



## **A Proposal for Archiving Context for Secondary Analysis**

**Libby Bishop**

University of Leeds

### **Abstract**

This paper seeks to provide a framework for thinking about context for secondary analysis, and to suggest minimal guidelines for providing context for qualitative research to enable its reuse. It addresses the challenge of what re-contextualisation of qualitative data looks like in detail and in practice. In the first section, we lay out a framework for thinking about context that entails looking in two dimensions: multiple levels of granularity and across time. The first dimension includes the levels of interaction, situation (setting and project) and culture/institutional. The longitudinal dimension highlights the temporal nature of contexts: there is context at the time of the original research and the context at the time of reuse. The later sections of the paper propose specific suggestions for what information should be included when researchers seek to provide context when depositing data for later use in secondary analysis. Our goal is to propose guidelines that are minimally burdensome to researchers preserving their materials (with ESDS Qualidata or elsewhere) while simultaneously being maximally helpful to researchers seeking to reuse data. Our recommendations are—as they must be—a conditional specification of context specific to data archiving and not a generic prescription.

### **Keywords:**

### **Introduction**

Context is a long-standing conundrum in qualitative research. The task is to make sense of utterances (or gestures, images, etc.) that are fundamentally deictic, or indexical. That is, they cannot be made sense of apart from the contexts in which they were produced. This interpretive difficulty is present in all qualitative research, but poses a particular challenge in secondary analysis when data produced in one context are then used in another.

At ESDS Qualidata, we are called upon to advise researchers preparing their data for deposit as to ‘how much’ and ‘what kind of’ context they should provide. Being only slightly flippant, two answers predominate in the literature: ‘it depends’ and ‘everything’ (Holstein and Gubrium, 2004). Neither answer is particularly helpful to researchers with very limited time and funding making a good faith effort to fully prepare their data for archiving. This paper seeks to provide a framework for thinking about context for secondary analysis, and to suggest minimal guidelines for providing context for qualitative research to enable its reuse. In doing so, we need to be very clear what we are and are not attempting. We are not saying that these guidelines are an absolute standard of context, suitable to all projects. Our recommendations are—as they must be—a conditional specification of context specific to data archiving.

Our goal is to propose guidelines that are minimally burdensome to researchers preserving their materials (with ESDS Qualidata or elsewhere) while simultaneously being maximally helpful to researchers seeking to reuse data.

Most researchers agree that the value of qualitative data for secondary analysis is dramatically enhanced when extensive context is provided (Fielding, 2004; Heaton, 2004). However, there is much confusion as to the purposes of that context. In some cases, the argument is put forward that secondary analysis is attempting to, and should, recreate the context of the original project (Mauthner, et al. 1998). As various approaches to re-using data have been put into practice (Savage, 2005; Thane, 2006; Johnson, [in press]), it [has become/is becoming] increasingly obvious that good secondary research is more nuanced and is not, in fact, recreating context at all. In her extremely useful reframing, Moore (2005) redefines context for secondary analysis: the objective is not to recreate the context of the original project, but rather to recontextualise the production of new data.

Thus secondary analysis is not the analysis of pre-existing data; rather ‘secondary analysis’ involves the process of re-contextualising data [T]hrough recontextualisation, the order of the data has been transformed, thus secondary analysis is perhaps more usefully rendered as primary analysis of a different order of data. [I’m not sure of the convention to use for indented quotes so I’ve used the one I’m familiar with i.e. Indented left and right with no quotation marks]

This paper supports this approach to secondary analysis as recontextualisation and then goes on to ask: what does recontextualisation look like in detail and in practice? In the first section, we lay out a framework for thinking about context that entails looking in two dimensions: multiple levels and across time. The later sections of the paper propose specific suggestions for what information should be included when researchers seek to provide context when depositing data for later use in secondary analysis

### **Multiple levels of context**

It is quite clear that there is not one context to account for, but many. One initial framing for these multiple contexts is along a continuum of scale or distance. There is a sense that at one end is the interaction itself and at the other, some cultural factors that are overarching, yet still deemed sufficiently related to the interaction as to be part of its context. Holstein and Gubrium (2004) use a framing of ‘proximal’ and ‘distal’ and mark three points: the conversation or interaction, the situation, and finally, the cultural or institutional. Working more explicitly in a framework of discourse analysis, van den Berg (2005) proposes a quite similar schema: intra-discursive, the conditions of discursive produce, and finally, extra-discursive (see Table 1 in appendix). Duranti and Goodwin (1992) use slightly different categories such as ‘setting’ and ‘extrasituational’, but still proffer a similar framing of near and far. The division into three is no doubt arbitrary—any given project might have more—but it seems to be a useful minimum. While finding some consensus in the idea of levels, it is also essential to reinforce Holstein and Gubrium’s (2004) point that no level is more important than another; they are ‘mutually constitutive’.

### **Contexts at different points in time**

The framing above clearly applies broadly to primary and secondary analysis. However, secondary analysis adds a longitudinal dimension to the framework. There are at least two points in time at which context needs to be considered: the period when the original project was done and the period for the reanalysis project. (See also Moore, this volume.) This section will examine issues that arise when considering each level of the contextual framework across time.

*Interactional (the conversation)*

What is the secondary researcher likely to discover about the conversation or interaction when looking back at the original project? Inevitably, the secondary researcher will have fewer materials available. How much less depends entirely on the completeness of the original preservation. In most cases, the key artifact preserved is an interview transcript. There can be many additional kinds of materials, ethnographic observations, video, etc. It is a practical reality that the majority of ESDS Qualidata's materials are interviews, so we are using these as the archetypical case, though we also hold numerous audio tapes as well. In many instances, social science researchers prefer working with transcripts, partly because of the time-consuming nature of listening to audio. Until recently, archiving technologies did not support preservation of more complex materials in digital form. Rapid advances are being made and we are now archiving a large collection of digital audio, and moving quickly into digital video, and web-archiving.

An addition to fewer records overall, there is the problem that the secondary researcher 'wasn't there'. Some argue that this gap is an irreplaceable bit of context that rules out most forms of data reuse (Parry and Mauthner, 2004). How should we think about the context afforded by being there? In the original project, there are two categories to consider, the 'event' itself, and then any recordings or transcriptions made as a record of those events. When, exactly, are data made? Does it happen while reading background literature, while creating a sample, during an interview, while reading of transcript, or while composing an analytical memo? Of course, data are constructed during all these phases. In secondary research, the interview experience itself is missing.

While in no way diminishing this gap, the fact is, once the interview, event, or fieldwork period has been completed, both primary and secondary researchers are in similar positions. The only possible relationship for construction of data is between the researcher and the record. Of course, the primary researcher will have memories of the event. However, no record of the event—memory, text, digital, even video—is 'the same' (Lyon, 2003). Any researcher, primary or secondary, must rely on whatever records were made and preserved. Data are co-constructed whether reading transcripts or doing an interview in real-time.

In summary, both primary and secondary researchers are constructing data, but they do have different contextual resources to work with. The primary researcher will have the context of having 'been there' with all the richness and depth that affords. The secondary researcher will lack that specific context, but will have any preserved artifacts. Both engage in constructing meaning based on the conversational contexts available to them.

*Situational*

Under the heading of 'situation', Holstein and Gubrium (2004) clearly have in mind a setting. In one of the examples they use at length, the conversation is between a patient and an examiner in a courtroom. They are suggesting that the relevant setting includes actors located in a bounded physical space and time. In addition to this definition of setting, it is also useful to distinguish 'project' as a special part of the situation. Because so much research is conducted in formal, funded projects, it is also helpful to distinguish a set of contextual features of a 'project' as a distinct subset of the overall setting.

*Setting*

The setting is usually what is often being talked about when people talk about the 'context' of qualitative work. This is the case because it captures a wide range of factors (who is present, how are they related, physical setting, appearances, body language, etc.). As an aside, it is an unfortunate unintended consequence of more rigorous research ethics that many of these authentic, visceral reactions of researcher are now deliberately not recorded for fears of violating research ethics.

Because so much of this kind of information is taken in tacitly, it is difficult for a researcher to explicitly recall and record most of it. When any record is made at all, it is often minimal: a sketch of a room or an interviewer's brief note on the appearance of a respondent. This is the area of contextual materials that is usually the most found wanting. In fact, a key aim for this paper is to better specify a minimal set of such information as a guide for researchers.

### *Project*

In most cases, especially where research was funded, a reasonably good record exists outlining a project's objectives, research questions and methodology, and final reports. Obviously, any published works augment this record. A secondary researcher has access to these materials; their relevance, as always, will be shaped by considerations of both the original research questions and the new ones.

Any project, however, has multiple narratives. There is one version that makes it to final reports and publications, and the other, 'messy' version that doesn't. In rare cases, there are multiple published versions of the 'real' story, or owning up accounts (Bell, 2004). But the fact is, no account, or even multiple accounts, can fully describe what happened. First, there is the fact that process as documented is always different from process as practiced. There is no way for any version of a research project to provide some kind of 'authoritative' account (Hammersley, 1997). Second, there are fears—and very legitimate ones given research review regimes—of showing any kind of messiness at all. For all these reasons, even when researchers have tried to document fully and faithfully, it is typically the case that the secondary researcher will have the official stories, but only a partial window into other versions.

### **Proposal for minimum context at three levels**

#### *Conversation*

At the risk of stating the obvious, any recording of context has to begin with preserving a record of the conversation. (In the following sections, we are referring to a proposed Model Transcript, see Figure 1.) The guidelines suggested here are specific with respect to archival purposes, not research. For example, the discussion of what kind of transcription method to use is beyond our scope. Whether complete Jeffersonian annotation is used (Sacks and Schegloff, 1974), or something simpler, is a decision for the researcher to make, according to their analytic intentions. However, we can specify minimal standards of transcription that relate directly to best practices for archiving. Of course, in many cases, there is considerable overlap with good research practice. A transcript must provide an accurate recording of audible utterances. In some cases, non-word sounds and gestures may be recorded, again, this is a matter for the researcher to decide. For any transcription, speaker tags (initials or a code indicating who is speaking) are essential. This is obvious for simple readability, however, it is even more important for some of the technical processes necessary to systematically display, search and retrieve data. If the identities of participants in an interview are not known, we will adopt a convention of I (interviewer) and R1 (respondent 1), R2 (respondent 2), and so on. Double spaces between utterances again enhances readability, but is also a requirement for some CAQDAS packages for auto-coding (further guideline are here [LINK\\*](http://onlineqda.hud.ac.uk/Step_by_step_software/CAQDAS_data_prep/index.php))

[http://onlineqda.hud.ac.uk/Step\\_by\\_step\\_software/CAQDAS\\_data\\_prep/index.php](http://onlineqda.hud.ac.uk/Step_by_step_software/CAQDAS_data_prep/index.php).)

#### *Situation*

Moving away from the conversation record itself, there much information about the data, or metadata, that is valuable to capture and record with the interview. Examples include the name of project, interviewer, and name or other identifier for each respondent. Another critical aspect of the situation is a detailed description of all participants, and most specifically not excluding researchers. We've proposed a minimal set of characteristics to capture in the model transcript, e.g., age, gender, marital status, employment status, and so on. More detail might be appropriate depending on the specific project. For our purposes, one of the main uses of an interview will be that a researcher will download copies for reuse. Especially if one is handling interviews from multiple sources, it is far too easy for interview pages to get confused unless essential

identifying information appears on every page. Hence, we have opted for key information about the study and the interviewee identifying information to appear on every page. The social characteristics of the interviewee appear only on the first page, partly because they are also available in accompanying databases. (See Figure 1 - Sample pages from model transcript)

In laying out further specifics for situational context, we have drawn very heavily on the work of van den Berg (2005). Most simply, we start with wanting to know the location and time of events. Time will most likely be specified as a day, month and year. Setting should include some standard geographical reference (city or town), specified in as much detail as possible. In some cases, of course, anonymisation procedures will have been applied so these details are not known, or have been replaced with pseudonyms. Furthermore, setting should include even more disaggregated locations, such as building and room descriptions.

There are three other areas that are a bit more subtle. First, it's desirable to understand why participants are in this setting. For formal interviews, this is a fairly straightforward element of selection (van den Berg, 2005). How were the participants selected or recruited and what do they know about that process? Some aspects of this might be covered in the project methodology (below), but here, it is also useful to consider this information from the standpoint of what each participant knows about why he or she is present, and the same for any other participants. Naturally occurring observations will have different principles about selection, but a brief summary of why participants are in this place at this time is still useful.

Second, it is useful to know how any participants are related and what they might know of each other in advance of the event. It can make a great deal of difference to the 'feel' of an encounter if a researcher is meeting someone for the first time, versus a visit that is the closing event of a long-term study that entailed many recurring visits. This sort of knowledge is also especially salient when the project relates to families or social networks.

Third, van den Berg (2005) mentions 'knowable or visible' characteristics. This is a delicate area involving the need to provide information about participants that is visible to each and from which it is plausible to assume inferences might be made. The obvious features here are gender, skin colour and aspects of personal appearance such as dress and hairstyle, etc. Again, the idea is to attempt to capture, fully but not exhaustively, the basic information one might have as a stranger walking into the interview scene.

### *Institutional/project*

Institutional and cultural contexts are in many ways the most difficult to work with, as in some measure, some aspect of culture can be found relevant to any given project. Given the nature of Qualidata's work, one useful step is to distinguish the relatively clearly defined area of project context from the broader cultural context.

In most cases, the qualitative data we are attempting to archive has been collected by researchers in higher education in the UK, most of whom have received funding for their projects. Because of the formal procedures of competition for funding, a comprehensive documentary record exists around a 'project'. At a minimum, we suggest that these official project records be included as project context. The guidelines from our website provide some more detail:

This should comprise details of the history of the project or process that gave rise to the dataset, in terms of the intellectual, financial and organisational origins and developments over time. For example, including the aims and objectives of the project, publications arising, policy developments to which it contributed and any other relevant contextual information.

(ESDS Qualidata, 2006)

For data, part of its context is an account of how the data came into being. For secondary analysis, part of that archaeology [British spelling] is, in fact, the methodology of the first study. This is critical contextual information even if the secondary research applies very different methods during reuse. Thus, a detailed

description of the methodology from the original project does constitute part of the context for re-contextualised secondary data. Guidelines from the Qualidata web site provide more details. (<http://www.esds.ac.uk/qualidata/create/documentation.asp>)

The information should describe the data collection process as well as the data collection method. If applicable, details of the sampling design should be included. It is also extremely useful to include information on any pilot research conducted. Details of the geographic and temporal coverage of the dataset, together with a record of the software and operating system on which the files were generated, the medium on which the data are stored and a complete list of all data files that make up the dataset should also be included. Much of this information is likely to be available from reports, working papers and other publications:

- a description of the originating project e.g. grant proposal, end-of-award reports or working papers
- a copy of the questionnaire if used
- consent forms and other materials regarding confidentiality
- a copy of any interviewer's instructions
- details about sampling, data collection methods and data management methods
- details of any anonymisation process undertaken to render the data confidential
- any reports or publications that provide additional information

In her research documenting what makes data more useful for secondary research, a large number of these items and categories are also covered by Heaton (2004).

### *Institutional/cultural*

Specific guidelines on what cultural context to archive are nearly impossible to specify because what is relevant is inseparable from the research questions of interest. As a practical matter, when archiving research projects done in the 1960s, we have found that grey literature (book reviews, newspaper clippings, commentaries, etc.) is something that secondary researchers have found very helpful. First, such literature reveals something about how the research was received. Second, it provides a window into the institutional and cultural milieu at the time. In the user guide for the study, *Education and the Working Class* (Marsden, 2006), we've included several reviews, both scholarly and from general news publications. Many of these situate the book with respect to contemporary debates: the decline of the working class, the role of grammar schools, the relationship between the Labour Party and the working class, and so on. No primary researcher can be expected to anticipate every cultural factor that might provide useful context to doing secondary analysis years later. However, where a study connects with larger social issues or policy debates, some formal documentation of that connection is very valuable.

We are suggesting this list, and the more detailed enumeration in Table 2 (in appendix), as a suggested minimum for archiving context. Of course, every project is different and this level of detail will not always be available, especially for smaller, individual projects with minimal support. However, in every situation where we take a sample of data for use in workshops or other training sessions, users always repeat the same comment: more context would be better. More is never enough, but more does seem to be better.

### **Conclusion**

Because secondary analysis is, in fact, re-contextualisation, as Moore (2005) suggests, this highlights how vital contextual information is to the process of reusing data. When the question is posed: 'what is the relevant context?' the only answer that is both general and correct is: 'it depends'. That said, we are attempting to advance the debate by exploring a middle ground of suggesting minimal guidelines, at least for the particular case of providing context for archiving qualitative data. Both primary and secondary researchers have the responsibility to be reflexive in a manner suited to their specific projects. In the case of secondary analysis, reflexivity requires consideration of both the contemporary context and that of the original project (Fielding, 2004). These challenges are significant, but not overwhelming; there may be lessons to learn from other disciplines. Historians do not lie awake nights agonising about not being able to do history because 'they

weren't there'. They interrogate historical artefacts, consider conditions that led to their production, while recognising that their own framework shapes what is seen (Scott, 1990). Thinking deeply about context is a useful reminder that even the most knowing subject is never all-knowing. A perspective in which 'data' is reflexively constructed, contextualised, and re-contextualised helps us to acknowledge the inherently social character of knowledge.

## References

Bell, C. (2004). 'Doing sociological research: the genre of "owning up"', *International Journal of Social Research Methodology*, 7(1), pp. 29-33.

Duranti, A. and C. Goodwin (1992). 'Rethinking context: an introduction' in Duranti, A. and C. Goodwin (eds.) *Rethinking Context*. Cambridge: Cambridge University Press.

ESDS Qualidata (2006). Create and Deposit web pages

LINK\* <http://www.esds.ac.uk/aandp/create/research.asp#descrip> and LINK\* <http://www.esds.ac.uk/qualidata/create/introduction.asp>

Fielding, N. (2004). 'Getting the most from archived qualitative data: epistemological, practical and professional obstacles', *International Journal of Social Research Methodology*, 7(1), pp. 97-104

Hammersley, M. (1997). 'Qualitative data archiving: some reflections on its prospects and problems', *Sociology*, 31(1), 131-42.

Heaton, J. (2004). *Reworking Qualitative Data*. London: Sage.

Holstein, J. and Gubrium, J. (2004). 'Context: working it up, down, and across' in Seale, C. et al. (eds.) *Qualitative research practice*. London: Sage.

J. Johnson, S. Rolph and R. Smith "'Revisiting 'The Last Refuge': Present day methodological challenges' (Chapter 4), in M. Bernard and T. Scharf (eds) *Critical Perspectives on Ageing Societies*. Bristol: Policy Press, [In Press].

Lyon, D. (2003). 'The Making of Careers: Women and men in business and politics in Britain, Belgium and France' Ph.D. thesis, Department of Social and Political Sciences, European University Institute, Florence.

Marsden, D. (forthcoming) 'Education and the Working Class' [computer file]. Colchester, Essex: UK Data Archive [distributor], SN: 5457.

Moore, N. (2005). '(Re)using qualitative data? Thoughts from the CRESC Qualitative Research Laboratory', paper presented at Reusing Qualitative Data Workshop, CRESC, Manchester, 28 September, 2005.

Moore, N. (2006). 'The contexts of context: Broadening perspectives in the (re)use of qualitative data', this volume.

Parry, O. and Mauthner, N. (2004). 'Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data', *Sociology*, 38(1), pp. 139-152.

**Figure 1- Sample pages from model transcript**

Study name:	Interview number:
Depositor:	Interview ID: Firstname Lastname
Interviewer:	Date of interview:

Information about interviewee

Date of birth:

Gender:

Marital status:

Occupation:

Geographic region:

I: Just one or two factual details first of all before we go on to your health and that.... how old are you?

FL: I'm 58 in June.

I: What schools did you go to? Can you remember that far back!

FL: Oh... the last school was at Longside.. aye, ken Longside?

I: No, where is that?

FL: Peterheid. Village... That was the last school.

I: Uh-huh, so you lived in Peterhead...

FL: No, Longside.

I: Longside. And, do you work at all? At the moment?

FL: Just look efter my grandchildren. Like that... well, my grand-daughter comes in at night.. well, her mither cleans the school... and I look efter my grandson whose mother works... aye, that kind of thing. I take him in an' keep an eye on him. Well, he's at the school but I give him his dinner an' look efter him at night till she comes an' picks him up.

I: Just one or two factual details first of all before we go on to your health and that.... how old are you?

FL: I'm 58 in June.



I: What schools did you go to? Can you remember that far back!

FL: Oh... the last school was at Longside.. aye, ken Longside?

I: No, where is that?

FL: Peterheid. Village... That was the last school.

I: Uh-huh, so you lived in Peterhead...

Study name:

Interview number:

Depositor:

Interview ID: Firstname Lastname

Interviewer:

Date of interview:

FL: No, Longside.

I: Longside. And, do you work at all? At the moment?

FL: Just look efter my grandchildren. Like that... well, my grand-daughter comes in at night.. well, her mither cleans the school... and I look efter my grandson whose mother works... aye, that kind of thing. I take him in an' keep an eye on him. Well, he's at the school but I give him his dinner an' look efter him at night till she comes an' picks him up.

[This part's repeated exactly]

## **Figure 2 - Summary List of all elements for archiving context**

### Transcript

- accurate recording of audible utterances
- speaker tags
- double spacing between utterances

### Situation

#### Collection level

- study name
- depositor

#### Interview level

- interviewer
- respondent
- interview number
- date/time of interview

#### For each participant:

- date of birth
- gender
- marital status
- occupation
- location (including building, room)

#### Other parts of setting

- why are participants there (sampling, selection)
- what do participants know about why they, and others, are there
- how are participants related
- other visible characteristics

#### Project level

- project history – intellectual, organisational, financial
- aims and objectives
- descriptions of methodology (sampling, pilots, etc)
- data collection processes and methods

Table 1.

<b>Holstein and Gubrium</b> (in Seale, 04)	<b>van den Berg</b> (FQS, 05)
Institutional, cultural	Extra-discursive
Situational	Conditions of discursive production
Conversational, interactional	Intra-discursive

Table 2

	<b>Original project</b>	<b>Current project</b>
<b>“Data” records</b>	transcripts, audio, etc.	transcripts, often no more
<b>Interview setting</b>	room, dress, appearance	often not documented
<b>Project</b>	original questions messy analysis	new questions ‘official methodology’
<b>Cultural, institutional</b>	what is relevant depends on the research question	what is relevant depends on the research question