In practice – managing and sharing research data from people/organisations

Louise Corti
UK Data Service
University of Essex

Essex Big Data and Analysis Summer School
University of Essex
24-26 August 2015
How to share data obtained from people in an ethical / legal way?

- **Obtain informed consent**, also for data sharing and long-term 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 (separately)
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.
Tensions

- Between protecting participants, researchers and institutional reputation
- Maximising the value of research data collected from public funding
- RECs/IRBs tendency to be risk averse, require data destructions etc.
## 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                             |
Aspects to consider

- Different forms of consent for different materials, e.g. audio recordings vs transcripts
- Right to withdraw – what to do with already collected data?
- Informed consent for ‘unknown future data uses’?

- Provide maximum information about reuse
  - who can access the data – authenticated researchers
  - purposes – research or teaching or both
  - confidentiality protections; agreement by future users
Medical research and biobanks

- Enduring, broad, open consent
- No time limits; no recontact required
- Unspecified hypotheses and procedures
- 99% consent rate (2500+ patients) – Wales Cancer Bank
In practice: wording in consent form / information sheet

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
As the ESRC is a publicly funded body, it has developed ways to share data among academic researchers (subject to strict conditions). To this end, we hope you will allow your anonymised transcript to be stored as part of the UK Data Archive (a service provider for the Economic and Social Data Service).
In practice: wording in consent form / information sheet

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

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

- Removing or replacing information in text can distort data, make them unusable, unreliable or misleading: A balance to preserve context
- Remove direct identifiers, or replace with pseudonyms – often not essential research information
- Avoid blanking out; use pseudonyms or replacements (identify replacements systematically)
- Plan or apply editing at time of transcription
- Consistency within research team / project
- Keep anonymisation log of all replacements or removals made; keep separate from anonymised data file
## Anonymisation log

### 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>20th 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>
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:
  - obtain consent to use and share data unaltered for research purposes
  - avoid mentioning disclosing information during audio recordings
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>
</table>

**Date of Interview:** 21/02/02

**Interview with:** Lucas Roberts, DEFRA field officer

**Date of birth:** 2 May 1965

**Gender:** Male

**Occupation:** Frontline worker

**Location:** Plumpton, North Cumbria

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: I'd like to start with your career history really. Perhaps you can start from when you finished school?

RES: First of all I'd 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.
Managing access to data

- Essential when anonymisation is ineffective or damaging
- Visual or audio data or disclosive microdata
- UK Data Service Access Policy has three tiers:
## Managing access to data

<table>
<thead>
<tr>
<th>Open</th>
<th>Safeguarded</th>
<th>Controlled</th>
</tr>
</thead>
</table>
| • available for download/online access under open licence without any registration | • available for download/online access to logged-in users who have registered and agreed to an End User Licence  
• Special agreements (depositor permission; approved researcher; embargo for fixed time period) | • available for remote or safe room (Secure Lab) access to authorised and authenticated users whose research proposal has been and who have received training |
Open about data with access restrictions

• Publish / advertise:
  • Which data exist
  • Where data are kept, e.g. which repository
  • Who can access them
  • For which purpose
  • Under which conditions
In practice: access conditions ReShare

Global Uncertainties: Security In an Africa of Networked, Multi-Level Governance


The programme of research centres on how the various institutions responsible for the production of security and the management of conflict in Sub-Saharan African societies do, could and should evolve in response to the presence of violent conflict. These institutions include: the armed forces, police, courts (civil as well as criminal), ‘traditional’ institutions of local governance, elections (and other methods of choosing leadership), international organisations (such as the United Nations, African Union, UN High Commission for Refugees, the World Court), and the community of international donor nations. The programme is built on the observation that all governance (especially in Africa) is multi-leveled and networked - from the village to the international organisation, and well beyond what is specified in formal government structures. Thus the focus will be not only on the ways in which key conflict-management institutions evolve themselves but also on the changing ways in which the networks in which they are embedded actually operate. This leading edge research challenges theories about the state, state formation, and sovereignty as well as the dynamics of violent conflict as presented in the dominant current work on Africa. The primary methods of research will be network analysis and qualitative interviewing.

<table>
<thead>
<tr>
<th>Creators:</th>
<th>Creator Name</th>
<th>Email</th>
<th>Affiliation</th>
<th>ORCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leonhard, David</td>
<td>Unspecified</td>
<td>Institute of Development Studies</td>
<td>Unspecified</td>
<td></td>
</tr>
</tbody>
</table>

| Research funders: | Economic and Social Research Council |
| Grant reference: | RES-071-27-0048 |
| Subjects: | Law, crime and legal systems, Politics |
| Date deposited: | 31 Jan 2014 19:00 |
| Last modified: | 22 Oct 2014 19:27 |
In practice: access conditions ReShare

--- Coverage and Methodology ---

Collection period: 1 April 2009 to 31 March 2013
Country: Sierra Leone
Data collection method: Group interviews in 39 villages and towns in rural Sierra Leone
Observation unit: Groups
Kind of data: Alpha-numeric, Textual
Type of data: Qualitative and mixed methods data
Resource language: English

--- Access and Administration ---

Data sourcing, processing and preparation:

Copyright holders: Leonard, David, unspecified
Affiliation: Institute of Development Studies
ORCID: unspecified

Contact:
Name: Leonard, David
Email: leonard@berkeley.edu
Affiliation: Institute of Development Studies
ORCID: unspecified

Notes on access: The names of individual respondents have been removed from the data set.
Publisher: Economic and Social Research Council
Last modified: 22 Oct 2014 19:27
In practice: access conditions ReShare

<table>
<thead>
<tr>
<th>AVAILABLE FILES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Accessible to:</td>
</tr>
<tr>
<td>File or bundle content:</td>
</tr>
<tr>
<td>File or bundle description:</td>
</tr>
<tr>
<td>File format:</td>
</tr>
<tr>
<td>License:</td>
</tr>
<tr>
<td>File size:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>+ Sierra_Leone_Security_%26_Networks__Coded.xlsx</td>
</tr>
<tr>
<td>+ Copy_of_Somalia_SC_peace_initiatives.xlsx</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Accessible to:</td>
</tr>
<tr>
<td>File or bundle content:</td>
</tr>
<tr>
<td>File or bundle description:</td>
</tr>
<tr>
<td>File format:</td>
</tr>
<tr>
<td>License:</td>
</tr>
<tr>
<td>File size:</td>
</tr>
</tbody>
</table>
New challenges: Big data and privacy

- Some big data analytics will involve processing personal data, from sources such as:
  - social media
  - loyalty cards
  - sensors in clinical trials

- New legal work on big data and data protection

- Those in the data supply and use chain must comply with their obligations under the Data Protection Act
Our data management guidance

- Online best practice guidance: [ukdataservice.ac.uk/manage-data.aspx](http://ukdataservice.ac.uk/manage-data.aspx)
- Helpdesk for queries: [ukdataservice.ac.uk/help/get-in-touch.aspx](http://ukdataservice.ac.uk/help/get-in-touch.aspx)
- Training: [www.data-archive.ac.uk/create-manage/advice-training/events](http://www.data-archive.ac.uk/create-manage/advice-training/events)
Exercise

• Instructions : Exercise-DisclosureReviewTreatment.doc

1. Wording in consent forms
Folder Consent_exercise/
Exercise - AssessingStatements_consentforms.doc

2. Anonymisation
Folder - Anonymisation_exercise/
ResearchSummary_HIVUganda
HIVUganda_Background__qual_data.doc
3 data extracts