

# Introduction to effective and practical research data management

Cristina Magder

Data Collections Development Manager



# Overview

- Research data lifecycle.
- FAIR data principles.
- Data management planning.
- Ethical and legal considerations.
- Data curation best practices.
- Data security, storage and backup.
- Data sharing strategies.
- Q&A session.

# Learning objectives

- Understand the basic concepts and importance of managing research data effectively throughout its lifecycle.
- Familiarise with key ethical and legal considerations when handling research data.
- Learn the fundamentals of making research data Findable, Accessible, Interoperable, and Reusable (FAIR) by following best practices for data curation processes, storage and dissemination.
- Understand appropriate data sharing strategies that respect ethical and legal considerations and enable FAIR data.

# Effective research data management

Effective research data management (RDM) practices ensures data are:

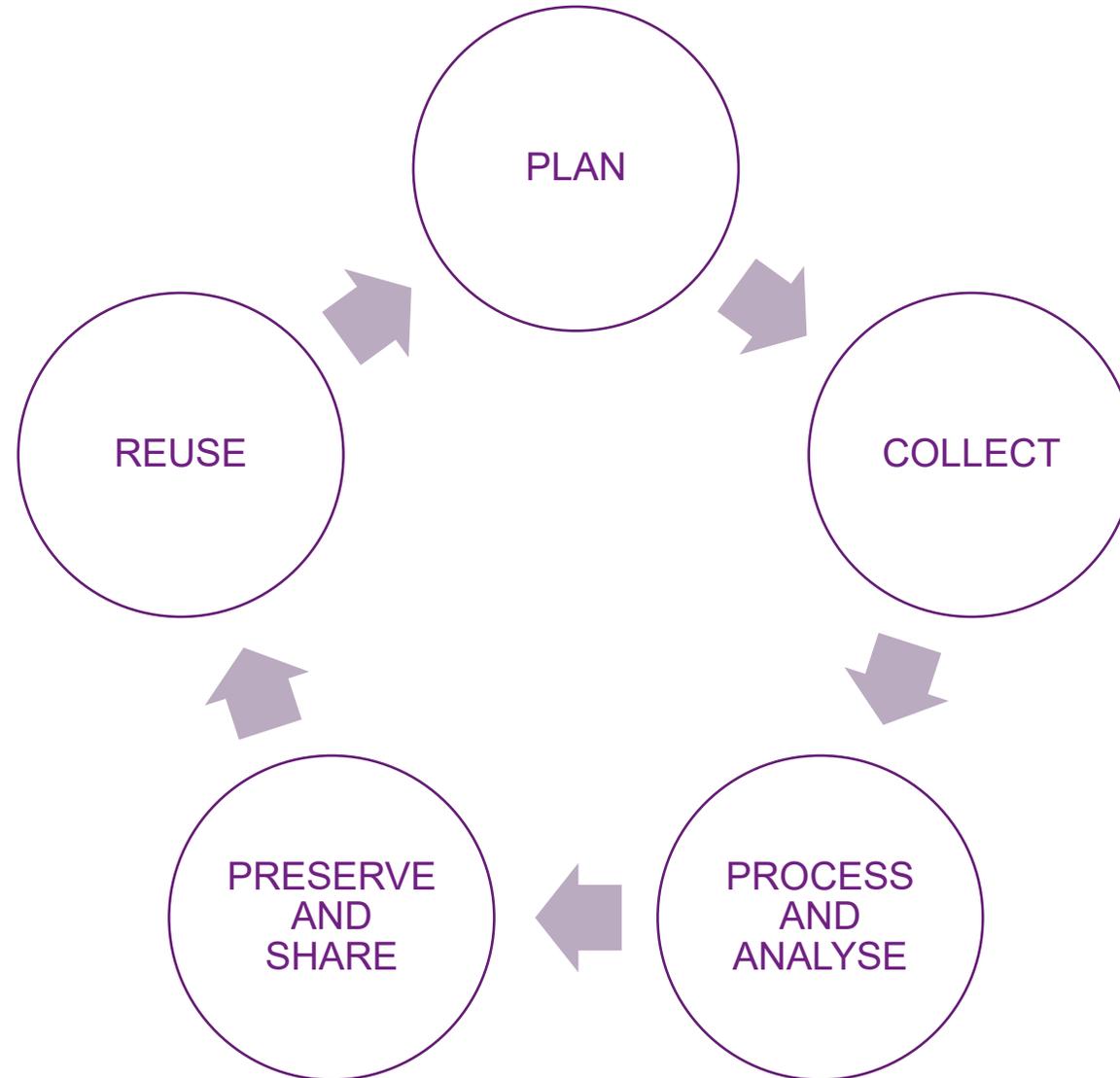
- compliant with ethical standards and applicable legislation.
- well-organised, quality controlled and well-documented
- safely stored, backed up, processed and analysed.
- responsibly archived and preserved, and appropriately shared for future reuse.

RDM practices safeguard the integrity of the research.

“UK Research and Innovation (UKRI) expects research data arising from its funding to be made as open as possible and as restricted as necessary. Good research data management practices should be followed throughout your project.”

[UKRI source](#)

# Research data lifecycle



# FAIR data principles

**F**indable

**A**ccessible

**I**nteroperable

**R**eusable

Published in 2016 in *Scientific Data*, FAIR principles are outlined as guidelines to help define what good data management looks like, emphasising machine-actionability.

Find out more about the [FAIR initiative](#).

# Making data FAIR: Findable

## Findable

- F1. (Meta)data are assigned a globally unique and persistent identifier.
- F2. Data are described with rich metadata.
- F3. Metadata clearly and explicitly include the identifier of the data they describe.
- F4. (Meta)data are registered or indexed in a searchable resource.

# Making data FAIR: Accessible

## Accessible

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol.
- A1.1 The protocol is open, free, and universally implementable.
- A1.2 The protocol allows for an authentication and authorisation procedure, where necessary.
- A2. Metadata are accessible, even when the data are no longer available.

# Making data FAIR: Interoperable

## Interoperable

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles.
- I3. (Meta)data include qualified references to other (meta)data.

# Making data FAIR: Reusable

## Reusable

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes.
- R1.1. (Meta)data are released with a clear and accessible data usage license.
- R1.2. (Meta)data are associated with detailed provenance.
- R1.3. (Meta)data meet domain-relevant community standards.

# How to ensure FAIR data?

- Share data using a responsible repository.
- Create rich metadata and documentation.
- Clearly define access procedures.
- Use standardise vocabularies, ontologies and schemas.
- Adopt open and standard formats for both data and metadata.

# Success stories

Using [Understanding Society](#), researchers looked at the way engaging with arts, culture and sports can [lead to greater satisfaction with life](#).

Studies like the [English Longitudinal Study of Ageing](#) can give insight into how [enjoying later life can be linked to living longer](#) and the impact of [social isolation and loneliness on mortality in older people](#).

The [Millennium Cohort Study](#) was used to [investigate the conditions associated with parental involvement with children](#) for policy recommendations.

[Family Resources Survey](#) to [analyse migrants' experiences of poverty and to compare them with the experiences of UK-born people](#).

# Data Management Plans (DMPs) overview



# DMP topics overview

From a generalised perspective DMPs should cover:

1. Data description: new and existing data.
2. Ethical and legal considerations and compliance.
3. Curation of data: organising, formatting, and documenting.
4. Data security, storage and backup.
5. Data sharing strategies.
6. Responsibilities and resources.

We provide in-depth guidance on our website for the [ESRC DMP](#).

Always remember to check funder requirements and that a DMP is a living document, as research evolves the DMP should be reviewed and updated as necessary.

# Why is data management planning essential?

- Anticipate and prepare.
- Keep on track.
- Secure necessary resources.
- Think ahead about storage and safeguard data.
- Share FAIR data and ensure reproducibility.
- Meet funders' expectations.

# DMPonline

DMPonline - web-based platform developed by the Digital Curation Centre to help researchers create, review, and share DMPs that meet institutional and funder requirements.

The screenshot shows the DMPonline web interface. At the top is an orange navigation bar with the DMPonline logo, 'My Dashboard', 'Create plans', 'Reference', and 'Help' menus. Below the navigation bar is a grey header area with 'University of Essex'. The main content area features a light blue info box with a message about SSO login upgrades. Below this is the 'Create a new plan' section, which includes a form for entering research project details, selecting a primary research organisation (University of Essex is selected), and selecting a primary funding organisation. The form has 'Create plan' and 'Cancel' buttons at the bottom.

**Info:**  
As part of our routine maintenance, we have upgraded our SSO login to enhance security.

- If your account was not linked to your institutional credentials, please log in as normal.
- If your account was linked to your institutional credentials, you will now need to re-link your account as part of this upgrade. To do this, please log in using your **DMPonline** email and password.
- Next, go to **Edit profile** > scroll down to the point **Institutional credentials**, and select the **Link your institutional credentials** option.
- After re-linking your account, you need to refresh your browser to complete the process. Remember to save your updated settings.

### Create a new plan

Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.

\* What research project are you planning?

mock project for testing, practice, or educational purposes

\* Select the primary research organisation

**Organisation**  
 - or -  No research organisation associated with this plan or my research organisation is not listed

\* Select the primary funding organisation

**Funder**  
 - or -  No funder associated with this plan or my funder is not listed

# Ethical and legal considerations overview

# Ethical considerations

- Maximise benefits and minimize risks.
- Voluntary participation and informed consent.
- Respect individual rights and dignity.
- Integrity and transparency.
- Clear responsibilities and independence.

# Legal considerations

- Lawful data handling.
- Data minimisation and anonymisation.
- Secure storage and access controls.
- Data retention and disposal.
- Data sharing protocols.
- Fairness, transparency and accountability.
- Intellectual property rights.

# How do ethical and legal considerations in data management affect the integrity and impact of research?

- Protect participants.
- Maintain and build trust.
- Enable sharing and collaboration.
- Ensure long-term impact.
- Enable appropriate risk management.

# Data curation: formatting and organising



# File formats strategy

When choosing the file format consider:

- What format is best suited for **data creation**?
- What format is best suited for **data analyses and other planned uses**?
- What format is best suited for **long-term sustainability and sharing of data**?
- Should you choose an **open versus a proprietary format**?
- Should the format be **lossy or not**?
- Is the format suitable for **conversion**?

# Recommended file formats

## Type of data

### Quantitative tabular data with extensive metadata.

A dataset with variable labels, code labels, and defined missing values, in addition to the matrix of data.

### Quantitative tabular data with minimal metadata.

A matrix of data with or without column headings or variable names, but no other metadata or labeling.

## Recommended formats

Proprietary formats of statistical packages e.g. SPSS (.sav), Stata (.dta), .sas7bdat.

Delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) containing metadata information.

Some structured text or mark-up file containing metadata information, e.g. DDI XML file.

Comma-separated values (CSV) file (.csv).

Tab-delimited file (.tab).

Including delimited text of given character set with SQL data definition statements where appropriate.

[UK Data Service recommended formats](#)

# Best practices for version control

Data management processes inevitably result in a number of edits to the data and documentation.

- Identify milestone versions to keep.
- Uniquely identify different versions.
- Record changes.
- Record relationships between items.
- Track the location of files.
- Regularly synchronise files.
- Identify a single location for the storage of milestone and master versions.

# File naming strategy

Common elements that should be considered when developing a file naming strategy are:

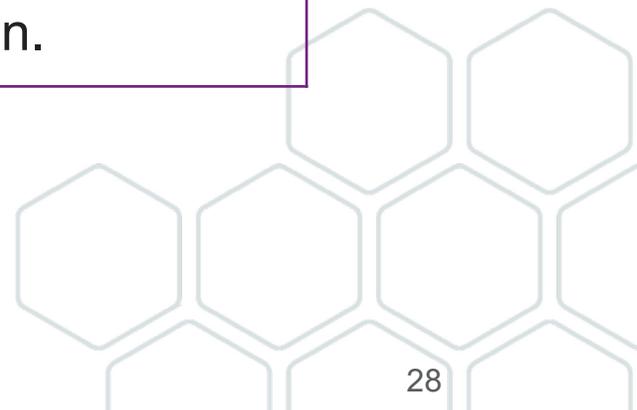
- Version number
- Date of creation
- Initials of creator
- Description of content
- Publication date
- Project number.

# Best practices for file naming

- Create meaningful but brief names.
- Use file names to classify types of files.
- Use dates in the format YYYY-MM-DD.
- Avoid capitalisation or 'camel case' where possible, as some computer platforms may be case-sensitive (e.g. Unix).
- Avoid using spaces, dots and special characters (& or ? or !).
- Use hyphens (-) or underscores (\_) to separate elements in a file name.
- Reserve the 3-letter file extension for application-specific codes of file format (e.g. .docx, .xlsx, .mov, .tif).
- Include versioning within file names where appropriate.
- Review file names for archived versions to ensure they do not contain any confusing, irrelevant information (e.g. versioning, misleading description).

# File naming examples

File name	Meaning
pn6614_ukhls_wave2_e10_2024-04-15.sav	Project number 6614, Wave 2 of the UKHLS SPSS data 10 <sup>th</sup> edition, last edited on 15 April 2024.
pn018_int127_js_v1_2024-03-02.rtf	Project number 18, transcript of interview with participant 127, conducted by JS on 2 March 2024, first version.
shes_21_dataset_documentation_v2.pdf	Scottish Health Survey 2021 dataset documentation second version.



# Variable formatting and measurement levels

Incorrect variable formatting can lead to incorrect data use and sometimes even reduces the usability of the data.

A very simple but important is to **determine whether the data are to be treated as string or numeric.**

Numeric variables also need to be checked to ensure the measurement level is correctly defined. Note that different software name and treat measurement levels differently.

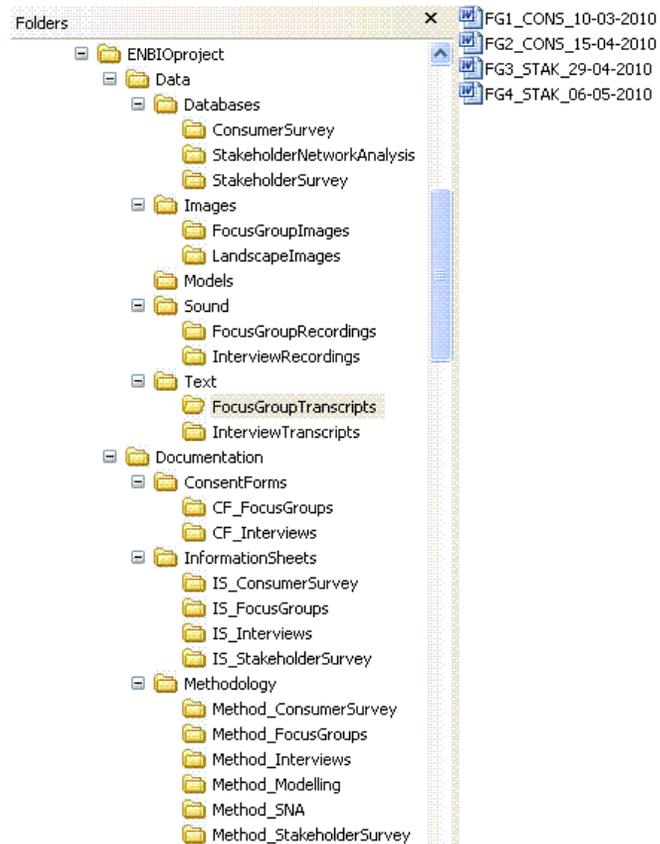
# Variable measurements examples

Variable	Stata	SPSS
<b>Ethnicity</b>	Categorical	Nominal
<b>Annual income (banded)</b>	Categorical	Ordinal
<b>Marital status</b>	Categorical	Nominal
<b>Age (banded)</b>	Categorical	Ordinal
<b>Monthly income (£)</b>	Continuous	Scale

STATA: Categorical, Continuous  
SPSS: Nominal, Ordinal, Scale



# Folder structure



It is important to think carefully how best to structure files in folders, in order to make it easy to locate and organise files and versions.

When working in collaboration with others, the need for an orderly structure is even greater.

The hierarchy for files will be project-specific but it is always useful to hold data and accompanying documentation in separate folders.

# Quality assurance

You must **check and document** any changes made to your data. This will provide a history, version control and provenance trail to help quality assure your data.

Quality assurance checks may include:

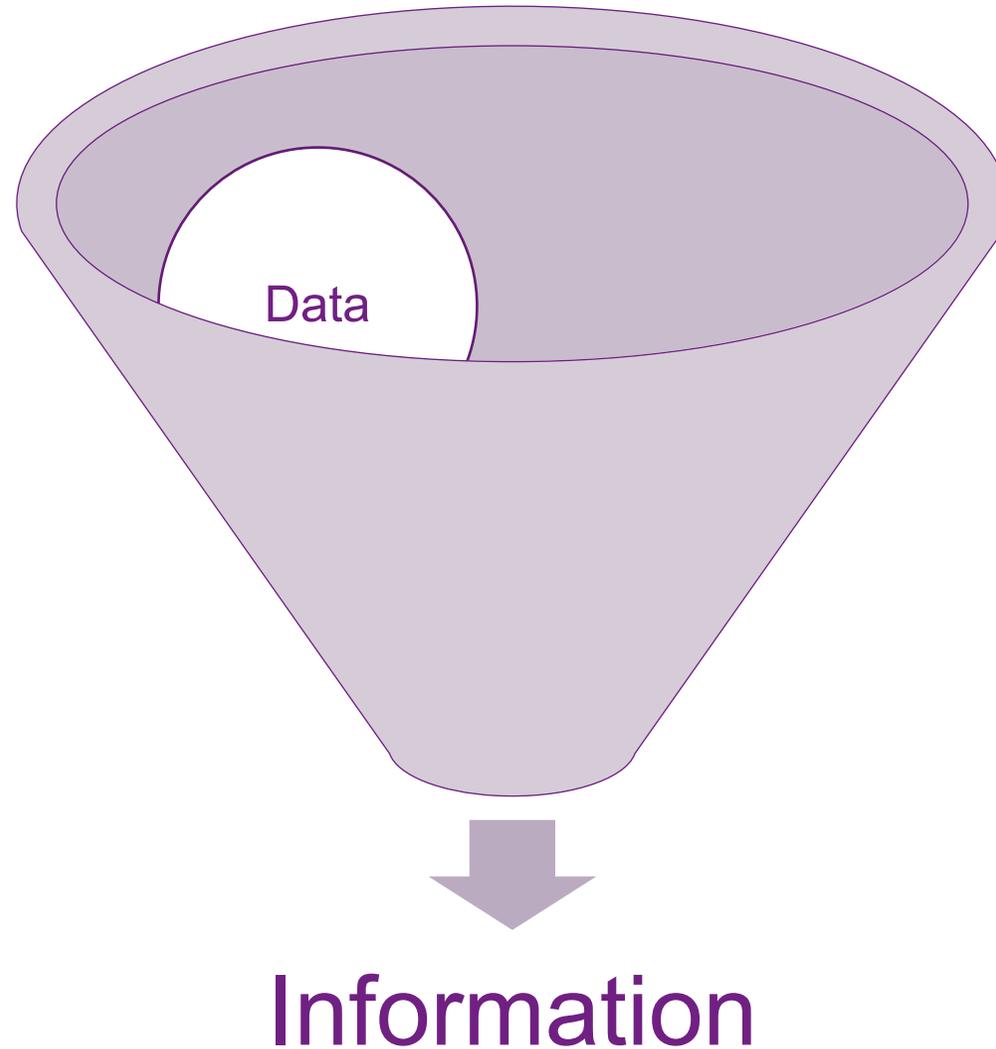
- Double-checking coding of observations or responses and out-of-range values.
- Checking data completeness.
- Adding variable and value labels where appropriate.
- Statistical analyses, such as frequencies, means, ranges or clustering to detect errors and anomalous values.
- For qualitative data, correcting errors made during transcription.

# Data curation: metadata and documentation

# Metadata

For data to become information, you need to understand the context in which the data are situated.

Metadata is what provides this essential context.



# Why is it essential to document data?

- Efficiency and accessibility.
- Ethical and accurate data reuse.
- Reproducibility and validation.
- Compliance and ethical standards.
- Long-term preservation.

# Types of data documentation

Data-level  
documentation

- provides information on the individual data objects, such as a variable in a data file or an interview transcript. It can be embedded in the data file, such as variable or value labels in a data file, or participant information added in the header or an interview transcript.

Study-level  
documentation

- provides high-level information on the research context and design, the data collection methods used, any data preparations and manipulations, plus summaries of findings based on the data.

# Data-level documentation: variable names

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: variable labels

Similar principles for variable labels:

- Be brief and concise, e.g. a maximum of 80 characters.
- Include unit of measurement where appropriate.
- Reference the question number of a survey or questionnaire:

e.g. variable 'q11bhexw' 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).

- Use coding or classification schemes e.g. Standard Occupational Classification 2010; ICD 10; NUTS.

The Office for National Statistics sets out a number of [classifications and standards](#) used in UK statistics and the Government Analysis Function department provides [harmonised standards and guidance](#) on a wide range of topic groups.

# Data-level documentation: value labels

For value labels:

- 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'.
- Be brief with a maximum of 120 characters. SPSS will truncate variable labels longer than 120 characters.
- Exclude non-ASCII characters that extend beyond the basic English alphabet and include symbols, accented letters, and characters from different languages.
- Use coding or classification schemes e.g. Standard Occupational Classification 2010; ICD 10; NUTS.

# Example for quantitative data

	Name	Type	Width	Decimals	Label	Values
1	serial_scr...	Numeric	8	2	Scrambled Individual Serial	None
2	BSA21_fina...	Numeric	8	2	British Social Attitudes Survey 2021 - final weight	{-9.00, Refu...
3	DVSex21	Numeric	3	0	What is your sex?	{-1, Not app...
4	Ragecat	Numeric	2	0	Age of respondent(grouped) <7 category> dv	{-1, Not app...
5	HHincomex	Numeric	8	2	What is the total income of your household from all sources before tax?	{-1.00, Not ...
6	RClassGP	Numeric	1	0	NS-SEC analytic classes (self-coded)	{-1, Not app...
7	hedqual2x	Numeric	8	2	Highest educational qualification attained	{1.00, Degr...
8	MarStat6x	Numeric	8	2	Marital Status	{1.00, Marri...
9	HhICHlGpdx	Numeric	8	2	Children in household (grouped)	{1.00, Yes}...

Value Labels

Spelling...

Value Labels:

Value	Label
1.00	Married/in a civil partnership/living with a partner
2.00	Separated/divorced/dissolved civil partnership
3.00	Widowed/surviving partner from a civil partnership
4.00	Single (never married/never in a civil partnership)

OK Reset Cancel Help

University of Manchester, Cathie Marsh  
 Institute for Social Research (CMIST), UK  
 Data Service. (2024). British Social Attitudes  
 Survey, 2021, Health Care and Equalities:  
 Open Access Teaching Dataset. [data  
 collection]. NatCen Social Research,  
 [original data producer(s)]. NatCen Social  
 Research. SN: 9236, DOI:  
<http://doi.org/10.5255/UKDA-SN-9236-1>

# Data-level documentation: Interview data

Can be embedded in the data file.

- Header or summary page.
- File properties.
- Within qualitative data analysis software.

Or be provided as a separate index file, known as a data list. This provides an overview of the whole data collection and allows relevant items to be easily identified.

# Example for interview data

Name / Interview ID	Year of Birth	Place of Birth	Gender	Interviewer	Place of Interview	Number of occasions interviewed	Date of Interview	Full interview/ Summary	No of Pages	Text File Name	Qualbank link to interview	YouTube Playlist of audio clips
Frank Bechhofer	1935	Germany	Male	Paul Thompson	Edinburgh	1	8 January 2001	Interview	78	6226int001	<a href="https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-a4e16904-7299-49f9-9b0a-4b601384a665">https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-a4e16904-7299-49f9-9b0a-4b601384a665</a>	<a href="https://youtu.be/UaKqnUC4IbA">https://youtu.be/UaKqnUC4IbA</a>
								Highlights	24	6226themext001	<a href="https://discover.ukdataservice.ac.uk/assets/qualibank/6226themext001.pdf">https://discover.ukdataservice.ac.uk/assets/qualibank/6226themext001.pdf</a>	
								Summary	8	6226intsum001	<a href="https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-d011d67f-63a0-436f-8e35-67cf677c59ca">https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-d011d67f-63a0-436f-8e35-67cf677c59ca</a>	
Daniel Bertaux	1941	France	Male	Paul Thompson	France	1	20 August 2002	Interview	82	6226int002	<a href="https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-d9af5999609749c59fa01b594f714397">https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-d9af5999609749c59fa01b594f714397</a>	<a href="https://youtu.be/dx4fz2zqzA">https://youtu.be/dx4fz2zqzA</a>
								Highlights	22	6226themext002	<a href="https://discover.ukdataservice.ac.uk/assets/qualibank/6226themext002.pdf">https://discover.ukdataservice.ac.uk/assets/qualibank/6226themext002.pdf</a>	
								Summary	14	6226intsum002	<a href="https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-2b777f6c915c422f867fe1c13683276b">https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-2b777f6c915c422f867fe1c13683276b</a>	
Mildred Blaxter	1925	Newcastle-on-Tyne	Female	Paul Thompson	Norwich	1	2 August 2002	Interview	57	6226int003	<a href="https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-1a09ab2d-e529-4629-9201-4a103644111c">https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-1a09ab2d-e529-4629-9201-4a103644111c</a>	<a href="https://www.youtube.com/watch?v=R6G5e15sG68&amp;list=PLaz84-Ixz3kiK5bOD5KBVAwe2eks_2GP">https://www.youtube.com/watch?v=R6G5e15sG68&amp;list=PLaz84-Ixz3kiK5bOD5KBVAwe2eks_2GP</a>
								Highlights	10	6226themext003	<a href="https://discover.ukdataservice.ac.uk/assets/qualibank/6226themext003.pdf">https://discover.ukdataservice.ac.uk/assets/qualibank/6226themext003.pdf</a>	
								Summary	6	6226intsum003	<a href="https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-5d1165f2-c369-4ecc-96ad-efca438a2077">https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-5d1165f2-c369-4ecc-96ad-efca438a2077</a>	
Pat Caplan	1942	Neston, Cheshire	Female	Paul Thompson	London	1	18 November 2009	Interview	66	6226int004	<a href="https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-51c8f358-2ce1-4012-b00f-87ce4bbe5a50">https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-51c8f358-2ce1-4012-b00f-87ce4bbe5a50</a>	<a href="https://www.youtube.com/watch?v=HUAfRpoWxEw&amp;list=PLaz84-Ixz3kiK5bOD5KBVAwe2eks_2GP">https://www.youtube.com/watch?v=HUAfRpoWxEw&amp;list=PLaz84-Ixz3kiK5bOD5KBVAwe2eks_2GP</a>
								Highlights	23	6226themext004	<a href="https://discover.ukdataservice.ac.uk/assets/qualibank/6226themext004.pdf">https://discover.ukdataservice.ac.uk/assets/qualibank/6226themext004.pdf</a>	
								Summary	9	6226intsum004	<a href="https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-84e8c183-29ce-481a-a5c6-b78f7ac92723">https://discover.ukdataservice.ac.uk/QualiBank/Document/?id=q-84e8c183-29ce-481a-a5c6-b78f7ac92723</a>	

Thompson, P. (2019). *Pioneers of Social Research, 1996-2018*. [data collection]. 4th Edition. UK Data Service. SN: 6226, DOI: <http://doi.org/10.5255/UKDA-SN-6226-6>

# Study-level 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).

# Study-level documentation example

## Scottish Crime and Justice Survey, 2019-2020

Details Documentation Resources [Access data](#)

### Documentation

Title	File name	Size (MB)
Scottish Crime and Justice Survey 2019-20: Disclosure Control Report	<a href="#">8799_scjs_ukda_disclosure_control_report_2019-20.pdf</a>	0.45
Scottish Crime and Justice Survey 2019-20: Questionnaire and User Notes	<a href="#">8799_scjs_questionnaire_and_user_notes_2019-20.pdf</a>	2.5
Scottish Crime and Justice Survey 2019-20: Technical Report	<a href="#">8799_scjs_technical_report_2019-20.pdf</a>	2.25
UK Data Archive Citation File for Study 8799	<a href="#">UKDA_Study_8799_Information.htm</a>	0
UK Data Archive Data Dictionaries	<a href="#">ukda_data_dictionaries.zip</a>	0.12
UK Data Archive ReadMe File for Study 8799	<a href="#">read8799.htm</a>	0

ScotCen Social Research. (2022). Scottish Crime and Justice Survey, 2019-2020. [data collection]. UK Data Service. SN: 8799, DOI: <http://doi.org/10.5255/UKDA-SN-8799-1>

# Metadata during the research data lifecycle

Metadata **should cover the whole research data cycle.**

- **Who created the data?** (Citation and information on data curators, data collectors, sponsors and contributors.)
- **What does the data file contain?** (Data format, file type, file structure, variable descriptions, missing data, weighting variables, software and keywords.)
- **When were the data created?** (Dates of fieldwork, time period covered, time dimension (e.g. cross-sectional or longitudinal).)
- **Where were the data created?** (Spatial units and geographical coverage.)
- **Why were the data created?** (Background to creation and planned use.)
- **How were the data created?** (Methodology of data collection, sampling and processing, data access information.)

**Structured metadata provides the answers to these questions in a structured format, which can be read by machines.**

# Structured metadata schemas examples

Metadata is produced and presented using standardised specifications. This again facilitates the re-use of data, data discovery, data access and the ability to share metadata between communities of researchers.

- **Dublin Core**: The first metadata standard for describing web content.
- **ISO 19115 -1:2014**: For describing geographic data and services.
- **Data Documentation Initiative (DDI)**: For describing the data in the social, behavioural, economic, and health sciences.
- **Statistical Data and Metadata eXchange (SDMX)**: An ISO standard designed to describe statistical data and metadata, normalise their exchange, and improve their efficient sharing across statistical and similar organisations.
- **DataCite metadata schema**: For the publication and citation of digital datasets with a persistent identifier

# Data Documentation Initiative (DDI)

DDI is a rich and detailed metadata standard, developed by the international [DDI Alliance](#).

It was originally designed for describing social, behavioural and economic sciences data. It is used by many social science organisations across the world.

There are several version of DDI.

- DDI Codebook (DDI-C) gathers basic information aimed at describing data collections and is generally used to compile catalogue records.
- DDI-Lifecycle (DDI-L) is a more complex schema, and contains additional elements used to describe survey questions and variables, as well as basic catalogue information.

# DDI metadata example 1

## Annual Population Survey, January - December, 2023

Details Documentation Resources [Access data](#)

### Details

Title:	Annual Population Survey, January - December, 2023
Alternative title:	APS
Study number (SN):	9248
Access:	These data are <a href="#">safeguarded</a>
Persistent identifier (DOI):	<a href="http://doi.org/10.5255/UKDA-SN-9248-1">10.5255/UKDA-SN-9248-1</a>
Series:	<a href="#">Annual Population Survey</a>
Data creator(s):	Office for National Statistics

Office for National Statistics. (2024). Annual Population Survey, January - December, 2023. [data collection]. UK Data Service. SN: 9248, DOI: <http://doi.org/10.5255/UKDA-SN-9248-1>

# DDI metadata example 2

## Coverage and methodology



Dates of fieldwork:	January 2023 - December 2023
Country:	United Kingdom
Spatial units:	Government Office Regions
Observation units:	Individuals
Observation unit location:	National
Population:	Persons resident in the UK in private households, and young people living away from the parental home in student halls of residence or similar institutions during term time.
Number of units:	146,419 cases
Method of data collection:	Face-to-face interview Telephone interview
Time dimensions:	Repeated cross-sectional study The survey has a panel element, in that the households at selected addresses are interviewed annually over four waves, in the same way as for the LFS boosts. They then leave the survey and are replaced by other households. The survey has a panel element, in that the households at selected addresses are interviewed annually over four waves, in the same way as for the LFS boosts. They then leave the survey and are replaced by other households.
Sampling procedures:	Multi-stage stratified random sample
Kind of data:	Numeric
Weighting:	Weighting used. See documentation for details

Office for National Statistics. (2024). Annual Population Survey, January - December, 2023. [data collection]. UK Data Service. SN: 9248, DOI: <http://doi.org/10.5255/UKDA-SN-9248-1>

# Data security, storage and backup

# Data security and storage

Data must be protected data from unauthorised:

- Access.
- Use.
- Change.
- Disclosure.
- Destruction.

# Data security strategy

- Control access to computers:
  - Use passphrases 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.
  - Always keep personal data separate, secure and encrypted.
  - 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.

# Digital back-up strategy

- Making backups of files is an essential element of research data management which ensures that original data files can be restored from backup copies, should they get damaged or go missing.

Three+ copies of the data, with at least one being stored offsite.

- Regular backups help protect against **accidental** or **malicious** data loss due to:
  - human error
  - hardware failure
  - software or media faults
  - virus infection or malicious hacking
  - power failure.

# Data disposal

- Simply deleting files and reformatting a hard drive will not securely erase information, meaning that it will still be possible to recover the data that was previously on the hard drive.
- Software is available to help erase files from hard disks, meeting recognised erasure standards. Example software is: [BCWipe](#), [WipeFile](#), [DeleteOnClick](#) and [Eraser](#) for Windows platforms; and [Permanent Eraser](#) for MacOS platforms.
- Shredders certified to an appropriate security level should be used for destroying paper and optical media.

# Data sharing strategies

# Data sharing strategies overview

There are various way of sharing research data including

- Domain specific repositories (data service providers, archives, centres).
- Institutional repositories.
- Self-preservation and dissemination.
- Commercial data sharing platforms.
- Direct submission with journal publications.

All data sharing methods have advantages and disadvantages.

# Key consideration for data sharing strategies

---

Purposes of data sharing

---

Data findability, accessibility, interoperability and reusability

---

Data sensitivity and confidentiality

---

Ethical and legal implications

---

Long-term preservation

---

Security measures

---

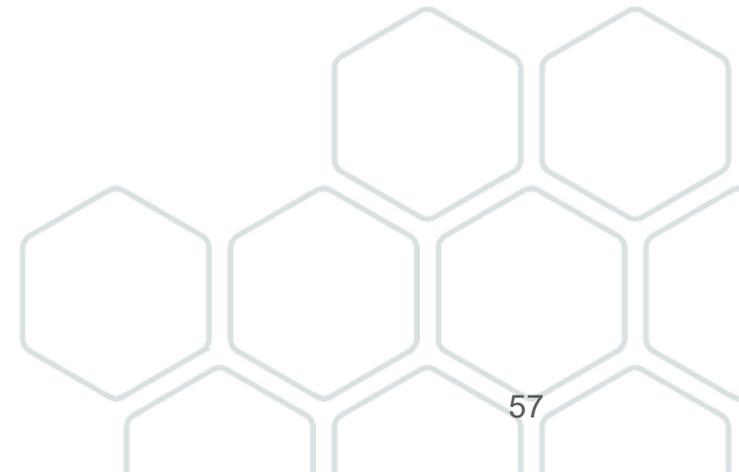
Costs and resources

---

Technology and infrastructure

---

Stakeholder engagement and support



# Responsible repositories

Responsible repositories are data service providers that host facilities that adhere to established standards and best practices in data management.

These ensure the integrity, preservation, and accessibility of the data they hold.

Responsible repositories play a critical role in the research ecosystem by providing reliable, secure, and accessible environments for storing and disseminating research data.



Source: [Research Data Alliance](#)  
[TRUST principles](#)

# Deposit licence agreements



Protect the rights of the data owner and the repository.



Ensure that data users are aware of their rights and responsibilities.



Facilitate ethical and legal sharing and use of data, enhancing its value to the research community.

# What if I use secondary data?

Always check the licence under which the data are made available.

While you might not be able to share derived data you can always share your code.

 code/syntax file are clean, well formatted and do not contain any data.

 avoid including unnecessary personal information.

 code/syntax file are well commented.

 always include the full citation (including the persistent identifier) for the data used.

 provide in-depth metadata describing the files and methods used.

 provide a ReadMe/Methods document for ease of use for secondary users.

# Registry of data repositories

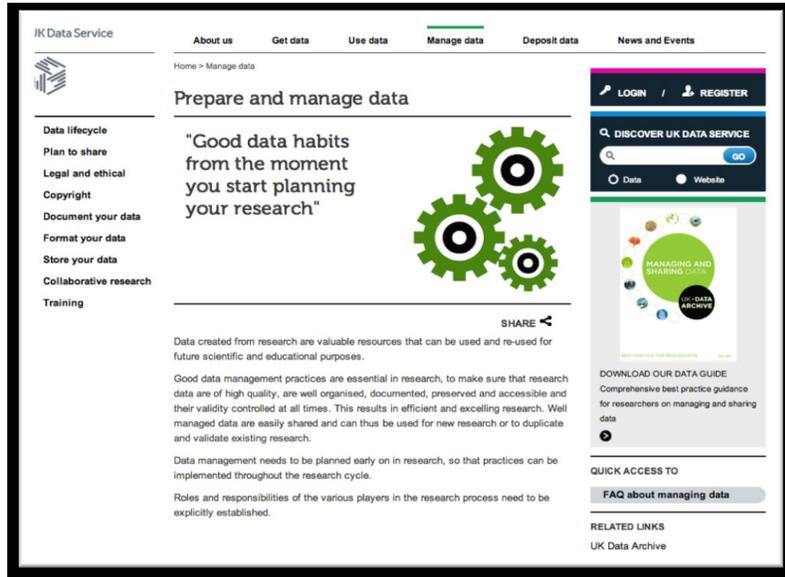
Re3data.org is a comprehensive registry of research data repositories that has been active for over a decade. It provides a curated index of more than 3,000 repositories worldwide, covering all academic disciplines.



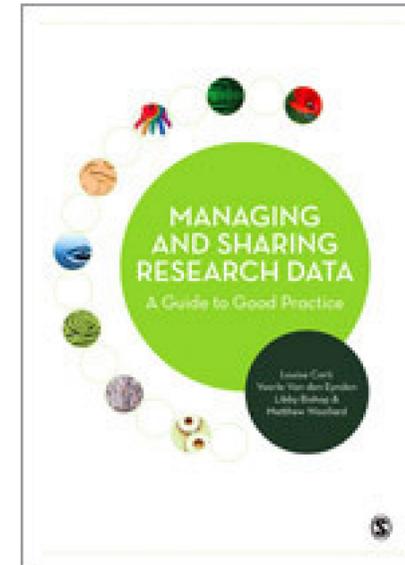
Source: [re3data](https://re3data.org)

# UKDS data management guidance

- [Best practice guidance](#).
- Managing and Sharing Research Data – a Guide to Good Practice (Sage Publications Ltd).
- [Training](#).
- Twitter: @UKDSRDM.



The screenshot displays the UK Data Service website interface. The main heading is "Prepare and manage data". A quote reads: "Good data habits from the moment you start planning your research". To the right of the quote are three interlocking green gears. Below the quote, there is a "SHARE" button and a paragraph of text: "Data created from research are valuable resources that can be used and re-used for future scientific and educational purposes." Further down, another paragraph states: "Good data management practices are essential in research, to make sure that research data are of high quality, are well organised, documented, preserved and accessible and their validity controlled at all times. This results in efficient and excellent research. Well managed data are easily shared and can thus be used for new research or to duplicate and validate existing research." A third paragraph follows: "Data management needs to be planned early on in research, so that practices can be implemented throughout the research cycle." The final paragraph reads: "Roles and responsibilities of the various players in the research process need to be explicitly established." On the right side of the page, there is a search bar with the text "DISCOVER UK DATA SERVICE", a "GO" button, and a "UK DATA ARCHIVE" button. Below the search bar, there is a "DOWNLOAD OUR DATA GUIDE" section with the text: "Comprehensive best practice guidance for researchers on managing and sharing data". At the bottom right, there is a "QUICK ACCESS TO" section with a button for "FAQ about managing data" and a "RELATED LINKS" section with a link to "UK Data Archive".



# Tools, templates and tutorials

- [Data management checklist](#)
- [Data management costing tool and checklist](#)
- [Model consent form and survey consent statement](#)
- [Transcription template](#) and [transcription instructions](#)
- [Data list template](#)
- [QAMyData](#)
- [Text anonymisation helper tool](#)
- [VeraCrypt](#)
- [AxCrypt](#)
- [BitLocker](#)
- [MD5summer.](#)

# Further resources

- [CESSDA Data Management Expert Guide](#)
- [CLOSER Training hub](#)
- [Guide to Social Science Preparation and Archiving](#) from the Inter-University Consortium for Political and Social Research.

# Get connected

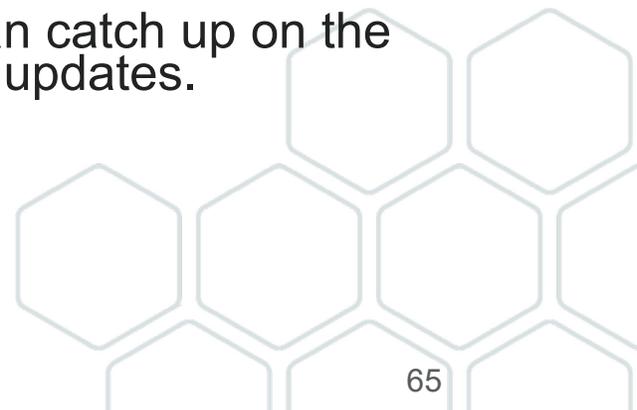
[UK Data Service](#)

[Jisc mail group](#)

[@UKDataService X \(formerly Twitter\)](#)

[UK Data Service YouTube channel](#)

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

`datasharing@ukdataservice.ac.uk`

<https://beta.ukdataservice.ac.uk/help>

