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

Libby Bishop

Producer Relations and Research Ethics

Third Swiss Methods Festival
Qualitative Research Methods
13 September 2014 - Basel

UK Data Service



Topics for today

- What is Qualitative Secondary Data Analysis?
- Key methodological challenges (and responses) to re-using qual data
 - Context
 - Fit (or sampling)
- Ethical challenges (and responses) to re-using qual data



Many ways of re-using data

- One team member using another's data
- Historians using documents
- Conversation analysis
- Textual data miners
- Defining features
 - Who produced the data
 - For what purposes
 - Have the data been analysed (or only collected)

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).

The context argument against reuse

- 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).

Limits on the construction of data

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/intuition 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 (2013?)–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



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 extend 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



Questions

Libby Bishop

ebishop@essex.ac.uk



Moral rights to 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)

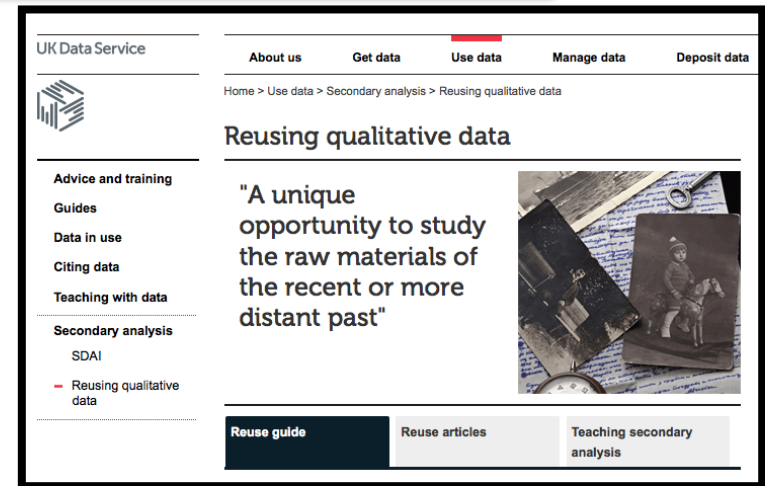
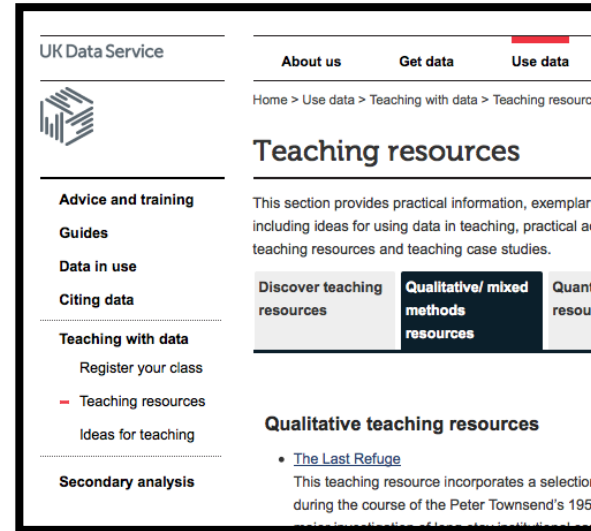
Context

... there may be a risk that undue emphasis on context and ‘presence’ itself encourages a particular mode of engagement with data, a mode which tends towards description, and a sense that the ‘answers’ reside within the specific data set, rather than in theorisation of how the data articulate evidence to address specific research questions.

Irwin & Winterton, SRO, (forthcoming)

Strategies to promote re-use

- Teaching
 - Guides and data
 - Post-grad & advanced
- Research
 - SA guides
- Case studies & workshops
- Collaborations
 - Digitise collections
 - Embed SA in curriculum





Qualitative interviews



Ethical reuse



Structured interviews



Semi-structured interviews



Unstructured interviews



Feminist interviews



Life story interviews



Oral history



Psychosocial interviews

- Open access
- PDFs
- Tiny datasets
- With teaching activities