Issues in Qualitative Secondary Data Analysis: Context, fit, and ethics

Libby Bishop
Producer Relations and Research Ethics

Fourth Swiss Methods Festival
Qualitative and Mixed Research Methods
7-8 September 2015 - Lausanne
Topics for today

• Reusing qualitative data – for and against
• An example of reusing qualitative data
• Many ways to reuse data
• Key methodological challenges
  • Context
  • Fit (and sampling)
• Ethics issues when reusing data
• Data exchange group exercise
• Optional: accessing data at UKDA
• Close
Reusing qualitative data

• FOR....

• AGAINST....
School Leavers Study

Original data – collected c. 1978  Ray Pahl

Teachers at a comprehensive school on the Isle of Sheppey were asked to set a particular essay to those pupils who were students in English lessons about ten days before they were due to leave school. The students were asked 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.

The data: 142 handwritten essays by school leavers aged 15 and 16 years old.

http://discover.ukdataservice.ac.uk/catalogue/?sn=4867&type=Data%20catalogue
Living and Working on Sheppey

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.

The data: 110 essays by school pupils and youth group members aged about 16 years old.
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.
Living and Working on Sheppey
What Sheppey’s young people said in 1978
141 essays (89 boys and 52 girls)

*Mundane & grounded jobs.*
*Gradual career progression.*
*Periods of unemployment.*

- “It was hard finding a job, I failed a few chances, but eventually got what I wanted locally, a craft apprenticeship” (Essay no.27, male)

- “I was on the dole for six months after leaving school, until I got a job in a garage” (Essay no.42, male)

- “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 no.104, female)
Living and Working on Sheppey

What Sheppey’s young people said about work in 2010
110 essays (55 boys and 55 girls)

Well-paid, instantaneous jobs.
Choice but uncertainty.
Influence of celebrity culture.

• “I was 20 now living the dream I had a amazing band...I had toured the world 3 times sold 4 million records” (Essay no.30, male)

• “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 no.40, male)

• “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 no.61, female)
Many ways of re-using data

• Description – literature review with data…
• Re-analysis – new questions of existing data
• Repurposing – e.g. keyword analysis of illness narratives
• Learning and teaching
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)
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 exercise - Gallwey
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**?
Consider: could it be unethical NOT to share data?

• Duty to scholarly community - to be transparent

• Duty to public – to be trustworthy custodians of public funding

• Duty to participants – not only to protect, but also to amplify, their voices
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
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
Resources

Reusing qualitative data. http://ukdataservice.ac.uk/use-data/secondary-analysis/reusing-qualitative-data


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

Timescapes Guide 19 – Qualitative Secondary Analysis, Irwin and Winterton
## Exercise: Reusing qual data

<table>
<thead>
<tr>
<th>Name</th>
<th>Topic</th>
<th>Name</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valentine</td>
<td>Amputee identity</td>
<td>Evi</td>
<td>Doctor/patient relationship</td>
</tr>
<tr>
<td>Laura</td>
<td>Construction work</td>
<td>Leon</td>
<td>Power lines/community</td>
</tr>
<tr>
<td>Kathrin</td>
<td>Gem stone trade</td>
<td>Angeline</td>
<td>Non personal identifiers</td>
</tr>
<tr>
<td>Patricia</td>
<td>Regional identity</td>
<td>Adrienne</td>
<td>Roma identity</td>
</tr>
<tr>
<td>Emilie</td>
<td>Goals of elderly</td>
<td>Brian</td>
<td>Borrowed data-love?</td>
</tr>
</tbody>
</table>

Silva, Anaid, Marieke and others without own data, pick a group to join.

### Instructions:
1. Read first extract – **DO NOT READ EXTRA INFORMATION**
2. Discuss questions: ideas for reuse, then role of context
3. Repeat with second extract
Questions

Libby Bishop

ebishop@essex.ac.uk