Data Management Basics 2: Ethical and legal issues in data sharing

Dr Hina Zahid
Senior Research Data Officer

UK Data Service Introductory Training Series

30th April 2021

Overview

Ethical obligations

- Key principles for ethical research
- Ethical research considerations and best practices

Legal compliance

- Duty of confidentiality
- Data protection considerations- GDPR
- Strategies for managing and sharing research data (disclosure assessment, anonymisation, consent, access controls)
- Copyright considerations
- Best practices for legal compliance

Further resources

Ethical obligations

Key principles for ethical research

- To maximise benefit 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; from planning and research design stage, data collection stage to the future use including publications, archiving, sharing and linking of data.
- 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

Legal compliance

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. Personal information or data is any information relating to an identified or identifiable natural person'

- 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

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
(the UK GDPR)

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. This invariably involves doing more than simply de-identifying the data, and often requires that data be further altered or masked. Anonymisation allows data to be shared ethically and legally while preserving confidentiality

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 for research ethics: provide information regarding study purpose, risks, benefits, voluntary participation

- Consent can also be used as a legal basis for the processing of personal data under GDPR

Conditions for consent when used as a legal base for processing personal data

- Must be freely given, informed, unambiguous, specific (granular)
- A clear affirmative action
- Cannot be inferred from silence, pre-ticked boxes or inactivity
- Participants can withdraw consent to process their personal data at any time
- Must be documented, i.e. recorded, written or oral
- An explicit consent is required to process special categories data (e.g. a person's race, ethnic origin, politics, religion, genetics, sex life, health)

  explicit = express statement of consent, e.g. written statement

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

Formats of consent

written consent	Advantages • More solid legal ground • Often required by Ethics Committees • Offers more protection for researcher	Disadvantages • Not possible for some cases: infirm • May scare people from participating
Verbal consent	Best if recorded	• Can be difficult to make all issues clear verbally • Possibly greater risks for researcher

Consent documentation

An information sheet should cover the following topics:

- Purpose of the research
- What is involved in participating
- Benefits and risks of participating
- Procedures for withdrawal
- Usage of the data during research, dissemination, storage, publishing and archiving
- Details of the research: funding source, sponsoring institution, name of project, contact details for researchers, how to file a complaint

Cont…

Consent documentation

Consent form should:

- Use simple language and free from jargon
- Allow the participant to clearly respond to points such as:
    - The participant has read and understood information about the project
    - The participant has been given the opportunity to ask questions
    - The participant voluntarily agrees to participate in the project
    - The participant understands that they can withdraw at any time without giving reasons and without penalty
    - Future uses (e.g. publications, share and reuse)
    - Signatures and dates of signing for the participant and the researcher

Cont…

Consent documentation

If personal information is collected:

- How personal information will be processed and stored and for how long
- Procedures for maintaining confidentiality
- Procedures for ensuring ethical use of the data

If the GDPR applies:

- The contact details of the data controller (DPO, REO, Researcher)
- Who will receive or have access to the personal data
- A clear statement on the right of the participant (right to access, correction or removal)

When to seek consent?

	Advantages	Disadvantages
One-off consent Used for taking part in the research project only once	Simple Least hassle to participants	Research outputs not known in advance Participants will not know all info they will contribute
Process consent Consent is requested continuously throughout the research project	Ensures ‘active’ consent	May not get all consent needed before losing contact Repetitive, can annoy participants

When to seek consent?

Retrospective consent

Consent for reuse of data can also be sought after their research contribution is complete

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 (*eg. 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

Copyright considerations

- Copyright is an intellectual property right assigned automatically to the creator
- Data owner (researcher) has copyright of research data
- Compiled datasets contain original copyright – seek permission to archive when collecting
- Data archives publish data – hold no copyright
- Information being in the public domain (e.g. online) does not mean copyright does not apply!

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

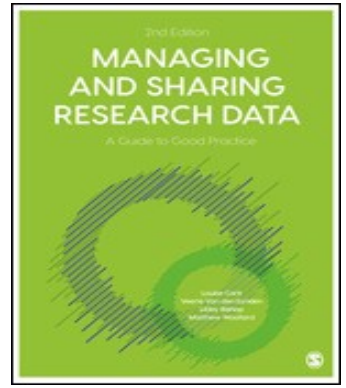
- If permission is not granted, need to remove copyrighted variables/material before publishing or sharing

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 practise



Future training events

- Depositing your data with Reshare
- Consent issues in data sharing

Past events

How to anonymised quantitative and qualitative data

Thank you!

- Are there any topics related to ethical & legal obligations in data sharing that you want us to include in our future training?
- Would you like to share what do you usually struggle with in the context of research ethics & legal compliance?

UKDS model consent form & real examples from the consent form

UKDS Model Consent Form

Informed Consent for [name of study]

Please tick the appropriate boxes **Yes** **No**

1. **Taking part in the study**

I have read and understood the study information dated [DD/MM/YYYY], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction. ☐☐

I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason. ☐☐

I understand that taking part in the study involves [……]. ☐☐

> Describe in a few words how information is captured, using the same terms as you used in the information sheet, for example: an audio-recorded interview, a video-recorded focus group, a survey questionnaire completed by the enumerator, an experiment, etc.].
>
> For interviews, focus groups and observations, specify how the information is recorded (audio, video, written notes).
>
> For questionnaires, specify whether participant or enumerator completes the form.
>
> For audio or video recordings, indicate whether these will be transcribed as text, and whether the recording will be destroyed.

If there is a potential risk of participating in the study, then provide an additional statement: I understand that taking part in the study has [……] as potential risk. ☐☐

UKDS Model Consent Form

2. Use of the information in the study

I understand that information I provide will be used for [.....].☐☐

List the planned outputs, e.g. reports, publications, website, video channel etc., using the same terms as you used in the study information sheet.
Consider whether knowledge sharing and benefits sharing needs to be considered, e.g. for indigenous knowledge

I understand that personal information collected about me that can identify me, such as my ☐☐
name or where I live, will not be shared beyond the study team.

At times this should be restricted to the researcher only.

If you want to use quotes in research outputs, add: I agree that my information can be quoted ☐☐
in research outputs.
If you want to use named quotes, add: I agree that my real name can be used for quotes. ☐☐

If written information is provided by the participant (e.g. diary), add: I agree to joint copyright ☐☐
of the [specify the data] to [name of researcher].

UKDS model consent form

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. de-identified (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 de-identified (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.

Example information sheet outline

- **What is the purpose of the study?**
- **Why have I been invited to take part in the study?**
- **Do I have to take part?**
- **What will happen if I take part?**
- **What will I have to do?**
  - *Walking interview and questionnaire*
  - *Living with sensors*
  - *Time use diary*
  - *Final interview and questionnaire*
- **What are the possible disadvantages or risks of taking part?**
- **What are the possible benefits of taking part?**
- **What happens when the research study stops?**
- **What if there is a problem?**
- **Will my taking part in the study be kept confidential?**
- **Who is organising and funding the research?**
- **Who has reviewed the project?**

Questions

Dr. Hina Zahid

hzahid@essex.ac.uk
Twitter: @hinazahid2016
@UKDataService
@UKDSRDM

UK Data Service