?

Makin' toast at the fire wis a great thing... you niver tasted toast at the fire that you will in a grill, it's nae the same taste. An' baked tatties in a fire. Used to sit aoun' an' bake tatties, or bake chestnuts... it wis somethin'... we wis happy sittin' singin'... even when the kids were little we used to say "Come on, kids, come on an' we'll hae a little concert". We used to dress them up an' they used to sing an' dance. It wis great, ken... nowadays they'd think we was feel! And now, you see, when they're up we could sit an' speak about this an' laugh aboot it . what we used to dee, an' this an' that, an' dress them up an' mak' them sing an'... Isabel wis that fat, but she wis goin' to be a ballet dancer an' she wis gaun aboot Like an elephant! Ken. We all laugh aboot this nowadays. They were happy, days.

Kids are nae happy nowadays. They're gettin' too much. They're never deprived of onythin', they get it eventually. Even wi' the school... they're nae feart at their teachers the wey we were. The wey they were, we used to come hame... "I got the strap" an' then get a punch up for gettin' the strap. Karen comes hame an' says to me "I got the strap today". I says "Oh well, ging back an' get anither een"... she thinks I should ging up an' say to the teacher "Dinna strap my daughter". Ken fit I mean? It's a difference... they're defiant wi' the strap they were feart at the strap. They come hame an' telt us they got the strap.. we wanted to know what for... an' then we got a punch up for... gettin' it.

CITATION OF DATA
Blaxter, M., *Mothers and Daughters: Accounts of Health in the Grandmother Generation, 1945-1978* [computer file]. Colchester, Essex: UK Data Archive [distributor], July 2004. SN: 4943. Retrieved from www.esds.ac.uk/findingdata/snDescription.asp?sn=4943

EXERCISE: CONTEXT FOR QUALITATIVE DATA

MOTHERS AND DAUGHTERS - CONTEXTUAL INFORMATION

The aim of Blaxter and Patterson's (1982) research was to study inter-generational transmission of deprivation using a sample of women in 58 three-generation families and was part of a larger ESRC programme on Transmitted Deprivation. Sampling was purposive: families that remained working class across two generations, grandmother-daughter co-location in a Scottish city, and continuing contact. The study addressed diverse factors, exploring whether health and social histories, attitudes, and health behaviours would affect the health experiences of the children and were possibly transmitted across generations. Nutrition was one of several topics addressed; others were orientations toward medicine, antenatal care, preventive behaviour, use of lay remedies, etc. The study used several types of data: information from longitudinal visits with the mothers, health visitor reports, etc. Other data, including the material archived at ESDS Qualidata, are semi-structured interviews that focused on attitudes and perceptions. The original study was intended to inform social policy.

A year after publishing their book, Blaxter and Patterson re-analysed their data to study the historical and moral significance of food. They reported on what constituted 'good food': specific foods were less important than a 'proper' meal, as contrasted with processed foods, or 'snackery'. They also used their rich intergenerational data to compare the different attitudes and behaviours between grandmothers and their daughters.

Interviews were done by two educated, white women. Patterson did the majority of the grandmother interviews. She was from the same area where the families lived. Blaxter praised her ability to gain rapport with the respondents. The study was presented to respondents as being about child-rearing and child-rearing beliefs and practices across generations. There had been regular visits to the families by either Blaxter or Patterson; mothers were interviewed at the end of the six-month study.

Biographical information and interviewer notes for the interview with Grandmother 19.

Date of interview: 1978; age of G19: 43.

[text has been edited slightly to remove identifiers.]

Upstairs flat in drab block of 4. Untidy. Back garden overgrown grass. A daughter with baby living with parents. Doesn't appear to be married. Another daughter who is pregnant was also present. Not sure whether she is living there also. A teen aged daughter also lives at home. The two daughters present looked gaunt and ill. Son-in-law came in later and left granddaughter - seemed to be about 4 or 5. G19 seemed quite forthcoming despite the presence of all these people. But when I was leaving she showed me to the door and confided that she and [someone of her daughter's generation] were very different: "Although she's a nice person, she was brought up on the good things of life. She likes to get out and enjoy herself, while I only thought of my family."

REFERENCES
The exercise is adapted from: Bishop, L. (May 2007) 'A Reflexive Account of Reusing Qualitative Data: Beyond primary/secondary dualism', Sociological Research Online [Online], Special Section on Reusing Qualitative Data, 12(3). Retrieved from www.socresonline.org.uk/12/3/2.html.

Blaxter, M. and Patterson, E. (1982) *Mothers and daughters : a three-generational study of health attitudes and behaviour*, London: Heinemann Educational Books.

CITATION OF DATA
Blaxter, M., *Mothers and Daughters: Accounts of Health in the Grandmother Generation, 1945-1978* [computer file]. Colchester, Essex: UK Data Archive [distributor], July 2004. SN: 4943. Retrieved from www.esds.ac.uk/findingdata/snDescription.asp?sn=4943

EXERCISE: CONTEXT FOR QUALITATIVE DATA

DOMESTIC COOKING - INTERVIEW EXTRACT

FS: …can you think of any techniques you use when you're trying to eat a healthy diet?

LA: …um…{thinks}…for a healthy diet well we don't…fry much…we…but then I've always grilled rather than fried stuff…for example sausages…it's the one sort of meat thing that we all quite like…and {L's older daughter} likes…we do have sausages…usually once a week…a nice quick meal…my liver and bacon equivalent…a kind of warm…easy to make meal and you've got nice gravy that goes into mashed potato…um so I would grill but I've always grilled…I don't think there's anything that I particularly change but then…

FS: …um…

LA: …I think also my mum was quite conscious of health even…twenty years…thirty years ago…she was quite advanced with that as a cook…to not fry…to grill most things…so it's more up-bringing I'd say then conscious healthy living…that has um…yeah I mean…you know with potatoes we're much more likely to have them mashed or boiled…we'll have the odd sort of…chips out of the freezer…but I'd never have a deep fat fryer and I'd never make chips…um I might do sliced fried potatoes with an omelette…I think that's a really nice combination…um…so…yeah…there's…the only concession is very little frying of food…

FS: …um…

LA: …we don't have lots of meat but again that's…I mean that is healthy {?}…we don't particularly like a lot of meat…um…not much fat…semi-skimmed milk…can't think of anything else…

FS: …the last section…in Britain today…etc?

LA: I think it's brilliant…I think um…my grandmother who I was talking about earlier and who's had a lifetime of preparing food and sick of it…she's also nearly blind…and she came to stay with me when {L's younger daughter} was first born so that was about eight months ago…and she's never been round a big supermarket…she lives in Shropshire…um…and they've got quite a good shop that sells quite a variety of foods for where it is…

FS: …um…

LA: …but I took her over to the Sainsbury's in Haringey…and we walked up and down the aisle…and I had to tell her what was on the shelves because her eyesight isn't good enough…and we spent half an hour just in the fruit and veg section…pointing out all the different pears and all the different…all the exotic…star fruit and…and it was all too much for her in a way because you know she's eighty five or something and she's had meat and two veg all her life and yet she's still sort of young enough at heart to want to try these different things…so she had pasta for the first time…the first time she'd ever had pasta was here…and I was just thinking we take it so much for granted now…but things are so different now…I think it's brilliant…I wonder where we can go from here…I suppose there'll just be more and more influx of foreign foods and foreign recipes and I'm all in favour of it all…I think it's a good thing…

FS: What about prepared and semi-prepared foods?

LA: …it's funny because this same Grandmother…if she knew the extent to which we buy ready prepared foods…she'd be absolutely horrified because of the war-time mentality…of everything must be…cooked…[edits] …and I…there's part of my background that tell me I shouldn't do it…

FS: …um…

LA: …but I also know that for today's living…it's impossible…[edits] I cannot think how they [mother and mother-in-law] managed it…but then I think our lives are different because we have a much bigger social like and…we're more house proud…we want to get our houses

looking nice and…I always remember our house being the same for years and years…but my Mum was always cooking our food…plus obviously most women I know work…with kids, so… you know you're both out of the house…[edits] so all in all…I think it's…it's…there's so much a place for it and I wish in a way it wasn't because it's cheating…I still feel it's cheating…I think I said earlier…

FS: …um…

LA: [edits]…in fact my sister makes me laugh 'cos her and George are the epitomy [sic] of the kind of nineties could both working really long hours in trendy jobs…and she said cooking for her now is heating up some pasta…some sauce…she says they've 'cooked' if they've done that… most of the time they're either eating out or getting a takeaway…so it's like 'oh should we cook tonight' and then they get ready prepared food and yet deep down…she's got the same sort of thing as me…she comes from a background where cooking food was important…um…and you feel you're cheating…[edits] …whereas because of the life I'm leading…you know at home with children…I've got more time here…more time to think about food and whatever…and more time to shop for it…but um..

FS: …um…

LA: there's still a huge part of our lives with ready prepared food…and I can't see that going to change for a while…I just can't see…because as the kids get older and get more time for me I'll probably end up working more…I can't see that suddenly I'm going to have two hours every day…to create wonderful food…and I just think back and wonder how my Mum did it…I don't think she went out much…she didn't go out much…didn't watch telly much…she didn't have hour long conversations on the phone with friends which I frequently have…I just think our lives are very different…I'm in favour of it I think.

CITATION OF DATA
Short, F., *Domestic Cooking and Cooking Skills in Late Twentieth Century England, 1996-1997* [computer file]. Colchester, Essex: UK Data Archive [distributor], August 2007. SN: 5663. Retrieved from www.esds.ac.uk/findingData/snDescription.asp?sn=5663

EXERCISE: CONTEXT FOR QUALITATIVE DATA

DOMESTIC COOKING - CONTEXTUAL INFORMATION

This research was conducted as part of Frances Short's work toward completing a Ph.D. in Food Policy in 2002. She also has a diploma from a recognised cooking school and has worked as a professional chef.

The aim of the research was to further understanding and debate by providing a systematically researched and theoretically based way of thinking about cooking and cooking skills. The research took the form of a two-stage study. (This data extract is from the first stage.) Both stages were based around semi-structured interviews but the first also included the keeping of 'cooking diaries'.

Couples aged between 30 and 50 from different social, financial and occupational backgrounds and household structures (for example: two couples had very young children living with them , another couple had no children, another a teenage daughter who came to stay at the weekends and so on) were used as data sources in the first stage of fieldwork. The interview schedule was designed around current areas of concern in regards to domestic food, eating and cooking practices but with sufficient room to allow exploration and points of interest develop. Topics discussed with the informants included 'childhood experiences of cooking and eating', 'current cooking practices', 'the role of ready-meals', and 'typically British food'.

In both stages, informants were selected opportunistically. For reasons of accessibility, both first and second stage informants came mostly from the Greater London area. The interviews took place in the informants' own homes or workplaces and were recorded. All the informants were given a shopping voucher worth fifteen pounds as a thank-you.

Biographical information and interviewer notes for the interview with L.A.

Interviewee description: Woman in her 30s, married with two daughters, white British, with polytechnic or university degree. She is not currently employed. Husband is a self-employed journalist. She receives over £500 per week in child benefits. She has lived with her husband for seven years in an owned home (with mortgage). Interview was done c2001.

Interviewer notes: "L.A. had a neat house, trendy but both she and her partner viewed their trendiness quite ironically and knowingly. I interviewed L.A. and her partner L.B. in turn on the same evening whilst they swapped over putting their daughters to bed."

REFERENCE
The exercise is extracted with minor editing from Short (2003) 'Domestic cooking practices and cooking skills: findings from an English study', Food Service Technology, Volume 3, Issue 3-4, pages 177–185, September/December 2003

CITATION OF DATA
Short, F., *Domestic Cooking and Cooking Skills in Late Twentieth Century England, 1996-1997* [computer file]. Colchester, Essex: UK Data Archive [distributor], August 2007. SN: 5663. Retrieved from www.esds.ac.uk/findingData/snDescription.asp?sn=5663

FORMATTING YOUR DATA

CREATE WELL ORGANISED AND LONGER-LASTING DATA

USING STANDARD AND INTERCHANGEABLE OR OPEN LOSSLESS DATA FORMATS ENSURES LONG-TERM USABILITY OF DATA. HIGH QUALITY DATA ARE WELL ORGANISED, STRUCTURED, NAMED AND VERSIONED AND THE AUTHENTICITY OF MASTER FILES IDENTIFIED.

RESOURCES OVERVIEW

AREAS OF COVERAGE

- File formats

- Data conversions

- Organising files and folders

- Quality assurance

- Version control and authenticity

- Transcription

RESOURCES AVAILABLE

- PowerPoint presentation on 'Formatting your data - data quality control, formats, software, versioning, organising and naming'

- PowerPoint presentation on 'Formatting your data – Qualitative data transcription'

- Exercise 1: File naming

- Exercise 1 Answers: File naming

- Exercise 2 Presenter's Guide: Qualitative data transcription

- Exercise 2a: Qualitative data transcription, Ph.D. study on ageing

- Exercise 2a Answers: Qualitative data transcription, Ph.D. study on ageing

- Exercise 2b: Qualitative data transcription, Ph.D. study on bulimia

- UK Data Archive model transcript

- Quiz: Formatting your data

- Quiz Answers: Formatting your data

- www.data-archive.ac.uk/create-manage/ training-resources
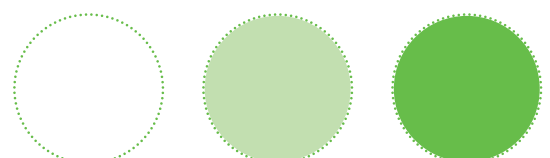
EXERCISE ONE

FILE NAMING

1.   Read through the following file names.
2.   If you returned to this data folder in a year's time do you think you would be able to recognise what each of these files contains?
3.   What information do you think you need in a file name in order to identify what is in the file's contents?

📄	Doc. 1	📊	My data
📄	IMPORTANT	📄	My Passwords
📄	Thesis Final final	📄	Thesis version 12
📄	My study	📊	Data chart for interviews
📄	Interview with Jane	📄	Int 1 (2)
📄	Interview with Janet	📄	My thesis (copy)
📄	Int. 1	📄	New doc.

FILE NAMING

Well-organised file names and folder structures make it easier to find and keep track of data files. A system needs to be practical and used consistently.

Good file names can provide useful cues to the content and status of a file, uniquely identify a file and help in classifying files. File names can contain project acronyms, researchers' initials, file type information, a version number, file status information and date.

Best practice is to:

• create meaningful but brief names
• use file names to classify broad types of files
• avoid using spaces and special characters
• avoid very long file names

Whilst computers add basic information and properties to a file, such as file type, date and time of creation and modification, this is not reliable data management. For example, when files are copied, for revision or as a template, the system does not account for their new purpose and treats them as a copy of the original file.

Examples:

• **FG1_CONS_12Feb10** is the file that contains the transcript of the first focus group with a study of consumers, that took place on 12 February 2010
• **Int024_AP_5June08** is an interview with participant 024, interviewed by Anne Parsons on 5 June 2008

Version numbering in file names is useful to indicate file revisions or edits, especially in collaborations. This can be through discrete or continuous numbering depending on minor or major revisions.

Example:

FILE NAME	CHANGES TO FILE
LC_Interviewschedule_1.0	Original document
LC_Interviewschedule_1.1	Minor revisions made
LC_Interviewschedule_1.2	Further minor revisions
LC_Interviewschedule_2.0	Substantive changes

PRESENTER'S GUIDE - QUALITATIVE DATA TRANSCRIPTION

Ask participants to read through the two transcription exercises based on real-life projects.

The first presents part of a transcribed interview from a Ph.D. project on ageing and midlife and the second presents a transcriber confidentiality agreement and transcription guidance notes from a Ph.D. project on recovery from bulimia nervosa.

Finally, hand out the UK Data Archive model transcript to show how the use of a consistent template can be useful across transcribers for a single project.

Ask participants to discuss and evaluate how good each is and in particular:

• what might they have done differently for the ageing transcript?
• what are their thoughts on the bulimia documents?
• had they considered giving guidance/confidentiality agreements to their own transcribers?

EXERCISE 2: DISCUSSION POINTS

EXERCISE 2A: AGEING AND MIDLIFE STUDY

The researcher gave the audio recordings to different transcribers with varying ranges of transcribing experience. Some transcribers were students and some were secretaries who all agreed to do the work informally, cheaply and quickly. The transcribers were not given any guidance by the researcher on prescribed transcription conventions or on how the transcript should look. As a result the transcripts were returned in varying formats using different conventions.

Discuss with the group the resulting types of problems for the researcher:

• transcripts were all formatted differently making the project appear messy, and perhaps less professional, when sharing the data with others
• turn-taking was recorded differently by different transcribers so some transcripts, where every 'umm' and 'ahh' were recorded, ended up being far longer than necessary - the use of recording these pause fillers may be appropriate in some research projects though
• spelling was not always correct so the transcripts needed careful proofing
• file naming was inconsistent
• descriptions recorded about the interview context were inconsistent
• no page numbers or line numbers were used, making referencing difficult
• there were confidentiality issues with regard to some transcribers where one in particular let her family members read through the transcripts and discuss the data - a confidentiality agreement for transcribers should have been used.

Other related issues that could be discussed with the group following this exercise:

• how to send data (e.g. audio recordings and transcripts) between the researcher and transcriber, such as sending encrypted files or USBs by recorded delivery

• the subject specificity of the transcription, for example conversation analysis would require a high level of annotation, while a psychosocial approach would require the researcher to do the transcription themselves (recording emotional and unconscious dynamics and reflections)
• formatting of the transcript, for example, if transcripts are formatted in two columns (one with speaker tags and the other with the utterances) it can be more difficult to import into a CAQDAS package.

EXERCISE 2B: RECOVERY FROM BULIMIA NERVOSA STUDY

The research incorporated a series of in-depth qualitative interviews, which contained highly sensitive data. The researcher asked her transcribers to follow a set of transcribing conventions that she clearly set out in a short guide. She issued a sample transcription and also asked her transcribers to sign a confidentiality agreement.

Points for discussion with a group could include:

• how important the experience and knowledge of the transcriber is - for example in the transcription for a project into drug use where 'ease' is written instead of 'Es' and 'heroine' instead of 'heroin'! (the Bulimia transcriber's guidance offers useful context)
• how it is useful to brief transcribers in advance of the work
• how even if transcribers follow conventions, errors or idiosyncrasies could still occur so double-checking is advisable
• the way that this researcher has considered security issues upfront, such as transfer of the data and storage of data
• how information about keeping research information 'secure' whilst in the possession of the transcriber could be explained, such as using locked filing cabinets, removing from the hard-drive after use, and ensuring secure transfer of data back to the researcher
• why it might be necessary to ask a transcriber to destroy data, at what stage and what precautions should be taken to avoid data loss - in this project, data was deliberately deleted and destroyed once the analysis took place so there was no opportunity for going back to the data, nor sharing or archiving

[Interview_2_with_Penny [1][1]]

Interview with Penny
Interview 2
In her office
6th November 2007

The first set of questions is about aging mostly and the experiences of aging. Um first of all, if you could tell me your date of birth actually?

Fifteenth of the fifth forty eight.

Ok could you start by describing yourself and telling me a few words about yourself

About myself, oh my god, I'm 55 and the mother of 3 daughters, um I therefore dye my hair constantly, I have still got one daughter still living at home and I am happily married. I have been married for 31 years, I know I don't seem like it [said in a breathy jokey voice].

Mmm

I was a late mother, I didn't have children till I was 30 and we were married for 6 years and I just didn't want to be a mother, but then again I didn't want to get married, so, and then all of a sudden I wanted it like yesterday. I work everyday, but not full time at the moment and I am very happy with my lot.

Sorry how old did you say you were?

55, botox is a wonderful thing

(Clears throat)

So starting wherever you like can you tell me about your life?

I had a charmed early life, my mum and dad, I was brought up in the east end.

Mmm

My mum didn't work, I had a sister and my sister is two years older than me and I had an idyllic childhood, even though I had a tragedy, I had a brother, John, who died in a car accident when he was 9 and obiously it affected my parents greatly because he was the boy, but we all got through it and I think that made me probably, I don't know if it affected my life in the way that I lived my life but I never felt that when someone said to me, Oh their Granny died it didn't affect me as greatly because they were quite old really, and so I suppose death affected me and how I thought about people dying affected me.

Mmm

If someone has lasted to this age I would still be very upset for that person but to be honest I wouldn't be upset that that person has gone because that person was old. But I have since lost my mum and she was 80, had a good life and all the rest of it,

File was labelled [Interview_2_with_Penny [1][1]]

Interview with Penny
Interview 2
In her office
6th November 2007

The first set of questions is about aging mostly and the experiences of aging. Um first of all, if you could tell me your date of birth actually?

Fifteenth of the fifth forty eight.

Ok could you start by describing yourself and telling me a few words about yourself

About myself, oh my god, I'm 55 and the mother of 3 daughters, um I therefore dye my hair constantly, I have still got one daughter still living at home and I am happily married. I have been married for 31 years, I know I don't seem like it [said in a breathy jokey voice].

Mmm

I was a late mother, I didn't have children til I was 30 and we were married for 6 years and I just didn't want to be a mother, but then again I didn't want to get married, so, and then all of a sudden I wanted it like yesterday. I work everyday, but not full time at the moment and I am very happy with my lot.

Sorry how old did you say you were?

55, botox is a wonderful thing

(Clears throat)

So starting wherever you like can you tell me about your life?

I had a charmed early life, my mum and dad, I was brought up in the east end.

Mmm

My mum didn't work, I had a sister and my sister is two years older than me and I had an idyllic childhood, even though I had a tragedy, I had a brother, John, who died in a car accident when he was 9 and obiously it affected my parents greatly because he was the boy, but we all got through it and I think that made me probably, I don't know if it affected my life in the way that I lived my life but I never felt that when someone said to me, Oh their Granny died it didn't affect me as greatly because they were quite old really, and so I suppose death affected me and how I thought about people dying affected me.

Mmm

If someone has lasted to this age I would still be very upset for that person but to be honest I wouldn't be upset that that person has gone because that person was old. But I have since lost my mum and she was 80, had a good life and all the rest of it,

Comment [b1]: The original file was labelled like this. This is poor version control. What does the [1][1] stand for? Might be better as Int.001_final.

Comment [b2]: Might an ID be useful too?

Comment [b3]: Is this the second interview with this particular respondent or the second interview in the study as a whole?

Comment [b4]: Would more contextual information be beneficial?

Comment [b5]: It would be easier to read dates if they were written numerically. It would also make it easier to search for dates in a transcript if written numerically.

Comment [b6]: Clear speech demarcation and speaker tags are crucial in transcripts. They are not used here.

Comment [b7]: How much description of intonation should be included and what are the rules for assigning it?

Comment [b8]: Are all sounds like 'mmm's going to be represented? Conversation analysis may be very interested in capturing them, and gaps.

These 'mmms' have been put into the turn taking sequence which makes the transcript overly long.

Comment [LC9]: Are round or square brackets to be used when noting sounds?

Comment [b10]: At what point should pseudonyms be added? If they are repeat interviews as these interviews were then it might be best to do it after fieldwork is completed. The researcher could brief interviewees to avoid using names when referring to people so it would reduce the burden on the anonymisation process.

Comment [b11]: Need to double check spelling.

but she had a good life. [cough]. Yes so I had a wonderful life and yeah I count myself very lucky.

You said something about the way it made you think about other people's deaths, but did it make you think about your own death?

No, not me no, no I wouldn't think about my own death.

Mmm

I don't think about death much, I have to say, because apart from my brother dying and my mum dying no death has affected me, you know, that's what I mean I don't think a death would affect me as much as my brother's did because he was so young and it was so tragic and it was my dad who backed a car into him, so he had to live with that and it was a pure accident…The car went that way and my brother instead of going that way [left or right] went backwards you know in a garage and um at the time my mum went a bit, I wouldn't say doolally, but she went so normal it was abnormal, you know so, at the time it was awful really and me and my sister would go out and my dad was so upset because he thought we blamed him, which we didn't, and we just felt for him you know. He had to live with that for all of his years, but he said thank god it was me, because if it had been anyone else he would have killed them, he would have had such anger. Yeah so that was the worse case we could have thought about.

Comment [b12]: Is it necessary to include additional noises?

Comment [b13]: What does the ellipse mean? Pause, missing words?

Comment [b14]: Here context is required. The original interviewer would have had to have recorded the body language here.

Comment [b15]: Need consistency with spelling certain words. Should it be capitalised or not?

Comment [b16]: Page number should be at the bottom. Possibly even line numbers.

A Qualitative Exploration of Recovery from Bulimia Nervosa:
Transcription Guidance Note

Background to the Research
Bulimia is very difficult to overcome, so how and why some people manage to recover is not very well understood. The term recovery also means different things to different people; clinicians have suggested various criteria but there is currently little attention to what recovery means to those who have actually lived and experienced the recovery process.

The purpose of my research is to look at recovery from the point of view of former sufferers of bulimia. After conducting a number of in-depth interviews, I hope to yield a rich collective account of the difficulties encountered and the insights gained during the journey of recovery.

I am conducting this research as part of a PhD in the Department of Sociology at the University of London, under the supervision of Professor Jane Woodstock. My research is being funded by the Economic and Social Research Council (ESRC) and has been granted full ethical approval by the Research and Enterprise Office (REO) at the University of London.

Due to the sensitive nature of this research, you will be required to sign a Confidentiality Agreement which confirms you will adhere to the principles of anonymity and confidentiality and will not speak about the content or nature of the interviews to anyone apart from myself.

Theoretical Approach: Verbatim Transcription
Decisions about how transcription should be carried out are intimately connected with the type of analysis that is intended. Transcription of speech is always a compromise: greater detail gives more material for interpretation, yet too much detail can slow up the reading of the text in an artificial manner. This project requires **full verbatim transcription.** Taking a full verbatim transcription approach means that as well as preserving the actual words which were spoken, extra verbal material captured on the recording – such as the speaker's use of intonation, pauses, rhythm and hesitation – is also preserved. This keeps some of the additional meaning that was conveyed in the original interview, thereby providing contextual information as to the manner in which words were spoken. In addition, verbatim transcription requires that the character of the conversational exchange is apparent, so the words of the researcher must also be included.

General Notes:
- Document should include a header on every page with the serial number of the interview on the left hand side and your name on the right hand side
- Insert page numbers at the bottom of each page, in the centre
- Use Times New Roman, font size 12, type what the interviewer says in **bold** and justify the text
- Identify the interviewer and the respondent separately and indicate the gender of the respondent. Use I: for the Interviewer and either F1: or M1: for the respondent depending on whether they are male or female (see attached example)
- Although I request that a record of what the interviewer says be included, the one exception to this concerns 'back channel utterances', i.e. where I can be heard in the background saying words such as "right", "yeah", "I see" or utterances such as "mmhhmm" whilst the interviewee is speaking. These function to encourage the

1

respondent to continue speaking and reassure them that they are being listened to. It is not necessary to break up the respondent's speech by including them
- Use punctuation as for normal written prose. Grammar should not be altered or "tidied up". Do not use 'eye spellings' (e.g. "enuff" for 'enough')

Things to Include in Full:
- Unfinished questions or statements that trail off – indicate these with ellipses (…), for example: "I never did understand her approach, the way she saw it, or…"
- False sentence starts
- Repeated phrases, words, statements or questions
- Discussion that continues after the interview appears to be 'formally' finished
- Non-lexical utterances or 'fillers' such as 'umms' and 'errs' and 'uhs'
- Hesitations and Pauses – indicate these with ellipsis (…), for example: "well…recovery to me…sort of means…err…"
- To indicate an exclamation of surprise, shock or dismay, use the standard exclamation mark
- Emphases – indicate any emphasis on a word or phrase by putting it in *italics*

Things to Include in Brackets
- Noises in background - for example (loud banging) or (door slams) or (muffled voices)
- The tone of the respondent. Here I am happy for you to include any comments on mood, feeling, passion, emotion and paralinguistics - for example: (laughs loudly) or (mumbles slowly) or (sounds angry) or (falters slightly) or (sighs)
- Unclear words or phrases must be marked where they occur within the text by placing the word "inaudible" in brackets and in bold e.g. **(inaudible)**. PLEASE DO NOT GUESS AT ANYTHING WHICH YOU CANNOT UNDERSTAND.

After Transcription
When you have completed transcribing an interview, please email it to me at (email address). On receipt of the transcript I will confirm that you can then delete and destroy both the interview recording and the interview transcript in all its forms of storage – e.g. CD, WAV file, Word Document.

Thank you for agreeing to take part in this research project. Attached is a sample interview, indicating how I would like the interviews for this research to be transcribed.

If there are large parts of the recording which you cannot understand or decipher, then please contact me immediately. Likewise, if you are concerned with any aspect of the transcription or these guidelines are unclear, then please do not hesitate to contact me on (mobile number) or at (email address)

Alice Jackson

2

<u>University of London</u>

An Exploration of Recovery from Bulimia Nervosa

Transcriber Confidentiality Agreement

This research is being undertaken by Alice Jackson, PhD candidate in the Department of Sociology, University of London. The purpose of the research is to explore recovery from bulimia nervosa from the point of view of former sufferers.

As a transcriber of this research, I understand that I will be hearing recordings of confidential interviews. The information on these recordings has been revealed by interviewees who agreed to participate in this research on the condition that their interviews would remain strictly confidential. I understand that I have a responsibility to honour this confidentially agreement.

I agree not to share any information on these recordings, about any party, with anyone except the Researcher of this project. Any violation of this and the terms detailed below would constitute a serious breach of ethical standards and I confirm that I will adhere to the agreement in full.

I, _____ agree to:

1. Keep all the research information shared with me confidential by not discussing or sharing the content of the interviews in any form or format (e.g. WAV files, CDs, transcripts) with anyone other than the Researcher.

2. Keep all research information in any form or format (e.g. WAV files, CDs, transcripts) secure while it is in my possession.

3. Return all research information in any form or format (e.g. WAV files, CDs, transcripts) to the Researcher when I have completed the transcription tasks.

4. After consulting with the Researcher, erase or destroy all research information in any form or format regarding this research project that is not returnable to the Researcher (e.g. CDs, information stored on my computer hard drive).

Transcriber:

_____   _____   _____
(print name)   (signature)   (date)

Researcher:

_____   _____   _____
(print name)   (signature)   (date)

This study has been reviewed and ethically approved by the Research and Enterprise Office (REO) at the University of London

EXAMPLE: UK DATA ARCHIVE MODEL TRANSCRIPT

Study Name: SN000
Depositor: Prof A Academic
Interviewer: JJ

Interview number: 002
Interview ID: Jane Doe
Date of interview: 1 Sept 2008

Information about interviewee
Date of birth: 1940
Gender: Female
Geographic region:

Marital status: Married
Occupation:

Y: I came here in late 1968.

I: You came here in late 1968? Many years already.

Y: 31 years already. 31 years already.

I: (laugh) It is really a long time. Why did you choose to come to England at that time?

Y: I met my husband and after we got married in Hong Kong, I applied to come to England.

I: You met your husband in Hong Kong?

Y: Yes.

I: He was working here [in England] already?

Y: After he worked here for a few years -- in the past, it was quite common for them to go back to Hong Kong to get a wife. Someone introduced us and we both fancied each other. At that time, it was alright to me to get married like that as I wanted to leave Hong Kong. It was like a gamble. It was really like a gamble.

I: You were very brave to think about going abroad as you were so young at that time.

Y: No. I was very innocent at that time. I was very obedient to my mum. The circumstance of my family was not very good. So I thought that if I married to someone whose financial situation was good, then it could help my family's finance.

I: Is that so? What rank are you?

Y: I am the eldest.

I: You are the eldest sister.

Y: Yes. I have two younger brothers and two younger sisters.

FORMATTING YOUR DATA

1. Organising my file names and folder structures is
   (choose all possible answers)

   a) a waste of time
   b) not necessary until I finish my project
   c) important for sharing and future use of the data
   d) good practice for my research project

2. Systematic and logical file naming
   (choose all possible answers)

   a) makes it easier to keep track of your data files
   b) provides useful cues to the content and status of a file
   c) can help in classifying files
   d) is not necessary as I don't need a system

3. In order to guarantee safe and long term access to your research data you should covert data into a standard format. Which standards are most appropriate for converting documentation?
   (choose all possible answers)

   a) Rich Text Format (.rtf)
   b) PDF/A or PDF (.pdf)
   c) OpenDocument Text (.odt)
   d) Notepad (.txt)

4. Proprietary software is software that is
   (choose one answer only)

   a) the best option for keeping my data safe
   b) going to be around forever
   c) a safe and stable way to store my data
   d) not recommended for long term storage of my data

5. Digital information can easily be copied, changed or deleted. How can you ensure your data are authentic?
   (choose all possible answers)

   a) keep master files of data
   b) regulate write access to master files
   c) assign responsibility for master files
   d) record changes to master files

6. Transcription of data
   (choose one answer only)

   a) should always be carried out by the researcher
   b) should always be carried out by a professional transcriber
   c) can be carried out by either a researcher or a professional transcriber as long as there is consistency within the transcripts
   d) should always be carried out in full

FORMATTING YOUR DATA

1. Organising my file names and folder structures is

   a) waste of time
   Incorrect. It's definitely not a waste of time. Having well organised file names and folder structures makes it easier to keep track of your data files, provides useful cues to the content and status of a file and can help in classifying files

   b) not necessary until I finish my project
   Incorrect. You should organise your folders along the way - as part of good research practice - then there will be very little reorganisation to be done at the end of a project.

   c) **important for sharing and future use of the data**
   Correct. Good organisation of your files and formats is essential for sharing your data and for understanding your own data in the future.

   d) **good practice for my research project**
   Correct. Make good file naming and folder structuring part of your research practice

2. Systematic and logical file naming

   a) **makes it easier to keep track of your data files**
   Correct. This is one advantage.

   b) **provides useful cues to the content and status of a file**
   Correct. This is one advantage.

   c) **can help in classifying files**
   Correct. This is one advantage.

   d) is not necessary as I don't need a system
   Incorrect. A good system really helps, especially when returning to data later on.

3. In order to guarantee safe and long term access to your data it is important to covert the data into a standard format. Which standard format is the most appropriate for converting documentation data into?

   a) **Rich Text Format (.rtf)**
   Correct. This is one option.

   b) **PDF/A or PDF (.pdf)**
   Correct. This is one option, but you won't have full access to any text.

   c) **OpenDocument Text (.odt)**
   Correct. This is one option.

   d) Notepad (.txt)
   Incorrect. This format will lose formatting and information.

4. Proprietary software is software that is

   a) the best option for keeping my data safe
   Incorrect.

   b) going to be around forever
   Incorrect. Proprietary software may be around for a long time, but there are no guarantees about how long.

   c) a safe and stable way to store my data
   Incorrect. It may well be safe and stable at the present moment and for the near future, but there are no guarantees. It is better to use non-proprietary software for long-term storage and security of your data)

   c) **not recommended for long-term storage of my data**
   Correct.

5. Digital information can easily be copied, changed or deleted. How can you ensure your data are authentic?

   a) **keep master files of data**
   Correct. This is one solution.

   b) **regulate write access to master files**
   Correct. This is one solution.

   c) **assign responsibility for master files**
   Correct. This is one solution.

   d) **record changes to master files**
   Correct. This is one solution.
   All the solutions above can help ensure the authenticity of your data

6. Transcription of data

   a) should always be carried out by the researcher
   Incorrect. Not necessarily. Researchers may want to carry out their own transcription to save money, or to immerse themselves in the data, or to use the transcription process as part of their methodology i.e. in conversation analysis, but it is also common to use professional transcribers.

   b) should always be carried out by a professional transcriber
   Incorrect. Not necessarily. Some researchers may want or need to do the transcription themselves.

   c) **can be carried out by either a researcher or a professional transcriber as long as there is consistency within the transcripts**
   Correct. It doesn't really matter whether the transcription is carried out by the researcher themselves or by a professional transcriber. Regardless of who does the task, the most important thing is to maintain consistency across the transcripts. This can be achieved through developing a short set of transcription guidelines that all transcribers should follow.

   d) should always be carried out in full
   Incorrect. Transcription does not always have to be carried out in full, but full transcripts are optimal for future sharing and re-use.

STORING YOUR DATA

**KEEP YOUR DIGITAL DATA SAFE,
SECURE AND RECOVERABLE**

LOOKING AFTER RESEARCH DATA FOR THE LONGER-TERM AND PROTECTING THEM FROM UNWANTED LOSS REQUIRES HAVING GOOD STRATEGIES IN PLACE FOR SECURELY STORING, BACKING-UP, TRANSMITTING, AND DISPOSING OF DATA. COLLABORATIVE RESEARCH BRINGS CHALLENGES FOR THE SHARED STORAGE OF, AND ACCESS TO, DATA.

RESOURCES OVERVIEW

AREAS OF COVERAGE

- Making back-ups

- Data storage

- Data security

- Data transmission and encryption

- Data disposal

- File sharing and collaborative environments

RESOURCES AVAILABLE

- PowerPoint presentation on 'Storing your data'

- PowerPoint presentation on 'Security breaches'

- Exercise 1: Data security breaches (no answers)

- Exercise 2: Data security breaches

- Exercise 2 Answers: Data security breaches

- Exercise 3: Checksum and encryption

- Quiz: Data storage and security

- Quiz Answers: Data storage and security

- www.data-archive.ac.uk/create-manage/
  training-resources
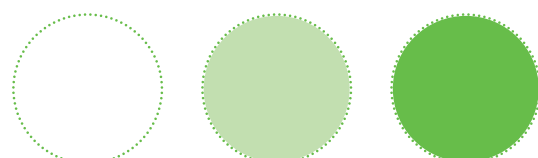
DATA SECURITY BREACHES

Think of five possible safety and security breaches which could pose a threat to your data. Fill in the table below with your answers.

	POTENTIAL SAFETY AND SECURITY BREACHES OF YOUR DATA
1	
2	
3	
4	
5	

EXERCISE: DATA SECURITY BREACHES

HINTS

Unfortunately data safety and security breaches are not uncommon. This can be a malicious act in which someone who is determined to gain access to secure, confidential information uses unscrupulous or illegal means to obtain that data such as through hacking computers and stealing laptops. It can be a careless act or breach on the part of the researcher in charge of the data such as leaving the data unattended or disposed of incorrectly. Or it could be an unforeseen event such as a fire or flood which destroys the data.

DATA SECURITY BREACHES

Read through the following real-life data safety and security breach scenarios. Think about what could have been done differently to prevent this happening? What preventative measures could the researcher or team have taken?

	SCENARIO	PREVENTATIVE MEASURES
1	Unshredded and unanonymised data transcripts are found on the street in a clear plastic rubbish bag. It was too time consuming to shred the large pile of documents with a basic office shredder so they were just thrown into the recycling bin.	
2	A senior lecturer stores personal and confidential data on the hard drive of her university computer. She is given a new computer by her department and the old one is given to research students to use in their office. The students are able to access both her personal and research data.	
3	A researcher has their laptop stolen whilst away on a conference trip. Vital research data was kept on the hard drive and was not backed up anywhere else.	
4	A researcher sends a set of audio cassettes for transcription in the post. The researcher writes the wrong address on the package and the tapes get lost in the post.	
5	Digital audio files are emailed to a transcriber who saves them to his computer desktop and also stores them in his email once received. The transcriber fails to delete the files from his email and from his computer once the transcription is completed and returned to the researcher. He later sells his computer on eBay.	
6	A researcher works in a shared office with four other researchers working on other projects. Whilst working on his project the researcher leaves paper copies of the signed consent forms loose on the desk. The other researchers have the opportunity to read through these confidential documents.	
7	A transcriber tells her friends and family about the 'interesting interview' that she is transcribing, giving details about the name and place of work of the interview participant.	
8	A fire at the university destroys a research office and all the paper copies of an important research collection which a researcher was preparing for archiving.	
9	Unanonymised data is inadvertently published on a project's website.	
10	A researcher encrypts his data folder, then forgets his password and can no longer access his data.	

DATA SECURITY BREACHES

	SCENARIO	PREVENTATIVE MEASURES
1	Unshredded and unanonymised data transcripts are found on the street in a clear plastic rubbish bag. It was too time consuming to shred the large pile of documents with a basic office shredder so they were just thrown into the recycling bin.	Ask your institution if there is an approved bulk shredding service available that can carry out the task instead of putting it out for recycling.
2	A senior lecturer stores personal and confidential data on the hard drive of her university computer. She is given a new computer by her department and the old one is given to research students to use in their office. The students are able to access both her personal and research data.	Do not presume IT services will clean the hard drive before passing on the computer. Always delete the data stored on a hard drive when disposing of a computer. Even then this might not be sufficient. Only 'scrubbing' or overwriting the data will sufficiently delete them from the machine.
3	A researcher has their laptop stolen whilst away on a conference trip. Vital research data was kept on the hard drive and was not backed up anywhere else.	Always keep a back-up of data. When travelling you could remove all unnecessary data from the laptop and back up what is on there onto a memory device and onto another hard-drive back at home/office. Better still, you can encrypt devices such as laptops and memory sticks. Under the Data Protection Act you have a legal obligation to protect personal data so it should be stored securely, especially when travelling.
4	A researcher sends a set of audio cassettes for transcription in the post. The researcher writes the wrong address on the package and the tapes get lost in the post.	Address packages carefully and if research data is sent in the post, either to a transcriber or back from a transcriber, always send via a trackable postal service. Royal Mail offer special delivery services.
5	Digital audio files are emailed to a transcriber who saves them to his computer desktop and also stores them in his email once received. The transcriber fails to delete the files from his email and from his computer once the transcription is completed and returned to the researcher. He later sells his computer on eBay.	Ask transcribers to sign an agreement to destroy their copies of the data once they have been returned and verify after transcription that this has indeed been done. It is best not to email files that you do not want to linger on other people's computer systems. You could encrypt the files before emailing them; or send them via secure transmission.
6	A researcher works in a shared office with four other researchers working on other projects. Whilst working on his project the researcher leaves paper copies of the signed consent forms loose on the desk. The other researchers have the opportunity to read through these confidential documents.	Always keep confidential documentation and data in a locked filing cabinet. Don't keep the key in the lock.
7	A transcriber tells her friends and family about the 'interesting interview' that she is transcribing, giving details about the name and place of work of the interview participant.	Ask transcribers to sign confidentiality agreements, preventing them from divulging confidential information and remind them of the confidentiality of information sent to them and their responsibility to keep that information confidential.
8	A fire at the university destroys a research office and all the paper copies of an important research collection which a researcher was preparing for archiving.	Always keep back-up copies of important data in an offsite location. With important paper documents, these can be scanned and kept digitally.
9	Unanonymised data is inadvertently published on a project's website.	This is unacceptable. Unanonymised data must be kept safely and only accessed by those permitted to use it. Always double check anonymisation has been carried out when data are published.
10	A researcher encrypts his data folder, then forgets his password and can no longer access his data.	The researcher could also have kept a copy of the password but in a separate location to the data folder.

EXERCISE THREE

CHECKSUM AND ENCRYPTION

CHECKSUMS
Checksums are a way to compute the integrity of your data and create a list of your files.

A checksum is a file verification tool, checking transfers and back-ups. Each time you run a checksum:

- it creates a number string for each file
- even if one byte of data has been altered or corrupted that string will change
- if the checksums match, the data has not altered

Have a go at the following checksum exercise using the free software, FastSum:

www.data-archive.ac.uk/create-manage/training-resources/storage

ENCRYPTION
Encryption transforms data to make it unreadable to anyone except those with a key. Encryption comes in strengths. A higher key size takes exponentially longer to crack. A key size of 8 takes 0 milliseconds to crack, a key size of 128 takes 150 trillion years to crack.

Encrypt anything you would not send on a postcard, and use for storing or moving files that contain personal information e.g. transcripts.

Have a go at the following encryption exercise which will create an encrypted storage space on your drive using the free software, SafeHouse Explorer:

www.data-archive.ac.uk/create-manage/training-resources/storage

DATA STORAGE AND SECURITY

1. The best place to store your current paper-based data while you are working on it is:
   (choose one answer only)

   a) on your desk
   b) in an unlocked drawer
   c) in a locked filing cabinet
   d) in a vault

2. Backing-up your data
   (choose all possible answers)

   a) protects against data loss or corruption
   b)| preserves it forever
   c) should be done regularly
   d) is the sole responsibility of your institution

3. How often should you copy or migrate your digital data onto new tape media?
   (choose one answer only)

   a) every day
   b) every month
   c) every year
   d) every 2-5 years
   e) every 20 years

4. What should you do with paper-based data to ensure the longevity of the information?
   (choose all possible answers)

   a) store it in a temperature/humidity-controlled environment
   b) store it in direct sunlight
   c) convert it to PDF/A format and store in a location fit for purpose
   d) store it in your attic

5. How should you ensure the security of personal information in a dataset?
   (choose all possible answers)

   a) by anonymising/reducing the precision of the data
   b) storing names/addresses separately from the other data
   c) encrypting data containing personal information
   d) requiring data users to sign the Official Secrets Act

6. Encrypting files on your mobile storage device ensures that your digital data are:
   (choose one answer only)

   a) open for anyone to see
   b) protected and only those with the encryption key can view them.
   c) completely deleted
   d) converted into a new format

7. What ways would be suitable for transferring transcribed research data to a colleague?
   (choose all possible answers)

   a) via an unencypted email attachment
   b) using Dropbox without encryption
   c) using Dropbox with encryption
   d) hand it in person to the recipient on a USB stick

8. How can you securely destroy data on a personal computer hard drive?
   (choose all possible answers)

   a) delete files and reformat the hard drive
   b) remove and shred the hard drive
   c) delete files and take the computer to the dump
   d) overwrite files using secure deletion software

10. How can you securely destroy data on a CD/DVD disc?
    (choose one answer only)

    a) put it through a CD shredder and throw it away
    b) break it in half and throw it away
    c) put it in the plastic recycling bin
    d) put it in the dishwasher on a hot cycle
    e) overwrite files on the disc using proprietary software

DATA STORAGE AND SECURITY

QUIZ
ANSWERS

1. The best place to store your paper based data while you are working on it is:

   a) on your desk
      Incorrect. This is not the ideal option if you are in a shared office. It may be less problematic if you are in a single, lockable office and the only person with the key.
   b) in an unlocked drawer
      Incorrect. Again this is not the ideal option if you are in a shared office. It may be less problematic if you are in a single, lockable office and the only person with the key.
   c) **in a locked filing cabinet**
      Correct. This is the ideal solution
   d) **in a high security bank vault**
      Correct. This would be a good, but not very realistic option for most people.

2. Backing up your data

   a) **protects against data loss or corruption**
      Correct.
   b) preserves it forever
      Incorrect.
   c) **should be done regularly**
      Correct.
   d) is the sole responsibility of your institution
      Incorrect. Your institution may back up data on your networked drives but it is not the sole responsibility of the institution, you should also take responsibility for backing up your data.

3. How often should you copy or migrate your digital data onto new tape media?

   a) every day
      Incorrect. Not necessary
   b) every month
      Incorrect. Not necessary
   c) every year
      Incorrect. Not necessary
   d) **every 2-5 years**
      Correct. This is recommended, e.g. migrate to new media as optical and magnetic media are subject to degradation
   e) every 20 years
      Incorrect. Not often enough

4. What should you do with paper-based data to ensure the longevity of the information?

   a) **store it in a temperature/humidity-controlled environment**
      Correct.
   b) store it in direct sunlight
      Incorrect.
   c) **convert it to PDF/A format and store in a location fit for purpose**
      Correct. This is a long-lasting digital format.
   d) store it in your attic
      Incorrect.

5. How should you ensure the security of personal information in a dataset?

   a) **by anonymising/**reducing the precision of the data
      Correct.
   b) **storing names/addresses** separately from the other data
      Correct.
   c) **encrypting** data containing personal information
      Correct.
   d) requiring data users to sign the Official Secrets Act
      Incorrect. Not necessary!

6. Encrypting files on your mobile storage device ensures that your digital data are:

   a) open for anyone to see
      Incorrect.
   b) **protected and only those with the encryption** key can view them
      Correct.
   c) completely deleted
      Incorrect.
   d) converted into a new format
      Incorrect.

7. What ways would be suitable for transferring transcribed research data to a colleague?

   a) via an unencypted email attachment
      Incorrect. This is insecure.
   b) using Dropbox without encryption
      Incorrect. This is insecure.
   c) **using Dropbox with encryption**
      Correct.
   d) **hand it in person to the recipient on a USB stick**
      Correct. This is fine, but make sure they look after it!

QUIZ ANSWERS: DATA STORAGE AND SECURITY

8. How can you securely destroy data on a personal computer hard drive?

    a) delete files and reformat the hard drive
        Incorrect. This is insecure.
    b) **remove and shred the hard drive**
        Correct.
    c) delete files and take the computer to the dump
        Incorrect. This is insecure.
    d) **overwrite files using secure deletion software**
        Correct. Make sure you use Government Standard software.

9. How can you securely destroy data on a CD/DVD disc?

    a) **put it through a CD shredder and throw it away**
        Correct.
    b) break it in half and throw it away
        Incorrect. This is not failsafe.
    c) put it in the plastic recycling bin
        Incorrect. This is not failsafe.
    d) put it in the dishwasher on a hot cycle
        Incorrect.
    e) overwrite files on the disc using proprietary software
        Incorrect. This is not failsafe.

ETHICS & CONSENT

SHARE SENSITIVE AND CONFIDENTIAL RESEARCH DATA ETHICALLY

A COMBINATION OF GAINING CONSENT FOR DATA SHARING, ANONYMISING AND REGULATING ACCESS TO DATA WILL INCREASE THE POTENTIAL FOR MAKING PEOPLE-RELATED RESEARCH DATA MORE READILY AND WIDELY AVAILABLE.

RESOURCES OVERVIEW

AREAS OF COVERAGE

- Definitions: personal, confidential data and sensitive personal data

- Legal and ethical issues

- Informed consent and data sharing

- Research Ethics Committees and data sharing

- Anonymising data

- Access control

RESOURCES AVAILABLE

- PowerPoint presentation on 'Ethics & Consent Overview'

- PowerPoint presentation on 'Consent – Practical'

- PowerPoint presentation on 'Anonymisation'

- Exercise 1: Consent - introduction

- Exercise 1a: Consent - qualitative data

- Exercise 1a Answers: Consent – qualitative data

- Exercise 1b: Consent - focus groups

- Exercise 1b Answers: Consent - focus groups

- UK Data Archive model consent form

- Exercises 2: Anonymisation - introduction

- Exercise 2a: Anonymisation – qualitative data

- Exercise 2a Answers: Anonymisation - qualitative data

- Exercise 2b: Anonymisation – quantitative data

- Exercise 2b Answers: Anonymisation – quantitative data

- www.data-archive.ac.uk/create-manage/training-resources

EXERCISE 1A: CONSENT - QUALITATIVE DATA

Mort, M., *Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003* [computer file]. Colchester, Essex: UK Data Archive [distributor], November 2006. SN: 5407. Retrieved from www.esds.ac.uk/findingData/snDescription.asp?sn=5407

BACKGROUND TO THE STUDY

The 2001 foot and mouth disease outbreak had an enormous effect on the economic, social and political life of rural areas in the UK. This research project, which was funded by the Department of Health, produced evidence about the human health and social consequences of the epidemic.

The study recruited a standing panel of 54 local people from the worst affected area (North Cumbria). This panel wrote weekly diaries over a period of 18 months describing how their lives had been affected by the crisis and the process of recovery they observed around them. The panel was recruited to reflect a broad range of occupations including: farmers and their families; workers in related agricultural occupations; those in small businesses including tourism, hotel trades and rural business; health professionals; veterinary practitioners; voluntary organisations; and residents living near disposal sites.

The panel members produced 3,200 weekly diaries of great intensity and diversity over the 18-month period. The data were supplemented by in-depth interviews with each respondent, focus group discussions, and 16 other interviews with stakeholders.

The research team gained consent from participants for primary participation in the project, but did not initially consider consent for sharing or archiving their data. When the project was finished they wanted to archive their data, so had to gain retrospective consent. They sought expert advice from copyright law specialists to help draft terms of agreement which would give respondents a series of options about how their diaries, copies or portions of diaries, and/or their audio material would be archived.

The exercise includes three consent forms to discuss. The first consent form is the initial consent that the researchers used. The second and third are the retrospective consent forms

EXERCISE 1B: CONSENT - FOCUS GROUPS

Winters, K., *Qualitative Election Study of Britain, 2010* [computer file]. Colchester, Essex: UK Data Archive [distributor], August 2011. SN: 6861. Retrieved from www.esds.ac.uk/findingData/snDescription.asp?sn=6861

BACKGROUND TO THE STUDY

The aim of the Qualitative Election Study of Britain, 2010 was to record and analyse the views and concerns of British citizens in their own words before and after the 2010 General Election.

The goal was to generate thick, rich qualitative data that could be used to provide insights into the opinions of citizens on: (inter alia) politicians, party leaders, political issues and perceptions of civic duty, political alienation, and campaigns both before and after the general election.

The project also aimed to generate qualitative data to facilitate analysis of the language used and meanings conveyed when participants articulated their assessments.

Through qualitative analysis this project investigated the sources of people's normative values, made explicit the tacit assumptions participants used to reach their judgements, and identified new research themes.

The Qualitative Election Study of Britain included focus groups and interviews.

The exercise includes a background information sheet and a consent form to discuss.

The Institute for Health Research
Lancaster University

CONSENT FORM

Name _____

Address _____

_____

_____

_____

I consent to participating in the health and social consequences of the 2001 foot and mouth outbreak project, which involves:

- Completing a weekly diary for a period of 18 months
- Competing a questionnaire about my quality of life (3 times during the study)
- Individual and group interviews about my life experiences, my work and my health.

I understand that:

1. Everything that I tell you will be held in the strictest confidence. Some of the information that I give you may be used in reports and articles, but my identity will remain anonymous (my name will not be given).

2. I am free to withdraw from the project at any time.

Signature _____ Date _____

Archiving Depositor Consent Form
(Diary –Economic and Social Data Service, UK Data Archive)

Terms of Agreement

Below are sets of statements that give you the depositor, a series of options in terms of how you wish your diary, copy of your diary, or portion of your diary, to be archived.   For each numbered statement, **please delete the part that is not applicable**. Please note that 'diary' refers to 'anonymised diary' (you are not named).

By signing below:

1.1   I agree to deposit **the whole/a portion** of my diary (or copy of my diary) with Economic and Social Data Service, UK Data Archive .

1.2   I agree that **my diary/ or the agreed portion** of my diary will be available **from………………………** to researchers and the public for scholarly and educational purposes.

By giving my permission I also:

2.   **I do/do not agree** that the Economic and Social Data Service, UK Data Archive may use this **diary/ portion of the diary**, including making a copy or copies of it or a part or parts of it, in any form or medium, and may authorise others to do so, without any further approval on my part.

3.   **I do/do not agree** that a part or parts of the **whole diary/ portion of the diary** may be **published from ……………………………………..**as long as I am not identified.

4.   I hereby assign ownership of the **diary/copy of the diary/ portion of the diary** to Economic and Social Data Service, UK Data Archive. I understand that I nevertheless retain copyright, subject to the rights which I have granted Economic and Social Data Service, UK Data Archive to make copies and publish and to grant permissions to others to do so.

Signature of depositor--   Date--------------------

Name of depositor (print)---

Witnessed By ………………………………………………….[name, print]

Signature of Witness…………………………………………………

Archiving Depositor Consent Form
(Audio material– Economic and Social Data Service, UK Data Archive)

Terms of Agreement

Below are sets of statements that give you the depositor, a series of options in terms of how you wish your audio material, or extract of your audio material, to be archived. For each numbered statement, **please delete the part that is not applicable**. Please note that 'audio material' refers to 'anonymised' audio material (you are not named).

By signing below:

1.1   I agree to deposit **the whole/ extracts of my audio material** with Economic and Social Data Service, UK Data Archive

1.2   I agree that **my audio material / or the agreed extracts of my audio material** will be **available from………………………..** to researchers and the public for scholarly and educational purposes.

By giving my permission I also:

2.   **I do/do not agree** that the Economic and Social Data Service, UK Data Archive may use this **audio material / the agreed extracts of my audio material,** including making a copy or copies of it or a part or parts of it, in any form or medium, and may authorise others to do so, without any further approval on my part.

3.   **I do/do not agree** that a part or parts of the **audio material / the agreed extracts of the audio material** may be **published from …………………………** as long as I am not identified.

4.   I hereby assign ownership of the **audio material / the agreed extracts of the audio material** to Economic and Social Data Service, UK Data Archive. I understand that I nevertheless retain copyright, subject to the rights which I have granted Economic and Social Data Service, UK Data Archive to make copies and publish and to grant permissions to others to do so.

Signature of depositor---   Date--------------------

Name of depositor (print)---

Witnessed By ………………………………………………….[name, print]

Signature of Witness…………………………………………………

CONSENT - QUALITATIVE DATA

HEALTH AND SOCIAL CONSEQUENCES OF THE FOOT AND MOUTH DISEASE EPIDEMIC IN NORTH CUMBRIA, 2001-2003 (SN 5407)

- When gaining retrospective consent the researchers spoke to their participants about archiving their data. Researchers often worry that participants will be reluctant to share their data, but when this research team went back to their participants, only seven (out of 54*) panel members declined archiving their data. Forty interview and diary transcripts are archived and available for re-use by registered users. Three interviews and five diaries are embargoed until 2015.

  * Figures here do not total exactly, due to interviewees also being diarists.

- The retrospective consent forms were well-thought through and the researchers used an individual approach to discussing archiving with respondents (one-to-one discussions on whether interviewees wanted their interviews and diaries archived or not), giving the respondents the opportunity to decide.

- The research team sought expert advice from a copyright specialist to help draft terms of agreement which would give respondents a series of options about how their diaries, copies or portions of diaries, and/or their audio material would be archived.

- A large dataset of sensitive, confidential information was archived with minimal anonymisation (only names, addresses, dates of birth and company names were removed).

- The drawbacks of using many different forms within a small-scale project are that:
  - additional paperwork puts a greater burden on respondents who have to complete them and on the research team who have to administer them
  - there is a danger of making the final collection difficult to access as a whole if the various elements have different consent agreements

- The second and third consent forms offer multiple choice text answers. This allows greater freedom for participants to record their wishes. However, too many permutations of consent could affect what is being made available and could potentially distort the final body of research that is being made accessible.

- On the diary consent form, respondents are given the option to select a date for their data to become available. We recommend that if this is to be offered then guidance should be given about setting a date - in case an arbitrary date is set. The date is usually best set by the research team as they will have a better sense of when the project will be completed and when the data should be accessible.

- These forms are very official and formal. This may put some participants off. They read more like legal documents. The first consent form was less formal, but did not gain consent for archiving the diaries and audio. The researchers later went back to respondents to gain retrospective consent, by which time the political and legal situation surrounding this issue had changed. There was more public awareness and sensitivity to the issue. They therefore employed legal experts to help design the consent form. This makes it very formal and perhaps intimidating. Some projects may benefit from a more rigorously laid out form. Others may require a more informal looking agreement.

- Respondents could have been offered an information sheet which explains archiving.

- The form states that the respondent 'assigns ownership' to the UK Data Archive. The Archive does not assert ownership of research data so instead this could be phrased 'I hereby license the diary....'

Understanding people's attitudes toward current events

A study funded by the British Academy

Seeking participants for focus group discussions about current events

£25 participation compensation

Research conducted by Dr. Kristi Winters
Birkbeck College, University of London
British Academy Postdoctoral Fellow
email....

The British Academy and Birkbeck College are committed to ethically conducted social research. We ask you to consider the following points **before** agreeing to participate.

- Your contribution to the research will take the form of a focus group participant. This will be digitally video recorded and transcribed.

- **Your name and any information which may directly or indirectly identify you will be altered to protect your anonymity.**

- Any recordings of the discussions will be kept securely, and only authorised to other researchers on the condition they preserve your anonymity.

- The transcriptions (*excluding* names and other identifying details) will be retained by the researcher and analysed as part of the study. They will also be deposited with the UK Data Archive which has strict regulations about accessing data for research and protecting participant confidentiality.

If you would like to discuss any aspect of the study, or the details of this form, please contact **Dr. Kristi Winters.**

Phone: .....

E-mail: ....

Web page: ....

Focus group details:

Date:

Time:

Location:

Other research in the media:

Drs. Rosie Campbell and Kristi Winters (2007) 'The myth of the homogeneous voter.' *Thinking Aloud* Podcast available at:
www.bbk.ac.uk/polsoc/research/rcampbellvoting/

EXAMPLE

Common Questions:

Q: What is this research about?

Local elections, an up-coming general election, economic worries, climate change, the future of public services, personal and national security…

At the start of the 21st century, Britain faces many challenges. The aim of this research is to talk to people about current events and issues and learn more about what people think about the challenges facing Britain.

What do I have to do?

Just talk and give your views.

The focus groups will consist of 5 to 10 other people and will be led by a moderator. You will first be asked to fill out a brief questionnaire. In order to get a good cross-section of the public not everyone who applies will be invited to participate in the focus groups, however you might be invited to participate in a one-on-one interview instead.

There will another focus group after the next general election which will review many of these topics again. If you are invited to the first focus group you may participate in the post-election focus group, and will again be eligible for £25 in participant compensation.

The focus group will take place at a location convenient for you. It will be recorded with a digital video recorder and transcribed.

The transcripts will be analysed to better understand the views of Britons today.

I also understand your time is valuable, and to help offset the costs you might occur travelling to and from the focus groups I am offering £25 compensation to all participants.

How long will it take?

The focus group will last about 90 minutes. The first 5 minutes will be devoted to reviewing and signing a consent form which lists your rights as a participant and to answer any other questions you might have before we begin. You may also request a copy of the consent form in advance of our meeting.

Participants will also be required to fill out a questionnaire after the focus group session.

What happens to the video files and written transcripts?

You will be assigned a participant ID number so your contact details will never be kept with the video or transcript files.

All files will be kept in a password-protected file on a password-protected computer. Files **will not** be sent by e-mail, only on password-protected data sticks and by registered post.

Transcribers will be required to sign documentation binding them legally to protecting your anonymity. During the course of transcription all direct identifiers (names, addresses, employment, streets, etc.) and indirect identifiers (workplace, organisations) will be removed or changed.

Any other researchers who use the data must comply with strict regulations about accessing data for research and protecting participant confidentiality.

The findings of this research will be written up and published. They may also be used for teaching and research training. The published works may include quotations from your discussions, but **individuals will never be named.** To aid in the use of your words in published works I request that you assign copyright to the researcher.

Can I change my mind about participation?

Your contribution is immensely valuable. However, if you wish to withdraw at any time I will respect your decision immediately.

How do I find out about publications which come out of this study?

Participants who would like to be notified of publications which come out of this research should identify how they would like to be notified on the participation consent form.

How do I find out more or become a participant in this study?

Just contact Dr. Kristi Winters by e-mail… or by phone …with any questions.

Understanding people's attitudes toward current events
Dr. Kristi Winters (Birkbeck College, University of London)
Funded by the British Academy

o   I have read and understood the project information brochure **Understanding people's attitudes toward current events**.

o   I have been given the opportunity to ask questions about the project. I also understand I may ask further questions about this research at any time.

o   I agree to take part in the project. Taking part in the project will include being interviewed and recorded (video).

o   I understand that my taking part is voluntary; I can withdraw from participating at any time and I will not be asked any questions about why I no longer want to take part. The focus group discussions I have participated in until my withdrawal may still be used.

o   I agree for the data I provided to be archived at the UK Data Archive and I understand that other researchers will have access to this data <u>only</u> if they agree to preserve my anonymity and confidentiality terms as specified in this form.

o   I understand that my words may be quoted in publications, reports, web pages, and other research outputs **but my name or other identifying details** will not be used.

o   I hereby assign the copyright of my contribution to Dr. Kristi Winters, so that my words may be quoted in publications, reports, web pages, and other research outputs.

o   I would / would not (delete as appropriate) like to be notified of any publications which are produced from this research. I would like to be notified of this by e-mail / text / phone call (delete as appropriate) and have provided the relevant contact information to Dr. Winters.

o   I understand my personal details such as phone number and address will not be revealed to anyone except Dr. Kristi Winters or her assistants.

o   I confirm that I have freely agreed to participate in this qualitative research project conducted by Dr. Kristi Winters. I have been briefed on what this involves and I agree to the use of the findings as described above. I understand that the material is protected by a code of professional ethics.

_____ _____ _____
Name of Participant   Signature   Date

Understanding people's attitudes toward current events
Dr. Kristi Winters (Birkbeck College, University of London)
Funded by the British Academy

o   I have read and understood the project information brochure **Understanding people's attitudes toward current events.**

> **Comment [a1]:** The researcher included an information sheet outlining all the background information for the project and the researcher's contact details.

o   I have been given the opportunity to ask questions about the project. I also understand I may ask further questions about this research at any time.

o   I agree to take part in the project. Taking part in the project will include being interviewed and recorded (video).

> **Comment [a2]:** The researcher could perhaps have offered separate consent forms for video/audio from the participation consent.

o   I understand that my taking part is voluntary; I can withdraw from participating at any time and I will not be asked any questions about why I no longer want to take part. The focus group discussions I have participated in until my withdrawal may still be used.

> **Comment [b3]:** The consent form for this project does not give the option to disagree with one particular statement and agree with others. This might have been a deliberate design due to the researcher's concern that people would withdraw from the focus group during or after it had happened.

o   I agree for the data I provided to be archived at the UK Data Archive and I understand that other researchers will have access to this data only if they agree to preserve my anonymity and confidentiality terms as specified in this form.

o   I understand that my words may be quoted in publications, reports, web pages, and other research outputs **but my name or other identifying details** will not be used.

> **Comment [a4]:** She could also have reserved the right to question withdrawal from the project and be able to explain the consequences for the project if they leave.

o   I hereby assign the copyright of my contribution to Dr. Kristi Winters, so that my words may be quoted in publications, reports, web pages, and other research outputs.

> **Comment [a5]:** The researcher deliberated and debated how to design this consent form, concerned about the consequences for the project and how it could be archived if someone withdrew. The researcher's strategy was to gain full consent for participation and archiving prior to the focus group. The problem is that respondents do not know beforehand what they will contribute. The researcher reserved the right to keep any prior contributions and respondents were only allowed to withdraw from future contributions. Is this coercive? How else could it have been done?

o   I would / would not (delete as appropriate) like to be notified of any publications which are produced from this research. I would like to be notified of this by e-mail / text / phone call (delete as appropriate) and have provided the relevant contact information to Dr. Winters.

> **Comment [a6]:** Mentions copyright which is good.

o   I understand my personal details such as phone number and address will not be revealed to anyone except Dr. Kristi Winters or her assistants.

o   I confirm that I have freely agreed to participate in this qualitative research project conducted by Dr. Kristi Winters. I have been briefed on what this involves and I agree to the use of the findings as described above. I understand that the material is protected by a code of professional ethics.

_____   _____   _____
Name of Participant   Signature   Date

Consent Form for [name of project]

Please tick the appropriate boxes	Yes	No
Taking Part		
I have read and understood the project information sheet dated DD/MM/YYYY.	☐	☐
I have been given the opportunity to ask questions about the project.	☐	☐
I agree to take part in the project. Taking part in the project will include being interviewed and recorded (audio or video).[1]	☐	☐
I understand that my taking part is voluntary; I can withdraw from the study at any time and I do not have to give any reasons for why I no longer want to take part.	☐	☐
Use of the information I provide for this project only		
I understand my personal details such as phone number and address will not be revealed to people outside the project.	☐	☐
I understand that my words may be quoted in publications, reports, web pages, and other research outputs.	☐	☐
*Please choose **one** of the following two options:*		
I would like my real name used in the above	☐	
I would **not** like my real name to be used in the above.	☐	
Use of the information I provide beyond this project		
I agree for the data I provide to be archived at the UK Data Archive.[2]	☐	☐
I understand that other genuine researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.	☐	☐
I understand that other genuine researchers may use my words in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.	☐	☐
So we can use the information you provide legally		
I agree to assign the copyright I hold in any materials related to this project to [name of researcher].	☐	☐

_____ _____ _____
Name of participant   [printed]   Signature   Date

_____ _____ _____
Researcher   [printed]   Signature   Date

Project contact details for further information: Names, phone, email addresses, etc.

Notes:
1. Other forms of participation can be listed.
2. More detail can be provided here so that decisions can be made separately about audio, video, transcripts, etc.

ANONYMISATION - INTRODUCTION

Here are two anonymisation exercises – one for a qualitative and one for a quantitative study. Both studies are from real-life collections at the UK Data Archive.

EXERCISE 2A
• Do you think this interview transcript can be archived and shared (yes/no)?
• If yes, indicate what anonymisation would be needed before sharing or archiving this transcript.

EXERCISE 2B
Compare the partial data file table, showing the data for only 10 respondents, with the survey questionnaire. How would the data files need to be anonymised before archiving?

EXERCISE 2A: ANONYMISATION - QUALITATIVE DATA

Mort, M., *Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003* [computer file]. Colchester, Essex: UK Data Archive [distributor], November 2006. SN: 5407. Retrieved from www.esds.ac.uk/findingData/snDescription.asp?sn=5407

BACKGROUND TO THE STUDY
The 2001 foot and mouth disease outbreak had an enormous effect on the economic, social and political life of rural areas in the UK. This research project, which was funded by the Department of Health, produced evidence about the human health and social consequences of the epidemic.

The study recruited a standing panel of 54 local people from the worst affected area (North Cumbria). This panel wrote weekly diaries over a period of 18 months describing how their lives had been affected by the crisis and the process of recovery they observed around them. The panel was recruited to reflect a broad range of occupations including farmers and their families, workers in related agricultural occupations, those in small businesses including tourism, hotel trades and rural business, health professionals, veterinary practitioners, voluntary organisations and residents living near disposal sites.

The panel members produced 3,200 weekly diaries of great intensity and diversity over an 18-month period. The data were supplemented by in-depth interviews with each respondent, focus group discussions, and 16 other interviews with stakeholders.

The research team gained consent from participants for primary participation in the project, but did not get consent for sharing or archiving their data. When the project was finished they wanted to archive their data, so had to gain retrospective consent. They sought expert advice from copyright law specialists to help draft terms of agreement which would give respondents a series of options about how their diaries, copies or portions of diaries, and/or their audio material would be archived. The research team gave respondents the choice to decide how their material was used. This collection also has a large dataset of sensitive, confidential information which was archived with minimal need for anonymisation (only names, addresses, dates of birth and company names were removed). These data are now available to other researchers.

The transcript used is based on an interview with a Department for Environment, Food and Rural Affairs (Defra) field officer, who was interviewed about his experiences (as a Defra field officer and as part of an affected farming family) during the foot and mouth crisis. The interviewee consented to the transcript of his interview being made available for research and educational purposes, as long as he would not be identified.

EXERCISE 2B: ANONYMISATION - QUANTITATIVE DATA
Anderson, B. and Trades Union Congress, *Polish and Lithuanian Workers: Opportunities and Challenges for Trade Unions, 2004-2006* [computer file]. Colchester, Essex: UK Data Archive [distributor], September 2009. SN: 6284. Retrieved from www.esds.ac.uk/findingData/snDescription.asp?sn=6284

BACKGROUND TO THE STUDY
This research project, conducted jointly by the Centre on Migration, Policy and Society (COMPAS) at the University of Oxford, and the Trades Union Congress (TUC), surveyed Polish and Lithuanian nationals working in the UK who had requested TUC leaflets on employment rights and the role of trades unions. The survey explored the kinds of difficulties experienced by Polish and Lithuanian workers in the UK labour market, and their potential for joining trades unions. More specifically, it addressed:

• Who is a member of or wants to join a trades union and why?
• What are the obstacles to joining a trades union?
• Where are prospective union members working?
• What are the kinds of difficulties that Polish and Lithuanian workers in the UK face in their employment relations and conditions?

**Health and Social Consequences of the Foot and Mouth Disease
Epidemic in North Cumbria, 2001-2003 (SN 5407)**
M. Mort, Lancaster University. Institute for Health Research

NOTE: all identifying info contained in this transcript is fictional

Interview with Lucas Roberts, DEFRA field officer

Date of Interview: 21/02/02

Date of birth: 2 May 1965
Gender: Male
Occupation: Frontline worker
Location: Plumpton, North Cumbria

Lucas was living at home with his parents, "but I'm hoping to move out soon" so we met at his parents' small neat house. We sat in a very comfortable sitting room with an open fire and Lucas made me coffee and offered shortbread. Although at first Lucas seemed a little nervous, quick to speech and very watchful he seemed to relax as we spoke and to forget about the tape.

I will just start by asking you to tell me a little bit about yourself and your background.

Well it is an agricultural background. I grew up on the farm where my brother is now. After I left school I did work on the farm but went to college and did exams, did land use recreation, sort of countryside/ environmental management course. So I obviously left agriculture, did the course and came back [to the farm] at weekends. During vacations, worked on the farm until I got my first placement in Suffolk. Went away in my early twenties, so wasn't a teenager and did that and then I took the Diploma, then took my studies further and did a degree in rural resources management in Bedfordshire and after that worked I on a community forestry project, actually based within a natural forest. I did that for 4 years and came back near the end of '95.

I think with countryside management, I think personally it's quite difficult to get in to, and as a job it's quite competitive. It's whether I kept my options a bit too broad or indeed simply gone into this like general management, I just wanted to stay connected with the environment. I did bits and pieces with CumbrExt which is now Milbro Ltd as has been privatised a few years. I had several contracts over the years with them off and on. Also did work for environmental consultancy agency, went to Germany, Slovakia, did various projects on a freelance self employed basis, and in between that I have a good friend who has a market garden business so I work there sometimes [laughs]. Do bits and pieces and obviously when he's gone on holiday, so I suppose I class myself as self employed. It's all continuous but it's very sort of like bits and pieces really. I did one office job as a research assistant, though it wasn't quite as [PAUSE]… with a traffic consultancy looking at local transport links and their problems. I found it quite being difficult being office based and stuck

behind a computer all the time, and sort of like used to being outdoors. Up until the out-break I had been working for Milbro Ltd doing some surveying work, and that finished there in February. I had the job in the market garden and then obviously it [foot and mouth] all flared up and we though well it will be nipped in the bud. The case in the abattoir in Essex and then I think it went to Devon. And then it was traced back to Heddon-on-the-Wall. There were several cases where I think [pause], it didn't reach Cumbria until March 3rd., I think, over a week later anyway, and even then [EMPHASIS], it obviously didn't seem quite serious but still thought they would eventually get on top of it., There seemed to a starting a sort of triangle. There was the one at Longtown auction mart and there was an earlier one near Maryport. It did appear to sort of close in and erm...[pause]

Can I just take you right back to your parents. Is it a mixed farm?

It's a dairy farm, we did do like a little bit of beef cattle, sort of stores, but no it's a dairy farm.

And there is you and your brother?

Well it is my brother's now [the farm], I just help out.

When you were growing up?

I forgot, I've got another brother, there is three of us. He worked on the farm for a year after he left school then he went to work on another farm and now he's a plasterer and a builder. So he left the farm as well, so it was only my eldest brother. What he did, he did 'A' levels and worked at a neighbour's farm for a year and then did a course at Newton Rigg and then stayed at the farm ever since.

So your parents….? You probably weren't old enough to remember the 67 out-break?

No I don't remember it all. In fact I wasn't really aware of it until they were making comparisons with this last one.

And have they [parents] said anything about the 67 out-break?

No, well it didn't really affect Cumbria, it got to Westmoreland, it just got there but Cumbria wasn't affected and they certainly weren't affected but I remember my parents didn't really speak about it. My Dad's in his seventies now and he doesn't really talk about it now, he sort of feels he's retired now and he just accepts that.

**Health and Social Consequences of the Foot and Mouth Disease
Epidemic in North Cumbria, 2001-2003 (SN 5407)**
M. Mort, Lancaster University. Institute for Health Research

NOTE: all identifying info contained in this transcript is fictional

Date of Interview: 21/02/02

Interview with Lucas Roberts, DEFRA field officer   **Comment [v1]:** Replace: Ken
Date of birth: 2 May 1965   **Comment [v2]:** Delete

Gender: Male

Occupation: Frontline worker
Location: Plumpton, North Cumbria   **Comment [v3]:** Delete

Lucas *was living at home with his parents, "but I'm hoping to move out*   **Comment [v4]:** Replace: Ken
soon" so we met at his parents' small neat house. We sat in a very
comfortable sitting room with an open fire and ***Lucas*** *made me coffee*   **Comment [v5]:** Replace: Ken
and offered shortbread. Although at first ***Lucas*** *seemed a little nervous,*   **Comment [v6]:** Replace: Ken
quick to speech and very watchful he seemed to relax as we spoke and
to forget about the tape.

I will just start by asking you to tell me a little bit about yourself and your
background.

Well it is an agricultural background. I grew up on the farm where my brother
is now. After I left school I did work on the farm but went to college and did
exams, did land use recreation, sort of countryside/ environmental
management course. So I obviously left agriculture, did the course and came
back [to the farm] at weekends. During vacations, worked on the farm until I
got my first placement in Suffolk. Went away in my early twenties, so wasn't a
teenager and did that and then I took the Diploma, then took my studies
further and did a degree in rural resources management in Bedfordshire and
after that worked I on a community forestry project, actually based within a
natural forest. I did that for 4 years and came back near the end of '95.

I think with countryside management, I think personally it's quite difficult to get
in to, and as a job it's quite competitive. It's whether I kept my options a bit too
broad or indeed simply gone into this like general management, I just wanted
to stay connected with the environment. I did bits and pieces with CumbrExt   **Comment [v7]:** Replace: [company]
which is now Milbro Ltd as has been privatised a few years. I had several   **Comment [v8]:** Replace: [company]
contracts over the years with them off and on. Also did work for environmental
consultancy agency, went to Germany, Slovakia, did various projects on a
freelance self employed basis, and in between that I have a good friend who
has a market garden business so I work there sometimes [laughs]. Do bits
and pieces and obviously when he's gone on holiday, so I suppose I class
myself as self employed. It's all continuous but it's very sort of like bits and
pieces really. I did one office job as a research assistant, though it wasn't

quite as [PAUSE]… with a traffic consultancy looking at local transport links and their problems. I found it quite being difficult being office based and stuck behind a computer all the time, and sort of like used to being outdoors. Up until the out-break I had been working for Milbro Ltd doing some surveying work, and that finished there in February. I had the job in the market garden and then obviously it [foot and mouth] all flared up and we though well it will be nipped in the bud. The case in the abattoir in Essex and then I think it went to Devon. And then it was traced back to Heddon-on-the-Wall. There were several cases where I think [pause], it didn't reach Cumbria until March 3rd., I think, over a week later anyway, and even then [EMPHASIS], it obviously didn't seem quite serious but still thought they would eventually get on top of it., There seemed to a starting a sort of triangle. There was the one at Longtown auction mart and there was an earlier one near Maryport. It did appear to sort of close in and erm...[pause]

Comment [v9]: Replace: [company]

Can I just take you right back to your parents. Is it a mixed farm?

It's a dairy farm, we did do like a little bit of beef cattle, sort of stores, but no it's a dairy farm.

And there is you and your brother?

Well it is my brother's now [the farm], I just help out.

When you were growing up?

I forgot, I've got another brother, there is three of us. He worked on the farm for a year after he left school then he went to work on another farm and now he's a plasterer and a builder. So he left the farm as well, so it was only my eldest brother. What he did, he did 'A' levels and worked at a neighbour's farm for a year and then did a course at Newton Rigg and then stayed at the farm ever since.

So your parents….? You probably weren't old enough to remember the 67 out-break?

No I don't remember it all. In fact I wasn't really aware of it until they were making comparisons with this last one.

And have they [parents] said anything about the 67 out-break?

No, well it didn't really affect Cumbria, it got to Westmoreland, it just got there but Cumbria wasn't affected and they certainly weren't affected but I remember my parents didn't really speak about it. My Dad's in his seventies now and he doesn't really talk about it now, he sort of feels he's retired now and he just accepts that.

EXERCISE 2B: ANONYMISATION - QUANTITATIVE DATA

TUC

This survey is confidential and purely statistical.
No details of individual respondents will be revealed.

W1. What is your age?

_____ years old

W2. Gender

1 ☐ male
2 ☐ female

W3. How would you assess your proficiency in speaking, reading, and writing English?

a. Speaking	b. Reading
1 ☐ Fluent	1 ☐ Fluent
2 ☐ Adequate	2 ☐ Adequate
3 ☐ Basic only	3 ☐ Basic only
4 ☐ None	4 ☐ None

W4. For how many years in total have you worked, and where?
(WRITE "O" FOR "NEVER")

a. Worked in the UK	_____ years
b. Worked in my home country	_____ years
c. Worked elsewhere	_____ years
TOTAL:	_____ years

W5. When did you first register with the Home Office?
Date (month and year): _____

W6. When did you last enter the UK to work (excluding holiday breaks)?
"LAST" REFERS TO THE "MOST RECENT" OCCASION WHEN WORKER ENTERED THE UK TO WORK

Date (month and year): _____

W7. How was your migration to the UK "organised" when you last entered the UK?
1 ☐ through private recruitment agency in my country

2

2 ☐ through private recruitment agency in the UK
3 ☐ through private travel agency in my home country
4 ☐ through private travel agency in the UK
5 ☐ through friends/relatives already working in the UK
6 ☐ through direct contact with employer in the UK
7 ☐ other, specify: _____

W8. When did you start your <u>first ever</u> job in the UK?
Date (month and year): _____

W9. Where are you <u>currently</u> working?

a. Name of your employer (who pays your wages or salary) _____
b. Is this an agency?
1 ☐ yes
2 ☐ no
c. Name of company where you are working (if different from above) _____
d. Postcode of workplace _____
e. What kind of business is your workplace (for example farm, hotel, food factory)? _____

W10. Approximately how many workers does the business that you currently work for employ, <u>at your usual workplace and in total</u>?

a. _____ workers employed at my usual workplace,
b. including **b.** _____ workers of my nationality
c. _____ workers employed in total by the business,
d. including **d.** _____ workers of my nationality

W11. In your current job, do you have a written employment contract?
1 ☐ yes
2 ☐ no

W12. Do you have a National Insurance number?
1 ☐ yes
2 ☐ no

W13. If NO, why not?
1 ☐ I have applied for a number and am waiting for it
2 ☐ I have not yet applied for a number
3 ☐ My employer has not provided me with the necessary documents
4 ☐ Some other reason (please specify) _____

W14. In your current job, what is your normal <u>gross pay</u> (i.e. your pay before any deductions)?

£ _____ ☐ per hour
☐ per week
☐ per month

W15. What do you mainly do in your current job (job title)? _____

W16. How many hours do you normally work per week (including overtime)?

_____

W17. Do you ever do any work which you would regard as "overtime"?
1 ☐ yes, paid at normal rate
2 ☐ yes, paid at more than the normal rate
3 ☐ yes, but it is not paid
4 ☐ no I do not work any overtime

W18. Do you have a British bank account?
1 ☐ yes
2 ☐ no

W19. How do you get your pay?
1 ☐ In cash
2 ☐ By cheque
3 ☐ By transfer to my bank account
4 ☐ Other (please specify)

W20. How many different businesses and employment agencies have you worked for since you first registered with the Home Office?
a. Number of businesses:
b. Number of employment agencies: _____

W21. How did you choose your current accommodation in the UK?
1 ☐ employer offered me accommodation
2 ☐ employer required me to live in a specific accommodation
3 ☐ friends helped me find accommodation
4 ☐ found accommodation by myself
5 ☐ other, specify:

W22. How would you rate your current living conditions?
1 ☐ very good
2 ☐ good
3 ☐ satisfactory/average
4 ☐ poor
5 ☐ very poor

W23. What problems have you had with your employer or employment agency in the UK?

_____
_____
_____

W24. Have you approached any of these organisations to get help with solving these problems?
☐ Citizens Advice Bureau
☐ Trade Union
☐ Polish club/organisation
☐ Church
☐ Police
☐ Others (please specify)

3

W25 What is your parents' education:
Mother's   Father's
☐ less than primary   ☐ less than primary
☐ primary   ☐ primary
☐ secondary   ☐ secondary
☐ tertiary   ☐ tertiary
☐ higher   ☐ higher

W26 Specify the type of your employment (including part-time work) just before you last came to the UK:
1 ☐ paid employee, specify
2 ☐ self-employed, specify
3 ☐ employer, specify
4 ☐ unpaid family labour
5 ☐ unpaid labour outside family
6 ☐ not working but looking for work
3 ☐ not working and not looking for work (e.g. student)

W27 Please tick the statements that apply to you:
1 ☐ I keep my ties to Poland because I was born and educated there.
2 ☐ I keep my ties to Poland because I have my own apartment there.
3 ☐ I keep my ties to Poland because my family and friends live there.
4 ☐ I keep my ties to Poland because of my language and culture.
5 ☐ I keep my ties to Poland because I can better develop my career there.
6 ☐ I keep my ties to Poland for other reasons. Specify
6 ☐ I don't feel related to Poland at all.
7 ☐ I feel related to Poland and I will probably return one day.

W28 Please tick the boxes which apply to you
1 ☐ I was a member of a trade union in Poland ☐ (please specify which union
2 ☐ I am a member of a trade union in Britain ☐ (please specify which union
3 ☐ I am interested in joining a union in Britain ☐ (please explain why

I am not interested in joining a union in Britain ☐ (please explain why

Please tick if you are interested in taking part in a more detailed survey ☐
Please tick if you would like to get in touch with a trade union ☐
If you have ticked either of the boxes above, please provide contact details:

Name

4

EXERCISE 2B: ANONYMISATION - QUANTITATIVE DATA

Polish and Lithuanian Workers: Opportunities and Challenges for Trade Unions, 2004-2006 (SN 6284), by B Anderson, University of Oxford.
NOTE: all identifying info contained in this sample file is fictional

	w1		w2	w3			w4a	w4b	w4c							w5		w6	
code	qutype	lang	c_age2	gender	engspea2	engroad2	mthswkuk	mthswkho	mthswkot	monwtot	retotmhw	yrswokuk	yrswkhom	yrswkoth	yrswtot	app4reg2	dateapr2	mylasten	ylasten
106p	4	3	33	2	3	3	8	18	0	26	3	0	0	0	0	1	23-Mar-2005	1-Jan-2005	2,005
	4	3	32	1	2	2	18	120	0	138	5	0	0	0	0	1	1-Aug-2004	1-Feb-2004	2,004
129p	4	3	31	1	2	2	0	0	0	0	0	0	0	0	0	1	21-Apr-2005	1-Mar-2005	2,005
15p	4	3	31	2	3	3	12	12	0	24	3	0	0	0	0	1	11-Sep-2004	1-Sep-2004	2,004
196p	4	3	29	1	2	1	12	36	0	48	3	0	0	0	0	1	1-Oct-2005	1-Oct-2004	2,004
202p	4	3	27	1	2	2	7	42	0	49	3	0	0	0	0	1	1-May-2005	1-Feb-2005	2,005
214p	4	3	27	1	3	3	0	60	0	60	3	5	5	0	5	1	1-Apr-2005	0-Jan-1900	888
433p	4	3	23	1	2	2	12	0	12	24	3	0	0	0	0	1	21-Sep-2004	1-Jun-2004	2,004
491l	4	2	22	1	2	2	0	180	0	180	5	0	0	0	0	1	1-Oct-2005	1-Aug-2005	2,005
111p	4	3		2	2	2	12	24	0	36	3	0	0	0	0	1	2-Jun-2004	1-Jun-2004	2,004

w7		w8				w9	w9b	w9d	w9d	w9e	w10a	w10b	w10c		w10d
migproc0	othmigp0	jobukfi2	jobukfi	myjbukfi	yrjbukfi	currwk	empag	wktown	wkcounty	typebus	numwowp1	natwowp1	numwobu1	propna	natwobu2
5		9-Feb-2005		1-Feb-2005	2,005	1	2	Gerrards Cross	Buckinghamshire	pizzeria	8	5	63	0.6250	40
5		0-Jan-1900		1-May-2004	2,004	1	2	Penrith	Cambs	chicken factory	20	11	100	0.5500	25
5		17-Mar-2005		1-Mar-2005	2,005	1	1	Trowbridge	Wiltshire	factory of packages for car parts	30	3	40	0.1000	888
5		4-Nov-2004		1-Nov-2004	2,004	1	2	Peterborough	Cambridgeshire	fruit factory	40	12	50	0.3000	12
5		0-Jan-1900		1-Oct-2004	2,004	1	2	Tunbridge Wells	Kent	restaurant, supermarket	20	4	888	0.2000	888
5		0-Jan-1900		1-Feb-2005	2,005	1	2	Cromarty	Ross Shire	scrap metal	20	5	20	0.2500	5
1		0-Jan-1900		1-Apr-2005	2,005	1	1	Leomister	Hereford	food factory	120	11	200	0.0917	11
5		11-May-2004		1-May-2004	2,004	1	2	Huntingdon	Cambridgeshire	food factory	300	200	300	0.6667	200
5		0-Jan-1900		1-Aug-2005	2,005	1	2	London	London	security		3	40	1.0000	25
7	through acquaintances	2-Jun-2004		1-Jun-2004	2,004	1	1	Cambridge	Cambs	kitchen in barracks	30	10	777	0.3333	777

empcont2 (w11)	nijob (w12)	whyni (w13)	whyniot	payhr (w14)	paywk	paymth	jobhrs (w16)	jobhrsgroup	jobtitle (w15)	overtime (w17)	bankacc (w18)	howpay (w19)	payoth	busemp (w20a)	agemp (w20b)	foundacc (w21)	accoth
2	2	3		0.00	0.00	1,046.00	70.00	4	topping pizzas	4	2	1		1	0	2	
1	1	999		0.00	234.00	0.00	50.00	4	bookings and sending goods to the shops	2	1	3		1	0	1	
2	1	999		5.50	0.00	0.00	40.00	3	welding, gluing, packing	1	1	3		2	2	4	
1	1	999		5.35	0.00	0.00	54.00	4	packing	1	3	3		1	1	3	
1	1	999		4.85	0.00	0.00	60.00	4	pizza chef (caffe uno) and cleaning operative (Tesco -- £6/h)	1	3	3		2	0	1	
2	1	999		4.85	0.00	0.00	50.00	4	almost everything	2	1	3		2	0	3	
1	1	999		0.00	0.00	1,915.16	42.00	3	Arbor Contract Engineers Ltd- machine engineer	3	1	3		0	1	1	
1	1	999		5.80	0.00	0.00	45.00	3	line runner	2	2	3		0	1	4	
1	1	999		6.00	0.00	0.00	50.00	4	security officer	1	1	3		1	1	3	
1	1	999		5.25	0.00	886.63	37.00	3	kitchen porter; sometimes waitress	1	1	3		1	1	2	

livecond (w22)	PROBWK (w23)	prob1	prob2	prob3	prob4	prob5
5	1	no work schedule	racism from the manager's side	mobbing	spreading fear	lack of contract or P60
3	1	overtime not paid, bank holidays not paid, no holiday pay, no sick pay	worse treatment			
2	2					
3	2					
4	2					
1	2					
2	2					
1	1	doesn't pay all the salary -- pay several hours to several days less	deduct for travel from the agency to the workplace and back -- £3.5 per day (			
1	1	had to fight to get the money back from agency	only when I said that I would call the police			
1	1	agency took £50 and documents for registration at HO, but they didn't send the money so I couldn't get registered, so I	when I gave the HO registration documents to the factory's office, they disappeared and then reappeared			
2	1					
4	1	farm -- constantly made mistakes when calculating salary; held passport	frequent mistakes with calculations of payment, we don't understand whether we get paid for overtime or not	payslip is difficult to read	don't inform about our rights and benefits; hide such information	

EXERCISE 2B ANSWERS: ANONYMISATION - QUANTITATIVE DATA

Polish and Lithuanian Workers: Opportunities and Challenges for Trade Unions, 2004-2006 **(study 6284 in UK Data Archive collection by B Anderson, University of Oxford)** *NOTE: all identifying info contained in this sample file is fictional*

code	qutype	lang	c_age2	gender	engspea2	engread2	mthswkuk	mthswkho	mthswot	monwtot	retotmhw	yrswokuk	yrswkhom	yrswkoth	yrswtot	app4reg2	dateapr2	mylasten	ylasten
106p	4	3	33	2	3	3	8	18	0	26	3	0	0	0	0	0	23-Mar-2005	1-Jan-2005	2,005
	4	3	32	1	2	2	18	120	0	138	5	0	0	0	0	1	1-Aug-2004	1-Feb-2004	2,004
129p	4	3	31	1	2	2	0	0	0	0	0	0	0	0	0	1	21-Apr-2005	1-Mar-2005	2,005
15p	4	3	31	2	3	3	12	12	0	24	3	0	0	0	0	1	11-Sep-2004	1-Sep-2004	2,004
196p	4	3	29	1	2	1	12	36	0	48	3	0	0	0	0	1	1-Oct-2005	1-Oct-2004	2,004
202p	4	3	27	1	2	2	7	42	0	49	3	0	0	0	0	1	1-May-2005	1-Feb-2005	2,005
214p	4	3	27	1	3	3	0	60	0	49	3	5	5	0	5	1	1-Apr-2005	0-Jan-1900	888
433p	4	3	23	1	2	2	12	0	12	24	3	0	0	0	0	1	21-Sep-2004	1-Jun-2004	2,004
491l	4	2	22	1	2	2	0	180	0	180	5	0	0	0	0	1	1-Oct-2005	1-Aug-2005	2,005
111p	4	3		2	2	2	12	24	0	36	3	0	0	0	0	1	2-Jun-2004	1-Jun-2004	2,004

Comment [LC1]: Although the questionnaire asked for month / year only; the full date was collected. This later had to be anonymised by removing date and month

migproc0	othmigp0	jobukfi2	myjbukfi	yrjbukfi	currwk	empag	wktown	wkcounty	typebu	numwowp1	natwowp1	propna	numwobu1	natwobu2
5		9-Feb-2005	1-Feb-2005	2,005	1	2	Gerrards Cross	Buckinghamshire	pizzeria	8	5	0.6250	63	40
5		0-Jan-1900	1-May-2004	2,004	1	2	Penrith	Cambs	chicken factory	20	11	0.5500	100	25
5		17-Mar-2005	1-Mar-2005	2,005	1	1	Trowbridge	Wiltshire	factory of packages for car parts	30	3	0.1000	40	888
5		4-Nov-2004	1-Nov-2004	2,004	1	2	Peterborough	Cambridgeshire	fruit factory	40	12	0.3000	50	12
5		0-Jan-1900	1-Oct-2004	2,004	1	2	Tunbridge Wells	Kent	restaurant, supermarket	20		0.2000	888	888
5		0-Jan-1900	1-Feb-2005	2,005	1	2	Cromarty	Ross Shire	scrap metal	20	5	0.2500	20	5
1		0-Jan-1900	1-Apr-2005	2,005	1	1	Leomister	Hereford	food factory	120	11	0.0917	200	11
5		11-May-2004	1-May-2004	2,004	1	2	Huntingdon	Cambridgeshire	food factory	300		0.6667	300	200
5		0-Jan-1900	1-Aug-2005	2,005	1	2	London	London	security	3	3	1.0000	40	25
7	through acquaintances	2-Jun-2004	1-Jun-2004	2,004	1	1	Cambridge	Cambs	kitchen in barracks	30	10	0.3333	777	777

Comment [LC2]: While the survey questionnaire needed to collect identifying information to help code occupations and industry, the disclosiveness of the data was not considered as it was entered directly into the SPSS spreadsheet

Comment [v3]: Replace with workplace region (Govt Office Region)

Comment [v4]: Replace with workplace region (Govt Office Region)

Comment [v5]: Replace with standard industrial classification code (SIC)

w1 w2 w3 w4a w4b w4c w5 w6

w7 w8 w9 w9b w9c w9d w9e w10a w10b w10c w10d

w11 w12 w13 w14 w15 w16 w17 w18 w19 w20a w20b w21

1

empcont2	nijob	whyni	whyniot	payhr	paywk	paymth	jobhrs	jobhr group	jobtitle	overtime	bankacc	howpay	payoth	busemp	agemp	foundacc	accoth
2	2	3		0.00	0.00	1,046.00	70.00	4	topping pizzas	4	2	1	1	1	0	2	
1	1	999		0.00	234.00	50.00		4	bookings and sending goods to the shops	1	3	1	1	0		1	
2	1	999		5.50	0.00	0.00		4	welding, gluing, packing	1	3	2	2	2	2	4	
1	1	999		5.35	0.00	0.00		4	packing	1	3	1	1	1	1	3	
1	1	999		4.85	0.00	0.00	60.00	4	pizza chef (caffe uno) and cleaning	1	3	2	2	0	0	1	
2	1	999		4.85	0.00	50.00		4	operative (Tesco – £6/h)	2	3	2	2	2	0	3	
1	1	999		0.00	0.00	1,915.16	42.00	3	almost everything	3	1	3	0	1	1	1	
1	1	999							Arbor Contract Engineers Ltd- machine engineer	3	1	3	0	1	1	1	
1	1	999		5.80	0.00	42.00	45.00	3	line runner	2	1	3	0	1	4		
1	1	999		6.00	0.00	50.00	50.00	4	security officer	1	1	1	3	1	1	3	
1	1	999		5.25	0.00	886.63	37.00	3	kitchen porter, sometimes waitress	1	1	3	1	1	2		

Comment [v6]: Replace with Standard Occupational Classification code (SOC)

w22

w23

livecond	PROBWK	prob1	prob2	prob3	prob4	prob5
5	1	no work schedule	racism from the manager's side	mobbing	spreading fear	lack of contract or P60
3	1	overtime not paid, bank holidays not paid, no holiday pay, no sick pay	worse treatment			
2	2					
2	2					
3	2					
4	2					
1	2					
1	1					
2	2					
1	1	agency took £50 and documents for registration at HO, but they didn't send the money so I couldn't get registered, so I had to fight to get the money back from agency	when I gave the HO registration documents to the factory's office, they disappeared and then reappeared only when I said that I would call the police			
2	1	doesn't pay all the salary -- pay several hours to several days less	deduct for travel from agency to the workplace and back – £3.5 per day (			
4	1	farm -- constantly made mistakes when calculating salary; held frequent mistakes with calculations of payment, we don't understand whether we get paid for overtime or not	payslip is difficult to read	don't inform about our rights and benefits; hide such information		

Comment [LC7]: All disclosive string data could have been coded rather than left as a string variable

2

COPYRIGHT

KNOW
HOW DATA
COPYRIGHT
WORKS

COPYRIGHT IS AN INTELLECTUAL PROPERTY RIGHT ASSIGNED AUTOMATICALLY TO THE CREATOR, THAT PREVENTS UNAUTHORISED COPYING AND PUBLISHING OF AN ORIGINAL WORK. COPYRIGHT APPLIES TO RESEARCH DATA AND PLAYS A ROLE WHEN CREATING, SHARING AND RE-USING DATA.
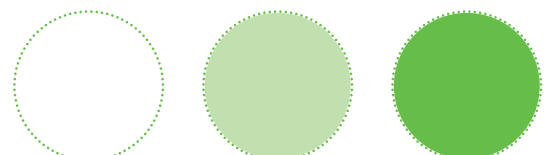
RESOURCES OVERVIEW

AREAS OF COVERAGE

- Definitions

- Copyright legislation

- Who owns copyright?

- Copyright and data sharing

- Secondary use of data and copyright

RESOURCES AVAILABLE

- PowerPoint presentation on 'Copyright'

- Exercise: Data copyright scenarios

- Exercise Answers: Data copyright scenarios

- www.data-archive.ac.uk/create-manage/ training-resources

DATA COPYRIGHT SCENARIOS

SCENARIO	DESCRIPTION
Scenario I: **Interviews with company directors**	A researcher has interviewed company directors about their careers and produced audio recordings and full transcripts. The researcher analyses the data and offers it to a data archive. The researcher did not get signed copyright transfers for the interviewees' words. What are the rights issues surrounding this offer of data?
Scenario II: **Transcription from a printed work into a spreadsheet**	A researcher has copied a series of statistical information from a printed work into a spreadsheet. The transcription is a direct copy with minimal alterations. The book is in copyright. What are the rights issues surrounding this research?
Scenario III: **Data in the public domain**	A researcher studies how health issues around obesity are reported in the media in the last 10 years. Freely available newspaper websites and library sources are used to obtain articles on this topic. Articles or excerpts are copied into a database and tagged according to various criteria for content analysis. Can the researcher use such public data without breaching copyright? Can the database be archived and shared with other researchers?
Scenario IV: **Archived data**	A researcher uses International Social Survey Programme (ISSP) data obtained from ZACAT/GESIS in Germany. These data are freely available to registered users. The researcher incorporates some of the ISSP data within a database containing his own research data. Can this database be placed on the researcher's website?
Scenario V: **Media database**	A researcher has collated articles about the British Prime Minister from *The Guardian* over the past ten years, using the LexisNexis database to source articles. They are then transcribed/copied by the researcher into a database so that content analysis can be applied. The researcher offers a copy of the database with the original transcribed text to a data centre.
Scenario VI: **Licensed data**	A researcher subscribes to access spatial AgCensus data from EDINA. These data are then integrated with data collated by the researcher. As part of the ESRC research award contract the data has to be offered for archiving at the UK Data Archive. Can such integrated data be offered?
Scenario VII: **Data obtained from the UK Data Archive**	A researcher has used the National Diet and Nutrition Survey (NDNS) data, obtained via the UK Data Archive. NDNS data are Crown Copyright. The researcher has processed the NDNS data (filtered, integrated and aggregated data across variables, while maintaining individual records) and used the processed data to model food chain risks. The researcher would like to archive the processed data that were used as input data for the modelling, as well as the modelling code, at the UK Data Archive.
Scenario VIII: **Survey questions**	A researcher wishes to re-use a set of questions from an existing survey questionnaire, to compare results between the newly proposed survey and the original.
Scenario IX: **Third party data**	The Stockholm Environmental Institute (SEI) has created an integrated spatial database, Social and Environmental Conditions in Rural Areas (SECRA). This contains a wide range of socio-economic and environmental characteristics for all rural Census 2001 Super Output Areas (SOAs) for England. Multiple third party data sources were used, such as Census 2001 data, Land Cover Map data and data from the Land Registry, Environment Agency, Automobile Association, Royal Mail and British Trust for Ornithology. Derived data have been calculated and mapped onto SOAs. The researchers would like to distribute the database for wider use.

DATA COPYRIGHT SCENARIOS

ANSWERS

SCENARIO	SOLUTION
Scenario I: **Interviews with company directors**	In this case the company directors hold the copyright over their words, whilst the researcher holds the copyright over the transcribed interviews. Quoting large extracts of the data, either in publications, or by archiving the transcripts, would breach the copyright of the interviewees in their words. Whilst many interviewees may not be concerned over such copyright, in this case where company directors share information about their careers, they may want to exert copyright over the content of the interviews (e.g. they may want to write their memoirs at some stage). If the researcher wants to publish large extracts of data, or archive transcripts, it is best to ask for transfer of copyright of the interview content from interviewees.
Scenario II: **Transcription from a printed work into a spreadsheet**	The researcher should technically have cleared copyright before transcription. If the work is for personal use only, this can be disregarded as fair dealing. If the newly constructed dataset is to be archived and disseminated, copyright clearance will need to be gained from the copyright holder.
Scenario III: **Data in the public domain**	Even though the articles obtained are in the public domain, they are still under copyright. Whilst such information can be used for personal research purposes (fair dealing), the articles cannot be archived, unless permission is obtained from the newspapers; otherwise this would breach copyright.
Scenario IV: **Archived data**	Although the ISSP data are available for free to all researchers, this does not mean that the data can be published on a website and through this made available to other people. The data can be incorporated into a database and used for personal analysis. Before this database is placed on a website, permission must be sought from the data owner.
Scenario V: **Media database**	Researchers cannot share either of these data sources as they do not have copyright in the original material. A data centre cannot accept these data as to do so would be a breach of copyright. The rights holders, in this case The Guardian and LexisNexis, would need to provide consent for archiving.
Scenario VI: **Licensed data**	The subscription agreement on accessing AgCensus data states that data may not be transferred to any other person or body without prior written permission from EDINA. Therefore, the UK Data Archive cannot accept the integrated data, unless the researcher obtains permission from EDINA. The researcher's partial data, with the AgCensus data removed, can be archived. Secondary users could then re-combine these data with the AgCensus data, if they were to obtain their own AgCensus subscription.
Scenario VII: **Data obtained from the UK Data Archive**	There is joint copyright over the processed data, shared between the researcher and the Crown (holding copyright over the NDNS data). The researcher must declare this joint copyright for the modelling data and requires no further permission from the Crown. The UK Data Archive End User Licence, which the researcher signed when obtaining the NDNS data from the UK Data Archive, specifically states: "offer for deposit any new data collections derived from the data supplied or created by the combination of the data supplied with other data." Thus the UK Data Archive can archive the processed data with a joint copyright declaration.

EXERCISE ANSWERS: DATA COPYRIGHT SCENARIOS

SCENARIO	SOLUTION
Scenario VIII: Survey questions	It should be assumed that all survey questions and instruments are copyright protected, with copyright residing with the organisation that commissioned, designed or conducted the survey. Our advice, therefore, is to contact the copyright holder directly for permission to reproduce questionnaire text for any new use. In our experience the copyright holder will almost always grant that permission. Some questionnaires contain measurement scales, batteries of questions or classifications. These particular instruments are copyright to the institution or company that produced them and must not be reproduced without permission. In many cases, the copyright for these instruments is printed on the relevant page of the questionnaire.
Scenario IX: Third party data	Whilst the database contains no original third party data, only derived data, there is still joint copyright shared between the SEI and the various copyright holders of the third party data. The researchers have sought permission from all data owners to distribute the data and the copyright of all third party data is declared in the documentation. The database can therefore be distributed.

UK•DATA
ARCHIVE