

# Data Management Planning for Social Sciences

UKDS Webinar

Cristina Magder & Hina Zahid

17 November 2022

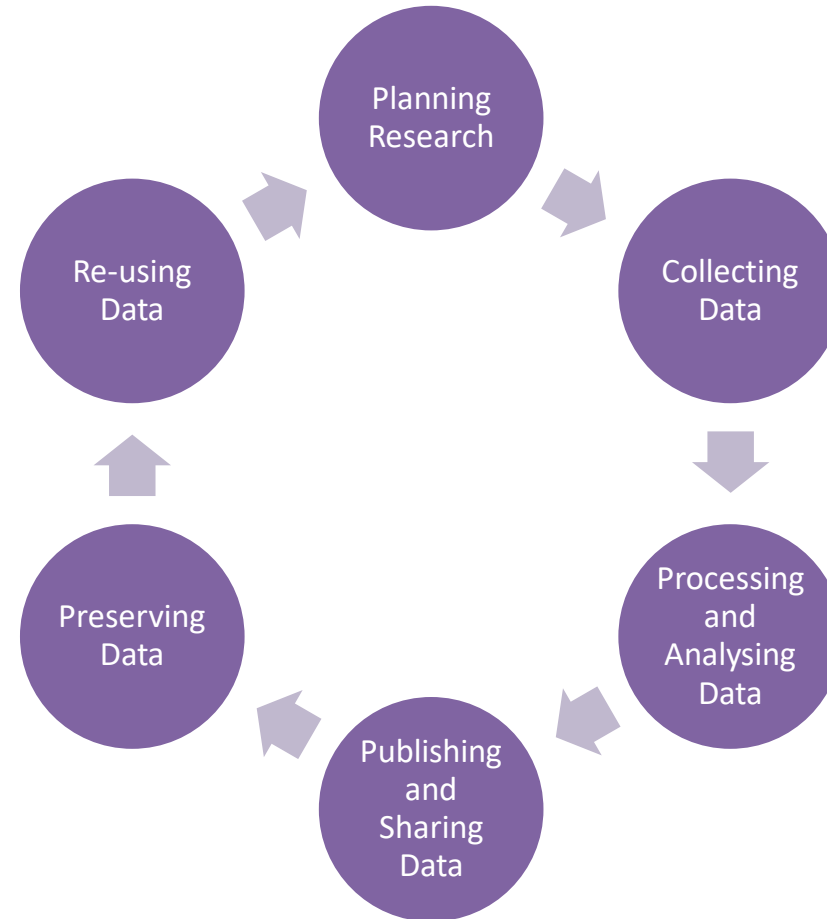


# Overview

- Data sharing and its' benefits.
- What is Data Management Planning (DMP) and why it is required?
- What is included in the Data Management Plan? Practical Example ESRC template.
- Resources, tools and templates.

# Data lifespan and data sharing

- data have a longer lifespan than the (research) project that creates them.
- data can be (re) used for various purposes when well managed, preserved and shared.
- who is benefiting from data sharing? researchers, public, funding bodies, scholarly community and research participants.

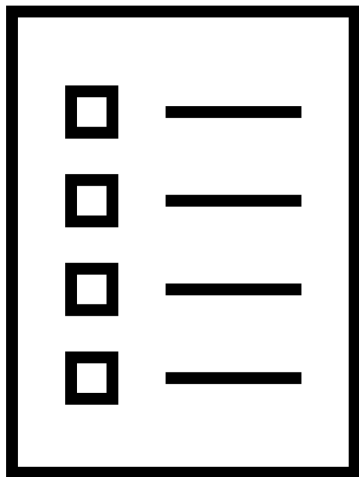


# Benefits of Sharing Data

Researchers	Funding Bodies	Public	Research Participants
increases visibility of scholarly work and increase citations (DOI)	make optimal use of publicly funded research	advance science to the benefit of society	allow maximum use of their contributed data/information
archiving provides long-term safe storage for data	avoid duplication of data collection	adoption of emerging norms – ‘open access’ publishing	minimise data collection on the hard-to-reach (e.g. ill, elites)
assists in implementing publishers’ data retention policies and enables collaborations	maximise return for investment	compliance with laws and regulations (transparency, openness, accountability)	enable participants’ experiences to be understood as widely as ethically possible

# Data Management Planning (DMP)

A data management plan (DMP) describes how the data will be



- Collected
- Organised
- Analysed
- Preserved
- Shared.

# Why we need data management planning?

- Plan ahead what to do with research data.
- Keep track of research data (e.g. staff leaving)
- Identify support, resources, budget, services needed.
- Plan storage, short & long-term.
- Plan security, ethical aspects.
- Plan long-term sharing.
- Shows accountability (funder, institution, partners)
- Makes data FAIRer: **F**indable, **A**ccessible, **I**nteroperable, **R**eusable.
- To fulfil funders requirement.

# FAIR Principles for publishing data

Findable

Accessible

Interoperable

Re-usable

Force 11 FAIR Principles

# Making Data FAIR

## Findable

- community-endorsed metadata (documentation) standards.
- persistent identifiers e.g. DOIs.

## Accessible

- availability statement (including restrictions)
- methods/tools to access the data.
- metadata preserved indefinitely.

## Interoperable

- standard vocabularies/ontologies.
- standard metadata schemas.

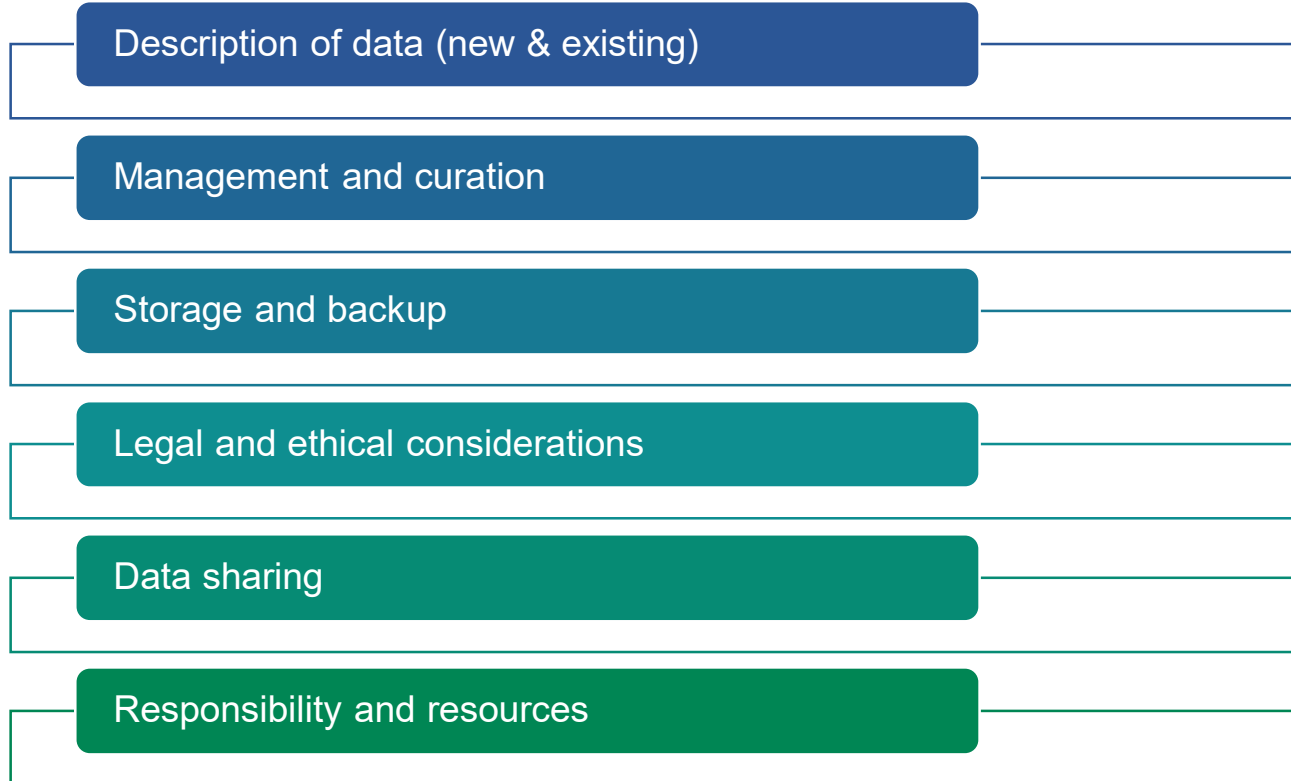
## Re-useable

- community-endorsed data licencing.
- established data quality assurance processes.
- long-term preservation.



# DMPs Structure

Data management plans can differ from one funder/institution to another - always consult your funder/host institution requirements



## Overview of UK research funders with data sharing policies and requirements for data sharing plans

Funder	DMP required?	Required at application	DMP topics
<a href="#">Arts and humanities research council (AHRC)</a>	Yes	Data Management Plan.	Data summary, data collection, short and long term data storage, data sharing, ethical and legal considerations.
<a href="#">Biotechnology and biological sciences research council (BBSRC)</a>	Yes	Data management and sharing plan.	Type, format, standards, sharing methods, restrictions, sharing timeframe.
<a href="#">Cancer research UK (CRUK)</a>	Yes	Data sharing plan.	Volume, format, standards, metadata, documentation, sharing method, timescale, preservation, restrictions.
<a href="#">Department for International development (DFID)</a>	Yes	Access and data management plan.	Repositories, limits, timescales, responsibilities, resources, access strategy.
<a href="#">Economic and social research council (ESRC)</a>	Yes	Data management plan.	Volume, type, quality, archiving plans, difficulties sharing, consent sharing, IPR, responsibilities.
<a href="#">Medical Research Council (MRC)</a>	Yes	Data management plan.	Collection, methods, documentation, preservation, standards, curation, security, confidentiality, sharing and access, timescale.
<a href="#">Natural environment research council (NERC)</a>	Yes	Outline data management plan.	Data management procedure, created data.
<a href="#">Welcome Trust</a>	Yes	Data management and sharing plan.	What data, when share, where share, how access, limits, how preserve, what resources.

# Example of a DMP

# ESRC DMP

## Following sections need to be completed:

1. Assessment of existing data.
2. Information on new data.
3. Quality assurance of data.
4. Backup and security of data.
5. Management and curation of data.
6. Difficulties in data sharing and measures to overcome these.
7. Consent, anonymization and re-use strategies.
8. Copyright/Intellectual property rights.
9. Responsibilities.
10. Preparation of data for sharing and archiving.

# Assessment of existing data

The section must provide:

- i) An explanation of the existing data sources that will be used by the research project, with references.
- ii) An analysis of the gaps identified between the currently available and required data for the research.

Consult UKRI [Gateway to Research](#) of past and present research grants and their outputs.

Example of FAIR Data Sources that can be consulted are:

- UK Data Service [Data Catalogue](#), with over 8,500 collections of key economic, social and historical data spanning many disciplines and themes.
- [CESSDA Data Catalogue](#) containing metadata for over 31,000 collections from European Archives.

# Information on new data

Provide structured information on the data that will be produced and/or accessed by the research project including:

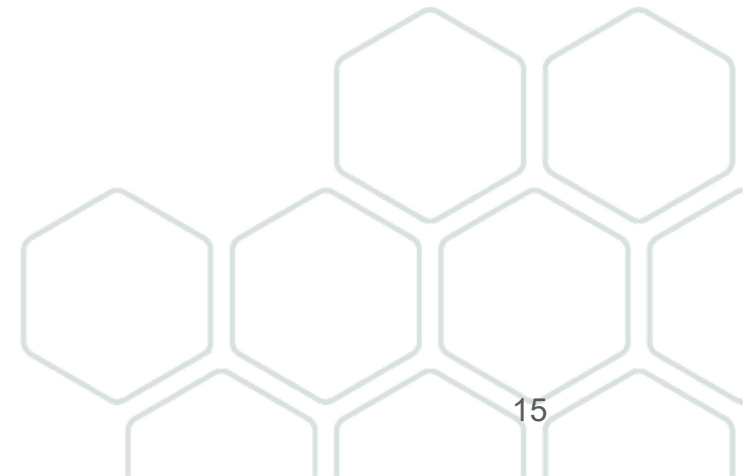
- data volume
- data type
- data quality and formats
- standards documentation and metadata
- methodologies for data collection and/or processing
- source and trustworthiness of third party data.

# Resources for information on new data

## Further information

### [Guidance on data formats](#)

- [Versioning](#)
- [Digitisation](#)
- [File formats](#)
- [Organising](#)
- [Quality](#)
- [Transcription](#)
- [Recommended formats](#)



# Quality assurance of data

This should include QA on data collected at the time of data collection, data entry, data checking and digitisation if applicable such as:

- Documenting the calibration of instruments.
- Taking duplicate samples or measurements.
- Standardised data capture, data entry or recording methods
- Data entry validation techniques including software used (e.g. [QAMyData](#)).
- Methods of transcription or methods of digitisation.
- Peer review of data.

[Please check our webpages for further information](#)



# Security and backup of data

Describe the data security and backup procedures that you will adopt to ensure the data and metadata are securely stored during the lifetime of the project

Always consider:

- Consulting your institution's policy on backups (policies vary from one institution to another so ensure information is up to date)
- If your data is sensitive (e.g. detailed personal data) you should discuss appropriate security measures which will be taken.
- Data security arrangements need to be proportionate to the nature of the data and the risks involved.

[Please check our webpages for further information](#)

# Management and curation of data

Outline your plans for i) preparing ii) organising and iii) documenting data always considering:

- Types of data e.g. primary, secondary, anonymised, pseudoanonymised etc.
- Metadata for data description, annotation and contextual information.
- Key documentation such as user guides, data dictionaries, interview schedules.
- Naming conventions for data files.
- Long-term preservation of data.

[Please check our webpages for further information](#)

# Study level documentation

## Scottish Crime and Justice Survey, 2018-2019

Details

Documentation

Resources

Access data

### Documentation

Title	File name	Size (MB)
Scottish Crime and Justice Survey 2018-19: Disclosure Control Report	<a href="#">8795_scjs_ukda_disclosure_control_report_2018-19.pdf</a>	0.45
Scottish Crime and Justice Survey 2018-19: Questionnaire and User Notes	<a href="#">8795_scjs_questionnaire_and_user_notes_2018-19.pdf</a>	2.44
Scottish Crime and Justice Survey 2018-19: Technical Report	<a href="#">8795_scjs_technical_report_2018-19.pdf</a>	2.12
UK Data Archive Citation File for Study 8795	<a href="#">UKDA_Study_8795_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 8795	<a href="#">read8795.htm</a>	0

# Data level documentation

	A	B	C	D	E	F	G
1	agrp	rsex	ethnicgrp	rnssecgp_6	adj_imd Quintile	rwcasual	snnolov
2	25-34	Male	White	Lower supervisory and technical occupations		3 Always wrong	Disagree
3	35-44	Female	White	Semi-routine and routine occupations		3 Not wrong at all	Agree
4	16-24	Female	White	Student in full-time education		4 Sometimes wrong	Agree
5	45-54	Female	Not white	Semi-routine and routine occupations	[most deprived]	Always wrong	Disagree strongly
6	25-34	Male	Not white	Managerial and prof occupations	[most deprived]	Always wrong	Disagree strongly
7	16-24	Female	Not white	Student in full-time education		4 Always wrong	Disagree strongly
8	25-34	Female	Not white	Lower supervisory and technical occupations	[most deprived]	Always wrong	Disagree strongly
9	16-24	Female	White	Semi-routine and routine occupations	[most deprived]	Mostly wrong	Agree
10	25-34	Female	White	Intermediate occupations	[most deprived]	Mostly wrong	Neither agree or disagree
11	35-44	Female	White	Semi-routine and routine occupations		3 Always wrong	Agree

<http://doi.org/10.5255/UKDA-SN-8735-1>

<http://doi.org/10.5255/UKDA-SN-6377-1>

**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	

UK Data Service

## Variable Information Log

**Introduction**  
 For datasets being deposited that include secondary data resources, researchers are advised to prepare a descriptive Variable Information Log describing these resources. The Variable Information Log should include the variable name, its source, how it was collected, a brief description, and any restrictions noted on its further use. (See the notes below)

**Notes**  
 These fields should be completed for the original data sources for each variable:

<b>Variable name:</b>	Provide a list of all the variables (name/number) used in the dataset.
<b>Variable label:</b>	A brief description necessary to identify the variable.
<b>Source:</b>	Source of the dataset/data owner or producer (e.g. World Bank data, IMF data, Penn World Tables data).
<b>Dataset version:</b>	Datasets keep evolving, so best practice is to indicate which version has been used.
<b>URL/DOI:</b>	Provide a persistent identifier or link of the source dataset used. Alternatively, if the data are not available online, provide a brief description of how they were obtained.

# Catalogue metadata

## Quarterly Labour Force Survey, January - March, 2021

Details	Documentation	Resources	Access data
---------	---------------	-----------	-------------

### Details

<b>Title:</b>	Quarterly Labour Force Survey, January - March, 2021
<b>Alternative title:</b>	QLFS
<b>Study number (SN):</b>	8806
<b>Access:</b>	These data are <a href="#">safeguarded</a>
<b>Persistent identifier (DOI):</b>	<a href="http://dx.doi.org/10.5255/UKDA-SN-8806-1">10.5255/UKDA-SN-8806-1</a>
<b>Series:</b>	<a href="#">Labour Force Survey</a>
<b>Principal investigator(s):</b>	Office for National Statistics, Social Survey Division Northern Ireland Statistics and Research Agency, Central Survey Unit

### Sponsors and contributors

### Citation and copyright

#### The citation for this study is:

Office for National Statistics, Social Survey Division, Northern Ireland Statistics and Research Agency, Central Survey Unit. (2021). *Quarterly Labour Force Survey, January - March, 2021*. [data collection]. UK Data Service. SN: 8806, <http://doi.org/10.5255/UKDA-SN-8806-1>

### Coverage and methodology

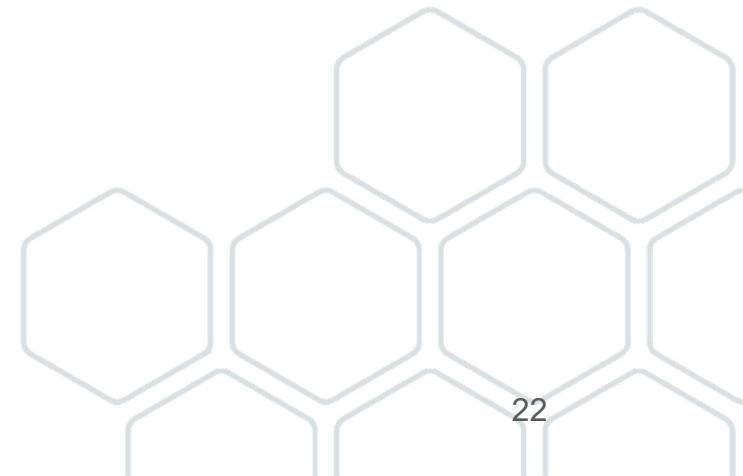
<b>Dates of fieldwork:</b>	January 2021 - March 2021
<b>Country:</b>	United Kingdom
<b>Spatial units:</b>	Government Office Regions Standard Statistical Regions
<b>Observation units:</b>	Individuals Families/households
<b>Observation unit location:</b>	National
<b>Population:</b>	All persons normally resident in private households in Great Britain and (from 1994) Northern Ireland.
<b>Number of units:</b>	82,015 cases
<b>Method of data collection:</b>	Face-to-face interview Telephone interview The first interview is conducted face-to-face, and subsequent interviews by telephone where possible.
<b>Time dimensions:</b>	Repeated cross-sectional study Data are collected quarterly
<b>Sampling procedures:</b>	Four sampling frames are used. See documentation for details.
<b>Weighting:</b>	Weighting used. See documentation for details

# Resources for documenting data

## Further information

### [Guidance on documenting data](#)

- [Data level documentation](#)
- [Study level documentation](#)
- [Metadata](#)
- [Other documentation resources](#)



# Difficulties in data sharing and measures to overcome these

- Identify potential obstacles & explain possible measures to overcome these.
- State explicitly which data may be difficult to share and why and if ethical issues could cause difficulties in data sharing, explain your strategies for dealing with these issues.
- The ESRC supports the position that most data can be curated and shared ethically provided researchers pay attention right from the planning stages of research to the following aspects:
  - when gaining informed consent, include consent for data sharing.
  - where needed, protect participants' identities by anonymising data.
  - address access restrictions to data in the data management and sharing plan, before commencing research.

# Consent, anonymisation and strategies to enable further re-use of data

Make explicit mention of the planned procedures to handle consent for data sharing for data obtained from human participants, and/or how to anonymise data, to make sure that data can be made available and accessible for future scientific research.

At UK Data Service we advise researchers to employ a three-pronged strategy:

- Consent.
- Anonymisation.
- Access Control.



# Consent

Consent is the process by which a researcher discloses appropriate information about the research so that a participant may make a voluntary, informed choice to accept or refuse to cooperate.

## Why to seek consent?

- Participants understand what they're signing up to making participation and research more effective .
- Research conducted is ethical.
- Compliance with data protection regulation .

# Three key areas to be addressed

Wording in consent forms and information sheets could be broken down in **three** key areas:

- Taking part in the study.
- Use of the information in the study.
- Future use and reuse of the information by other.

# UKDS Model Consent Form template

## Informed consent form

### Addressing future reuse of research data

This template can be used by researchers to gain informed consent to conduct research that collects data from people using questionnaires, observations, interviews, diaries, focus groups, video recordings, etc. It pays particular attention to ensure that research data can be curated and made available for future use, as well as addressing all standard requirements of a consent form. Also, this version of our template is consistent with our current knowledge of the requirements of the General Data Protection Regulation (GDPR) which comes into effect from 25 May 2018.

- This is *a template* to assist researchers in the design of their informed consent form. You must adapt this template to the requirements of your particular study, using the notes and suggestions provided.
- Before using this template, check whether your organisation provides a template consent form and if so, incorporate their requirements into the form (e.g. with regards to data protection).

Available at <https://ukdataservice.ac.uk/ukdamodelconsent/>

# Best practices

- Explicitly mention planned procedures to handle consent for data sharing.
- Ensure that consent procedures inform participants correctly about data sharing intentions.
- Do not preclude or unnecessarily limit sharing of data.
- Do not set a time limit on the retention of the data.
- Avoid stating that all data will be destroyed at the end of the project.
- Avoid statements such as 'data will not be shared outside of the project.'

[Guidance on consent and ethics](#)

# Duty of confidentiality and data sharing

- Exists in UK common law and may apply to research data.
- Disclosure of confidential information is lawful when:
  - the individual to whom the information relates has consented.
  - disclosure is necessary to safeguard the individual, or others, or is in the public interest.
  - there is a legal duty to do so, for example a court order.
- Best practice is to avoid very specific promises in consent forms.

# Data protection considerations

- If personal information about people is collected or used in research, then the data protection regulations applies.
- Data protection act 2018 (DPA), General data protection regulation 2018 (GDPR) & the UK GDPR 2021.
- DPA (2018) & the UK GDPR (2021) applies when
  - ✓ a researcher based in the UK collects personal data about people anywhere in the world.
  - ✓ a researcher outside the UK collects personal data on UK citizens.
- DPA (2018), EU GDPR (2018) & the UK GDPR (2021) applies when
  - ✓ a researcher based in the UK collects personal data about people across Europe.

# Anonymisation

- Anonymisation allows data to be shared, whilst preserving privacy. For example, identifiers are changed in some way, such as being removed, substituted, distorted, generalised or aggregated.
- Make explicit mention of how to anonymise data.
- If you are unsure of how issues of confidentiality are to be addressed to facilitate data sharing, please [get in touch](#) for advice.

[Guidance on anonymising data.](#)

# Access control

- Address access restrictions to data in the data management and sharing plan, before commencing research.
- Researcher should always consider if dissemination embargo is needed, e.g. for publication purposes.



# Licencing and access framework at UKDS

## Open

- No real disclosure risk. Under open licences; almost no restrictions on reuse.

## Safeguarded

- Zero to low real disclosure risk. Requires authentication and authorisation e.g. registered user and End User Licence Agreement.

## Controlled

- Real disclosure risk. Requires project approval, user vetting and training; access via a safe setting; output checking.

# Copyright and intellectual property ownership

- Copyright is an intellectual property right assigned automatically to the creator.
- Data owner (researcher) has copyright of research data & data archives do not hold copyright ownership.
- Compiled datasets contain original copyright – seek permission to archive when collecting.
- State who will own the copyright and IPR of any new data that you will generate.

# Best practices when using secondary data

Question to ask:

- Who the copyright holder of the datasets is?
  - Are you allowed to use them and in what way?
  - Are you allowed to archive and publish them in a data repository?
- 
- If not, you may need to seek for further permission to distribute material you do not own - copyright clearance.
  - If permission is not granted, need to remove copyrighted variables/material before publishing or sharing.

## Useful links

### Further information

- [Guidance on ethical issues](#)
- [Guidance on rights in data](#)
- [Guidance on Data Protection Act](#)
- [Guidance on other rights including Freedom of Information Act](#)

# Responsibilities

Indicate who within your research team will be responsible for data management, metadata production, dealing with quality issues and the final delivery of data for sharing or archiving

- If several people will be responsible, state their roles and responsibilities
- For collaborative projects explain the coordination of data management responsibilities across partners

[Further information on data management roles and responsibilities](#)

[Further information on how to cost data management](#)

# Preparation of data for sharing and archiving

Must consider:

- i) plans for preparing and documenting data for sharing and archiving are appropriate.
- ii) enough evidence was provided that the data will be well documented during research to enable high quality secondary research.

# Data management checklist (UKDS)

# Tools and templates

- Funder template for DMP
  - [ESRC DMP requirements in data policy](#) and [DMP guidance](#)
  - [MRC DMP guidance and template](#)
  - [AHRC DMP requirements](#)
  - [NERC DMP guidance and template](#)
- DCC's [DMPonline](#) tool





# Thank you.

UK Data Service

Email: [datasharing@essex.ac.uk](mailto:datasharing@essex.ac.uk)

Twitter: UKDataService

