UK Data Service

Data management basics: Ethical and legal issues in data sharing

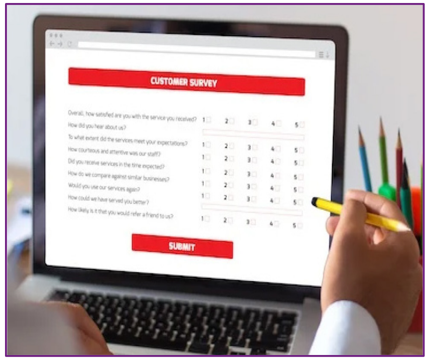
Presenter: Dr Hina Zahid

Senior Research Data Officer

Economic
and Social
Research Council

Research data

Primary data



Secondary data

Overview

- Key considerations in primary data use and sharing.
- Key considerations in secondary data use and sharing.

UK Data Service

Key considerations in primary data sharing and use

- Principles of ethical research.
- Duty of confidentiality.
- Data protection considerations.
- Consent in research.

Key principles for ethical research

- To maximise benefits for individuals and society & minimise risk and harm.
- The rights and dignity of individuals and groups should be respected.
- Voluntary and appropriately informed participation.
- Research should be conducted with integrity and transparency.
- Clearly defined lines of responsibility and accountability.
- Independence of research should be maintained and where conflicts of interest cannot be avoided they should be made explicit.

Ethical considerations in data sharing

- Clear guidance designed by the National Statistician's Data Ethics Advisory Committee's (NSDEC).
- UKSA Ethics Self-Assessment Tool.

NSDEC ethics principles

Public Good	The use of data has clear benefits for users and serves the public good.
Confidentiality, data security	The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.
Methods and Quality	The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.
Legal Compliance	Data used and methods employed are consistent with legal requirements such as Data Protection Legislation, the Human Rights Act 1998, the Statistics and Registration Service Act 2007, and the common law duty of confidence.
Public views & engagement	The views of the public are considered in light of the data used and the perceived benefits of the research.
Transparency	The access, use and sharing of data is **transparent**, and is communicated clearly and accessibly to the public.

Best practices for ethical sharing of research data

- Ethical obligations should be considered throughout the research lifecycle.
- Be knowledgeable about relevant research organisations own standards and requirements.
- Comply with relevant laws.
- Avoid social and personal harm.
- Data centres facilitate ethical and legal re-use of research data, protection of participants and safeguarding of personal data.

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.

Personal information

- Personal information or data is any information relating to an identified or identifiable natural person (name, address, postcode).

- It also includes sensitive personal data or special category data such as ethnicity, political or religious beliefs, biometric data, health.

Data protection considerations

If personal information about people is collected or used in research data protection regulations apply.

EU GDPR (2018) & country specific laws:

- a researcher based in a country from the European Union (country that adheres to EU GDPR) collects personal data about people living in a country from the European Union.

e.g. DPA (2018) & the UK GDPR 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 .

e.g. DPA (2018), EU GDPR (2018) & the UK GDPR applies when

- a researcher based in the UK collects personal data about people across Europe.

Data subjects rights

- The right to be **informed.**
- The right of **access.**
- The right to **rectification** (correction).
- The right to **erasure** (right to be forgotten).
- The right to **restrict processing.**
- The right to data **portability.**
- The right to **object.**

Principles of processing personal data

(DPA and the UK GDPR)

All data must be:

- be processed lawfully, fairly and transparently.
- be kept to the original purpose.
- be minimised (only the personal data that is necessary is collected).
- be removed if not necessary.
- have the accuracy upheld.
- be kept confidential and integrity maintained.

Legal bases to process personal data

Legal base	Example
Consent	Survey to capture public opinion, whereby email addresses are collected to contact respondents at a later stage. Qualitative study on a sensitive topic, e.g. violence against women, where respondents may be identifiable from the collected information. Oral history project where people's real names are used.
Public interest / public task	Longitudinal study of people living with dementia and their carers, to identify how people would like to be supported. Findings inform and support the caring strategy and public advocacy
Legitimate interest	Research project funded and undertaken by a private corporation to look at the effects of smoking on car passengers.
Protect vital interests	Unlikely in research. Hospital treating a patient after a serious road accident can search for his/her ID to find previous medical history or to contact his next of kin.
Legal obligation	Unlikely in research. Processing personal data as part of a health and safety report or incident.
Performance of a contract	Unlikely in research. Processing personal data as part of an employment contract.

GDPR and research

- Principles.
- Rights of data subjects.
- Processing grounds for processing personal data.
- Emphasis on transparency, clear information, clear documentation.
- Reuse for research allowed with safeguards.

Strategies for managing and sharing research data obtained from people

- Protection of identities when promised (anonymisation, de-identification).
- Processing ground for personal data (consent).
- Regulated access where needed (open, safe guarded, controlled).

Disclosure assessment

- **Direct identifiers**: e.g. name, address, postcode, telephone number, biometrics data.
- **Indirect identifiers**: e.g. occupation, geography, marital status, educational qualification, unique or exceptional values (outliers) or characteristics.

De-identification & anonymisation

- **De-identification** – refers to a process of removing or masking direct identifiers in personal data.
- **Anonymisation** - refers to a process of ensuring that the risk of somebody being identified in the data is negligible.

Anonymising quantitative data

- Remove direct identifiers
  e.g. names, address, institution, photo.

- Reduce the precision/detail of a variable through aggregation
  e.g. birth year instead of date of birth, occupational categories rather than jobs; 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, age.

Further info

Anonymising qualitative data

- Plan or apply editing at time of transcription except: longitudinal studies.
- Avoid blanking out; use pseudonyms or replacements.
- Avoid over-anonymising – removing / aggregating information in text can distort data, make them unusable, unreliable or misleading.
- Consistency within research team and throughout project.
- Show replacements, e.g. with [brackets].
- Keep a log of all replacements, aggregations or removals made – keep separate from de-identified data files.
- Further information.

What if anonymisation is impossible?

- Obtain consent for sharing non-anonymised data.
- Regulate or restrict user access.

Consent in research

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.

How to seek consent?

- Consent can be gained in **written** or **oral** form.
- Format depends on the kind of research.
- Important to document how it has been gained, what information has been provided to the participants and what they have agreed to.

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.

Challenges in obtaining informed consent

- Participant perception and expectations.
- Children & vulnerable people.
- Participant’s poor awareness of their rights.
- Failure to provide adequate information.
- Absence of consideration of participant’s background such education, culture.
- Use of jargon.
- Skeptical of confidentiality issues.

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 and who have received training.

Handling personal data in research

- Will the research project collect personal data or special categories of data?
- Who will be the data controller for the research project?
- Will the research involve collaboration with other partners.
- Who will have access to the personal data collected?
- What ground will be used for processing the personal data in a project?
- What information needs to be communicated to participants?
- How and where will the personal data be stored?

Key considerations in secondary data use and sharing

- Rights in research data
- Using secondary data

What rights might there be in research data?

Copyright

Database rights

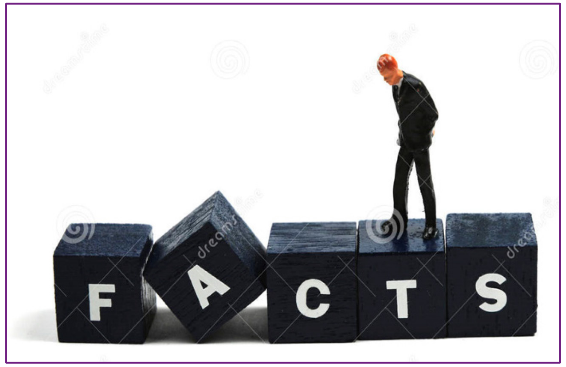
Further information on [rights in research data]() and [different scenarios]()

Rights in research data

- Right granted to creators and owners of works that are the result of human intellectual creativity.
- Assigned automatically to the creator.
- When data is shared, original owner is the right holder.

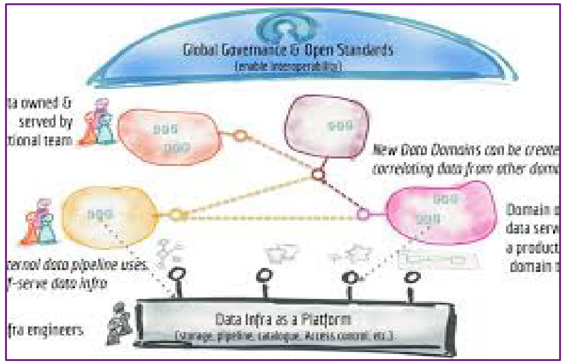
Further information on rights in research data and different scenarios

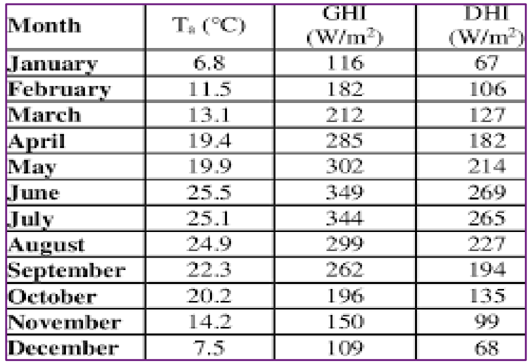
When my research data be protected by copyright?



""""

✗





Month	T_a (°C)	GHI (W/m^2)	DHI (W/m^2)
January	6.8	116	67
February	11.5	182	106
March	13.1	212	127
April	19.4	285	182
May	19.9	302	214
June	25.5	349	269
July	25.1	344	265
August	24.9	299	227
September	22.3	262	194
October	20.2	196	135
November	14.2	150	99
December	7.5	109	68

✓

Best practices when using secondary data

Questions 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.

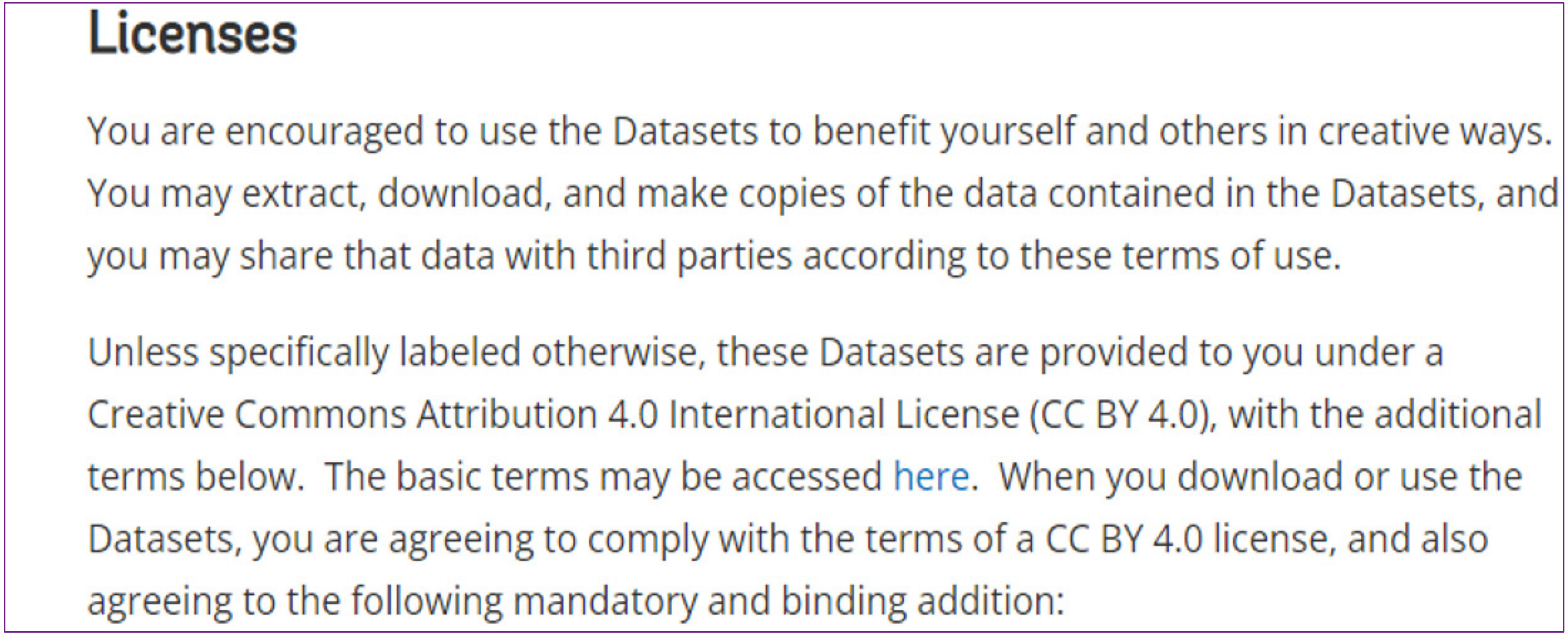
- If permission is not granted, need to remove copyrighted variables/material before publishing or sharing.

Copyright scenario

A researcher has used secondary data sources for a research project. He plans to share his data. The sources he used were:

- World Bank open data.
- Microsoft Academics.

World Bank open data



Licenses

You are encouraged to use the Datasets to benefit yourself and others in creative ways. You may extract, download, and make copies of the data contained in the Datasets, and you may share that data with third parties according to these terms of use.

Unless specifically labeled otherwise, these Datasets are provided to you under a Creative Commons Attribution 4.0 International License (CC BY 4.0), with the additional terms below. The basic terms may be accessed here. When you download or use the Datasets, you are agreeing to comply with the terms of a CC BY 4.0 license, and also agreeing to the following mandatory and binding addition:

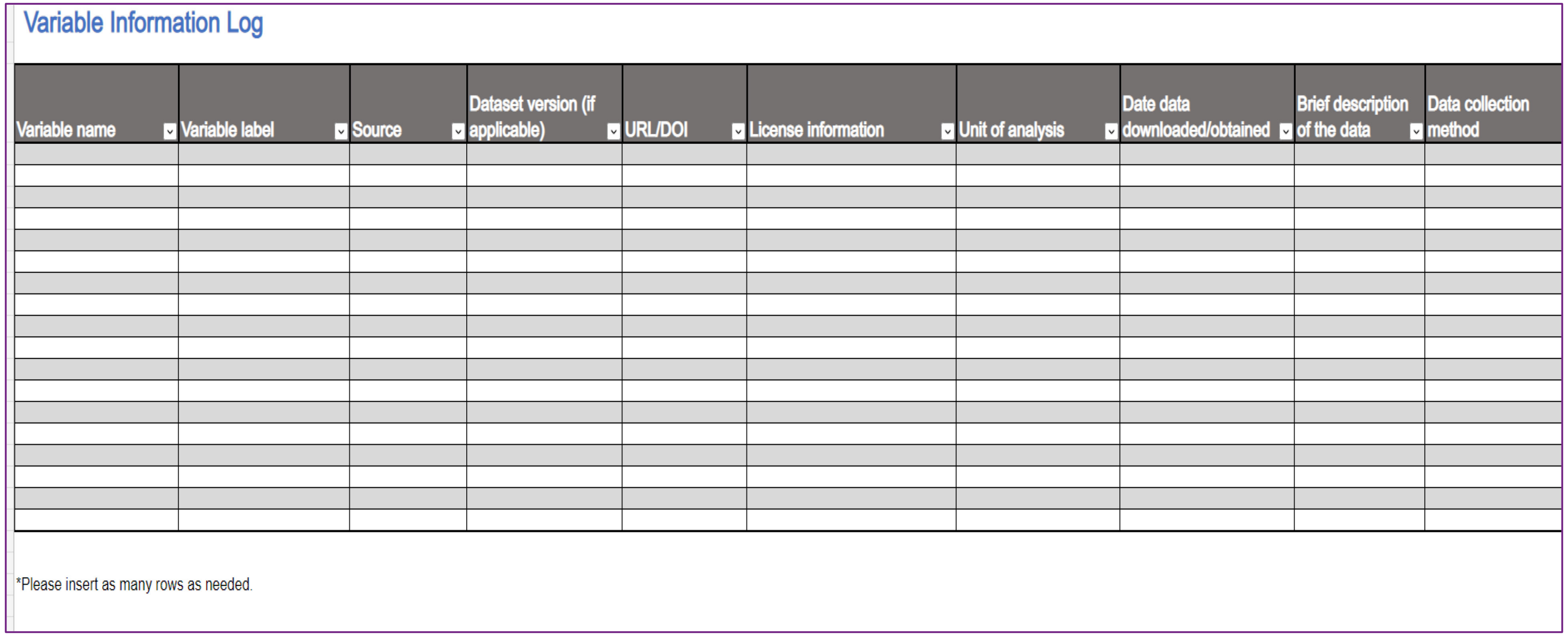
Microsoft Academics

Personal and Non-Commercial Use Limitation

Unless otherwise specified, the Services are for your personal and non-commercial use. You may not modify, copy, distribute, transmit, display, perform, reproduce, publish, license create derivative works from, transfer, or sell any information, software, products, or services obtained from the Services.

Resources

- Copyright (UK Data Service)
- Variable information log template for secondary data users



Variable Information Log

Variable name	Variable label	Source	Dataset version (if applicable)	URL/DOI	License information	Unit of analysis	Date data downloaded/obtained	Brief description of the data	Data collection method

*Please insert as many rows as needed.

Best practice for legal compliance

- Investigate early which laws apply to your data, including cross-country collaborative working.
- Do not collect or keep personal or sensitive data if not essential to your research.
- Plan early on; seek advice from your research office.
- Ensure that you check participants know how this data will be used.
- Remember: not all research data are personal (e.g. anonymised data are not personal).

Further resources

- UK Data Service
- UKDS Model Consent Form
- Example Consent Forms
- Example Information Sheet
- Consent for data sharing
- DARIAH ELDAH Consent Form Wizard | CFW
- Rights when using secondary data sources
- Regulating access to data
- Managing and sharing research data: A guide to good practice

UK Data Service

Thank you.

Dr Hina Zahid

hzahid@essex.ac.uk

Twitter: @hinazahid2016