Data protection and research ethics

Veerle Van den Eynden
UK Data Service
University of Essex

Managing and sharing research data for transparency and FAIRness

24 January 2020

University of Essex

UK Data Service

Scenario

- Research on asylum seekers and refugees' experiences of forced labour, using interviews. These participants can be considered vulnerable
- We want to use the collected research data (interviews) for analysis and share afterwards with other researchers
- Interviews are likely to contain personal information

- Is it ethical to share the research data? What are pros and cons?
- How should we protect participant's anonymity?

- Read about this case study

UK Data Service

What the researchers did

- Address ethical issues before the research started
- Discuss principles of ‘doing no harm’, anonymity and informed consent iteratively in the team as issues arose throughout the project
- Stress independence of research from authority
- Provide written consent at the end of the interviews, so interviewee knew and could reflect on what they had shared
- Hold interviews in places convenient for interviewees
- Include interviewee in interpreter selection
- Collect biographical data after interview
- Not record any official identifying data (e.g. Home Office numbers)
- Let interviewee choose a pseudonym

Ethical obligations in research

- Research with human participants requires ethical review (Research Ethics Committee)
- Avoid social and personal harm
- Uphold scientific standards
- Comply with relevant laws
- Data repositories such as UK Data Service facilitate ethical re-use of research data, protection of participants and safeguarding of personal data:
  - data anonymisation
  - regulate data access
- Data sharing is NOT violation of data privacy or research ethics

Key principles of research ethics

- Research should aim to maximise benefit for individuals and society and minimise risk and harm
- The rights and dignity of individuals and groups should be respected
- Wherever possible, participation should be voluntary and appropriately informed
- Research should be conducted with integrity and transparency
- Lines of responsibility and accountability should be clearly defined
- Independence of research should be maintained and conflicts of interest made explicit

ESRC Framework for Research Ethics

Duty of confidentiality

- Duty of confidentiality exists in UK common law and may apply to research data
- information given in circumstances where it is expected that a duty of confidence applies, cannot normally be disclosed without the information provider's consent
- Disclosure of confidential information is lawful when:
  - the individual to whom the information relates has consented – consent for data sharing
  - 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

The General Data Protection Regulation (GDPR)

- Applies to ‘personal data’: any information relating to an identifiable (living) person who can be directly or indirectly identified in particular by reference to an identifier
- Living persons
- Anonymised data is NOT personal data so the GDPR does NOT apply
- Applies to:
  - any EU researcher (data controller) who collects personal data about a citizen of any country, anywhere in the world
  - A data controller or data processor based outside the EU but collecting personal data on EU citizens

The GDPR principles for processing personal data

1. Process lawfully, fair and transparent
    Inform participant of what will be done with the data, process accordingly
2. Keep to the original purpose
    Collect data for specified, explicit and legitimate purposes
    Do not process further in a manner incompatible with those purposes
3. Minimise data size
    Personal data collected should be adequate, relevant and limited to what is necessary
4. Uphold accuracy
    Personal data should be accurate and kept up to date
5. Remove data which aren't used
6. Ensure data integrity and confidentiality
    Protection against unauthorised or unlawful processing, accidental loss, destruction or damage, using appropriate technical or organisational measures

UK Data Service

Data subject 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
- Rights in relation to automated individual decision-making and profiling

Grounds for processing personal data

One of these must be present to process a data subject's personal data:

- Consent of the data subject
- Necessary for the performance of a contract
- Legal obligation placed upon controller
- Necessary to protect vital interests of the data subject
- Carried out in the public interest or is in the exercise of official authority
- Legitimate interest pursued by controller

Examples of legal bases

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.
Performance of a contract	Unlikely in research. Processing personal data as part of an employment contract.
Legal obligation	Unlikely in research. Processing personal data as part of a health and safety report or incident.
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.
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.

What's new with GDPR

- Principles, rights of data subjects and processing grounds for processing personal data are largely the same as under DPA, but much more explicit
- People have more control over their personal data
- Emphasis on transparency, clear information, clear documentation, accountability
- Extra rights of subjects: data portability
- Reuse for research allowed with safeguards
- Easier enforcement through national bodies

UK Data Service

Strategy for managing research data obtained from people

1. Obtain informed consent, also for data sharing and preservation or curation
2. Protect identities e.g. anonymisation, not collecting personal data
3. Regulate access where needed (all or part of data) e.g. by group, use or time period
4. Securely store personal and sensitive data

UK Data Service

Consent is needed across the data life cycle

- Engagement in the research process
- Dissemination in presentations, publications, the web
- Data sharing and archiving
  - consider future uses of data

Always dependent on the research context – special cases for covert research, verbal consent, etc.

[UKDS template consent form]

UK Data Service

Consent for personal data - GDPR

- When consent is the legal basis for collecting and processing personal data in accordance with the GDPR, this consent for the use of personal data should be distinguished from other consent requirements (ethics, procedural obligation)
- Consent for collecting and processing personal data needs to be freely given, informed, unambiguous, specific (granular) and a clear affirmative action
- Consent cannot be inferred from silence, pre-ticked boxes or inactivity
- Participants can withdraw consent to process their personal data at any time
- Consent must be documented, i.e. recorded, written or oral
- When special categories data are processed (e.g. a person's race, ethnic origin, politics, religion, genetics, sex life, health,…) – and the processing grounds for this is consent – then this must be based on explicit consent
  - explicit = express statement of consent, e.g. written statement, two-stage verification of consent

Consent form / information sheet when collecting - processing personal data under GDPR

If consent is processing ground, then provide as information:

- Contact details of the researcher, data controller (the entity that determines the reason for processing personal data), and the Data Protection Officer
- Who will receive or have access to the personal data, including information on any safeguards if the personal data is to be transferred outside the EU
- Right of the participant to request access to their personal data and the correction (rectification) of removal (erasure) of such personal data
- Reminder that the participants have the right to lodge a complaint with the information Commissioner's Office (ICO)
- Period of retention for holding the data or the criteria used to determine this. (If data are to be archived for re-use, then the retention period should be indefinite)

Promising ‘anonymity’

- Once ‘anonymised’, data falls out of data protection legislation
- But, bear in mind that not all research data can be fully or easily anonymised/de-identified
  - Combinations of unique key attributes
  - Rich textual data
  - Combining data from different sources

Consent in practice

- Inform participants about the purpose of the research
- Discuss what will happen to their contribution (including the future archiving and sharing of their data)
- Indicate the steps that will be taken to safeguard anonymity and confidentiality
- Outline the right to withdraw from the research
- Need to balance
  - As simple as possible
  - Complete for all purposes: use, publishing and sharing
  - Avoid excessive warnings
  - Easy language

Timing and form of consent

	Advantage	Disadvantage
One-off consent: participant is asked to consent to 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 : participant's consent is requested continuously throughout the research project	Ensures 'active' consent	May not get all consent needed before losing contact Repetitive, can annoy participants

	Advantage	Disadvantage
Written consent	More solid legal ground, e.g. participant has agreed to disclose confidential info Often required by Ethics Committees Offers more protection for researcher (as they have written documentation of consent)	Not possible for some cases: infirm, illegal activities May scare people from participating (or have them think that they cannot withdraw their consent)
Verbal consent	Best if recorded	Can be difficult to make all issues clear verbally Possibly greater risks for researcher (in regards to adequately proving participant consent)

Types of material and consent

Different data sharing consent agreements may be applied to different types of research data, e.g. less sensitive (survey) vs. highly sensitive (medical)

- Text and transcripts:
  - Can be anonymised
- Images, audio/video recordings:
  - Data more likely to reveal identities
  - Less usable after anonymising (distortion or blurring)
  - Anonymising costly
  - Consent or access control may be better alternatives than anonymisation

Identity disclosure

A person's identity can be disclosed through:

- **direct identifiers**

  e.g. name, address, postcode, telephone number, voice, picture

often NOT essential research information (administrative)

- **indirect identifiers** – possible disclosure in combination with other information

  e.g. occupation, geography, unique or exceptional values (outliers) or characteristics

Anonymise quantitative data

- Aggregate categories to reduce precision

  e.g. birth year vs. date of birth, 5-year age groups, occupational categories, area vs. village name
- Restrict upper or lower ranges of a variable to disguise outliers

  e.g. income, expenditure, age (>65)
- Use standard coding frames – e.g. SOC2010 for employment
- Generalise meaning of detailed text

  e.g. occupational expertise

Anonymise qualitative data

- plan or apply editing at time of transcription

  except: longitudinal studies – deidentify when data collection complete (linkages)
- 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

What if anonymising is impossible?

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

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

Tools and templates

- Model consent form (doc) that takes into account consent for data sharing and future data reuse
- Sample survey consent statement (doc) that considers consent for data sharing and future data reuse
- Text anonymisation tool (zip) to help you find disclosive information to remove or pseudonymise in qualitative textual data files.

UK Data Service

Exercises

- Consent
- Anonymisation

Questions

Veerle Van den Eynden

veerle@essex.ac.uk