ETHICAL AND LEGAL ISSUES IN DATA SHARING

A COMBINATION OF GAINING CONSENT FOR DATA SHARING, ANONYMISING AND REGULATING ACCESS TO DATA WILL INCREASE THE POTENTIAL FOR MAKING PEOPLE-RELATED RESEARCH DATA MORE READILY AND WIDELY AVAILABLE

AREAS OF COVERAGE
• Legal and ethical aspects
• Informed consent for data sharing
• Anonymising data
• Controlling access to data
• Working with Research Ethics Committees
ETHICAL ARGUMENTS FOR ARCHIVING DATA

- Store and protect data securely
- 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
- Enable fullest ethical use of rich data

In each, ethical duties to participants, peers and public may be present
DUTY OF CONFIDENTIALITY AND DATA SHARING

• Duty of confidentiality exists in 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 living individual
  - individual can be identified from those data or from those data and other information
  - include any expression of opinion about the individual

- **Requirements for handling personal data**
  - processed fairly and lawfully
  - obtained and processed for a specified purpose
  - adequate, relevant and not excessive for the purpose
  - accurate
  - not kept longer than necessary
  - processed in accordance with the rights of data subjects, e.g. right to be informed about how data will be used, stored, processed, transferred, destroyed, …; right to access info and data held
  - kept secure
  - not transferred abroad without adequate protection

- **Personal data can be disclosed only with consent**
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

If data are anonymised (personal identifiers removed) then DP laws will not apply as these no longer constitute ‘personal data’

DPA is not intended to, and does not, inhibit ethical 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)
OPTIONS FOR SHARING CONFIDENTIAL DATA

• Obtain informed consent, also for data sharing and preservation / curation

• Protect identities e.g. anonymisation, not collecting personal data

• Regulate access where needed (all or part of data) e.g. by group, use, time period

• Securely store personal or sensitive data
INFORMED CONSENT FOR ETHICAL RESEARCH

• What does it mean for consent to be “informed”?
  • purpose of the research
  • what is involved in participation
  • benefits and risks
  • mechanism of withdrawal
  • data uses – primary research, storing, processing, re-use, sharing, archiving,…
  • strategies to ensure confidentiality of data where this is relevant – anonymisation, access restrictions…

• RCUK expects data to be accessible for other uses
  RCUK Common Principles on Data Policy
  http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx

• Now a requirement for ESRC awards:
  “Where research data are considered confidential or contain sensitive personal data, award holders must seek to secure consent for data sharing or alternatively anonymise the data in order to make sharing possible.” ESRC Research Data Policy 2010 2.4(32)
DO PARTICIPANTS CONSENT TO SHARE DATA?

- Timescapes
  - data on personal relationships
  - 95%+ consent rate
- Foot and mouth disease in N. Cumbria
  - sensitive community information
  - UK Data Archive consultation; pilot with 4 participants
  - 40/54 interviews; 42/54 diaries; audio restricted
- Finnish research on consent
  - Re-contact project: life stores, gender, etc.
  - 165/169 (98%) agreed
- Bereaved relatives want others to benefit from their data
INFORMED CONSENT FOR UNKNOWN FUTURE USES

- It is possible to provide much information about reuse
  - who can access the data – only authenticated researchers
  - purposes – research or teaching or both
  - confidentiality protections, undertakings of future users
  - general consent (similar to consent with emergent research topics)

- Medical research and biobank models – enduring, broad, open consent
  - no time limits; no recontact required
  - unspecified hypotheses and procedures
  - 99% consent rate (2500+ patients) – Wales Cancer Bank

*ESRC expects that others will also use it [data], so consent should be obtained on this basis and the original researcher must take into account the long-term use and preservation of data.* (ESRC Framework for Research Ethics, 1.17.5.1)
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,…) where this is relevant

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

- UK Data Archive model consent form
  [http://www.data-archive.ac.uk/media/210661/ukdamodelconsent.doc](http://www.data-archive.ac.uk/media/210661/ukdamodelconsent.doc)
## WHEN TO ASK FOR CONSENT

<table>
<thead>
<tr>
<th></th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-off</td>
<td>Simple</td>
<td>Research outputs (even questions, not known in advance)</td>
</tr>
<tr>
<td></td>
<td>Least hassle of participant</td>
<td>Participants will not know all content they will contribute</td>
</tr>
<tr>
<td>Process</td>
<td>Most complete for assuring active consent</td>
<td>Might not get consent needed before losing contact</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Repetitive, can annoy participant</td>
</tr>
</tbody>
</table>
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*
KEY POINTS FOR ANONYMISING

- never disclose personal data - unless consent for disclosure
- reasonable/appropriate level of anonymity
- maintain maximum meaningful information
- where possible replace rather than remove
- identifying information may provide context, do not over-anonymise
- re-users of data have the same legal and ethical obligation to NOT disclose confidential information as primary users
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
Spatial references (point coordinates, small areas) may disclose position of individuals, organisations, businesses

Remove spatial references - prevents disclosure; also all geographical and related information lost

Better

- reduce precision - replace point co-ordinates with larger, non-disclosing geographical areas  
  *e.g. km² area, postcode district, ward, road*

- reduce precision - replace point coordinate with meaningful variable typifying the geographical position; or summary statistics of location  
  *e.g. catchment area, poverty index, population density*

- keep spatial references and impose access restrictions on data
ANONYMISING QUALITATIVE DATA

• not collect disclosive data unless necessary
• 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
• xml mark-up can be used for anonymisation
  
  `<seg type="anonymised">word to be anonymised</seg>`
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</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p1</td>
<td>Spain</td>
<td>European</td>
</tr>
<tr>
<td>p1</td>
<td>E-print Ltd</td>
<td>Printing</td>
</tr>
<tr>
<td>p2</td>
<td>20\textsuperscript{th} June</td>
<td>June</td>
</tr>
<tr>
<td>p2</td>
<td>Amy</td>
<td>Moira</td>
</tr>
<tr>
<td>Int2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p1</td>
<td>Francis</td>
<td>my friend</td>
</tr>
</tbody>
</table>
REGULATING ACCESS DATA

- Essential when anonymisation ineffective or damaging to quality
  - visual or audio data
  - disclosive microdata

- UK Data Archive has gradation of access controls
  - small number of studies are open (no registration)
  - majority require registration
  - data users sign legally binding End User Licence – e.g. not identify any potentially identifiable individuals
  - stricter regulations for certain types of data:
    - Special Licences
    - Approved researchers
    - require data access authorisation from data owner prior to data release
    - embargo for given time period
    - Secure Data Service (no direct data access)

- Multiple AC can apply to different data types within one study
DATA SHARING AND RESEARCH ETHICS COMMITTEES

• RECs 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.

• UKDA tries to ensure that RECs know:
  • anonymised data are not subject to DPA
  • 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.