

# Data Management Basics: Introduction to data management and sharing

Anca Vlad

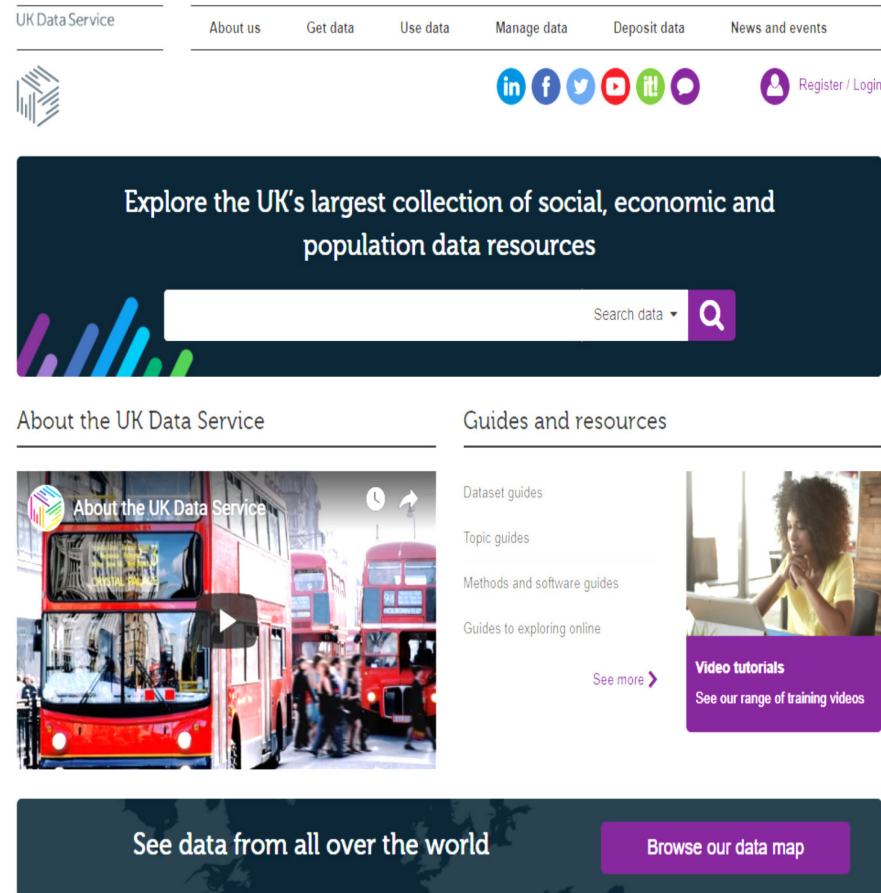


# Overview


- UK Data Service – About.
- Managing your data – background, why and how.
- Data Management Plans.
- Documentation.
- File formats.
- Data Security Strategy.
- Security, Encryption, Backups.
- Useful resources.
- Quiz.
- Questions.

# 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.
- Website, data catalogue:  
<https://www.ukdataservice.ac.uk/>
- More webinars available:  
<https://www.ukdataservice.ac.uk/news-and-events/events.aspx>




# 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 publishers now specify requirements for data handling, including the formulation of a data management plan.
- 
- A decorative purple gradient shape is located on the right side of the slide, starting from the middle and extending to the bottom right corner.



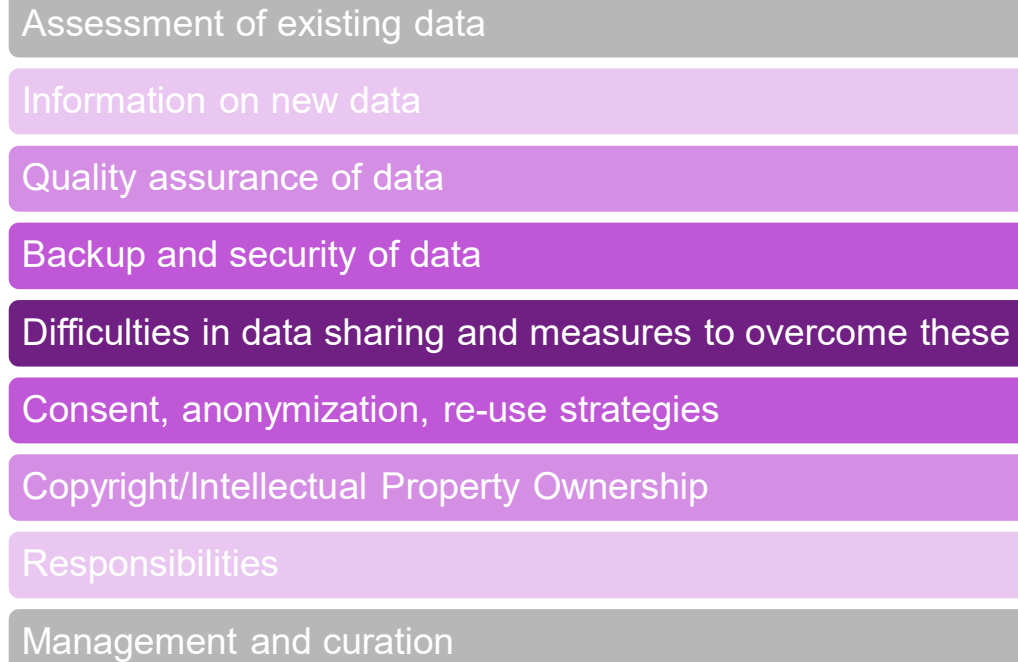
# Why is data management important?

- 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.
- 
- A decorative graphic element on the right side of the slide, consisting of a purple-to-white gradient shape that tapers to a point at the top right.

# Practical steps researchers can take

- Write a data management plan.
- Make sure data are shareable and can be understood:
  - What legal gateway will you use if you are collecting/using personal data.
  - If you use consent, ensure it covers data sharing.
  - Do not disclose identities without consent.
  - Use open source and standard formats.
  - Provide context and documentation.
  - Protect your data at all stages (secure storage, encryption).


# 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, anonymization, re-use strategies
- Copyright/Intellectual Property Ownership
- Responsibilities
- Management and curation

[Data management plan guidance](#)

# 3 Step Approach for Protecting Participants

1. Seek **informed consent**, also for data sharing and long-term preservation and curation.
  2. Protect identities e.g. **anonymization**, and (or) not collecting personal data (only collect data that is necessary).
  3. Regulate **access** where needed (all or part of data) e.g. by group, use or time period.
- 
- A decorative graphic element on the right side of the slide, consisting of a purple-to-white gradient shape that tapers to a point at the top right.

# Documenting your data

- Enables you to understand the data if/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 and research instruments.
  - Ensure accurate processing and archiving.
  - Create a catalogue record for a published data collection.

# What to include as documentation?

- Data collection methodology and processes: sampling, sample size, fieldwork protocol, experiment protocol, interviewer instructions.
- Codebook, user guide (for quantitative data).
- Information sheet, consent form (blank versions).
- Questionnaires, show cards, topic guides.
- Transcripts: header with context information: data and place of interview, interviewer, interviewee details (in line with consent form) etc.
- Data list: overview of key information about each interview, a map of the data collection (for qualitative data).
- Links to reports and publications (preferably DOIs where possible).



# Data-level documentation

- All structured, tabular data should have adequate variable names, variable and value labels.
- Variable names might include:
  - Question number system matching questions in the questionnaire used e.g. Q1a, Q1b, Q2, Q3b.
  - 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'.
  - For interoperability across platforms, variable names should not be longer than 8 characters and without spaces.

# Data-level documentation contd...

Similar principles for variable labels:

- Be brief and concise, maximum 80 characters.
- Include unit of measurement where appropriate.
- Reference the question number of a survey or questionnaire:  
e.g. variable 'q11hexw' with label 'Q11b: hours spent taking physical exercise in a typical week' – the label gives the unit of measurement and a reference to the questions number (Q11b).
- Coding or classification schemes used, with a bibliographic reference:  
e.g. Standard Occupational Classification 2000; ISO 3166 alpha-2 country codes.

For value labels:

- 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(skipped)', '96= not known', '95=error'.

# In practice: user guide and documentation

A user guide should contain variety of documents that provide context: interview schedule, methodology, study findings, consent procedures, transcription notes, codebook etc.

User guide for Mort, M. (2006). *Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003. [data collection]. UK Data Service. SN: 5407*  
<http://doi.org/10.5255/UKDA-SN-5407-1>.

# In practice: data list

Study Number 6377

Integrated Floodplain Management, 2006-2008

Morris, J.

## Floodplain farm survey

Interview ID	Farmer code	Age	Farm scheme	Farm type	Size of farm (hectare)	Number of holdings	Date of interview	Interviewer name	No of pages	Text file name	Audio file name
1	Be1	35-45	Beckingham	Beef	360	1	04.12.2006	Helena	28	6377int001	6377int001
2	Be2	45-55	Beckingham	Arable	364	1	05.12.2006	Helena	21	6377int002	6377int002
3	Be3	45-55	Beckingham	Arable	372	2	06.12.2006	Helena	22	6377int003	6377int003
4	Be4	45-55	Beckingham	Arable	194	3	06.12.2006	Helena	18	6377int004	6377int004
5	Be5	55-65	Beckingham	Arable	108	1	07.12.2007	Helena	21	6377int005	6377int005
6	Be6	45-55	Beckingham	Arable	1254	2	01.02.2008	Helena	19	6377int006	
7	Bu1	55-65	Bushley	Mixed	101	2	13.02.2007	Quentin	29	6377int007	6377int007
8	Bu2	>65	Bushley	Mixed	97	1	15.02.2007	Quentin	15	6377int008	6377int008
9	Bu3	>65	Bushley	Arable	194	4	13.02.2007	Quentin	21	6377int009	6377int009
10	Bu4	55-65	Bushley	Mixed	202	1	15.03.2007	Helena	19	6377int010	6377int010
11	Cu1	35-45	Cuddychurch	Dairy	64	1	08.05.2007	Helena	19	6377int011	6377int011
12	Cu2	55-65	Cuddychurch	Dairy	189	2	08.05.2007	Helena	18	6377int012	6377int012
13	Cu3	55-65	Cuddychurch	Mixed livestock	76	1	08.05.2007	Helena	13	6377int013	6377int013
14	Cu5	45-55	Cuddychurch	Mixed livestock	198	1	09.05.2007	Helena	24	6377int014	6377int014
15	Cu6	55-65	Cuddychurch	Dairy	89	1	09.05.2007	Helena	14	6377int015	6377int015
16	Cu7	>65	Cuddychurch	Mixed livestock	190	4	11.05.2007	Helena	20	6377int016	6377int016
17	Cu8	55-65	Cuddychurch	Mixed livestock	109	2	11.05.2007	Helena	22	6377int017	6377int017
18	Id1	55-65	Idle	Arable	158	3	07.02.2007	Quentin	17	6377int018	6377int018a
18	Id1	55-65	Idle	Arable	158	3	07.02.2007	Quentin	17	6377int018	6377int018b
19	Id1b	55-65	Idle	Arable	158	3		Quentin	22	6377int019	
20	Id2	45-55	Idle	Dairy	150	1	08.02.2007	Quentin	17	6377int020	6377int020

# 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: data, place, interviewer name, interviewee details, etc.

Other considerations:

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

# In practice: transcript format

Study Name:

Depositor:

Interview ID:

Date of Interview:

Information about interviewee:

*(e.g. Age, Gender, Occupation, Marital Status, Geographic region, etc. as relevant /appropriate)*

R= Respondent/Interviewee *(if more than one respondent, use R1, R2, etc.)*

I=Interviewer

R: I came here in late 1968.

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

R: 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?

R: 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?

R: Yes.

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

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

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

Model Interview Transcript:

<https://www.ukdataservice.ac.uk/media/622380/ukdamodeltranscript.pdf>



# 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:

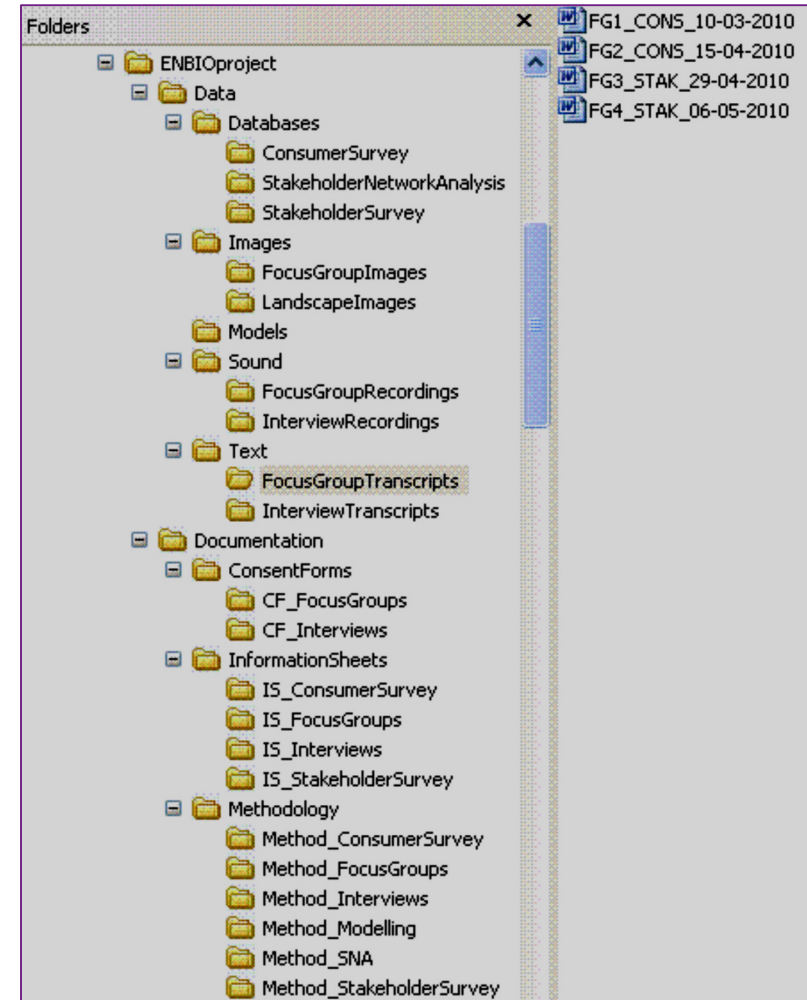
- **standard, interchangeable and open.**
- [UK Data Service optimal file formats](#) for various data types.
- [Digital Preservation Coalition](#) guidance on preservation formats.

# Organising data

- Plan in advance how to best organise data (project specific).
- 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.



# 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...



# 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.
  - Restrict access to sensitive materials e.g. consent forms and patient records.
  - Personal data need more protection – always keep them separate and secure.
  - Utilise encryption:
    - on all devices: desktops, laptops, memory sticks and mobile devices.
    - at all locations: work, home and travel.
- Control physical access to buildings, rooms and filing cabinets.
- Properly dispose of data and equipment once the project is finished.

# 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](#)



# Video tutorials

- VeraCrypt: <https://www.youtube.com/watch?v=Ogm9QHQPfQU>
- AxCrypt: <https://www.youtube.com/watch?v=ACcRInsoYZg>
- FileVault 2: <https://www.youtube.com/watch?v=JIZ9EFMS0ic>
- BitLocker: <https://www.youtube.com/watch?v=y4losu-Yfsw>
- Time Machine: <https://www.youtube.com/watch?v=hlsQaVj7WtA>
- MD5summer: <https://www.youtube.com/watch?v=VcBfkB6N7-k>



# 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."

# 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?)
    - More secure options? [Mega.nz](https://mega.nz) [SpiderOak](https://spideroak.com) [Tresorit](https://tresorit.com)
- Assess risks of using cloud storage.



# 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;



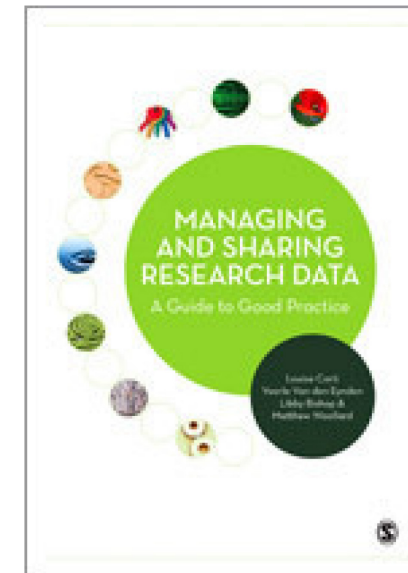
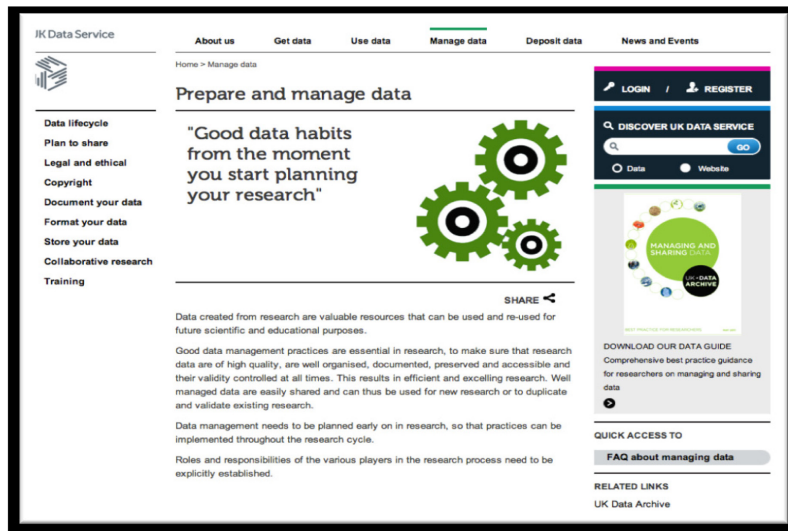
**AxCrypt** - free open source file and folder shredding.

- Integrates into Windows well, useful for single files.



# UKDS data management guidance

- Best practice guidance: [www.ukdataservice.ac.uk/manage-data.aspx](http://www.ukdataservice.ac.uk/manage-data.aspx)
- [CESSDA Data Management Expert Guide](#)
- Managing and Sharing Research Data – a Guide to Good Practice (Sage Publications Ltd)
- Training: [www.ukdataservice.ac.uk/news-and-events/events](http://www.ukdataservice.ac.uk/news-and-events/events)
- Twitter: @UKDSRDM



# Tools and templates

- Model consent form and survey consent statement:  
<https://ukdataservice.ac.uk/learning-hub/research-data-management/ethical-issues/consent-for-data-sharing/>.
- Transcription template:  
<https://dam.ukdataservice.ac.uk/media/622380/ukdamodeltranscript.pdf>
- Transcription instructions:  
<https://ukdataservice.ac.uk/app/uploads/ukda-example-transcription-instructions.pdf>.
- Transcription confidentiality agreement:  
<https://dam.ukdataservice.ac.uk/media/622354/ukda-transcriber-confidentiality-agreement.pdf>.
- Data list template:  
[https://ukdataservice.ac.uk/uk\\_data\\_archive\\_data\\_listing\\_template](https://ukdataservice.ac.uk/uk_data_archive_data_listing_template).

# Further resources

- [Research data management](#) - UKDS Guidance.
- [Data Management Expert Guide](#) – CESSDA.
- [DMPOnline](#) - Digital Curation Centre.
- [Anonymising Research Data](#) - ESRC National Centre for Research Methods, Working Paper 7/06.
- [Guide to Social Science Preparation and Archiving](#) from the Inter-University Consortium for Political and Social Research.
- [UKDS Data management costing tool](#).



# Get connected

<http://ukdataservice.ac.uk/about-us/contact.aspx>

<https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=UKdataservice>

<https://twitter.com/UKDataService> @UKDataService

<https://www.facebook.com/UKDataService>

<https://www.youtube.com/user/UKDATASERVICE>

\*\*\*Powerpoint slides will be available on our website in due course and you can catch up on the recording on our Youtube channel.  
Check out our Twitter for more updates.\*\*\*

# Thank you.

Anca Vlad  
[advlad@essex.ac.uk](mailto:advlad@essex.ac.uk)