Legal and Ethical Issues in Sharing Data

Scott Summers and Veerle Van den Eynden
UK Data Service
University of Essex

Cancer Research UK Data Sharing and Management
24th May 2017
Ethical and legal issues in data sharing

• Legal and ethical aspects
• Informed consent for data sharing
• Anonymising data
• Controlling access to data
• Working with Research Ethics Committees (RECs) and Institutional Review Boards (IRBs)
Ethical arguments for archiving data

- Not burden over-researched, vulnerable groups
- Make best use of hard-to-obtain data, e.g. elites, socially excluded, over-researched
- Extend voices of participants
- Provide greater research transparency

In each, ethical duties to participants, peers and public may be present
Ethical obligations and data sharing

- Research with human participants usually requires ethical review (Research Ethics Committee)
- Ethical conduct in research and protection of safety, rights and well-being of research participants – ‘do no harm’
- Data archives such as UK Data Archive 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
Legal Compliance
Duty of confidentiality and data sharing

• Duty of confidentiality exists in UK common law and may apply to research data

• If participant consents to share data, then sharing does not breach confidentiality

• Public interest can override duty of confidentiality

  • May need to give up data for court subpoena or to police

  • Best practice is to avoid vague or general promises in consent forms
Data Protection Act 1998

- Personal data:
  - relate to a living individual
  - individual can be identified from those data or from those data and other information
  - include any expression of opinion about the individual

- Only disclose personal data if consent given to do so, and if legally required to do so

Handling personal data:
- processed fairly and lawfully
- obtained and processed for specified purpose
- adequate, relevant and not excessive for purpose
- accurate
- not kept longer than necessary
- processed in accordance with the rights of data subjects
  - e.g. right to be informed how data will be used, stored, processed, transferred, destroyed
  - e.g. right to access info and data held
- kept secure
- not transferred abroad without adequate protection
Data Protection Act and research

- Exceptions for personal data collected as part of research:
  - can be retained indefinitely, if needed
  - can be used for other purposes in some circumstances
  - people should still be informed
  - for anonymised data (personal identifiers removed) DP laws will not apply as these no longer constitute ‘personal data’

- EU Data Protection Directive will be replaced by the General Data Protection Regulation on May 25th 2018
  - directly binding on all member states (not via national legislation) – includes the UK
  - key changes possible in: consent; rights of data subjects; international data transfer; sanctions; reuse for research
‘Sensitive data’

Data regarding an individual's race or ethnic origin, political opinion, religious beliefs, trade union membership, physical or mental health, sex life, criminal proceedings or convictions (DPA, 1998)

• Can only be processed for research purposes if:

  • explicit consent (ideally in writing) has been obtained; or
  • medical research by a health professional or equivalent with duty of confidentiality; or
  • analysis of racial / ethnic origins for purpose of equal opportunities monitoring; or
  • in substantial public interest and not causing substantial damage and distress
Best practice for legal compliance

- Investigate early which laws apply to your data
- Do not collect personal or sensitive data if not essential to your research
- Seek advice from your research office
- Plan early in research
- If you must deal with personal or sensitive data
  - inform participants about how their data will be used
  - remember: not all research data are personal (e.g. anonymised data are not personal)
Our advice to researchers

- Do not collect personal or sensitive data if not essential to your research
- Plan early in research
- If you collect personal or sensitive data, inform participants how their data will be used
- Not all research data are personal, e.g. anonymised data are not personal
Options for sharing confidential data

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
   - **Securely store** personal or sensitive data
Informed Consent
Consent needed across the data life cycle

- Engagement in the research process
  - decide who approves final versions of transcripts

- Dissemination in presentations, publications, the web
  - decide who approves research outputs

- Data sharing and archiving
  - consider future uses of data

Always dependent on the research context – special cases for covert research, verbal consent, etc.
A good information sheet & consent form

• Meets requirements of Data Protection laws
  • purpose of the research
  • what is involved in participation
  • benefits and risks
  • mechanism of withdrawal
  • usage of data – for primary research and sharing
  • strategies to ensure confidentiality of data (anonymisation, access etc.) where this is relevant

• Need to balance
  • as simple as possible
  • complete for all purposes: use, publishing and sharing
  • avoid excessive warnings

• UK Data Archive model consent form
  http://www.data-archive.ac.uk/media/210661/ukdamodelconsent.doc
## Timing and form of consent

<table>
<thead>
<tr>
<th></th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-off</td>
<td>Simple</td>
<td>Research outputs not known in advance</td>
</tr>
<tr>
<td></td>
<td>Least hassle to participants</td>
<td>Participants will not know all info they will contribute</td>
</tr>
<tr>
<td>Process</td>
<td>Ensures ‘active’ consent</td>
<td>May not get all consent needed before losing contact</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Repetitive, can annoy participants</td>
</tr>
</tbody>
</table>

| Written        | More solid legal ground, e.g. participant has agreed to disclose confidential info |
|                | Often required by Ethics Committees                                |
|                | Offers more protection for researcher                              |
|                | Not possible for some cases: infirm, illegal activities            |
| Verbal         | Can be difficult to make all issues clear verbally                 |
|                | Possibly greater risks for researcher                              |
|                | Best if recorded                                                   |
Right to withdraw

- Right to withdraw – one of key features of consent

- What about already collected data?
  - not usually allowed, at least in most surveys

- What if project is longitudinal?
  - permit withdrawal
In practice: wording in consent form / information sheet – interviews, photos

We expect to use your contributed information in various outputs, including a report and content for a website. Extracts of interviews and some photographs may both be used. We will get your permission before using a quote from you or a photograph of you. After the project has ended, we intend to archive the interviews at ….. Then the interview data can be disseminated for reuse by other researchers, for research and learning purposes.

The interviews will be archived at ……. and disseminated so other researchers can reuse this information for research and learning purposes:

- I agree for the audio recording of my interview to be archived and disseminated for reuse
- I agree for the transcript of my interview to be archived and disseminated for reuse
- I agree for any photographs of me taken during interview to be archived and disseminated for reuse
In practice: wording in consent form / information sheet – focus group

Any personal information that could identify you will be removed or changed before files are shared with other researchers or results are made public.

We ask you to consider the following points before agreeing to participate.

- Your contribution to the research will take the form of a focus group participant. This will be digitally video recorded and transcribed.

- Your name and any information which may directly or indirectly identify you will be altered to protect your anonymity.

- Any recordings of the discussions will be kept securely, and only authorised to other researchers on the condition they preserve your anonymity.

- The transcriptions (excluding names and other identifying details) will be retained by the researcher and analysed as part of the study. They will also be deposited with the UK Data Archive which has strict regulations about accessing data for research and protecting participant confidentiality.

ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/consent-forms.aspx
Dear,

[Introduction explaining the research]

Thank you very much for agreeing to participate in this survey.

The information provided by you in this questionnaire will be used for research purposes. It will not be used in a manner which would allow identification of your individual responses.

Anonymised research data will be archived at the UK Data Archive in order to make them available to other researchers in line with current data sharing practices.

Yours,

[Name, institution and contact details of researcher]

http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/confidentiality/conf-language.html
Audio-visual data

Digital manipulation of audio and image files can remove personal identifiers

e.g. voice alteration, image blurring (e.g. of faces)

Labour intensive, expensive, may damage research potential of data

Better:

• to obtain consent to use and share data unaltered for research purposes

• to avoid mentioning disclosing information during audio recordings
Special cases of consent

Children
- Aged 16 and above can give their own consent
- If minor is competent, need consent from child, and parent/guardian
- Gillick principle – even children under 16 can consent to medical treatment, without parental consent

Employees
- Employee may owe duty of confidentiality to employer

Vulnerable participants, disabilities of any kind
- Need to balance protection from harm with right to participate

Criminal activities
- Usually no obligation to disclose, unless investigation is active

Internet, blog, social media – “New social media, new social science?”
- [http://aoir.org/reports/ethics2.pdf](http://aoir.org/reports/ethics2.pdf)

Retrospective consent; covert research, observational experiments
Anonymisation
Why anonymise research data?

Ethical reasons
- protect people’s identity (sensitive, illegal, confidential info)
- disguise research location

Legal reasons
- not to disclose personal data (DPA)

Commercial reasons

Discuss with your research participants
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*
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 vs. date of birth, occupational categories, 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

- **combining** variables
  - *e.g.* creating non-disclosive rural / urban variable from place variables
Anonymising qualitative data

- plan or apply editing at time of transcription
  *except: longitudinal studies - anonymise 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.

- identify replacements, e.g. with [brackets]

- keep anonymisation log of all replacements, aggregations or removals made – keep separate from anonymised data files
Anonymising qualitative data

**Example: Anonymisation log interview transcripts**

<table>
<thead>
<tr>
<th>Interview / Page</th>
<th>Original</th>
<th>Changed to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int1 p1</td>
<td>Spain</td>
<td>European</td>
</tr>
<tr>
<td>Int1 p1</td>
<td>E-print Ltd</td>
<td>Printing</td>
</tr>
<tr>
<td>Int1 p2</td>
<td>20\textsuperscript{th} June</td>
<td>June</td>
</tr>
<tr>
<td>Int2 p2</td>
<td>Amy</td>
<td>Moira</td>
</tr>
<tr>
<td>Int2 p1</td>
<td>Francis</td>
<td>my friend</td>
</tr>
</tbody>
</table>

P31. Joan $\rightarrow$ Mary
P97. Carol $\rightarrow$ {Mother}
P34. Colchester $\rightarrow$ {Town in S.E.England}
P65. Welshpool High School $\rightarrow$ @@#High School##@@
**In practice: example anonymisation**

<table>
<thead>
<tr>
<th>Ex 1. Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003 (study 5407 in UK Data Archive collection) by M. Mort, Lancaster University, Institute for Health Research.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Date of Interview:</strong> 21/02/02</td>
</tr>
<tr>
<td><strong>Interview with [Lucas Roberts] DEFRA field officer</strong></td>
</tr>
<tr>
<td><strong>Date of birth:</strong> 2 May 1965</td>
</tr>
<tr>
<td><strong>Gender:</strong> Male</td>
</tr>
<tr>
<td><strong>Occupation:</strong> Frontline worker</td>
</tr>
<tr>
<td><strong>Location:</strong> [Humpton], North Cumbria</td>
</tr>
</tbody>
</table>

**[Lucas] was living at home with his parents, “but I’m hoping to move out soon” so we met at his parents’ small neat house. We sat in a very comfortable sitting room with an open fire and [Lucas] made me coffee and offered shortbread. Although at first [Lucas] seemed a little nervous, quick to speech and very watchful he seemed to relax as we spoke and to forget about the tape.**

**I will just start by asking you to tell me a little bit about yourself and your background.**

Well it is an agricultural background. I grew up on the farm where my brother is now. After I left school I did work on the farm but went to college and did exams, did land use recreation, sort of countryside/environmental management course. So I obviously left agriculture, did the course and came back [to the farm] at weekends.
In practice: example anonymisation

Yeah. So is part of your job to look for funding bids and to write funding bits or is that separate?

No. That was what P3 used to do and then it sort of passed down to... really it's with P1 and P4. But I don't actually think there's anything out there at the moment. I think at the moment, because there's all this money saving and things, there's nothing to... there isn't actually anything to access.

M27UK

INT: Id like to start with your career history really. Perhaps you can start from when you finished school?

RES: First of all Id like to say that I do not think that my career is very traditional for someone who comes and works abroad. I graduated from mathematics in 'Country F'. That was equivalent to what is now BSc and MSc and then I did a separate MSc in Computer Science which at that time was called like Specialisation in computer science. And then I started working at the 'Country F' University as a teacher.

INT: Was that in 'City D'?

RES: No, my Specialisation in Computer Science was in 'City D' and my first degree, my mathematics degree was in another university in 'Country F' it is called [Text Cut]

So after I finished in 'City D' my specialisation in computer science I went back to [Text Cut] where I did my first degree and I started working there as an assistant Professor.
What if anonymising is impossible?

- Obtain consent for sharing non-anonymised data
- and / or
- Regulate or restrict user access
Access Controls
Managing access to data

**Open**
- available for download / online access under open licence without any registration

**Safeguardsed**
- 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
Open about data with access restrictions

- Publish or advertise:
  - which data exist
  - where data are kept, e.g. which repository
  - who can access them
  - for which purpose
  - under which conditions
In practice: data with access conditions

Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003 (study 5407 in UK Data Archive collection) by M. Mort, Lancaster University, Institute for Health Research.

- Interviews (audio and transcript) and written diaries with 54 people
- 40 interview and diary transcripts are archived and available for re-use by registered users (Safeguarded)
- 3 interviews and 5 diaries were embargoed until 2015 (Safeguarded – Embargoed)
- Audio files archived and only available by permission from researchers (Safeguarded – Special Agreement)

[Links to data and user guide]
Working with RECs / IRBs
Data sharing and research ethics committees

• REC / IRBs are responsible for safeguarding participants from harm and ensuring ethical research and protecting home institutions, but are not always informed about sharing

• There can be perceived tensions between data sharing and protection

• We try to ensure that REC / IRBs understand that:
  • anonymised data are not subject to data protection laws
  • most funders require or encourage data to be shared
  • most research data can be shared
  • procedures (consent, anonymisation, regulating access) are available to enable ethical sharing
  • data archives ensure ethical re-use of research data, protection of participants and safeguarding of personal data
Conclusion

- Always consider anonymisation of research data together with consent agreements and access restrictions
- Regulating or restricting user access may offer better solution than anonymising
- Avoid collecting data that need anonymisation
  \( \textit{e.g. do not ask for full names if they later need to be removed from data} \)
- Maintain maximum information
- Retain unedited versions of data for preservation
- Plan anonymising at start of research, not at the end
Questions

Contact details:

Collections Development and Producer Relations team
UK Data Service
University of Essex
ukdataservice.ac.uk/help/get-in-touch