

DATA MANAGEMENT RECOMMENDATIONS FOR RESEARCH CENTRES AND PROGRAMMES

LAURENCE HORTON, VEERLE VAN DEN
EYNDEN, LOUISE CORTI AND LIBBY BISHOP
28 MARCH 2011

.....
T +44 (0)1206 872001
E name@data-archive.ac.uk
www.data-archive.ac.uk
.....



UK DATA ARCHIVE
UNIVERSITY OF ESSEX
WIVENHOE PARK
COLCHESTER
ESSEX, CO4 3SQ

.....
WE ARE SUPPORTED BY THE **UNIVERSITY OF ESSEX**, THE **ECONOMIC AND SOCIAL
RESEARCH COUNCIL**, AND THE **JOINT INFORMATION SYSTEMS COMMITTEE**

Contents

| | | |
|-----------|---|-----------|
| 1. | Introduction | 3 |
| 2. | Data Policy Context | 3 |
| 3. | Definitions | 4 |
| | 3.1. Data collection..... | 4 |
| | 3.2. Data | 4 |
| | 3.3. Documentation | 4 |
| | 3.4. Metadata..... | 4 |
| 4. | Participating ESRC investments..... | 4 |
| 5. | Methodology..... | 5 |
| 6. | General findings..... | 6 |
| 7. | Data management planning recommendations..... | 6 |
| | 7.1. Overall data management strategy | 7 |
| | 7.2. Ethics and consent | 7 |
| | 7.3. Copyright | 8 |
| | 7.4. Anonymisation..... | 8 |
| | 7.5. Documentation..... | 9 |
| | 7.6. Re-use..... | 9 |
| | 7.7. Data formatting..... | 10 |
| | 7.8. File formats | 10 |
| | 7.9. Version control | 11 |
| | 7.10. Security and back-up..... | 11 |
| | 7.11. Storage and file sharing..... | 12 |
| | 7.12. Transmission | 12 |
| | 7.13. Destruction of data and documentation | 13 |
| | 7.14. Develop a Data Management Resources Library | 14 |
| | 7.15. Data management coordination | 15 |
| | 7.16. Create a Data Inventory | 15 |
| | 7.17. File structure..... | 15 |
| 8. | Summary of data management recommendations for research centres and programmes | 17 |



1. Introduction

The Data Management Planning for ESRC Research Data-Rich Investments project (DMP-ESRC)¹ is funded by the Joint Information Systems Committee (JISC) under the Managing Research Data Programme. It aims to:

- evaluate existing data management practices amongst researchers in the social sciences community, in particular in large research investments
- help develop and integrate effective data management planning procedures and tools into the research lifecycle
- increase the data management and sharing capability of research hubs within the social sciences through support and training

The project is coordinated by the Research Data Management Support Service at the UK Data Archive in collaboration with the Economic and Social Research Council (ESRC) Policy and Resources Directorate.

The project has been working closely with the following ESRC research investments:

- New Dynamics of Ageing (NDA) Programme
- Centre for Research on Socio-Cultural Change (CRESC)
- Centre for Social and Economic Research on Innovation in Genomics (Innogen)
- Third Sector Research Centre (TSRC)

Research Centres and Programmes were selected based on the stage of research, research domains and re-use potential of resulting data.

Presented in this report are generic data management planning recommendations for research centres, based on information gathered on current data management practices and needs from researchers, centre managers and directors in the four collaborating research centres and programme.

2. Data Policy Context

The ESRC is at the forefront of data sharing in the United Kingdom and its Research Data Policy states that research data should be openly available to the maximum extent possible, through long-term preservation and high quality data management.²

The 2010 Research Data Policy introduces the concept of data management and sharing planning to promote a structured approach to managing data throughout the research life cycle. This enables better quality data that is ready for depositing and sharing. The DMP-ESRC project has provided valuable practical input in how data management planning strategies can work for large investments.

ESRC award holders are required to offer all research data resulting from research grants to a designated data centre, the UK Data Archive. The Archive supports ESRC applicants and award holders in enabling data sharing for both quantitative and qualitative data through the Economic and Social Data Service (ESDS). ESDS also ensures preservation and dissemination of archived research data in order to make them available to the research, learning, and teaching communities.

From the spring of 2011, the ESRC will require award applicants whose research will involve creating data to develop a data management plan as part of their application. The ESRC expects the plan to address consent and plans for data reuse, anonymisation and data security, data management roles and responsibilities, and strategies for addressing data management difficulties.³ A poor quality data management plan could influence the success of an application.⁴ Furthermore, the ESRC expects annual reporting on the

¹ UK Data Archive, Data Management Planning for ESRC Centres and Programmes, available at www.data-archive.ac.uk/create-manage/projects/JISC-DMP accessed 24 March, 2011

² Economic and Social Research Council (2010) Research Data Policy, available at http://www.esrc.ac.uk/images/Research_Data_Policy_2010_tcm8-4595.pdf accessed 08 March, 2011

³ Economic and Social Research Council (2010) Research Data Policy, available at http://www.esrc.ac.uk/images/Research_Data_Policy_2010_tcm8-4595.pdf accessed 08 March, 2011 (par 10, p.3)

⁴ idem, 2011 (par 25, p.6)

implementation of data management plans⁵.

3. Definitions

These definitions are not taken to be incontestable, but are included to illustrate the operational definitions applied by the UK Data Archive to concepts used in this report. The source of these definitions is the UK Data Archive's Preservation Policy document.⁶

3.1. Data collection

A data collection is typically comprised of three components: data, documentation and metadata. Occasionally, a fourth component exists: code. Data collections are typically organised by reference to a particular survey or research topic and cover a specific geographic area and time period.

3.2. Data

Data are all the material, regardless of format, which are intended to be analysed. As part of datasets, they are the primary element of a data collection. More precise definitions of data vary according to context. Quantitative data may refer to just the matrices of numbers or words that comprise a data file, but may also cover other information (metadata) held within a statistical package data file, such as variable labels, code labels and missing value definitions. Qualitative data might include interview transcripts as well as audio and video recordings (analogue or digital).

3.3. Documentation

Documentation is that portion of a data collection that is required in order to re-use data. It commonly covers the subjects of sampling design, methods of data collection, questionnaire/interview design, structure of the data files, lists of variables and coding schemes, details of weighting, confidentiality and anonymisation, and provenance of any secondary data used. It also includes licence arrangements and all materials obtained through the original negotiation and data deposit, as well as post-deposit information created during preservation and ingest activities. The terms metadata and documentation are often used interchangeably and there is overlap between the two, though metadata are typically more structured and documentation tends to have a structure that is specific to each data collection.

3.4. Metadata

Metadata are a subset of core data documentation which are highly structured. Catalogue metadata provides standardised structured information explaining the purpose, origin, time references, geographic location, creator, access conditions and terms of use of a data collection. Metadata are typically used for resource discovery, providing searchable information that helps users to find existing data, as a bibliographic record for citation, or for online data browsing.

Data-level metadata can include file names and descriptions for variables, records and their values. They can also document codes and classification schemes used. Additional metadata for qualitative data can include document annotation such as textual mark-up.

4. Participating ESRC investments

The ESRC Centre for Social and Economic Research on Innovation in Genomics (Innogen)⁷ runs from 2002 to 2012. It is part of the ESRC Genomics Network and studies the evolution of genomics and life sciences and their social and economic implications and is based at the University of Edinburgh and the Open University. Law, economics and social sciences researchers engage in research projects in the UK, Africa, China and India.

The Centre for Research on Socio-Cultural Change⁸ (CRESC), 2004-2014, is based at the University of Manchester and the Open University. Its mission is to analyse socio-cultural change in the context of socio-technical innovation, economic insecurity, and cultural diversity, with the intention to recognise different

⁵ Economic and Social Research Council (2010) Research Data Policy, available at http://www.esrc.ac.uk/images/Research_Data_Policy_2010_tcm8-4595.pdf, accessed 08 March, 2011 (par 22. p.6)

⁶ UK Data Archive (2010), UK Data Archive Preservation Policy, pp.14-15, available at http://www.data-archive.ac.uk/media/54776/ukda062-dps_preservationpolicy.pdf, accessed 22 December, 2010

⁷ Centre for Social and Economic Research on Innovation in Genomics, available at www.genomicsnetwork.ac.uk/innogen/, accessed 24 March, 2011

⁸ Centres for Research on Socio-Economic Change, available at www.cresc.ac.uk/, accessed 24 March, 2011

definitions and approaches to culture in its interface with processes of social change. CRESC research covers quantitative reuse of secondary data (e.g. longitudinal survey analysis) and qualitative research (ethnography, interviewing, audio and visual data).

The Third Sector Research Centre⁹ (TSRC), 2008 - 2013 is a collaboration across the Universities of Birmingham, Southampton, Kent and Middlesex. TSRC will bring together experts from a range of disciplines to develop a research programme that will lead to improved understanding of the key patterns, processes, and impacts of developments in the sector. This will strengthen the evidence base for policy towards the sector. TSRC will collaborate with, and offer a wide variety of services to, those working in and supporting the voluntary sector. In addition, TSRC co-ordinates the work of three Capacity Building Clusters which will support and enhance research capacity within the sector.

The New Dynamics of Ageing¹⁰ (NDA) programme is a cross-council collaboration seeking to improve the quality of life of older people. It runs from 2005 until 2012 and is funded by the Economic and Social Research Council (ESRC), Engineering and Physical Sciences Research Council (EPSRC), Biotechnology and Biological Sciences Research Council (BBSRC), Medical Research Council (MRC) and the Arts and Humanities Research Council (AHRC). The emphasis is strongly on multidisciplinary and cross-disciplinary research. Its funding is £22 million, with 43 percent contributed by the ESRC. The programme currently consists of 35 projects across 62 UK higher education institutes. NDA projects cover 47 disciplines, the main ones being psychology, sociology, health sciences and primary care.

5. Methodology

The recommendations in this document are the result of interviewing researchers on 26 projects across collaborating ESRC centres and programme between April and September 2010. Researchers within each hub were selected on the basis of presenting challenging issues of data management, methodological diversity, cross-institutional (and international) collaboration, varying stages of research completion and representation of a centre's research diversity. Additional one-off interviews were held with the director and/or a staff members in the Centre for Competition Policy at the University of East Anglia, the Centre for Economic and Social Aspects of Genomics at Cardiff University, and the Centre on Migration, Policy and Society at the University of Oxford.

Introductory meetings were held with directors and core staff of the participating hubs. At TSRC, researchers representing ten of the centre's core research programmes at the Universities of Birmingham, Southampton and Middlesex were interviewed. For CRESC six researchers from either the University of Manchester or Open University were interviewed. Six Innogen researchers at the University of Edinburgh or Open University were also interviewed.

Researchers from four projects in the New Dynamics of Ageing programme with a significant social science focus were interviewed: ProjectMIMA, TACT3, HALcyon, and SomnIA. Interviews were conducted with principal investigators on these projects and other members of the project were also often present and contributed to these interviews, or were available to answer questions.

All interviews were undertaken face-to-face and followed a data inventory questionnaire designed by the DMP-ESRC project based on and adapted from the Digital Curation Centre's Data Asset Framework¹¹, the Rural Economy and Land Use Programme Data Support Service Data Management Plan¹², and the Australian National University Information Literacy Programme's Data Management Plan.¹³ This inventory questionnaire also acted as an interview schedule allowing scope for discussion and supplementary questions where appropriate. Topics covered included:

- ethics and consent
- copyright
- anonymisation
- documentation
- reuse and archiving
- data formatting

⁹ Third Sector Research Centre, available at www.tsrc.ac.uk/ accessed 24 March, 2011

¹⁰ New Dynamics of Aging, available at www.newdynamics.group.shef.ac.uk/ accessed 24 March, 2011

¹¹ Digital Curation Centre, (2010) Data Asset Framework, available at <http://www.dcc.ac.uk/resources/tools-and-applications/data-asset-framework> accessed 24 March, 2011

¹² Rural Economy and Land Use Programme/UK Data Archive, (2010) Data Management Plan, available at <http://relu.data-archive.ac.uk/plan.asp> accessed 24 March, 2011

¹³ Australian National University (2010) Information Literacy Programme's Data Management Plan, available at http://ilp.anu.edu.au/dm/dmp_template.pdf accessed 24 March, 2011

- file formats
- data validation
- version control
- data security/storage/back-up
- transmission
- destruction of data and documentation

Interviews were recorded using a digital voice recorder for note taking purposes and interview notes written up immediately following the interview. Projects also occasionally provided documentation from the project as examples to illustrate data management practices.

All information gathered was organised into a spreadsheet for each hub. From this centre or programme specific data management recommendations were created. Each hub received a copy and was invited to comment on the suitability of recommendations. Copies were also provided to relevant ESRC case officers and the ESRC's Principal Policy Manager (Research Resources) Policy and Resources Directorate.

6. General findings

From the interviews it was apparent that there exist many good data management practices within these research centres and programme. Researchers are capable data managers, although they may not refer to the research and project management practices they implement as 'data management' or 'DM planning'.

Projects generally have an open attitude to data reuse, either during or at the end of a project – and most have explicit commitments to sharing either formally or informally. For some, obstacles to data sharing are a perceived lack of interest in researchers reusing data and this has at times been reinforced by the decision of an archive not to accept an offered data collection. This can be a deterrent to preparing future data to be shared.

7. Data management planning recommendations

Good data management a foundation of good research. It is about how to plan, propose and enable future data sharing across the research life cycle. Good data management also enhances the value of data for researchers. Making data available for reuse significantly increases citation rates – especially when that data is available through open access¹⁴, and even if data are not collected with the aim to be shared or contain barriers to reuse, well managed data help create an efficient research project. Data management planning can help by either obliging issues to be addressed early in the research design, or by stimulating thought as to how to best attend to data.

These recommendations ensure that data management practices will conform with the requirements of the ESRC Research Data Policy, and are based on best practice guidance of the UK Data Archive for data sharing, data preservation and long-term validity of data. They are not intended to be universally adopted by all centres or programmes. Instead they reflect a range of suggestions – either specific advice, general points of awareness, or indicating potential gaps that need to be addressed in cooperation with host institutions - that hopefully assist effective data management planning for research hubs.

¹⁴ JISC/SURF (2008) Changing the Rights Culture in HEIs with Open Access:

An information resource for academic authors, HEI managers and librarians/repository managers, available at <http://www.lboro.ac.uk/departments/dis/disresearch/poc/pages/academicguide-summary.html> accessed 08 March, 2011

7.1. Overall data management strategy

Choosing the level of centre- or programme-level coordination 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.

Centres and programmes cooperating with this project operated degrees of decentralisation in approaches to data management. The level and extent of decentralisation towards data management where hubs transferred authority and responsibility for data management to researchers varied.

Research hubs may operate a centralised or devolved approach to data management, either strongly coordinating how data are handled and who owns them, or giving researchers autonomy and responsibility over their research data. Factors that may influence which approach is taken can be size of the research hub, whether a single or cross-institutional entity and how much methodological and discipline diversity it covers.

A centralised approach to data management has as benefits:

- researchers can share good practice and data management experiences with each others, therefore building capacity for the centre
- establish a uniform approach to data management and central policies on various data aspects
- a hub can establish who owns data and keep track of them in time, especially when researchers leave the hub
- data can be stored at a central location
- ensure that all researchers and staff are aware of duties, responsibilities and funder requirements regarding research data, with easy access to relevant information

At the same time, researchers need to equally take responsibility for data management.

7.2. 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.

The ESRC expects data created in projects it funds to be available to other researchers for reuse and preservation.¹⁵ Consequently, it expects consent for reuse to be sought and not inferred.¹⁶

However, researchers can be flexible according to the nature of the research, kind of data gathered, format of data collection, and intended use. For instance, consent can be considered and negotiated at different stages of research. For research where formal written consent is problematic, consent can be obtained verbally. For data collected as audio or video recordings, consent can be recorded. It is obviously to the researcher's discretion how to approach issues of reuse and/or copyright, either up-front or negotiated ongoing consent. However, within a centre all projects should take into account future uses of data, including sharing and preservation, and that consent form language should not prohibit data reuse and preservation.

Suggestions for advice on the wording of written consent forms and information sheets for qualitative and quantitative research, including specific attention to consent for data sharing, are available from the UK Data Archive.¹⁷

¹⁵ Economic and Social Research Council (2010) Research Ethics Framework, http://www.esrc.ac.uk/images/Framework_for_Research_Ethics_tcm8-4586.pdf accessed 08 March, 2011 (section 1.17.5 p.24)

¹⁶ idem (par 32, p.7)

¹⁷ UK Data Archive (2010) Consent Overview, available at <http://www.data-archive.ac.uk/create-manage/consent-ethics/consent>

Research hubs should provide a link to the ESRC's Research Ethics Framework (REF)¹⁸ and should create a centralised resource of relevant legal information, such as the Data Protection Act (1998)¹⁹, The Freedom of Information (Fol) Acts in England and Wales (2000)²⁰ and Scotland (2002)²¹ and where applicable, the Environmental Information Regulations (2004).²² Resources and advice should also be based on institutional recommendations and consultation. Further Fol resources for researchers have been developed by JISC.²³

7.3. 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.

It is important there is clarity on the ownership rights of primary and reused data to avoid problems if multiple funders exert copyright and when meeting requirements to offer data for reuse. Agreements on data sharing in externally funded or co-funded projects depend on the funding agency. UK funding councils all have either a commitment to funded data being offered for reuse (or a case to be made as to why data can not be reused), or strongly encourage researchers to make data available for reuse. Summaries to UK funding council policies are provided by the Digital Curation Centre.²⁴

Projects should identify sources of existing data and any conditions placed upon their use, such as ownership, copyright constraints, and redistribution rights. Researchers reusing data need to be clear on copyright or licences under which data are used and/or offered for archiving, and should state conditions under which use of data are permissible in, for example, a data inventory .

A useful resource would be a Q&A on use of licences and copyright, drawing on institutional policies and other advice. In academia, in theory, the employer is the first owner of copyright in a work. However, many academic institutions waive copyright and grant ownership to researchers. Nonetheless a centre statement on intellectual property rights should be consistent with institutional and/or funding body policy to clarify guidelines for researchers. For example, the ESRC's policy is that unless stated otherwise, intellectual property rights rests with the organisation carrying out the research.²⁵

7.4. 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.

Like the ESRC²⁶ we would like centres to encourage researchers to plan and anonymise data early in the research. Doing this avoids the pressure to anonymise while writing up publications, organising the collection

accessed 10 January, 2011

¹⁸ Economic and Social Research Council (2010) Research Ethics Framework, available at <http://www.esrc.ac.uk/about-esrc/information/research-ethics.aspx> accessed 02 February, 2011

¹⁹ Data Protection Act 1998, available at <http://www.legislation.gov.uk/ukpga/1998/29/contents>

²⁰ Freedom of Information Act 2000, available at <http://www.legislation.gov.uk/ukpga/2000/36/contents>

²¹ Freedom of Information (Scotland) Act 2002, available at <http://www.legislation.gov.uk/asp/2002/13/contents>

²² Environmental Information Regulations 2004, available at <http://www.legislation.gov.uk/uksi/2004/3391/contents/made>

²³ Andrew Charlesworth, Chris Rusbridge (2010) Freedom of Information and Research Data: Questions and answers, available at <http://www.jisc.ac.uk/publications/programmerelated/2010/foiresearchdata.aspx> accessed 10 January, 2011

²⁴ Digital Curation Centre (2010) Funders' Data Policies, available at <http://www.dcc.ac.uk/resources/policy-and-legal/funders-data-policies> accessed 10 January, 2011

²⁵ Economic and Social Research Council (2010) Research Data Policy, available at http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf accessed 08 March, 2011 (par 35, p.7)

²⁶ Idem (par 32, p.7)

to offer for archiving, and preparing for the next project. Sensitive and confidential data can be shared ethically provided researchers address anonymisation issues early.²⁷

We would encourage all projects to create an anonymisation log of all replacements, aggregations or removals made - care should be taken to store such a log separately from the anonymised data files. Cases exist especially in small, individual research projects where only the principal investigator has familiarity as to what has and has not been anonymised and in their absence or without a written a log of what anonymisation has occurred, the reusability of the data could be diminished. Examples of anonymisation logs are available. A qualitative example is available from the UK Data Archive²⁸. Quantitative suggestions are also available.²⁹

7.5. 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.

Documentation allows researchers within a project or as re-users to make the best use of the data. Generating documentation is a necessary aspect of the research process, but the range, extent and quality of documentation can vary according to the project. Nonetheless, it is an expectation of the ESRC that research data be accompanied by documentation that facilitates reuse.³⁰

Centres should encourage projects to create and sustain comprehensive documentation during the research process explaining how data were created, its content, structure, validation, and any manipulations undertaken. Any additional value added through variable derivation should be recorded and syntax kept as documentation.

Although strong awareness of specialised terms exists within projects this awareness almost never transfers to other potential users in the absence of good descriptive material. This material is often already present in the data – either the research design, or descriptions within data itself, and need not represent an extraordinary effort above and beyond data creation to capture as documentation. We recommend that projects create definitions of specialist terms and an acronym list for data files or to be explained within data files.

7.6. Re-use

Plans for sharing data and future use of data.

Know all relevant funding council data sharing policies.
Know data centres and institutional repositories to preserve your data and distribute them for re-use.

All ESRC funded research grants carry a contractual obligation to make data available for reuse.³¹ Where this is perceived not to be possible by the researcher, a valid case covering all data types created must be presented to the ESRC.³²

For ESRC centres, data archiving is required only at the end of the centre contract and if the centre secures a second phase of funding, then ten years can elapse before data has to be offered for reuse. Many centres also acquire external funding, which may be bound by different data policies.

²⁷ Economic and Social Research Council (2010) Research Data Policy, available at http://www.esrc.ac.uk/images/Research_Data_Policy_2010_tcm8-4595.pdf accessed 08 March, 2011 (par 33, p.7)

²⁸ UK Data Archive (2010) Anonymisation/Quantitative, available at <http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation?index=2> accessed 10 January, 2011

²⁹ UK Data Archive (2010) Anonymisation/Quantitative, available at <http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation?index=1> accessed 10 January, 2011

³⁰ Economic and Social Research Council (2010) Research Data Policy, available at http://www.esrc.ac.uk/images/Research_Data_Policy_2010_tcm8-4595.pdf accessed 08 March, 2011 (par 7, p.3)

³¹ Idem (par 6, p.3)

³² Idem (par 20, p.6)

7.7. 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.

Producing a coversheet for qualitative interviews can provide contextual information about the interview consistent within, and where possible, across projects. Coversheets should provide space for recording the data collection event: date, place, interviewer name, interviewee details. Guidance for researchers and transcribers includes a unique identifier, participant ID, specified uniform layout across the research project – font, spacing, page numbers, header information, and marking sensitive text. These characteristics can be held in a spreadsheet and mail merged into the documents.

Ideally, for interviews we recommend guidance for researchers and transcribers on systematic procedures and uniform conventions to be applied. Always try to maintain a high level of consistency between transcripts and the conventions used in those transcripts, particularly when multiple people are carrying out transcription on the same research project. Transcripts should also include speaker tags to identify the question and response sequence. For other textual material we encourage the use of uniform elements. These should be integral at the stage of data collection and the conversion of material into written text to avoid the stresses of retrospective formatting when pressure on resources is at its highest. Furthermore, formatting at the stage of analysis and research can be good data management because it improves the quality of the data.

Establishing a confidentiality agreement with administrative staff and transcribers covering the handling of confidential data will provide additional reassurance that participant's data are being handled securely. It should outline the nature of confidentiality, storage procedures and returning and retention of data.

Useful examples and guidelines on transcriptions available from the UK Data Archive³³ and Timescapes.³⁴

7.8. File formats

Formats and software in which research data are created, analysed, stored or preserved.

Encourage use of standard or open lossless formats.

Within many research hubs there appears to be few set procedures for the use of particular types of data formats. All digital data are susceptible to damage or loss, endangered by the obsolescence of hardware and software environments as time passes. Researchers need to be aware that while default formats in commonly used research-oriented software may be easy to use, they can increase the risk of problems for long-term usability and accessibility of data. Working formats and preservation formats are not mutually exclusive and we would not seek to prohibit use of proprietary formats, just to encourage usage of standard, interchangeable formats and/or open lossless formats to be used for data collection where possible and where convenient. The UK Data Archive has a list of suggested formats for sustainability.³⁵

³³ UK Data Archive (2010) Transcription available at <http://www.data-archive.ac.uk/create-manage/format/transcription> accessed 24 March, 2011

³⁴ ESRC Timescapes (2008) Interview Transcription Guidelines and Model, available at http://www.timescapes.leeds.ac.uk/assets/files/timescapes/Transcription_guidelines_and_model_23July08_current.doc accessed 11 January, 2011

³⁵ UK Data Archive (2010) Data Formats Table, available at <http://www.data-archive.ac.uk/create-manage/format/formats-table> accessed 10 January, 2011

7.9. 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.

Significant elements of data management may be integrated into the research process. One good example is version control. While projects and individual researchers have their own informal forms of version control that may be effective in small teams or in their own research, informal forms are vulnerable to confusion and data loss. Different copies or versions of files should be tracked by a version controlling procedure. Projects should establish checks and procedures to ensure which version of a data file or documentation is the agreed shareable version, or master version, for example the most current. This is critical where projects are collaborating across different locations. Minimal best practice for version control is to identify files using a systematic naming convention, with unique file names ending in a date (YYYY-MM-DD) or version number (00-01). Alternatively, data files can include a version control table or file history in a data file (see example below). Data and documentation should have versions and statuses recorded (interim, final) and all should have an 'owner' responsible for their status and changes. Virtual research environments (VRE) or off-the-shelf versioning software can ease the burden of manual version control.

Example of version control table

| | | | |
|-----------------------|--------------------|--|---------------------|
| Title: | | Vision screening tests in Essex nurseries | |
| File Name: | | VisionScreenResults_00_05 | |
| Description: | | Description of the data files | |
| Created By: | | Chris Wilkinson | |
| Maintained By: | | Sally Watsley | |
| Created: | | 04/07/2007 | |
| Last Modified: | | 25/11/2007 | |
| Based on: | | VisionScreenDatabaseDesign_02_00 | |
| | | | |
| Version | Responsible | Notes | Last amended |
| 00_05 | Sally Watsley | Version 00_03 and 00_04 compared by SW | 25/11/2007 |
| 00_04 | Vani Yussu | Entries checked by VY, independent from previous | 17/10/2007 |
| 00_03 | Steve Knight | Entries checked by SK | 29/07/2007 |
| 00_02 | Karin Mills | Test results 81-120 entered | 05/07/2007 |
| 00_01 | Karin Mills | Test results 1-80 entered | 04/07/2007 |

7.10. Security and back-up

Steps to securely store data and back-up procedures to make copies to restore originals in case of data loss.

Know your institutional IT security arrangements.

Know your institutional procedure and regularity of data back-up, especially for remote and cross-institutional working.

The ESRC adheres to the UK government's requirements on protecting personal data³⁶. It expects researchers to protect the confidentiality of participants, where necessary, by ensuring data is safely and securely stored.³⁷ Institutions are expected to provide support and facilities to enable data security, but researchers should show due diligence in their efforts to ensure that data is protected. This includes verifying back-up procedures to ensure data is recoverable.

Centres should be aware of institutional IT security and back-up policies, linking to a policy where possible.

³⁶ Cabinet Office (2008) Data Handling Procedures in Government: Final Report, available at http://www.cesg.gov.uk/products_services/iatp/documents/data_handling_review.pdf, accessed 08 March, 2011

³⁷ Economic and Social Research Council (2010) Research Data Policy, available at http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf accessed 08 March, 2011 (par 5, p.2)

This can provide researchers with awareness of if and/or how their data are being looked after securely.

7.11. Storage and file sharing

Strategy for sharing and storing research data and providing shared access.

For cross-institutional collaborative research, assess the feasibility of adopting a virtual research environment or file sharing and storage system.

Data storage capacity has not emerged as an issue in this project because greater storage is now both affordable and accessible. Instead the ability to share files and collaborate on data has emerged as a key challenge for research hubs.

One of the themes emerging from the Data Management Planning for ESRC Centres and Programmes project is the problem researchers have with the practicalities of collaboration within and especially across institutions. No centre interviewed in this project operated an effective virtual research environment or collaborative file sharing space. Those that have implemented a system have found them limited, impractical, difficult to use, and a burden. Consequently, it is administrative staff who have been the central point of guidance for providing information about policy and guidance documents that are relevant to researchers.

Researchers often rely on shared drive spaces, portable storage media, or email for data storage and transfer and, while using memory sticks may be a useful short-term solution, it is certainly not a safe method of storage. Researchers need to be aware of the risk of data loss due to the high probability of failure with such devices.

Virtual research environments do exist, and we have found projects in the New Dynamics of Ageing programme that use them and claim to do so satisfactorily. They can enable collaborative work spaces, storage, automated version control, encrypted security and controlled access. Options include the market leaders - Microsoft SharePoint 2010 (adopted in many UK HE institutions) and the open-source product Sakai (adopted by some of the JISC's Virtual Research Environment programme projects³⁸). There are also numerous kinds of wikis and online spaces. If a team is considering adopting one product consultation with institutional IT services is essential. We recommend that research hubs assess the feasibility of adopting a virtual research environment or file sharing/storage system.

7.12. Transmission

Methods of moving data files from one location to another.

Only send anonymised or non-sensitive data in unencrypted form.
 Encrypt confidential or personal data before transmission or password protected files.
 Explore secure storage and sharing files and folders with others across the Internet using file synchronization, virtual research environments or file sharing software.

Relying on email to transmit data and documentation, even internally, could be a vulnerable point in preserving the confidentiality of research participants and data as anything that ends up on email persists in numerous exchange servers (the sender, the receiver and others in-between).

We recognise that transferring large files is problematic for researchers. Third party commercial file sharing services (yousendit.com³⁹, Microsoft SkyDrive⁴⁰, Google Docs⁴¹, Dropbox⁴²) exist to facilitate the movement of files. However, file sharing services are not necessarily permanent or secure, or covered by UK law and in potential violation of UK law - particularly in relation to the UK Data Protection Act (1998) which states data should not be transferred to other countries without adequate protection.

We recommend that implementing a policy whereby only anonymised/non-sensitive data can be sent in

³⁸ JISC (2011) Virtual Research Environment Programme available at <http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx> accessed 24 March, 2011

³⁹ <http://www.yousendit.com> accessed 10 January, 2011

⁴⁰ <http://explore.live.com/windows-live-skydrive> accessed 10 January, 2011

⁴¹ <http://www.google.com/google-d-s/tour1.html> accessed 10 January, 2011

⁴² <http://www.dropbox.com/features> accessed 10 January, 2011

unencrypted form or files are at least password protected. For encryption the UK Data Archive recommends the use of Pretty Good Privacy standard technology, this is available in open source or commercial software, for example Utimaco⁴³, TrueCrypt⁴⁴, GnuPG⁴⁵. However, consultation with the institution IT service is advised.

We also suggest centres explore options for secure storage and sharing files and folders with others across the Internet using file synchronization with the institution's IT service. As mentioned in 7.11, VREs do exist, for example, integrating workspaces into SharePoint Workspace 2010⁴⁶, in addition to other options like MediaWiki⁴⁷. These platforms provide an encrypted shared workspace for data files and documents in group collaboration. Users can create workspaces, add, and invite members to a workspace. Each member has a privately editable copy of the workspace. Users interact and collaborate in the common workspace which is a private virtual location. Changes are tracked by SharePoint 2010, sent to all members and all copies of the workspace are synchronised via the network in a peer-to-peer manner.

Technology solutions aside, a change in habits and culture is needed for people to adopt online storage and workspaces. If a form of VRE is adopted it should cover version control and security (encrypted data). Otherwise, at a minimum, password protected data should be introduced.

7.13. Destruction of data and documentation

Procedures for safe and secure erasing of data at the conclusion of research project or when needed.

Policy on retention and disposal of data.

A strategy for the secure erasing of data and documentation which require destruction can be relevant at multiple points in time: for example, during research by destroying copies of data no longer needed, at the conclusion of a project, after completion and publishing or archiving your data. Researchers should be aware of ways and means for the safe and secure erasing of data and documentation so it is rendered impossible to recover. Sole copies of original research data that have not been saved for sharing/archiving should not be destroyed unless there is a compelling security reason.

We recommend a central policy on the retention and disposal of data and documentation consistent with respective institutional policies: this includes the retention of data and documentation at the end of a project, but may also cover destruction of material no longer needed. Destruction of data and documentation should be properly carried out using paper shredding, rewriting software, or specialist CV/DVD shredders. Free software is available containing methods for securely erasing files from hard disks that meet recognised standards of overwriting to adequately scramble sensitive files. For total disc wiping - scrambling everything on a machine, Darik's Boot and Nuke⁴⁸. For scrambling selected files, BC Wipe⁴⁹, Wipe File⁵⁰, DeleteOnClick⁵¹, Eraser⁵² all work with Windows PC's. For Mac users the standard 'secure empty trash' option is usually sufficient. However, there is no surer way to dispose of data and documentation than physical destruction.

Shredders should match the standard the UK government uses, a minimum standard of DIN 4, which ensures a cross cut particles of at least 2x15mm. A risk-adverse approach for all digital data storage devices is to encrypt devices on installing operating software, and when data and documentation is in a position to be disposed, seek to physically destroy the drive using an approved secure destruction facility, usually a commercial service.

⁴³ www.utimaco.com accessed 10 January, 2011

⁴⁴ www.truecrypt.org accessed 10 January, 2011

⁴⁵ www.gnupg.org accessed 10 January, 2011

⁴⁶ <http://office.microsoft.com/en-us/sharepoint-workspace> accessed 10 January, 2011

⁴⁷ <http://www.mediawiki.org/wiki/MediaWiki> accessed 10 January, 2011

⁴⁸ <http://www.dban.org/> accessed 10 January, 2011

⁴⁹ <http://www.jetico.com/wiping-bcwipe/> accessed 10 January, 2011

⁵⁰ <http://www.gaijin.at/en/dlwipefile.php> accessed 10 January, 2011

⁵¹ <http://www.2brightsparks.com/onclick/doc.html> accessed 10 January, 2011

⁵² <http://eraser.heidi.ie/> accessed 10 January, 2011

7.14. Develop a Data Management Resources Library

Develop a Data Management Resources Library.

Develop a Data Management Resources Library as either an intranet site or website, as a one-stop place where a centre or programme can hold all its relevant data management and data management planning guidance documents for researchers to use. Initially a folder could be created called 'data management' that pulls together policies, guidance documents, exemplars and templates, and links to key external resources. An ethical review folder could hold relevant national and local guidance and ethical review documents.

The library could contain:

Internal documents

- 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
- Transcription coversheets
- Information on the Data Protection Act (1998), the Freedom of Information (FoI) Acts in England (2000) and/or Scotland (2002), the Environmental Information Regulations (2004)
- Local statement on copyright
- Local statement on retention and destruction of data

External resources

- The ESRC Data Policy (2000)⁵³
- The ESRC Research Data Policy (2010)⁵⁴
- The ESRC Framework for Research Ethics⁵⁵
- The UK Data Archive's Guidance on Managing and Sharing Data⁵⁶
- UK Data Archive advice on preservation and analysis formats
- Cabinet Office Data Handling Procedures in Government: Final Report (2008)⁵⁷
- JISC Freedom of Information and research data: Questions and answers (2010)⁵⁸
- Summary of UK funding council policies on data reuse: Digital Curation Centre⁵⁹

⁵³ Economic and Social Research Council (2000) Data Policy available at http://www.esrc.ac.uk/_images/DataPolicy_tcm8-4594.pdf accessed 24 March, 2011

⁵⁴ Economic and Social Research Council (2010) Research Data Policy available at http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf accessed 24 March, 2011

⁵⁵ Economic and Social Research Council (2010) Research Ethics Framework, available at http://www.esrc.ac.uk/_images/Framework_for_Research_Ethics_tcm8-4586.pdf accessed 08 March, 2011

⁵⁶ UK Data Archive (2011) Create and Manage Data, available at <http://www.data-archive.ac.uk/create-manage> accessed 24 March, 2011

⁵⁷ Cabinet Office (2008) Data Handling Procedures in Government: Final Report, available at http://www.cesg.gov.uk/products_services/iatp/documents/data_handling_review.pdf, accessed 08 March, 2011

⁵⁸ Andrew Charlesworth, Chris Rusbridge (2010) Freedom of Information and Research Data: Questions and answers, available at <http://www.jisc.ac.uk/publications/programmerelated/2010/foiresearchdata.aspx> accessed 10 January, 2011

⁵⁹ Digital Curation Centre (2010) Funders' Data Policies, available at <http://www.dcc.ac.uk/resources/policy-and-legal/funders-data-policies> accessed 10 January, 2011

7.15. Data management coordination

Allocate clear roles and responsibilities for all data management aspects; designate data management coordinator for each project.

From 2011, the ESRC expects individual grant holders to report on the on-going implementation of data management.⁶⁰ This can also be applied to centre and programme projects to ensure data management strategies are being implemented and any arising problems are addressed.

Each project should have a data management coordinator or someone responsible for addressing data management issues within a project. This need not necessarily be the Principal Investigator, but someone close to the data. It is useful for research hub to have a designated Data Manager, someone who is charged with keeping the Data Management Resources Library and the Data Inventory (see below) up to date.

7.16. Create a Data Inventory

Create a Data Inventory.

Creating a Data Inventory of data collections created and/or acquired by researchers can help with three aspects of data management: ensuring licensed or disclosive data are held appropriately, raising awareness of local data resources, and helping administer data management throughout the research process. The Inventory can act as an initial data management plan, which can be periodically updated and at the end of the project will contain key information that can make offering data for archiving more straightforward. Even if a data collection is not suitable for archiving, having it recorded in a inventory maintains a record for the hub so it can better keep track of ownership, management and contents of projects undertaken by its researchers.

An example of how this inventory could operate is by using a MS Access database. The data management contact can use a Data Collection Inventory form in the database to supply contextual information about the project. A Data form and a Documentation form are designed to create an inventory of the project's data and documentation. The advantage of a database is it allows more sophisticated integration and analysis of information on the hub's projects than using an excel spreadsheet. Furthermore, it allows the hub to control vocabulary where this is useful, for example types of formats, rather than one researcher using the term 'Word doc', and another using 'MS word document'. Controlling vocabulary makes the process of information querying/retrieval that much easier, for example how many researchers are using Word documents as a data format type? Of course, there has to be flexibility in other areas, for example, describing consent processes and these are left as open text fields.

7.17. File structure

Recommended file structures for projects.

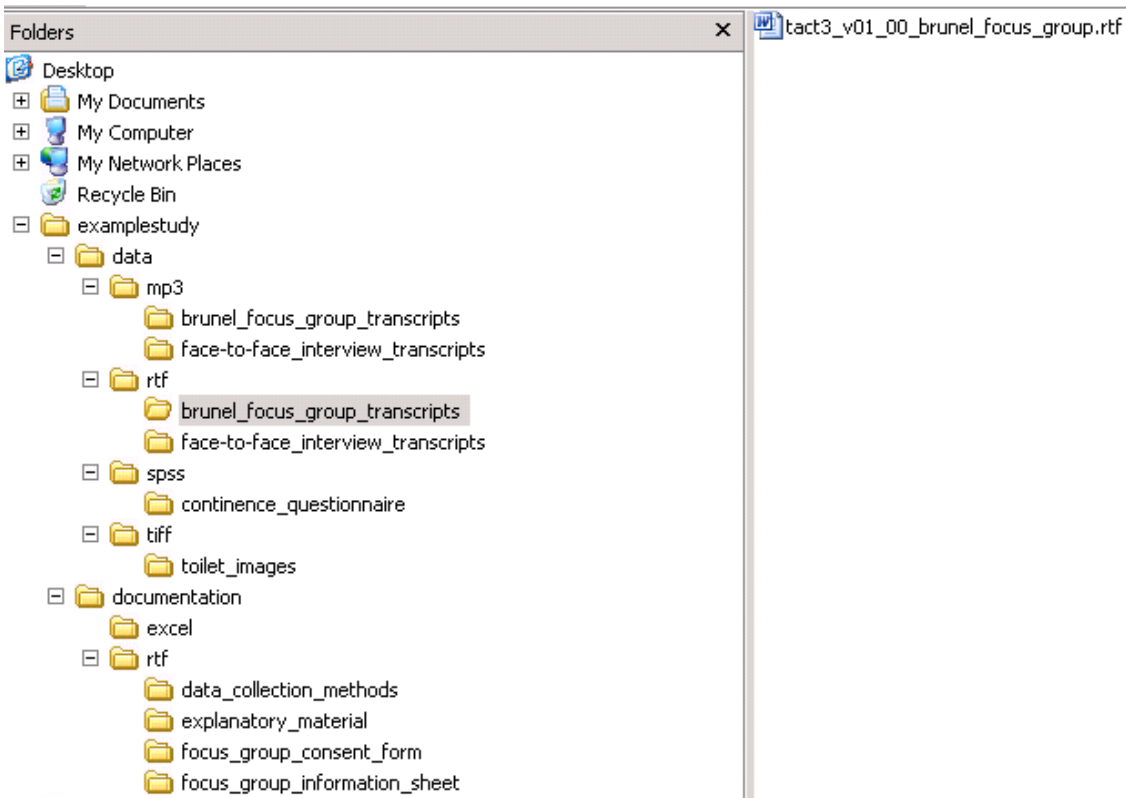
A clear file structure can provide an efficient way of storing, accessing and disposing of data and documentation.

In this example, documentation and data are separate folders under which files are organised by format and then data creation type or documentation purpose. It helps restrict the folder hierarchy to a depth of three to four levels. Ideally there should be no more than ten folders in each list – depending on the nature of the project. In this example, neither data nor documentation is nested too deeply in the structure.

Hubs and projects should decide on whether a deep or a shallow hierarchy is preferable, which will suit their own data types. Although structure is dependent on the type of project and the preferences of individual the researchers, it is a good idea to consider hierarchy at the time of developing the structure.

⁶⁰ Economic and Social Research Council (2010) Research Data Policy, available at http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf accessed 08 March, 2011 (par 22, p.6)

Example file structure for a project



8. Summary of data management recommendations for research centres and programmes

| Topic | Recommendation |
|---|---|
| <p>DATA MANAGEMENT STRATEGY</p> <p><i>Choosing the level of centre- or programme-level coordination and decision-making in planning and implementing management and sharing of research data</i></p> | <p>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</p> <p>Allocate clear roles and responsibilities for all data management aspects; designate data management coordinator for each project</p> <p>Develop a data management resources library</p> <p>Create a data inventory</p> <p>Recommended file structures for projects</p> |
| <p>ETHICS AND CONSENT</p> <p><i>Informed consent for people to participate in research and for use of information collected</i></p> | <p>Consent not prohibiting data sharing</p> <p>Gain consent for sharing, preservation and re-use of research data</p> |
| <p>COPYRIGHT</p> <p><i>Intellectual property rights of data</i></p> | <p>Identify source and ownership of third party data</p> <p>Establish conditions of use, copyright constraints and redistribution rights</p> <p>Statement on intellectual property rights of generated data</p> |
| <p>ANONYMISATION</p> <p><i>Procedures to protect the identify of individuals, organisations or businesses and avoid disclosure</i></p> | <p>Plan and anonymise data early in research</p> <p>Create anonymisation log of edits, replacements, removals or aggregations made</p> |
| <p>DOCUMENTATION</p> <p><i>Material explaining how data are created, what they mean, their content and structure, and alterations done</i></p> | <p>Create and sustain comprehensive documentation of data creation methodology</p> <p>Define specialist terms and acronyms used</p> |
| <p>RE-USE</p> <p><i>Plans for sharing and future use of data</i></p> | <p>Know all relevant funding council data sharing policies</p> <p>Know data centres and institutional repositories to preserve your data and distribute them for re-use</p> |
| <p>DATA FORMATTING</p> <p><i>Layout, editing and structuring of data</i></p> | <p>Guidance for researchers and transcribers on transcription procedures and use of uniform editing and layout</p> <p>Use coversheet for transcripts of qualitative interviews</p> <p>Use non-disclosure agreement for administrative staff and</p> |

| | |
|---|---|
| | <p>transcribers accessing confidential data</p> <p>List of recommended transcribing services</p> |
| <p>FILE FORMATS</p> <p><i>Formats and software in which research data are created, analysed, stored or preserved</i></p> | <p>Encourage use of standard or open lossless formats</p> |
| <p>VERSION CONTROL</p> <p><i>Checks and procedures to establish which version of a file is the most current</i></p> | <p>Establish procedures to control which version of a file is the agreed shareable version, or master version</p> |
| <p>SECURITY AND BACK-UP</p> <p><i>Steps to securely store data and back-up procedures to make copies to restore originals in case of data loss</i></p> | <p>Know your institutional IT security arrangements and capacity for data storage</p> <p>Know your institutional procedure and regularity of data back-up, especially for remote and cross-institutional working</p> |
| <p>FILE SHARING AND STORAGE</p> <p><i>Strategy for sharing and storing research data and providing shared access</i></p> | <p>For cross-institutional collaborative research, assess the feasibility of adopting a virtual research environment or file sharing and storage system</p> |
| <p>TRANSMISSION</p> <p><i>Methods of moving data files from one location to another</i></p> | <p>Only send anonymised or non-sensitive data in unencrypted form</p> <p>Encrypt confidential or personal data before transmission or password protected files</p> <p>Explore secure storage and sharing files and folders with others across the Internet using file synchronization, virtual research environments or file sharing software</p> |
| <p>DESTRUCTION</p> <p><i>Procedures for safe and secure erasing of data at the conclusion of research project or when needed</i></p> | <p>Policy on retention and disposal of data</p> |