Managing and Sharing Research Data

Louise Corti
Research Data Management Team

REC session
NUI Galway
14 October 2014
Overview of this session

Presentation

- About the UK Data Service
- Managing and sharing data – an overview
- Formatting, documenting, and organising data
- Training materials available

Exercises

- ‘Fun’ quizzes on documenting, formatting, organising, and storing your data
What is the UK Data Service?

- a comprehensive resource funded by the Economic and Social Research Council (ESRC)
- a single point of access to a wide range of secondary social science data
- support, training and guidance throughout the data life cycle
- listen to our recorded webinars at

http://ukdataservice.ac.uk/news-and-events/videos.aspx
Welcome to the UK Data Service

Your resource for quality social research data

A unified point of access to data from ESDS, Census Programme, Secure Data Service and others

Integrates ESDS, Survey Question Bank and Census.ac.uk
What does the UK Data Service do?

- put together a **collection of the most valuable data** and enhance these over time
- **preserve data in the long term** for future research purposes
- make the **data and documentation** available for reuse
- **provide data management advice** for data creators
- **provide support** for users of the service
- information about **how data are used**
- easy access through **website**
Who is it for?

- academic researchers and students
- government analysts
- charities and foundations
- business consultants
- independent research centres
- think tanks
- citizen scientists, where skills enable analysis
Our approach to data sharing

• Our organisation has 45 years experience in selecting, ingesting, curating and providing access to social science data

• Extensive experience of supporting researchers and data creators of social science data and related disciplines

• Been managing data sharing for the ESRC Data Policy since 1995, including cross council programmes

• Our best practice approaches to making data shareable based on:
  • challenges faced by researchers to share data
  • handling research data – quantitative and qualitative
Our managing and sharing data resources

- Online best practice guidance: [ukdataservice.ac.uk/manage-data.aspx](http://ukdataservice.ac.uk/manage-data.aspx)
- Sage Managing and Sharing Research Data – a Guide to Good Practice: [www.uk.sagepub.com/books/9781446267264](http://www.uk.sagepub.com/books/9781446267264)
- Busy training programme
Key skills that enhance methods training

- Policy landscape and data sharing
- Writing and implementing a data management plan
- Documenting and contextualising data
- Formatting and organising data
- Storing and transferring data, incl. encryption and security
- Legal, ethical issues in handling and sharing data – consent, anonymisation and access control
- Rights relating to research data
- Publishing and citing research data
Why 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.

Research benefits:

- think what to do with research data, how collect, how look after
- keep track of research data (e.g. staff leaving)
- identify support, resources, services needed
- plan storage, short & long-term
- plan security, ethical aspects
- be prepared for data requests (FoI, funder)
ESRC research data policy

Research data should be openly available to the maximum extent possible through long-term preservation and high quality data management. (ESRC Research Data Policy, 2010)

• ESRC grant applicants planning to create data during their research include a data management plan with their application, as an attachment to the Je-S form

• ESRC award holders offer their research data to the ESRC Data Store (managed by UK Data Service) within three months of the end of their grant, to preserve them and to make them available for new research.

Researchers who collect the data initially should be aware that ESRC expects that others will also use it, so consent should be obtained on this basis and the original researcher must take into account the long-term use and preservation of data. (ESRC Framework for Research Ethics, 2012)
## ESRC data management plan

<table>
<thead>
<tr>
<th>Topic</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment of existing data</td>
<td>Blue</td>
</tr>
<tr>
<td>Information on new data</td>
<td>Green</td>
</tr>
<tr>
<td>Quality assurance of data</td>
<td>Green</td>
</tr>
<tr>
<td>Backup and security of data</td>
<td>Green</td>
</tr>
<tr>
<td>Expected difficulties in data sharing</td>
<td>Yellow</td>
</tr>
<tr>
<td>Copyright / Intellectual Property Right</td>
<td>Blue</td>
</tr>
<tr>
<td>Responsibilities</td>
<td>Green</td>
</tr>
<tr>
<td>Preparation of data for sharing and archiving</td>
<td>Green</td>
</tr>
</tbody>
</table>

[ESRC DMP guidance](https://www.esrc.ac.uk/dmp)
Data life cycle intervention

- Sign off consent form
- Agree data & metadata templates/organisation
- Data sharing protocols
- Licensing, terms and conditions for sharing, formal documentation
- Data formats, data migration
Key planning issues

- Know your legal, ethical and other obligations towards research participants, colleagues, research funders and institutions
- Know your institution’s policies and services: storage and backup strategy, research integrity framework, IPR policy, institutional data repository
- Assign roles and responsibilities to relevant parties
- Incorporate data management into research cycle
- Implement and review management of data during project meetings and review
How to cost data management

**STEP 1**
- check data management activities in table and *tick* what applies to your proposed research; we propose 18 essential RDM activities

**STEP 2**
- for each selected activity, estimate / calculate additional time and/or resources needed and cost this

**STEP 3**
- add data management costs to your research application; coordinate resourcing and costing with your institution, research office and institutional IT services
Can you understand/use these data?

SrvMthdDraft.doc
SrvMthdFinal.doc
SrvMthdLastOne.doc
SrvMthdRealVersion.doc
Formatting for high quality data:

- use consistent templates for the same kind of data
- are well organised – folders
- are suitably named
- are properly versioned
- have the authenticity of master files identified
  - tools available for versioning and synchronising files
File formats

• Choice of software format for digital data:
  • hardware used e.g. audio capture
  • discipline-specific customs and planned data analyses
  • software availability/cost

• Digital data endangered software/ hardware obsolesce

• Best formats for long-term preservation are standard, interchangeable and open formats
  • tab-delimited, comma-delimited (CSV), ASCII
  • SPSS portable, XML
  • RTF, OpenDocument format, PDF/A,
  • See Recommended formats

• Beware of errors/losses of data when converting!
**Format conversion**

MS Excel (.XLSX) format using colour highlighting for annotation

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Timber volumes in m³</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Logging private land</td>
<td>20346.345</td>
<td>47005.223</td>
<td>26001.754</td>
<td>11468.897</td>
</tr>
<tr>
<td>5</td>
<td>Logging forest reserves</td>
<td>4060.567</td>
<td>1777.783</td>
<td>804.997</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>Logging state land</td>
<td>0.000</td>
<td>1200.000</td>
<td>559.162</td>
<td>2077.567</td>
</tr>
<tr>
<td>7</td>
<td>Total</td>
<td>61119.912</td>
<td>87065.006</td>
<td>64802.913</td>
<td>51354.464</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td><strong>Data missing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td><strong>Estimate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>¹ temporary volumes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tab-delimited text format, and loss of colour annotation
Organising data

- Plan in advance how best to organise data
- Use a logical structure and ensure collaborators understand

Examples

- hierarchical structure of files, grouped in folders, e.g. audio, transcripts and annotated transcripts
- survey data: spreadsheet, SPSS, relational database
- interview transcripts: individual well-named files
Recommendations for audio transcription

- adopt a uniform layout throughout the research project

- compatibility with import features of Computer Assisted Qualitative Data Analysis Software (CAQDAS)

- role of transcription varies by discipline. What to transcribe?
  - verbal and non-verbal?
  - turn-taking?
  - ‘interruptions’

- who does it – researcher, service? Need rules

- implications of technologies – video, multiple camera, screen capture, webcams
Documenting data - what is useful?

Any useful documentation such as:
- final report, published reports, user guide, working paper, publications, lab books, questionnaires, consent forms

Information on dataset structure
- inventory of data files
- relationships between those files
- records, cases…

Variable-level documentation
- labels, codes, classifications
- missing values
- derivations and aggregations
- in social science we use the Data Documentation Initiative (DDI)
Example documentation quantitative study

- Single user guide or many documents presented separately, e.g. Understanding Society study
Example data list qualitative study

- Data listing provides an at-a-glance summary of interview sets

<table>
<thead>
<tr>
<th>Respondent ID</th>
<th>Population Group</th>
<th>Date of Birth</th>
<th>Gender</th>
<th>Occupation</th>
<th>Interview summary</th>
<th>Place of Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM02</td>
<td>Group 6: Animal / Human Health Professionals</td>
<td>1975</td>
<td>M</td>
<td>Veterinary Surgeon</td>
<td>Family and background, career and work arrangements during FMD epidemic and perceptions of situation</td>
<td>North Cumbria, respective home</td>
</tr>
<tr>
<td>PM03</td>
<td>Group 6: Animal / Human Health Professionals</td>
<td>1990</td>
<td>F</td>
<td>Veterinary Surgeon</td>
<td>Family and background, career and work arrangements during FMD epidemic and perceptions of situation</td>
<td>North Cumbria</td>
</tr>
<tr>
<td>PM07</td>
<td>Group 6: Animal / Human Health Professionals</td>
<td>1994</td>
<td>F</td>
<td>Veterinary practice manager</td>
<td>Family and background, career and work arrangements during FMD epidemic and perceptions of situation</td>
<td>North Cumbria, respective home</td>
</tr>
</tbody>
</table>
Example documentation qualitative study

A user guide could contain a variety of documents that provide context: interview schedule, transcription notes, even photos.
Storing data safely

- Looking after research data - protecting them from damage and loss
- Strategies in place for:
  - backing-up
  - transmission
  - secure storage
  - disposal

**News Technology**

8 December 2010 Last updated at 11:43

**Nasa sells shuttle PCs without wiping secret data**

US space agency Nasa has been left red-faced after selling off computers without ensuring that highly sensitive data had been removed.

An internal investigation found 10 cases where PCs were sold despite failing data removal procedures.

Another four PCs - which were about to be sold - were found to contain data restricted under arms control rules.
Stuff happens: hard drive failure

Digital back-up strategy

Consider

- what’s backed-up? - all, some, just the bits you change?
- where? - original copy, external local and remote copies
- what media? - CD, DVD, external hard drive, tape, etc.
- how often? – assess frequency and automate the process
- for how long is it kept? Data retention policies that might apply?
- verify and recover - never assume, regularly test a restore

Backing-up need not be expensive

- 1Tb external drives are around £50, with back-up software

Consider non-digital storage too!
Encryption and security

Always encrypt personal or sensitive data:
• when moving data files
• when or storing files

Free softwares that are easy to use:
• encrypt hard drives, partitions, files and folders
• encrypt portable storage devices such as USB flash drives
• Safehouse, Truecrypt, Axcrypt

Protect data from unauthorised access, change, disclosure, destruction
• control access to all computers devices
• control physical access to buildings, rooms, cabinets
• restrict access to sensitive materials e.g. consent forms
Storage services

• Online or ‘cloud’ services increasingly popular
  • Google Drive, DropBox, Microsoft SkyDrive etc.
  • Accessible anywhere
  • Background syncing
  • Mirror files
  • Mobile apps available
  • Very convenient

• Everyone uses them, and that’s ok BUT precautions must be taken..
• Consider if appropriate, as services can be hosted outside the EU (DPA for personal data)
• Encrypt anything sensitive or avoid services altogether
Other storage options

- Your university or department may have options available e.g.
  - Secure backed up storage space, VRE
  - VPN giving access to external researchers
  - Locally managed Dropbox-like services such as ownCloud and ZendTo
  - Secure file transfer protocol (FTP) server

- Data repository or archive
  - A repository acts as more of a ‘final destination’ for data
  - Many universities have data repositories now catering to its researchers
  - UK Data Service has its own recently launched service called ‘ReShare’, for social science data of any kind
Data destruction software

- **BCWipe** - uses ‘military-grade procedures to surgically remove all traces of any file’
  - Can be applied to entire disk drives

- **AxCrypt** - free open source file and folder shredding
  - Integrates into Windows well, useful for single files
  - If in doubt, physically destroy the drive using an approved secure destruction facility
  - Physically destroy portable media, as you would shred paper

*Note: take caution when installing AxCrypt to ensure that you uncheck options to install extra ‘bundled’ software*
Disposal

Proper disposal of equipment and media
- even reformatting the hard drive is **not** sufficient
- special software available
Quizzes

Some light-hearted questions covering:

• Formatting and organising. Exercise 5.2 p. 74

• Storage, backing up and transferring data. Exercise 6.3. p. 100
Contact

Collections Development and Producer Relations team
UK Data Service
University of Essex
ukdataservice.ac.uk/help/get-in-touch.aspx