Data Management Basics Webinar

The webinar will begin at 15:00

- You now have a menu in the top right corner of your screen.

- The red button with a white arrow allows you to expand and contract the webinar menu, in which you can write questions / comments.

- We will answer your questions at the end.

- If we do not get to a question, we will reply later by email.

- You will be on mute throughout the webinar – we need to do this in order to ensure a high quality recording.

UK Data Service

Data Management Basics

Dr Scott Summers

UK Data Service

University of Essex

Webinar
6th December 2018

UK Data Service

Overview of this session

- UK Data Service

- Managing your data – background, why and how
  - The GDPR
  - Consent
  - Anonymisation
  - Access controls
  - Documentation
  - Security
  - Encryption
  - Backups

- Your questions

UK Data Service

Data Management at the UK Data Service

- Support and training for data creators with accessing, managing, and using data

- One-stop-shop for social science data

https://discover.ukdataservice.ac.uk/

- More webinars (and events) available

https://www.ukdataservice.ac.uk/news-and-events

Some statistics about the UK Data Service

- **7,277** datasets in the collection

- **1034** qualitative and mixed methods collections

- **400** new datasets added each year

- **219** case studies of data reuse

- **25,000** registered users

- **60,000** downloads worldwide per year

- **4000+** user support queries per year

UK Data Service

Background

- Data sharing is fast becoming a new paradigm in research across all disciplines, providing benefits to individual researchers, institutions, funders and more

- Good research data management habits are essential to creating data that are suitable for sharing and reuse

- Many funders and academic publishers now specify requirements for data handling, including the formulation of a data management plan

UK Data Service

Why is it important to manage research data well?

- Data creation in research is often expensive

- Data is the cornerstone of research

- Good quality data leads to good quality research

- Data underpins published findings

- Enables compliance with ethical codes, data protection laws, journal requirements and funder policies

- To protect data from loss, destruction and potential exposure

UK Data Service

Practical steps researchers can take

- Write a data management or a sharing plan

- Make sure data are shareable and can be understood:
  - Obtain consent to share
  - Do not disclose identities without consent
  - Use open and standard formats
  - Provide context and documentation
  - Protect your data at all stages

UK Data Service

ESRC data management plan

Assessment of existing data

Information on new data

Quality assurance of data

Backup and security of data

Difficulties in data sharing and measures to overcome these

Consent, anonymisation, re-use strategies

Copyright / Intellectual Property Ownership

Responsibilities

Management and curation

ESRC DMP guidance

UK Data Service

Multiple tools for protecting participants

1. Seek informed consent, also for data sharing and long-term preservation and curation

2. Protect identities e.g. anonymisation, and (or) not collecting personal data for admin

3. Regulate access where needed (all or part of data) e.g. by group, use or time period

UK Data Service

The GDPR – basics

- The General Data Protection Regulation (GDPR) is the new EU-wide data protection regulation

- Modernises and unifies European data protection rules

- Creates new rights and strengthens and enhances previous rights

- Is about **clarity**, **transparency** and **accountability**

UK Data Service

The GDPR – basics

- Applies to personal data, pseudonymised data and living persons only

- Personal data are 'any information relating to an identified or identifiable natural person'

- Note that not all research data is personal data

- Also note there may still be ethical reasons for wanting to protect this information though!

UK Data Service

The GDPR – processing grounds

- There are **six** grounds for the processing of personal data, and one of these must be present in order to process a data subject's personal data:

1. Consent
2. Contract
3. Legal obligation
4. Vital interests
5. Public interest (public task)
6. Legitimate interest

UK Data Service

Consent for sharing – one more small step

- Engagement in the research process
  - What activities are involved in participating in the project?

- Dissemination in presentations, publications, the web
  - Consent for use of quotes for articles and video publicity

- Data sharing and archiving
  - Consider future uses of data

Consent is *always* dependent on the research context – special cases of covert research and verbal consent

UK Data Service

In practice: wording in consent forms / information sheets

Break down into **three** key areas:

I.   Taking part in the study;

II.   Use of the information in the study;

III.   Future use and reuse of the information by others.

In practice: wording in consent forms / information sheets

3. Future use and reuse of the information by others

I give permission for the [specify the data] that I provide to be deposited in [name of data repository] ☐ ☐
so it can be used for future research and learning.

> Specify in which form the data will be deposited, e.g. anonymised transcripts, audio recording, survey database, etc.; and if needed repeat the statement for each form of data you plan to deposit.
>
> Specify whether deposited data will be anonymised, and how. Make sure to describe this in detail in the information sheet.
>
> Specify whether use or access restrictions will apply to the data in future, e.g. exclude commercial use, apply safeguarded access, etc.; and discuss these restrictions with the repository in advance.

We expect to use your contributed information in various outputs, including a report and content for a website. Extracts of interviews and some photographs may both be used. We will get your permission before using a quote from you or a photograph of you.

After the project has ended, we intend to archive the interviews at …. Then the interview data can be disseminated for reuse by other researchers, for research and learning purposes.

UK Data Service

Anonymising quantitative data - tips

- Remove direct identifiers

  e.g. names, address, institution and photos

- Reduce the precision / detail of a variable through aggregation

  e.g. birth year instead of date of birth; occupational categories rather than job; and, area rather than village

- Generalise meaning of detailed text variable

  e.g. occupational expertise

- Restrict upper lower ranges of a variable to hide outliers

  e.g. income and age

- Combining variables

  e.g. creating non-disclosive rural / urban variable from place variables

UK Data Service

Anonymising qualitative data

- Plan or apply editing at time of transcription
  Except: longitudinal studies - anonymise when data collection complete (linkages)

- Avoid blanking out; use pseudonyms or replacements

- Avoid over-anonymising – removing / aggregating information in text can distort data or make it misleading
  (https://www.ukdataservice.ac.uk/deposit-data/stories/gush)

- Consistency within research team and throughout project

- Identify replacements, e.g. with [brackets]

- Keep an anonymisation log of all replacements, aggregations or removals made and keep it *separate* from anonymised data files

UK Data Service

Audio-visual data

- Digital manipulation of audio and image files can remove personal identifiers

    e.g. voice alteration and image blurring (e.g. of faces)

- Labour intensive, expensive, may damage research potential of data

- Better alternatives:

    - Obtain consent to use and share data unaltered for research purposes

    - Avoid mentioning disclosing information during audio recordings

UK Data Service

In practice: example anonymisation

Ex 1. Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003 (study 5407 in UK Data Archive collection) by M. Mort, Lancaster University, Institute for Health Research.

Date of Interview: 21/02/02

Interview with Lucas Roberts, DEFRA field officer

Date of birth: 2 May 1965

Gender: Male

Occupation: Frontline worker

Location: Plumpton, North Cumbria

Comment [v1]: Replace: Ken

Comment [v2]: delete

Comment [v3]: delete

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 abut the tape.

Comment [v4]: Replace: Ken

Comment [v5]: Replace: Ken

Comment [v6]: Replace: Ken

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.

UK Data Service

Managing access to data

Open
- available for download / online access under open licence without any registration

Safeguarded
- available for download / online access to logged-in users who have registered and agreed to an End User Licence (*e.g. not identify any potentially identifiable individuals*)
- special agreements (depositor permission; approved researcher)
- embargo for fixed time period

Controlled
- available for remote or safe room access to authorised and authenticated users whose research proposal has been vetted and who have received training

In practice: data with access conditions

Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003 (study 5407 in UK Data Archive collection) by M. Mort, Lancaster University, Institute for Health Research.

- Interviews (audio and transcript) and written diaries with 54 people
- 40 interview and diary transcripts are archived and available for re-use by registered users (Safeguarded)
- 3 interviews and 5 diaries were embargoed until 2015 (Safeguarded – Embargoed)
- Audio files archived and only available by permission from researchers (Safeguarded – Special Agreement)

discover.ukdataservice.ac.uk/catalogue/?sn=5407
doc.ukdataservice.ac.uk/doc/5407/mrdoc/pdf/q5407userguide.pdf

UK Data Service

Documenting your data

- Enables you to understand data when you return to it!
- Sufficient information for future researchers to understand and use the data

- If using your data for the first time, what would a new user need to know to make sense of it?

- The UK Data Archive uses data documentation to:
  - Supplement a data collection with documents such as a user guide(s) and data listing
  - Ensure accurate processing and archiving
  - Create a catalogue record for a published data collection

UK Data Service

Include as documentation

- Data collection methodology and processes: sampling, sampling size, fieldwork protocol and interviewer instructions

- Information sheet / consent form

- Questionnaire, showcards and question lists

- Transcripts: header with context information: date and place interview, interviewee name, etc.

- Data list: overview of key information about each interview, as 'at-a-glance' summary of the data collection

- Links to reports and publications

UK Data Service

Data-level documentation: variable names

- All structured, tabular data should have cases / records and variables adequately documented with names, labels and descriptions

- Variable names might include:
  - Question number system related to questions in a survey / questionnaire *e.g. Q1a, Q1b, Q2, Q3a*
  - Numerical order system *e.g. V1, V2, V3*
  - Meaningful abbreviations or combinations of abbreviations referring to meaning of the variable

    e.g. 'oz%=percentage ozone', 'GOR=Government Office Region', 'moocc=mother occupation', 'faocc=father occupation'
  - For interoperability across platforms - variable names should be max 8 characters and without spaces

UK Data Service

Data-level documentation: variable labels

- Similar principles for variable labels:
  - Be brief, maximum 80 characters
  - Include unit of measurement where applicable
  - Reference the question number of a survey or questionnaire

    e.g. variable 'q11hexw' with label 'Q11: hours spent taking physical exercise in a typical week' - the label gives the unit of measurement and a reference to the question number (Q11b)

- Codes of, and reasons for, missing data
  - Avoid blanks, system-missing or '0' values

    e.g. '99=not recorded', '98=not provided (no answer)', '97=not applicable', '96=not known', '95=error'

- Coding or classification schemes used, with a bibliographic ref

    e.g. Standard Occupational Classification 2000; ISO 3166 alpha-2 country codes

UK Data Service

Embedded data-level metadata in SPSS file

In practice: user guide and documentation

- A user guide could contain a variety of documents that provide context: interview schedule, transcription notes, even photos
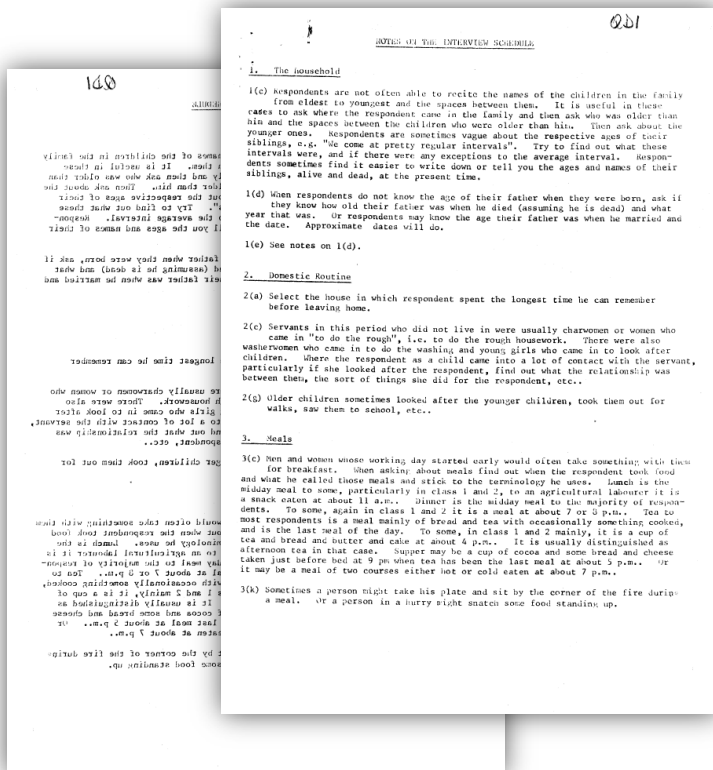
In practice: data list

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

Study Number 5407
Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001
Mort, M.

The panel respondents for the study were divided into six population groups. The data list for the diary and interviews has been colour-coded accordingly for clarity, using the depositor's original colours:

Group 1: Farmers	Group 2: Rural Business	Group 3: Agricultural related occupations	Group 4: Frontline Workers	Group 5: Community	Group 6: Animal / Human Health Professionals

1. Interviews

Respondent ID	Population Group	Date of Birth	Gender	Occupation	Interview summary	Place of Interview
PM02	Group 6: Animal / Human Health Professionals	1975	M	Veterinary Surgeon	Family and background, career and work, arrangements during FMD epidemic and perceptions of situation	North Cumbria, respo home
PM03	Group 6: Animal / Human Health Professionals	1966	F	Veterinary Surgeon	Family and background, career and work, arrangements during FMD epidemic and perceptions of situation	North Cumbria
PM07	Group 6: Animal / Human Health Professionals	1964	F	Veterinary practice manager	Family and background, career and work, arrangements during FMD epidemic and perceptions of situation	North Cumbria, respo home
					Family and background, career and work,	

UK Data Service

Transcription template

Should:

- Possess a unique identifier
- Adopt a uniform layout throughout the research project
- Make use of speaker tags - turn-taking
- Carry line breaks
- Be page numbered
- Carry a document header giving brief details of the interview: date, place, interviewer name, interviewee details, etc.

Other considerations:

- Cover page
- Compatibility with import features of Computer Assisted Qualitative Data Analysis Software (CAQDAS)

UK Data Service

In practice: transcript format

Study Name: Interview number:
Depositor: Interview ID: Firstname Lastname
Interviewer: Date of interview:

Information about interviewee
Date of birth: Marital status:
Gender: Occupation:
Geographic region:

Y=Interviewee

I=Interviewer

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?

File formats

Choice of software format for digital data:

- Planned data analyses

- Software availability / cost

- Hardware used – e.g. audio capture

- Discipline-specific standards and customs

Digital data is software dependent, so endangered by obsolescence of software / hardware

Best formats for long-term preservation:

- Standard, interchangeable and open

- *e.g. tab-delimited, comma-delimited (CSV), ASCII, RTF, PDF/A, OpenDocument format and XML*

- UK Data Service optimal file formats for various data types

- Digital Preservation Coalition guidance on preservation formats

UK Data Service

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

UK Data Service
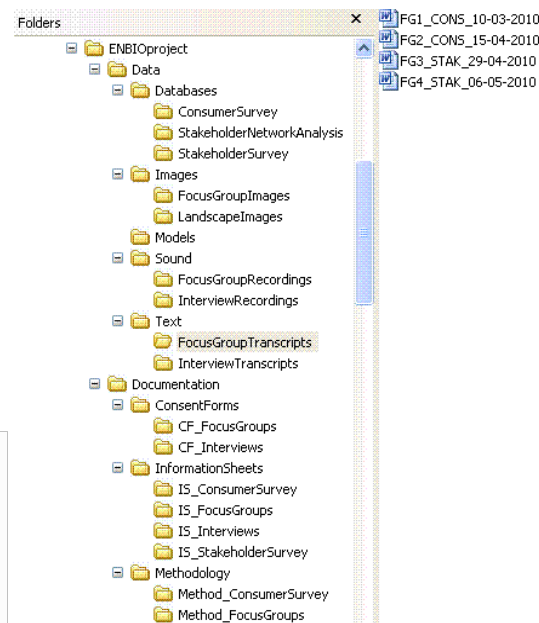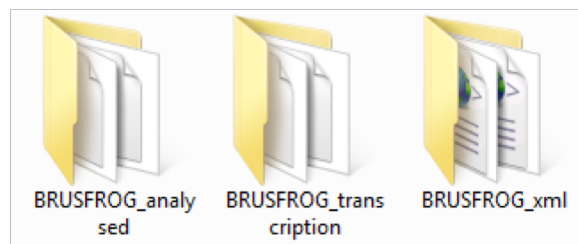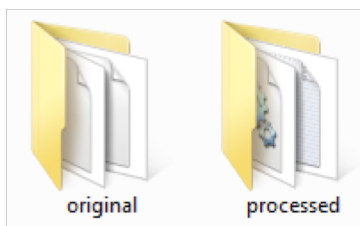
Data security and storage

Protect data from unauthorised:

- Access
- Use
- Change
- Disclosure
- Destruction

Who knows who is watching, listening or attempting to access your data…

UK Data Service

Data security strategy

- Control access to computers:
  - Use passwords and lock your machine when away from it
  - Run up-to-date anti-virus and firewall protection
  - Power surge protection
  - Utilise encryption
    - on all devices: desktops, laptops, memory sticks and mobile devices
    - at all locations: work, home and travel
  - Restrict access to sensitive materials e.g. consent forms and patient records
  - Personal data need more protection – always keep them separate and secure

- Control physical access to buildings, rooms and filing cabinets

- Properly dispose of data and equipment once the project is finished

UK Data Service

Encryption software

Encryption software can be easy to use and enables users to:

- Encrypt hard drives, partitions, files and folders
- Encrypt portable storage devices such as USB flash drives

[VeraCrypt](#)

[BitLocker](#)

[Axcrypt](#)

[FileVault2](#)

UK Data Service

Digital back-up strategy

Consider
- **What's backed-up?** - all, some or just the bits you change?
- **Where?** - original copy, external local and remote copies
- **What media?** - DVD, external hard drive, USB, Cloud?
- **How often?** - hourly, daily, weekly? Automate the process?
- **What method / software?** - duplicating, syncing or mirroring?
- **For how long is it kept?** - data retention policies that might apply?
- **Verify and recover** - never assume, regularly test and restore

Backing-up need not be expensive
- 2Tb external drives are around £50, with back-up software

Also consider non-digital storage too!

"We back up our data on sticky notes because sticky notes never crash."

UK Data Service

File sharing and collaborative environments

Sharing data between researchers
- Too often sent as insecure email attachments

Other options:
- Virtual Research Environments
  - MS SharePoint
- Locally managed; ownCloud and ZendTo
- File transfer protocol (FTP)
- Physical media
- Cloud solutions
  - Google Drive, DropBox, Microsoft OneDrive and iCloud (insecure?)
  - Securer options? - Mega.nz, SpiderOak and Tresorit

- Assess risks of using cloud storage

ARE YOU SURE THIS IS HOW WE GET DATA INTO THE CLOUD?

By David Fletcher
http://www.cloudtweaks.com/2011/05/the-lighter-side-of-the-cloud-data-transfer/

UK Data Service

Data Disposal

Proper disposal of equipment and media
- Even reformatting a hard drive is *not* sufficient
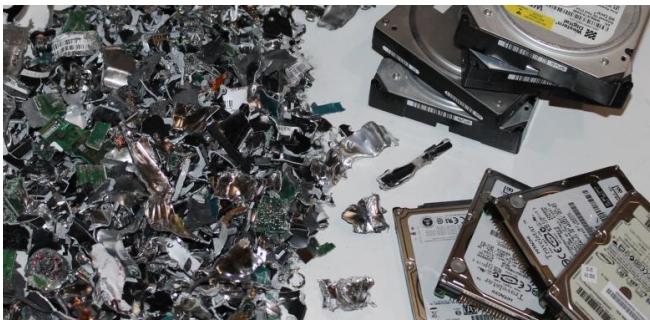- If in doubt, physically destroy the drive

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

UK Data Service

Our data management guidance

- Online best practice guidance: ukdataservice.ac.uk/manage-data.aspx

- Managing and Sharing Research Data – a Guide to Good Practice: (Sage Publications Ltd)

- Helpdesk for queries: ukdataservice.ac.uk/help/get-in-touch.aspx

- Training: www.ukdataservice.ac.uk/news-and-events/events
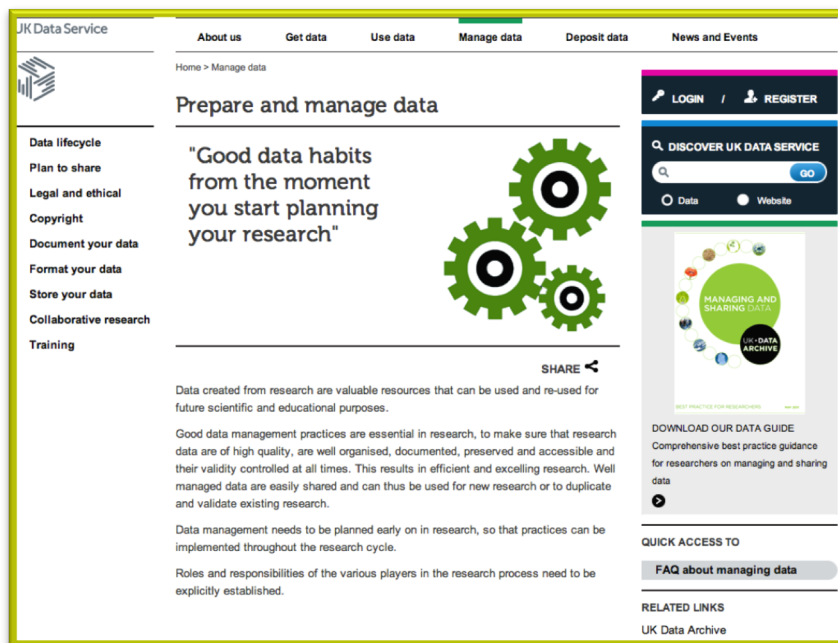
UK Data Service

Keep connected

- Subscribe to UK Data Service list:
  www.jiscmail.ac.uk/cgi-bin/webadmin?A0=UKDATASERVICE

- Follow UK Data Service on Twitter: @UKDataService

- Follow our RDM account on Twitter: @UKDSRDM

- Facebook

- Youtube: www.youtube.com/user/UKDATASERVICE

UK Data Service

Questions

Contact Details:

Scott Summers

UK Data Service

University of Essex

ukdataservice.ac.uk/help/get-in-touch