Sharing Qualitative Data: Challenges and Opportunities

Libby Bishop
UK Data Service
UK Data Archive, University of Essex

Lancaster University
5 April 2017
Plan for the day

• Introductions
• Sharing qualitative data – pros and cons
• Ethical issues – consent for sharing
• Anonymising data
• Sharing practicalities
  • Documentation for qualitative data
  • Considerations about where to share data
How Britain Dies is a research project run by the think tank Demos and funded by Help the Hospices. One focus of our work is to look at the views of dying people and their families around what makes a good death and where people are dissatisfied with how they and their loved ones die.

I am interested in accessing Oral Interviews… to analyse the responses of psychiatric nurses to changes in their profession … I believe they will give me an unrivalled opportunity to bring the voice of the nurse to the foreground.

This data will be used to pilot test an innovative method for qualitative data analysis using crowd sourcing technology.
“But no one reuses qualitative data…”

Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003 (SN5407)

- secondary analysis to study families and food;
- policy briefing of the economic cost of animal health diseases with aim of considering how UK may be better prepared to deal with outbreaks like this;
- use transcripts from previous focus groups with farmers to study biosecurity on dairy farms in the UK;
- the data will be used as teaching material for medical students for interview skills;
- building a speech recognition engine to automatically transcribe interviews for qualitative research.
Re-use purposes of qualitative data downloaded from UK Data Service, 2002-2016

Source: Bishop & Kuula-Luumi, Sage Open 2016
Sharing Data – pros and cons

• Group One: generate reasons for why researchers should/might want to share their data.
  o What are the benefits of sharing data?
  o Who does sharing data benefit? And how?

• Group Two: generate reasons for why researchers should not/might not want to share their data.
  o What are some of the concerns associated with sharing data?
  o What are some of the impediments to sharing data?

• Take about 5 minutes in groups, then we will discuss.
Qualitative data – challenges for sharing

- Strong relationships of trust, commitments to confidentiality
- Participant identity difficult to conceal
  - Audio and visual data
- Research locations potentially identifiable
- Difficult to anonymise data without reducing research value
- Research may investigate illegal activities

- But potential benefits of data sharing make it imperative to face these challenges
Arguments for sharing and open data

• Duties to participants – protect and
  • Empower – give voice
  • Avoid burdensome replication

• Duties to scholarly community
  • Transparency
  • Research integrity

• Duties to public
  • Use public funds wisely

Data underpin facts, and facts underpin reason...
Benefits to researchers of data sharing

• increases visibility of your scholarly work
• may enhance your reputation
• may increase citations of your publications
• provides long-term safe storage for data
• satisfies publishers’ data access policies
• enable collaborations on related themes and new topics
Funder policies

- Largely based on the [OECD Principles and Guidelines for Access to Research Data from Public Funding](http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies)

- UK: variety of models
  - Research Councils UK: *Publicly funded research data…should be made openly available with as few restrictions as possible*
  - Data management plans and recommendation only
  - Dedicated data centres

- Europe (European Research Council/Horizon 2020)
  - [Data management guidelines](http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies) for Horizon 2020 (~ policies)
Ethics and sharing

- Common worry about onward sharing of data covering ‘sensitive topics’, but all data are not sensitive
- Consider both legal and ethical duties
- Deal with personal sensitive data properly
- Discuss consent status and implications
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 with consent (and if legally required)

- Sensitive - race/ethnicity, pol. opinion, religion, union, health, sexual life, criminal offence

- 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 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
Three tools 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

*Use in different combinations depending on data*
Consent needed across the data life cycle

- Engagement in the research process
  - Explain the nature of research participation

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

- Data sharing and archiving – consent for sharing
  - Consider future uses of data
  - Balanced view of risks and benefits
  - Participants should be able to choose

Always dependent on the research context – special cases for covert research, verbal consent, etc.
In practice: wording in consent form / information sheet

<table>
<thead>
<tr>
<th><strong>Use of the information I provide beyond this project</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I agree for the data I provide to be archived at the UK Data Archive.²</td>
<td>☐ ☐</td>
</tr>
<tr>
<td>I understand that other authenticated researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.</td>
<td>☐ ☐</td>
</tr>
<tr>
<td>I understand that other authenticated researchers may use my words in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.</td>
<td>☐ ☐</td>
</tr>
</tbody>
</table>

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).

Complete for all purposes: use, publishing, **sharing**

Examples of consent forms:

https://www.ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/consent-forms
In practice: 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

Case study

SN 5407 Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003 (SN5407 at UK Data Archive)  Maggie Mort, Lancaster University

Discover.ukdataservice.ac.uk/catalogue

- funded by Department of Health
- recruit panel of 54 local people in affected area at time of FM crisis: farmers, agricultural professionals, small businesses, health professionals, vets, residents
- weekly diaries for 18 months describing how their life was affected by the crisis and process of recovery observed around them (handwritten)
- in-depth interviews and group discussions (audio recordings, transcripts)
- at start of research – researchers believed no data could or should be archived
Consent – exercise – in groups

• Read handout – “Assessing statements in consent forms”

• Use a project from your group, or the foot and mouth case

• Write a draft of a consent statement seeking permission for data sharing – just the portion about data sharing…. 
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
- Consistency within research team and throughout project
- Identify replacements, e.g. with [brackets]
- Keep separate anonymisation log of all replacements, aggregations or removals made
## Sample 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 country</td>
</tr>
<tr>
<td>p1</td>
<td>E-print Ltd</td>
<td>Printing company</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>
Anonymisation – exercise – in groups

• Read the interview extract from the foot and mouth project

• Decide what elements you think need anonymisation, and how would you make the necessary changes
In practice: example anonymisation

<table>
<thead>
<tr>
<th>Interview with</th>
<th>Date of Interview: 21/02/02</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lucas Roberts</td>
<td>Date of birth: 12 May 1965</td>
</tr>
<tr>
<td>DEFRA field officer</td>
<td>Gender: Male</td>
</tr>
<tr>
<td>Date of birth: 2 May 1965</td>
<td>Occupation: Frontline worker</td>
</tr>
<tr>
<td>Location: Plumpton, North Cumbria</td>
<td>Location: Plumpton, 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.
## Access summary – three tiers

<table>
<thead>
<tr>
<th>Tier</th>
<th>Access Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Open</strong></td>
<td>• available for download/online access under open licence without any registration</td>
</tr>
<tr>
<td><strong>Safeguarded</strong></td>
<td>• available for download/online access to users who have registered and agreed to an End User or Special Licence</td>
</tr>
<tr>
<td><strong>Controlled</strong></td>
<td>• available for remote or safe room access to authorised authenticated users</td>
</tr>
</tbody>
</table>
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 + transcript) and written diaries with 54 people
- 40 interview and diary transcripts are archived and available for re-use by registered users
- 3 interviews and 5 diaries were embargoed until 2015
- audio files archived and only available by permission from researchers

[link to catalogue](https://discover.ukdataservice.ac.uk/catalogue/?sn=5407)
[link to user guide](https://doc.ukdataservice.ac.uk/doc/5407/mrdoc/pdf/q5407userguide.pdf)
Documenting qualitative data

**Why (bother) documenting your research project?**
- Enables you to understand/interpret data when you return to it
- Needed to make data reusable
- Helps avoid incorrect use/misinterpretation by others

**What kinds of materials count as documentation?**
- **Study level**: research design, funding proposals, questionnaires, methods sections, reports
- Transcription/translation protocols
- Anonymisation records
- **Data level**: characteristics of individuals, other units of analysis
  Should not place unreasonable burden on primary researchers (and funding is available…)}
Preparing qualitative data

- Consider license and access conditions
- Assemble documentation – methods/publications
- Prepare study description catalogue record
- Prepare a data list
- Data
  - Convert formats?
  - Amend layout?
  - Check data-level documentation
  - Correct typos (do not remove, but use tags)
  - Disclosure review – anonymise? Potentially libelous/ scandalous content
Useful documentation

- Interview schedule or topic guide
- Observation templates/grids
- Diary template
- Stimuli e.g. scenarios, photos, images
- Field notes
- Outputs e.g. reports
- Consent agreement
- Errata
What to keep from analysis software

• Discipline-specific customs/ planned data analyses
• Proprietary nature of software
  • CAQDAS, e.g. NVIVO, Atlas-ti, Max-QDA
• What to keep?
  ✓ Clean transcripts, anonymised where needed
  ✓ Final coding frame – any open format
  ✓ Data list – data items categorised (finding aid)
  ✓ Export of any quantitative (coded) data tables
  ✓ Data ordered time-stamped memos (research notes), like a research diary
Catalogue

UK Data Service data catalogue record for:

Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003

TITLE DETAILS

SN: 5407
Title: Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003
Alternative title: Health and Social Consequences of the 2001 Foot and Mouth Disease Epidemic
Persistent identifier: 10.5255/UKDA-SN-5407-1
Depositor: Mort, M., Lancaster University. Institute for Health Research
Principal investigator(s): Mort, M., Lancaster University. Institute for Health Research
Sponsor(s): Department of Health
Grant number: 121/7499

SUBJECT CATEGORIES

Community and urban studies - Society and culture
Rural life - Society and culture
Qualitative study – user guide and doc

- A user guide could contain a variety of documents that provide context: interview schedule, transcription notes, even photos.
Qualitative study – data listing

- Data listing provides an at-a-glance summary of interview sets

<table>
<thead>
<tr>
<th>Respondent ID</th>
<th>Population Group</th>
<th>Date of Birth</th>
<th>Gender</th>
<th>Occupation</th>
<th>Interview summary</th>
<th>Place of Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM02</td>
<td>Group 6: Animal / Human Health Professionals</td>
<td>1975</td>
<td>M</td>
<td>Veterinary Surgeon</td>
<td>Family and background, career and work, arrangements during FMD epidemic and perceptions of situation</td>
<td>North Cumbria, responder's home</td>
</tr>
<tr>
<td>PM03</td>
<td>Group 6: Animal / Human Health Professionals</td>
<td>1966</td>
<td>F</td>
<td>Veterinary Surgeon</td>
<td>Family and background, career and work, arrangements during FMD epidemic and perceptions of situation</td>
<td>North Cumbria</td>
</tr>
<tr>
<td>PM07</td>
<td>Group 6: Animal / Human Health Professionals</td>
<td>1964</td>
<td>F</td>
<td>Veterinary practice manager</td>
<td>Family and background, career and work, arrangements during FMD epidemic and perceptions of situation</td>
<td>North Cumbria, responder's home</td>
</tr>
</tbody>
</table>
You can publish data nearly anywhere, but...

- Web sites
- Generic repositories – Zenodo, figshare
- Institutional repositories – University of Lancaster
- Domain Data Centers
  - UK Data Service https://www.ukdataservice.ac.uk/
  - ReShare – self archiving for social science data
- Journals (not many take data but increasing)
- Cloud
- Jisc UK Research Data Discovery Service project
- Find a repository:
  - Registry of Research Data Repositories-http://www.re3data.org/
  - Registry of Open Access Repositories (ROAR)
    - http://roar.eprints.org/
Consider trusted institutions (FAIR*)

- **Discoverable**: Publicly available and freely searchable
- **Usable**: Established machine-actionable digital formats
- **Meaningful**: Offer metadata and documentation to facilitate data re-use
- **Citable**: Assign persistent, globally resolvable, machine-actionable identifiers linked to specific versions of data
- **Secure**: Maintain policies and procedural controls to protect confidentiality and personal privacy as required by law and research ethics standards
- **Durable**: Ensure long-term preservation of and access

*Findable; Accessible; Inter-operable; Re-usable*
One option for sharing - ReShare

- Self-deposit repository for social research data-open to all

- Features:
  - Embargo option for publication
  - Set permissions at the file-level

- UKDS staff review data collections:
  - Confidentiality and copyright
  - Documentation and file formats

- reshare.ukdataservice.ac.uk
Easy to publish and upload data
ReShare in Discover - FAIR

Discover

Search and browse our data collections, support guides, case studies, and related publications.

Type

Subject

Date

Data type

Results per page: 10

Displaying 1-10 of 15 results

1. SN 851150 Investigating Adolescent Violence towards Parents
   Rachel Condry, University of Oxford

2. SN 850417 Contested development?: intimate partner violence and women's employment in urban and rural Tanzania

Sorted by: Relevance

Download | DDI XML | Similar data collections
Data sharing outcome – Foot & Mouth

- sought advice from copyright specialist re. terms of agreement for archiving
- met with UK Data Service for advice data archiving
- developed separate consent forms for written and audio material, with opt in / opt out and an embargo option
- piloted discussion on data archiving with 4 panel members to explore:
  - feelings re. data anonymisation, confidentiality, copyright, ownership
  - understanding of archiving by participants and information required
  - user options of archived data - scholarly / educational purposes
- discussed archiving individually with each panel member
  - 7 panel members declined archiving their data
  - 40 interview and diary transcripts were made available for re-use by registered users
  - 3 interviews and 5 diaries were embargoed until 2015
  - audio files archived and only available by permission from researchers

Detailed information: [www.esds.ac.uk/findingData/snDescription.asp?sn=5407](www.esds.ac.uk/findingData/snDescription.asp?sn=5407)
Open access data

It is our explicit long-term goal to work with data owners to identify and remove all unnecessary barriers to access.

An increasing number of our datasets are available without registration or authentication using open data licences described in our Data Access Policy. These data are for use with an open licence and are not classified as personal. We are also working to gain open data certification via the Open Data Institute.

We also provide links to other open data resources that may be of interest.

<table>
<thead>
<tr>
<th>Census data</th>
<th>International microdata</th>
<th>Qualitative data</th>
<th>Survey data</th>
</tr>
</thead>
</table>

Qualitative and mixed methods data available under a Creative Commons Licence.

- SN 4867 School Leavers Study, 1978
  Access via Qualidata | More about this dataset
  Access via Qualidata | More about this dataset
- SN 7465 Moral and Home Intelligence Reports, 1941-1949
  Access via Qualidata | More about this dataset
- SN 5252 British Oral Archive of Political and Administrative History, 1920-1930
  Access via Qualidata | More about this dataset
- SN 2000 Family Life and Work Experience Before 1918, 1670-1973
  Access via Qualidata | More about this dataset

We also make data extracts from some qualitative collections available via these open online teaching resources:

- Teaching resource: Interview methods
- Teaching resource: Non-interview methods

---

Maggie Mort

"Gaining consent retrospectively"

Owners and producers

ORCID

- Depositor stories
  - Pat Caplan
  - Maggie Mort
  - Jane Elliott
  - Sheila Henderson

Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003

About the research

The 2001 foot and mouth disease epidemic had an enormous impact on the economic, social and political life of Cumbria. A 30-month research project, funded by the Department of Health, was designed to produce evidence specifically about the human health and social consequences of the epidemic.
Our data management guidance

• Online best practice guidance: ukdataservice.ac.uk/manage-data.aspx
• Managing and Sharing Research Data – a Guide to Good Practice: (Sage Publications Ltd)
• Anonymisation tool - http://data-archive.ac.uk/curate/standards-tools/tools
• Helpdesk for queries: ukdataservice.ac.uk/help/get-in-touch.aspx
• Training: www.data-archive.ac.uk/create-manage/advice-training/events
…how do we design systems that make use of our data collectively to benefit society as a whole, while at the same time protecting people individually?…This is it: this is the fundamental issue of the information age.”

Bruce Schneier 2015 Data and Goliath
Questions

Libby Bishop

ebishop@essex.ac.uk