Secondary Analysis of Qualitative Data: Sharing and Reusing Data

Libby Bishop
UK Data Archive and GESIS

Maureen Haaker
UK Data Archive and University of Suffolk

Essex Social Science Summer School
10-14 July 2017 – Wivenhoe Park
Plan for the week

- **Day 1** – Introduction to SAQD
- **Day 2** - Key issues in SAQD
  - Context
  - Ethics
- **Day 3** – Methods
  - Sampling
  - Learning methods from data
- **Day 4** – Data exchange exercise
- **Day 5** – Accessing, using, and sharing data
Day 1 – Introduction to SAQD

• Pros and cons of reusing qual data
• Whys and ways of reusing qual data
• A case study of data reuse
• How might you reuse examples of data?
Reusing Qual Data – pros and cons

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

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

• Take about 10 minutes in groups, then we will discuss.
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...
### Why engage in re-use of “old” data?

<table>
<thead>
<tr>
<th>Why engage in re-use of “old” data?</th>
<th>May prevent unneeded data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>new questions of existing data</td>
<td>repetitive</td>
</tr>
<tr>
<td>methodological study of original research</td>
<td>burdensome on vulnerable populations</td>
</tr>
<tr>
<td>teaching</td>
<td></td>
</tr>
<tr>
<td>comparative</td>
<td></td>
</tr>
<tr>
<td>across time</td>
<td></td>
</tr>
<tr>
<td>across locations</td>
<td></td>
</tr>
<tr>
<td>cross-sectional sub-samples</td>
<td></td>
</tr>
<tr>
<td>make better use of painstakingly collected, rich, extensive data</td>
<td></td>
</tr>
<tr>
<td>cost/time savings(?)</td>
<td></td>
</tr>
</tbody>
</table>

| May allow more energy for           |                                     |
| research questions                  |                                     |
| causal mechanisms                   |                                     |
| methods development                 |                                     |
| theory development                  |                                     |
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
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.
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.
Who would re-use data?

- Mike Savage (York)
  - Class and class identity

- Jane Elliott (soon to be at Exeter)
  - On being a good mother

- Dawn Lyon (Kent) and Graham Crow (Soton)
  - Revisiting Ray Pahl’s Sheppey studies

- Jo Haynes (Bristol)
  - Reusing data for PG methods teaching

- Julia Brannan (IoE)
  - Food and meal practices

ESRC Secondary Data Analysis Initiative – Phase 3
Re-use purposes of qualitative data downloaded from UK Data Service, 2002-2016

Source: Bishop & Kuula-Luumi, Sage Open 2016
Examples of re-using “old” data

- assess the credibility of new research or the generalizability of small studies (Hammersley 1997)

- supplement one’s own primary data, e.g. as exploratory analyses prior to new data collection (Hinds et al. 1997)

- provide rich descriptive information, e.g. an historical perspective (Bornat 2005, Gillies & Edwards 2005)

- reveal new methodological insights (Mauthner et al 1998, Savage 2005; Bornat 2010)

- generate new findings by analysing ‘old’ data from a ‘new’ research context (Holland & Thomson 2009; Bornat 2010; Walkerdine and Lucey 1989)

- gain insight on hard to reach populations or sensitive topics without further intrusion into vulnerable populations (Fielding & Fielding 2000)

- teaching (Haynes 2011) (Irwin & Winterton, 2010)
Explosion of qualitative data archives

- UK Data Service – over 1000 data collections
- Germany, Finland, France, Switzerland, et al.
- ARK – N. Ireland - The Troubles Archive
- Australian Qualitative Data Archive
- Irish Qualitative Data Archive
- UCL – Human communication AV archive
- Timescapes – qual longitudinal data on families
- Qualitative longitudinal data in Europe – 14 countries

and “analysis of secondary qualitative data” a recognised methodology by the National Centre for Research Methods in 2015
Many ways of re-using data

• Description – literature review with data…
• Research design and methodology
• Re-analysis – new questions of existing data
  • When might it be an advantage for the topic for reuse to NOT be the same as the primary research?
• Repurposing – e.g. keyword analysis of illness narratives
• Learning and teaching
In 1978, teachers at a comprehensive school on the Isle of Sheppey were asked to set an essay about 10 days before pupils were due to leave school. The essay asked students to imagine that they were nearing the end of their life, and that something had made them think back to the time when they left school. They were then asked to write an imaginary account of their life over the next 30 or 40 years.
How can the data be used?

Living and Working on Sheppey, Dawn Lyon and Graham Crow

1. Digitised the original 1978 handwritten data.
2. Coded and compared the data across time

Together, these two sets of essays shed light on the aspirations of Sheppey’s young people (and young people more generally) and cover a range of topics including health, education, career, family and leisure.
School Leavers Re-Study: Imagining the Future

Re-users: Graham Crow and Dawn Lyon
Living and Working on Sheppey

1978: 141 essays (89 boys and 52 girls)


2009-10: The exercise was repeated by the Living and Working on Sheppey project and 110 essays (55 boys and 55 girls) were gathered from school pupils and members of youth groups on the Isle of Sheppey to compare to the earlier ones.

Together, these two sets of essays shed light on the aspirations of Sheppey’s young people (and young people more generally) and cover a range of topics including

- Health
- Education
- Career
- Family and leisure
School Leavers Re-Study

Essay instructions 2010: Imagining the Future:
I want you to imagine that you are towards the end of your life. Look back over your life and say what happened to you. Don't write a very exaggerated story, just tell the straightforward story of your life as it might really be. Of course you cannot know what is going to happen to you, but you can describe the sort of thing that could happen if things go as you expect or hope. Spread your story over your whole life from the time of leaving school. Continue on another sheet as necessary.

Sheppey: Essay 30
Group: Boys

My Past

I was sitting in front of the fire smoking my pipe, just thinking of the past. I was nearly seventeen when I left school, I had an apprenticeship in the Tudor Glass, an electrical apprenticeship lasting four years, I served my four years and got my diploma for an electrical engineer. I was now twenty and courting a nearly lovely girl. I had to work in London in my next job and came back home every weekend. Jobs were scarce so you had to take what you can get, I didn’t want to go to London, but we did the only thing we could, I was earning $70 but not much $35 a week.
<table>
<thead>
<tr>
<th>Year</th>
<th>Description</th>
<th>1978</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1978</td>
<td>Mundane &amp; grounded jobs; gradual career progression; periods of unemployment.</td>
<td>“It was hard finding a job, I failed a few chances, but eventually got what I wanted locally, a craft apprenticeship.” (Essay 27, male)</td>
<td>“I could not decide what to do. Then it came across my mind that I should be a body builder, and be good enough to enter the Mr Olympia contest. After that I went to an athletics club, within a few months I was scouted for the Olympics for the 200 metre sprint.” (Essay 10, male)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“I was on the dole for six months after leaving school, until I got a job in a garage.” (Essay 42, male)</td>
<td>“I was 20 now living the dream I had a amazing band...I had toured the world 3 times sold 4 million records.” (Essay 30, male)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“When we found a house it was a semi-detached in Sittingbourne. I wanted to live in Italy but that was asking for too much.” (Essay 56, male)</td>
<td>“I arrive at my 3-bedroom luxury villa; I land my helicopter on my own heli-pad and walk inside. I grab my keys and jump in my Bentley Continental GTS.” (Essay 40, male)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“I longed for something exciting and challenging. But yet again I had to settle for second best. I began working in a large clothes factory.” (Essay 104, female)</td>
<td>“In my future I want to become either: a dance teacher, hairdresser, or a Professional Show Jumper/horse rider. If I do become a dancer my dream would be to dance for Beyoncé or someone really famous.” (Essay 61, female)</td>
</tr>
</tbody>
</table>
Welcome to the Living and Working on Sheppey: Past, Present and Future project website.

Living and Working on Sheppey explores the recent history and changes in working lives on Sheppey in the last decades of the 20th century and into the 21st. It concentrates on the period since the closure of the dockyard at Sheerness in 1960.

The site contains materials from older people and their memories of work in or around the dockyard and everyday life in Blue Town, the area of Sheerness closed to the dockyard. You can listen to clips of these interviews on different themes including the closure of the Dockyard, various trades, and changes in family life across different generations, and you can read the interviews in full (you need to register to do that).

You can read and listen to short essays written by young people in 2009-10 on what they imagine their futures have in store for them. They talk about their hopes and aspirations for their working lives, families, and travel as well as some of the difficulties they foresee.

You can watch the two videos made by the artists, Tea, as part of the Living and Working on Sheppey project. In ‘Back and Forth on High Street Blue Town’ you can see images of Blue Town High Street past, present – and future! And you can hear older people’s memories of everyday life on the High Street through the nineteenth century. In ‘Sheerness Port’, you can go on a journey around the site of the former Dockyard today.

The Living and Working on Sheppey project was funded by the South East Coastal Communities Programme from 2009 to 2011. See Peter Hatton and Jenny Hurkett and discuss the project, the Blue Town Heritage Centre and the Isle of Sheppey in the video clip below.

http://www.livingandworkingonsheppey.co.uk/
School Leavers Study: publications

How could you reuse these data?


How would you reuse this data?

Consider:

• What was the original question of the research project?
• What are 2 new questions you might ask of the data?

When brainstorming new questions, you might consider the following types of reuse:

• Re-study
• Repurpose/re-analysis
• Description
• Teaching and learning
• Methodological advancement
• Comparative study
Day 2

• The Context Debate
Constructing knowledge/reusing data

What is distinctive about qualitative research?

From this point of view, data cannot be first collected and analysed and then ‘re-used’ by other researchers for the purposes of ‘secondary’ analysis. Indeed, data cannot even be ‘collected’ in the first place because they are always constructed, as Bateson (1984) pointed out long ago in the context of survey research (Hammersley 2010).
“For one thing, in the process of data collection researchers generate not only what are written down as data but also implicit understandings and memories of what [primary researchers] have seen, heard, and felt, during the data collection process. For instance, not everything that is experienced in the course of participant observation or in-depth interviewing will be or can be written down. Nevertheless, it will often be drawn on tacitly, and perhaps sometimes consciously, in the course of analysis, perhaps playing an important role in making sense of the data that have been recorded” Hammersley (2009, para 3:4)

But do these insights preclude others from their own substantive insights?
The context argument

• Data do not exist independently of the contexts in which they were produced or (co)constructed or generated.

• That context is—by definition—inaccessible to subsequent researchers.

• Anyone conducting SA lacks the context known by the primary researcher(s), “head-notes”.

• Thus all SA is inevitably limited to providing methodological, but not substantive, insights because of this lack of contextual information (Mauthner-several).
4.3 However, there is a counter-intuitive quality to this argument that ought to trouble us. After all, surely we do not and should not *make up* our data? I take this to be true whether or not one is a realist or even a ‘latent positivist’ (Mauthner et al 1998:736 and 743; Moore 2006:11; Moore 2007:3.3), though there are no doubt some who would challenge the point.[5] What this means is that the data must in some ways constrain what inferences we make and the conclusions we reach, rather than being freely constructed in and through our inferences. And this implies that they must, in some sense, exist prior to and independently of the research process (Hammersley 2010).
“Being there” is not the “be-all and end-all”

- Depends on primary research design and secondary research questions
  - Ethnography vs. semi-structured interview
  - Content analysis
- Even primary researchers miss features of context that later prove salient
- Presence/closeness may conceal as well as reveal
  - What primary researcher “knows” is not always right
- Distance may reveal new understandings
  - “Some forms of interpretation are possible only from a distance” (Mason 2007)
  - Wilson (2014)—youth interviews—proximity & distance
The context debate: response

• Hammersley (2010): data are both given and constructed

  • **Data**: that which is collected or generated in the course of research; but cannot be *completely* constructed.

  • **Evidence**: the analysed data which provides the grounds for inference and for the descriptive and explanatory claims which are built on the data.

  • There is temporal and conceptual overlap, but evidence is more constructed than data (my wording).

• How does this connect with context again?

  • access to context (‘head notes’) may give primary researcher more privileged relationship to some “data as given”, but

  • does NOT imply privileged relationship to “data as evidence”, *interpretation* (Irwin and Winterton 2011)
Summary of the arguments

• Primary researchers have more privileged knowledge of, and access to, primary data but both primary and secondary analysts will construct data as evidence in the service of some empirically grounded set of arguments and knowledge claims.

• How effectively such arguments are made can be judged against the criteria of social scientific explanatory adequacy. *Presence at the point of data generation is not a final arbiter.*

• Overplaying the significance of proximate context relative to other salient factors may risk privileging description over explanation

• Theorising and analysing context needs to be part of a critical secondary analysis.
“Primary analysts have a privileged relationship to the data they have generated, but do not necessarily have a privileged claim on the arguments which can be made from that data. Sociological data will support different theoretical understandings, and ‘being there’ is not the final arbiter of the adequacy of such understandings”

(Irwin and Winterton 2010)
Day 2

- The context debate
- Ethical questions
Ethical questions about data re-use

- Can *consent* for unknown future purposes be informed?
- does sufficient *anonymisation* for re-use damage data quality?
- does archiving data increase risk of *misuse*?
Participants share their data more than we predict

- 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 (Arja Kuula, IASSIST Quarterly)
  - re-contact project: life stores, gender, etc.
  - 165/169 (98%) agreed

- even 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

- 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
Consent, anonymisation, and access

- Ask for consent to share – researchers must be informed about risks and benefits of data sharing
- Anonymise – only if damage to data is minimal (not images)
- Regulate access
  - End User Agreement (UK Data Archive)
  - Embargo
  - for selected sensitive or disclosive data – registered users; permission from data depositor

These strategies enable most data to be shared
Anonymising qualitative data

- plan or apply editing at time of transcription except: longitudinal studies - (linkages)

- 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

- avoid blanking out; use pseudonyms or replacements

- avoid over-anonymising - removing/aggregating information in text can distort data, make them unusable, unreliable or misleading

Controlling access a better option than over-anonymising
Data access spectrum

- Open
- Safeguarded
  - End User Licence
  - Special Conditions
    - Online agreement
    - Depositor permission
    - Special Licence
- Controlled – secure remote access
- Possibility for multiple versions
- UK Data Archive - certified for its secure data handling procedures under the international ISO 27001 standard for information security
Risks of mis-use in re-using data

• Researchers’ reputations (senior and junior)
• Harms to participants
  • Disclosure of information
  • Their views or opinions misrepresented
    • what if another researcher interprets “my” participants’ words differently?
    • Consider argument in light of all kinds of participants: terrorists, paedophiles, Ku Klux Klan, other hate groups
• Comes back to role of researcher
  • Respect participants, represent their views
  • But not unreflexively, not uncritically, always as part of analytic work
  • Interpretations must be adjudicated openly, in publications, and (where possible) based on shared data
Who owns the data?

“But it’s also the notion of intellectual property, isn’t it? Whose intellectual property is that stuff there? We say it’s – we put our stamp on it, it’s our intellectual property.”

“I’m sorry [laughs] I don’t agree with that. I think there’s – I mean I see research as being a public benefit. It’s publicly funded; it’s for public benefit. I also see research as being intrusive and demanding of the participants and so therefore what the participants record … is of value and I think archiving it, even if it had a 30 year embargo on it, is actually paying respect to what people have said and building up a stock of the world’s knowledge.” (Broom et al. 2009)
Day 3 – sampling and other methods issues

- Sampling – Jefferys and Bornat as background
- Sampling – Gallwey – read background then do exercise
- Other methods – using data to learn methods with UK Data Service resources
Fit, lack of fit, and sampling

- Lack of fit (sample not suited for RQ)
  - It is a problem in much primary research as well
  - But even more so in QSA (e.g., no ability to probe)
- Some tools for sampling are available, but limited
  - No unified portal for data archives
  - Search possible only of metadata, or at collection level
- Growing list of exemplary practices
  - Bornat et al. (2012) geriatrics OH interviews
  - Gallwey (2013) single motherhood
Sampling – adding new data to existing

- One solution – collect new data to match, or fit with what you know already exists
- Bornat, et al. (2012) Revisiting the archives: a case study from the history of geriatric medicine
- Primary data – Margot Jeffries – origins of geriatrics
Sampling – rationale for 1ary and 2ary samples

- Jeffreys project – 1991, origins of geriatric medicine
- Primary Sample
  - 72 people – many roles – politicians, medical civil servants,
  - 54 geriatricians –
    - 18 born before 1914, oldest was 92
    - Qualified at doctors before WW2
    - All but one were white
- Bornat’s project – experience of S Asian doctors and meaning for geriatrics, her sample..
  - 60 geriatricians from S Asia, worked in England or Wales
  - 2/3 entered labour market in 1976 or before
  - All but 5 male
  - Snowballing used – any issues?
  - S Asian doctors less willing to deposit in British Library-why?
Case Study: Dr April Gallwey

The rewards of using archived oral histories in research: the case of the Millennium Memory Bank

by April Gallwey

Abstract: This article is about using archived oral histories in research. It advocates greater exploration of oral history archives in the UK in light of burgeoning research into the history of post-war society and the social sciences. Drawing on the author’s experience of utilising interviews from the Millennium Memory Bank (MMB), it offers a critical reflection on the process of using this particular archive from the perspective of a social historian. Although the challenges of working with the MMB are noted, its unprecedented scale, and incorporation of the ‘life history’ method, demonstrate its value as a source for historical enquiry. The author highlights the importance of summaries and supporting literature attached to interviews to assist with accessibility and contextualisation, envisioning the archive as an ongoing process, whereby original and subsequent researchers ensure its legacy.

Key words: oral history archives, social history, life history interview, Millennium Memory Bank, re-use, secondary analysis, single motherhood

During the course of the 20th century, increasing amounts of qualitative data collected by social scientists – field notes, interview transcripts, letters, drafts and the like – have been collected and archived. [...] Until recently, these documents were largely ignored: social scientists did not know how to use them since they did not conform to standardized norms from which trends could be inferred. Historians did not have the experience of using such sources and in many cases were ignorant about them.¹

In this statement, Mike Savage captures how the trail of the social scientist alters our gaze onto the historical landscape of the twentieth century. At the same time he foregrounds how a reluctance amongst social scientists to conduct secondary analysis of qualitative data and the negation of such sources by many historians, has until relatively recently, left these sources under-utilised.² Savage has argued for greater inclusion of archived qualitative data and demonstrated how social research archives, such as the Mass Observation Archive (MOA), profoundly illuminate historical shifts at the socio-cultural level. Savage’s most recent work places the social science archive at the heart of his project for a historical sociology of the post-war years.³ Debate about re-use of data has been circulating amongst sociologists for some time, generating a very substantial body of literature.⁴ Although there has been a recent turn to archived social scientific data amongst historians – Paul Thompson has highlighted the multi-
Resources

Reusing qualitative data. [http://ukdataservice.ac.uk/use-data/secondary-analysis/reusing-qualitative-data](http://ukdataservice.ac.uk/use-data/secondary-analysis/reusing-qualitative-data)


[http://www.data-archive.ac.uk/about/staff?sid=ebishop](http://www.data-archive.ac.uk/about/staff?sid=ebishop)

Questions

Libby Bishop

ebishop@essex.ac.uk