METHODOLOGICAL INNOVATIONS Online

Methodological Innovations Online (2006) 1(2) 33-46

Ethnography and Data Re-use: Issues of Context and Hypertext

Bella Dicks[a], Bruce Mason[a] , Matt Williams[a] and Amanda Coffey[a]

[a]School of Social Sciences, University of Cardiff

Abstract

This paper seeks to open up debate around context and the re-analysis of stored qualitative data. How to ensure that subsequent users of deposited datasets can appreciate and be guided by the context of the original study? The paper introduces the idea of hypertext as one way of facilitating this. We discuss how 'context' might be thought through in particular relation to ethnography, where it is frequently difficult to distinguish between data and context, and highlight some of the inherent problems in the notion of archiving ethnographic context. In a discussion focusing in on multimedia, we draw attention to the different kinds of contextual information that are necessary to interpret data in different media forms. The paper's starting position is that originators of data and re-users have in front of then a qualitatively different kind of knowledge-base, due to the fact that data and data-records are not the same thing. This doesn't rule out re-use but does imply that quite full and careful kinds of documentation are necessary to try and make it sufficiently rigorous, a demand which also, however, has to be balanced against the dangers of information overload. These challenges lead us to question whether the traditional archiving model is the most suitable way of communicating context to re-users; we present some of our insights from work on hypertext to explore the potential of the hyperlink as a key contextualising tool.

Introduction

Whilst the dilemmas of qualitative data re-use are well documented (see Holstein and Gubrium, 2004), the issues associated with preparing data for later use by third parties are yet to be thoroughly conceptualised. This paper argues that qualitative data cannot be understood in isolation from the contexts in which they were generated, and so addresses the inseparability of context and text in relation to data re-use. Hence, we question the utility of storing, sharing and re-using qualitative data (records) that are devoid of context. There are a number of dimensions of context – including media-related, substantive and methodological aspects – that need to be communicated to re-users, and this paper addresses the whys and hows of such an approach. We also contend that the artifactual nature of data-records, in that they represent a particular intervention in the field of action, further underlines the importance of documenting context. Our discussion focuses on the use of digital resources in preparing data for re-use. These resources, we argue, can be used at all phases of research, including data-generation, analysis and authoring. Our arguments draw on our experience of using multi-media, hypermedia and hypertext in ethnography, and, in particular, we advance the idea that the hyperlink can help bring context and text into closer alignment in the representation of research findings. We discuss, for example, the potential of using collage and montage on screen, via the hyperlink, as a way of indicating contextual information. Further, we also question the separation of data (records) from the original

Correspondence: Dr Bella Dicks, 0.57 Glamorgan Building, University of Cardiff, UK. Email: DicksB@cf.ac.uk

ISSN: 1748-0612online
DOI: 10.4256/mio.2006.0010

analysis, promoting our use of an Ethnographic Hypermedia Environment which provides a 'one-stop-shop' for representation, access to substantive/methodological context and the original multi-media data-set. Concrete exemplars of using hypertext to bridge the connections between context, data and original analysis show the importance of multiple contexts in interpreting secondary qualitative data.

Whilst the field has until recently been dominated by the now routine re-use of quantitative datasets, there has been for a while an increasing push towards the archiving and sharing of qualitative work. The Economic and Social Research Council stipulates that all 'award holders are required to offer their computer-readable data for deposit, prepared to a standard which may be used by a third party, within three months of the end of an award' (see ESRC Data Policy April 2000). For many, this push to archive qualitative data has been contentious, as Louise Corti discusses in the editorial for this journal issue. It is the issue of context that has particularly fuelled debate, since the idea of re-use suggests that qualitative data can be taken out of context as though data were neutral and transferable. Many have pointed out that the context of any qualitative study is necessarily multiple and complex (see Holstein and Gubrium, 2004), and that data are generated within, and cannot be abstracted from, its particular conditions. Some fear that archiving establishes a distance between analyst and data that pushes qualitative data in positivist directions (Mauthner et.al. 1998).

This paper will not engage with these debates, as this has been done elsewhere (Moore 2005; Bishop 2005a and b; Heaton, 2004; Corti & Thompson 2004). We are not interested here in the dilemmas of how, if at all, to re-use data but how datasets can be prepared for re-use. Nevertheless, we do not accept the idea of re-use uncritically. We would endorse the observation made by Mauthner and colleagues (1998), keen critics of qualitative archiving, that qualitative data are 'inextricably bound to the unique conditions of their production' (Mauthner et al, 1998:740). We do not however share their conclusion that re-use is impossible; instead, we argue that it has to be very carefully and reflexively documented for reasons discussed in more depth below, and that – due to this complexity– archives may not be the best way to facilitate sharing. We do therefore hold that familiarity with the original context – both substantive and methodological - in which the data were generated is essential to their significance and hence subsequent interpretation. However, it should be noted that there are many different kinds of uses to which stored data can be put - from systematic reviews to interrogations of existing findings: Corti and Thompson (2004) also mention description, reanalysis of various kinds, research design, verification and teaching and learning. It is clear that not all of these purposes share the same imperative to get fully to grips with the context of the original study. It is really in the case of the in-depth *analysis* of existing data – the focus of this paper – that contextual knowledge is so vital.

We start by distinguishing between two kinds of methodological information about context that is usually documented for archiving purposes:

1. Cataloguing information: this is often referred to as 'metadata' and includes supplementary information (such as biographical identifiers) that allows the data records to be identified, searched and navigated. Such information can potentially be standardised, possibly in XML form, and hence searched for online;

2. Holistic dimensions of context: What information to give about a) the methodological approach taken as a whole and b) the substantive context of one's project.

Our focus is on the second kind of information. Rather than metadata, this type is more appropriately thought of as a level of knowledge – knowledge that can only be conveyed in narrative form (since it cannot be disaggregated into individual records) and concerns the original study as a whole and its unique conditions of production.

The paper will proceed by discussing what contextual knowledge re-users of qualitative research need to access in order to make sense of the original data. We will address three aspects of context:

1. The **media context** of a study: the implications of the choices of media-form used in data-records, and their differing requirements in terms of contextual clarification;
2. The **substantive context** - the myriad kinds of knowledge about the local field of study that the originator needs to draw upon in order to interpret and analyse the data;
3. The **methodological context** - the guiding frameworks, means and processes through which the study was conducted.

We start, however, by considering the problems posed by seeing context as something separate from the data themselves by discussing the particular case of ethnography.

Ethnography and complexity

The first point we wish to make concerns the necessarily complex nature of ethnographic fieldwork and the multiplicity of the empirical resources it uses. Ethnography is a very messy process; it doesn't necessarily suit interview schedules or particularly neat kinds of data generation. Mostly it involves just being there as a person and interacting with people, writing up notes and having discoveries of a fairly serendipitous nature. Ethnographic fieldwork is planned and systematic, of course, but most of the best insights come from the interactions that can't be planned, as they emerge from the mix of everyday predictability and unpredictability that is the very object of its understanding. So, ethnography has an interest in multi-layered description, i.e., in representing the complexity of the field of action and not reducing it to codes and matrices, easy summaries or neat explanations. In our own work, we have made use of hypermedia resources that succeed better than print, in our view, at managing and representing this complexity, enabling its interpretation, and authoring the results. This work is based around the central importance of the 'hyperlink' (the electronic linking of two or more items of digital information on computer screen to enable readers to click from one to the other). It is the hyperlink that allows data to be easily embedded in contextual information. We shall return to discuss its important role later.

Clearly, how to convey a 'thick description' of complexity to subsequent re-users represents quite a challenge. In ethnography, context cannot be minimal, it has to be maximal; we are trying to produce a thick description of the field that allows all its dimensions to come through. How do you share this idea of complex context? How would it fit in with the idea of the ESDS Qualidata catalogue or User Guide, and how could it be made comparable across all those different datasets (ESDS Qualidata 2006)? Would we want to provide just 'the facts' alone (bracketing out for a moment the debates and controversies this idea conveys), i.e. just hand over the data records together with summaries of contextual information? This would be the 'just facts' approach. Or do we want to keep it as complex as possible, so if you want to use the data you've got to read your way through and around them linking back all the time to the contingencies of data-generation and field relationships? This would be the 'messy' approach. Alternatively, there's always the established default option of never letting anyone see the either the facts or the messiness, and keeping the data under lock and key. This seems to us to be an undesirable position, especially since digital technologies make sharing information so much easier. We subscribe to the position of Thompson (2000) and others (such as Bell, 2004) who believe it is right to allow others access to one's data and the means of understanding the complex and necessarily fluid processes of interpretation out of which findings emerge.

In spite of everything that has just been said about the contingencies of fieldwork and interpretation, the ethnographer still has the task of being reflexive about the process of research, that is, not only coming to understand the social world under investigation but also to understand what he or she is doing with that information at any one time. In other words, one doesn't just dive in with a fieldnote book or a camera; one thinks about what one observes, how one is observing it, how to represent it and how to recall at a later time the variety of explanations that may be possible for any one event or action. Only as fieldwork and interpretation progress does this set of available explanations gradually coalesce into a pattern. This involves thinking reflexively about fieldwork relationships. For example, how do fieldworkers and participants inter-

subjectively communicate such that the knowledge produced is always a co-construction? Note can and should be taken of these kinds of possibilities. This brings us to our second point, which concerns the distinction between data and data records.

Data, data records and multimedia

Our starting point is the observation that when qualitative researchers make records of phenomena under investigation, they are producing *artifacts*. Making a record of some aspect of the field means carrying out an interpretative action (for example, we can say that taking a photograph is an authored act). This has implications for re-use, since re-users cannot access the original data; they can only access the available *data-records*. This means that the choices made (e.g. of recording medium) by the data-originators have an important role to play in how the data are represented and hence made available for subsequent re-use. Re-users can only access the materials archived, which are inevitably the result of processes of translation, selection and reduction during the original fieldwork (every researcher has to decide what to record, and how). Data, to illuminate this point further, are what we learn from observing, making sense of, interacting with and eliciting information about the social world. They are not the 'world in itself' (Bauer et. al., 2000); neither are they coterminous with the physical materials produced when researchers make records of the field. It is true that many researchers refer to these records as 'data' in and of themselves; however, we suggest that this usage tends to obscure the representational transformations effected when field interactions are recorded. The record is a representation of data rather than being the same thing[1]. To be usable in analysis, data have to be recorded, or more properly represented, in some way. The data record – be it a field-note, an interview recording, a photograph, etc. – is the fieldworker's attempt to produce a permanent record of the observations, interpretations and elicitations that he or she has made, in a process of co-construction with the research participants. The data record is inevitably, therefore, a reduction and abstraction of the 'field of action' itself (ibid.) and of the situated relationships pertaining therein between researcher(s) and participants. In this sense, the process and means of making records is a crucial aspect of what is commonly referred to as 'data-generation'. Below, we shall attempt to show how recording-medium choices during fieldwork have important implications for how the field of action is communicated to re-users.

Those who analyse data-records they have constructed themselves have the advantage of being able to situate them within their own multiple *recollections* of the field of action. This is an important capacity in making sense of them when one is no longer in the field. Data re-users do not have access to these recollections of the field of action, except where they have been recorded by the original researcher. Hence, re-users necessarily have less information available to them than those who participated in the original fieldwork. They are of necessity more distanced from the data themselves. As Bishop (2005a: 15) cautions, we must recognise what she calls the 'inescapable problem of "not having been there"'. We feel it is important to acknowledge these difficulties and to think them through, rather than to argue – as some appear to – that there is no qualitative distinction between primary and secondary analysis (e.g. Moore, 2005). We need to be aware of the dangers of adopting a positivist approach to data due to this distance pertaining between secondary analyst and data. We therefore need to approach re-use with extreme care and reflexivity.

A consideration of multimedia brings these issues clearly to the fore. Multimedia is often considered a methodological issue only as a feature of the data-records that ethnographers construct (e.g. the photographic images or video footage that they take), rather than as an inherent feature of the worlds they study. But multimedia within the field is not the same thing as multimedia within data-records. We perceive data through all of our senses, including sight, hearing, touch, smell and even taste and they are by their very nature composed of diverse media (likely to include sounds, objects, visual designs, people's actions and bodies, etc.). Yet the media that appear in our data-records are necessarily more restricted. Video footage, for example, limits the information recorded to that amenable to audio-capture and camera-work. Field-notes limit it to writing. In comparison to the semiotically rich, multi-sensory field of action, both writing and video employ more restricted 'modes' of meaning (or semiotic resources – see Kress and Van Leeuwen, 2001). Hence, data

records are never 'multimodal' to the same degree as the field of action. Emmison and Smith (2002) make this point in relation to visual methods:

> Stated in its bluntest form our reservations about an image-based social science rest on the view that photographs have been misunderstood as constituting forms of data in their own right when in fact they should be considered in the first instance as *means of preserving, storing or representing information*.
>
> (Emmison and Smith 2002: 2 our italics)

Hence, data are not 'what the camera can record but […] what the eye can see' (ibid: 4), and we would add, what the ear can hear, the nose smell, the mouth taste and the body touch and feel. Accordingly, in making records or representations of that field, ethnographers need to think about how the field's 'multimodality' has been reduced, or re-produced, through the recording medium chosen (Dicks et. al. 2005).

This argument has implications for re-use. What re-users will be faced with is the data-record – the photograph, the video footage, the fieldnote – and these are already highly partisan selections and reductions of the field under study. It seems likely that researchers will increasingly use audio-visual media as time goes on, due to the expanding opportunities afforded by cheap, usable and portable technology. But the increasing ease of representing data in image-form poses dilemmas for later re-use. Quite aside from the considerable ethical issues involved in archiving multimedia data, which we do not go into here but deal with elsewhere (Coffey et. al. 2005) , there are important issues of 'missing context' that image-making raises. We shall explore some of these in the next section.

Dilemmas of media context

There are two major ways in which multimedia data-records pose a challenge for re-users. The first concerns the above-mentioned reduction of multi-sensory data into restricted-media data-records. A video camera can record sound and the moving image, but not taste, smell, touch. Hence, we have to rely on memory to flesh out what a photograph shows in order to recall the full perceptual context in which it was produced (Emmison and Smith, 2000). This means that re-users need to recognise the modes that are lost to them through relying on data-reductions. Of course it should be noted that the original fieldworker may also forget, or fail to note, the multi-sensory nature of the setting too – so 'being there' does not guarantee that particular insights will be drawn. We can illustrate this from our own dataset based on a science 'discovery' centre. We took some video footage of one of the interactive exhibits that some children we observed were playing with; it was called the Kugel Ball (a vastly heavy black granite sphere sitting on a bed of water that can be effortlessly revolved in spite of its weight). When we first video-recorded two girls playing on it we got the impression that they were thinking of it as a 'big planet'; it was only later, during a focus group held with them, that we discovered they had actually been using it as a 'wishing ball'. The point is that video data can never be relied upon as an objective record; everything depends on when and where you point the camera and how.

The second point concerns the inherent ambiguity of images (something that is well documented; see Banks, 2000; Pink, 2001). Even though the camera footage we took of the science discovery centre gives us a good sense of the action, it does not allow us to represent the modes of information on which the exhibits depend. That is, we cannot understand the point of the Kugel Ball exhibit if we rely on visual and aural information alone. This is because the important signifying element in the exhibit is its *weight*; the ball cannot be understood unless one has actually noted its wetness, felt its extreme weight and contrasted this with its ability to move even when barely touched. Only then can one appreciate the didactic point of the exhibit as outlined in the accompanying (written) label – to demonstrate the extreme forces generated by the bed of pressurised water upon which it is sitting. Our point here is that audio-visual materials need to be contextualised: they need supplementary written explanation in order for their significance to be correctly interpreted. Here is an extract from a fieldnote that referred to the Kugel Ball, jotted down by one of the fieldworkers:

> 'By the side of each exhibit are written instructions, showing you how to activate the machine. When you do, things happen. For example, at the back of the hall is a huge granite ball, glistening with water, sitting on a plinth. It is obviously massively heavy. But, amazingly, when you touch it, it revolves around. You read the instructions, and find out that it is not sitting directly on the plinth at all, but on a thin bed of water under intense pressure. All around you are the sounds of children shouting, adults talking and there is movement everywhere.'

The field-note does not mention many of the elements that our photographs or the video footage would effortlessly capture: the colour of the ball, for example, or its position in relation to other exhibits around it. Writing employs primarily one mode, verbal language[2], but this mode is a particularly communicatively rich one in that it allows other modes to be described in their *absence*. So, in using verbal language:

1) the modes of weight, sound, shapes, textures as well as actions and movement can be conveyed through linguistic description;

2) subjects can be *contextualised* more fully – so our written field-notes describe the Kugel ball in relation to the field of action in which it is located. Visual images provide context, too, through representing the space within which subjects are located, but this context is restricted to that provided by the camera's position, even in the case of portable video cameras that afford mobile perspectives. Though the camera's[?] gaze appears to be neutral, it is obviously providing a restricted view due to the framing of action by the lens. By contrast, writing can jump back and forth between multiple 'views', affording an impression of both movement and engaged interaction. The point is not that writing is more objective (in fact it is arguably much more selective); rather, it introduces a subjective, embodied, conceptually rich narrative. This is informed by the writer's multiple and mobile senses, differing from the frame-generated perspective of camera-derived information;

3) The subjective reflections of the observer can arguably be conveyed less ambiguously by field-notes than by camera work: the field-note above, for example, affords a particular, immersed perspective - that of a narrator who is 'in' the scene being described selecting out particular elements for our attention[3]. It presents the writer's subjective assessment of the scene and description of emotional responses ('amazingly', 'excitedly')[4] Film, by contrast is more opaque and open to different readings.

For these reasons, we suggest that image-based data that are going to be archived cannot stand alone. They have to be carefully logged and supplemented with written records, which can capture meaningful elements not easily represented through audio-visual or photographic media. At the very least, contextual information needs to be provided to explain the conditions under which visual records were made, such as:

1) What was the situational context in which the images were produced (what else was happening in the field of action at the time?)? Video logs and other written descriptions are essential here;

2) When, where and by/with whom was the image/footage produced? This requires metadata attached to the images;

3) Why was this particular photograph/footage made? What was I trying to show? This requires some reflexivity on the part of the image-producer, to try and interrogate what particular 'biases' he or she has brought into play (e.g. my views on gendered interaction might have led me to take lots of images of boys running around and of girls talking)[5].

We argue that this level of contextual documentation is essential in that it allows images to be understood. But how can it easily be produced for archives so that it can be conveyed to re-users? If the footage is only to be used for purposes of preserving a broad-brush historical record of environments, rather than ethnographic analysis, then the detail of documentation required is arguably not so great. But if it is to be used for asking questions of the type, 'Just what ***is*** going on in this interaction?', then a range of supplementary information is clearly required. Video logging, incorporating reflexive notes recording the conditions of camera-work, is one way of ensuring that visual materials are given contextual fleshing -out.

We not mean to suggest that data re-use renders images inferior to field-notes. Indeed, the issue of historical record introduces another aspect of video and re-use. Although there are all the above-discussed problems of the incomplete, partisan, ambiguous nature of images, equally, however, there is no doubt that the provision of sound and image-based records enhances the semiotic modes available to re-users – in comparison to written media alone. It allows all kinds of spatially and temporally organised information to be effortlessly depicted rather than having laboriously to be described in writing (or, more often, simply excluded from the records). The later re-user who has access to datasets that contain both written, sound and image-based records has a much more extensive and detailed informational resource than those comprising transcripts/fieldnotes alone. In this sense, it might be argued that data depositors should be encouraged to use, with ethical considerations in mind, a camera as well as a pen in the recording of data from fieldwork projects – always with the proviso that the images need to be satisfactorily contextualised and logged.

Dilemmas of methodological context

We now turn to the question of methodological context. How do we define context of this kind and how should we treat it in relation to re-use? There is a consensus that the provision of full documentation about a project's methodology is essential to its usefulness as archived data (Corti 2000; Heaton 2003). Our view is that methodological information needs to be fully documented, although, as already mentioned in the introduction, this does not mean that we see the matter as a simple one of 'handing over' the context. We accept, for example, that there are problems implicit in the idea of 'providing' context as if it were a static set of factors that could be straightforwardly described. In what follows we try to highlight the complexity of context, whilst at the same time acknowledging the difficulty of deciding how much information can practicably be documented for re-users given the dangers of information overload.

In relation to originators' datasets, it is clear that only substantive context is likely to be fully analysed and documented in a routine way since it is something that they need to examine in order to make sense of their data. Methodological context, on the other hand, is less likely to be fully documented or examined. The extent to which it is will depend on how *reflexive* originators are being about their methods. We argue that it is possible for a detailed 'warts and all' account of methodological decisions and reflections to be made available for deposition; we recognise, however, that it is more likely that methodological information will be confined to standard, 'official' descriptions of the project's methodology of the kind provided in final reports. These often read along fairly well established lines that tick the relevant boxes (such as defining the field, sampling, negotiating access, gaining consent, and so forth). It is clear, however, that these 'public' accounts may provide little in the way of insight into the messy actualities of fieldwork and research relationships. Yet it is this kind of detailed methodological information that is of particular value to re-users; in fact, we would argue that it is essential.

For example, it's clear that methodological context is not only about the whys and wherefores of technique and process, but also about the case-specific and complex relationships that are negotiated between the researcher and the participants during the course of research. These include consideration of differentials in power and status, the contingencies of planned and chance interactions, the responses of participants to the research process, along with many other variables. When we archive our data for others, we cannot 'account for' this necessarily complex and particular aspect of context in any simple or straightforward way. We cannot, certainly, wish it away, for the quality and nature of the dataset is to a large degree dependent on its specifics.

That is, any qualitative data-set will necessarily be generated within the parameters of particular research relationships, such that the data cannot be seen as detached from or independent of them.

All of this suggests that context does not comprise a set of static circumstances that the originator is 'surrounded by' in the course of study. Holstein and Gubrium (2004) prefer to see it as 'a fluid, socially emergent constellation of contingent factors that are "worked up" – not just encountered - in the course of everyday interaction'. Context, in this view, is about the flux and flow of the everyday, not (only) about predictable influences and processes. All originators can do is to ensure this elasticity is 'stopped in its tracks momentarily to allow for description and analysis' (ibid.: 309), through providing a rich account of relevant factors such that subsequent re-users can understand the conditions under which the data were generated and the conclusions drawn. As we will now discuss, the existing guidelines on providing documentation do not in our view recognise this fluid quality of context sufficiently clearly.

ESDS Qualidata guidelines on the Qualidata webpages list the following as common types of documentation that can be usefully and usually straightforwardly provided by the original research investigators:

- original grant application
- end of award report
- description of methodology
- interview schedule(s)/topic guide
- questionnaire
- observation checklist
- interviewer instructions/prompt cards
- communication with informants relating to confidentiality
- written consent forms
- matrices
- tree diagrams
- information of equipment used (e.g. recording equipment)
- other background information
- details of missing information
- correspondence
- speaker markers in text, typically associated with internal metadata; question or thematic markers in text; cross-reference of text to audio material
- references to publications and reports based on the study

Quite a substantial amount of documentation of both types – both individual records and holistic information - is required of the depositor in order to conform to these guidelines. Nevertheless, these have to be considered as minimal, in our view, as they do not adequately represent certain kinds of methodological information that are important to include with a deposited dataset. They suggest, for a start, that contextual information is external to the data-set, and can be treated as 'informing' the dataset or needing to be 'factored in' to it. We suggest that context needs to be seen more reflexively than this, as part and parcel of the dataset itself. It should be acknowledged of course that ESDS Qualidata welcome and invite actively all kinds of contextual information, including extensive bespoke notes and materials that are unique to every project. Our concern, however, is where one draws the line. The complexity of context is such that almost everything 'can' be understood as context.

As a start to discussions about the extent of methodological information needed, we suggest the following as important dimensions of qualitative methodology that also need to be documented:

1. nature and evolution of researcher-participant relationships and interaction;

2. epistemological, ontological and political frameworks of the project, and how these worked out in practice;

3. processes of arriving at the interpretation of meaning, and the type of approach(es) utilised in data-analysis.

Under these headings, originators can provide much invaluable information for subsequent re-users; however, it is clear that such reflections need to take place, and be recorded, during the data-generation phase of a study. They cannot easily be added in afterwards. We recognise that – whilst in our own case – we had a good incentive to do this (since our project had an overtly methodological remit), others may not. We are, therefore, bound to be talking about an ideal-case scenario. Further, the question remains of how best such narrative information (which might include researcher-generated vignettes, memos, notes, images, and so forth) can be presented in a way that shows how they help to illuminate particular data-records. In our view, the hyperlink has a key role to play here; we return to this below.

Substantive context

Broadly speaking, substantive context refers to all the local factors that a researcher needs to take into account in order to make sense of the data that s/he is collecting. There are many aspects of substantive context that a researcher would normally need to consider.

Holstein and Gubrium (2004) make a distinction, for instance, between:

a. 'bottom up' or 'proximal' context – the situated, interactional contingencies of talk and action that characterise the data generation phase of research

b. 'top-down' or 'distal' context – the cultural, socio-economic, political and institutional meanings, discourses, relations and forms of organisation that condition the field of action.

It is essential to bring both levels into dialogue with each other. This means, in our view, studying context at both levels (it may be better to reverse their order):

a. investigating what the cultural and social aspects of the field of action comprise and understanding their power and effectiveness. For example, in our study, we would need to consider how the leisure market has become intertwined with educational functions such as the communication of science. How is this market constituted and how has it changed over time? How insulated can visitor attractions, such as museums, remain from the imperatives of marketing and business competition? How do other science centres position themselves within it?

b. simultaneously examining the minutiae of data to see whether and in what form traces of these wider aspects show up in talk and interaction, and how they make their influence felt.

There is a logical problem inherent to the notion of archiving context as though it were separated from 'raw data'. It is possible to argue, in a strictly interactionist perspective, that the significance of context can only be specified by analysing the data themselves. Hence can one really 'provide' it for re-users who are then going to commence their own analysis of what they see as 'the raw data'? Can substantive context really be separated from data? One may have a good knowledge of distal contextual factors such as socio-historical contingencies and political-economic relations. But these remain at an abstract level unless their actual effects are traced within the data at hand. Holstein and Gubrium argue this point forcefully:

> Social structure, social class, social integration, social disorgnisation, and other overarching constructs are commonly invoked without empirical specification or description of just what these social 'things' might amount to in the situation being examined. While a singular definition of context may not be possible, those empirical manifestations that one chooses to consider must be demonstrated if context is to be brought on board in service to analysis. This requires the analyst to demonstrate empirically the linkage between action and contextual effects.
>
> (Holstein and Gubrium 2004: 308).

In other words, in their view, context can only be specified in terms of its actual manifestation within data.

This argument requires some qualification. Holstein and Gubrium insist that what is knowable emerges in the details of organised social conduct. They see context as 'a practical accomplishment' (2004: 309). They eschew a structuralist perspective that starts with the bigger picture and works down to ground level. This is not, however, to suggest they subscribe to a naïve empiricism, in which only what is directly manifest in objective reality is taken as fact. The definition of what counts as evidence is crucial here. We argue that the effects of 'distal' structures such as the market, for example, may not be manifest in *what* people actually do or say, but they may be powerful precisely because they work in more latent or opaque kinds of way. For example, they may affect the 'wheres, whys, whens and hows' of what they say. Clearly, participants may not be aware or conscious of this effectiveness of context. Discourse analysis takes this latency as axiomatic and proceeds by exploring how traces of wider cultural and socio-economic formations make their effects felt by organising what people can and cannot say, and how they say it.

With these qualifications in mind, however, we would agree with Holstein and Gubrium that consideration of context (at a number of levels) means examining how it works out *in everyday practice*. The ways in which market relations, for example, operate in everyday life can only be identified by analysing the data at hand. It follows, then, that data are not separable from context in any meaningful sense. Indeed, the data cannot be said to stand apart from it. Hence, this poses a problem for storing and sharing. To take the most frequently deposited dataset – interview transcripts – to what extent can these be considered as 'raw data' when they are redolent with these traces and shaped to a large extent by them? If you as an originator recognise these traces, you do not see your data as 'raw'. The job of interpreting them becomes one of carefully considering their significance in the light of this context. You may feel quite uneasy about depositing them as though they were amenable to other kinds of analysis less sensitive to these traces.

At the heart of this dilemma is whether qualitative data can ever be treated as 'raw data'. In our view, they cannot, for all of the reasons so far specified (including the role of recording media). But, this still begs the question: how much of a steer should originators provide? Or should they leave the data to 'speak for themselves'? Here, we turn to our own data for an illustration of some of the issues at stake.

Exemplar

Let's take the interactional level (b.) as an example. Consider this extract from an interview with the Director of the science centre we studied:

> Edutainment, I mean, I have, I wouldn't use it particularly comfortably but it's, it's not a word I would - I wouldn't refuse to be described as being involved in edutainment, at least certainly not in the, in the holidays we certainly are.

This extract shows quite clearly a level of ambivalence or even anxiety from the Director about how best to articulate his stance towards the word 'edutainment'. In the larger extract, including the question asked, it is

clear that the speaker is framing his response in terms of a discourse of authenticity - attempting to distinguish the science centre from the unreality of Disney and American models. This anxiety is confirmed too in the audio-track of the interview – which shows how the speaker's voice falters and hesitates, then abruptly shuts off signifying a determination to say no more on this particular topic. The extract shows how the Director attempts to hedge his bets and in the end settles for an answer using a 'when' frame (we're only about edutainment in the holidays). Rather than using the question as an opportunity to articulate the happy conjunction of education and entertainment (as he has done already earlier in the interview), he responds to the interviewer's specific question about edutainment as though it were a challenge, and reacts somewhat defensively.

The interview both in its parts and its entirety is traced through with context, in that the wider political-economic relations of the Science centre position it uneasily between a commercial model on the one hand and a publicly-funded model on the other. Wider contextual knowledge about how the Centre is funded and the political 'problematic' at the heart of it (should tax-payers' money be used to fund an attraction that *might* be suspected of providing more in the way of entertainment rather than education?) enables us to understand both speakers' circumspection and the interviewee's hesitation. In choosing to answer the question about edutainment through making a temporal distinction between *when* the centre is and when it isn't a place of edutainment, the interviewee shows an impulse both to excise edutainment, but also to acknowledge its presence. In this way he manages to present the Centre as both about edutainment and not about it, and erect an effective temporal barrier between the two. Yet this is not accomplished in a smooth or confident way, suggesting the ambivalences and instabilities at the heart of this issue remain unresolved.

How useful and/or indeed essential is this kind of distal contextual knowledge for re-users to have? We would argue that it is part and parcel of the data. If re-users elect to interrogate the dataset in order to answer different research questions - for example, perhaps to evaluate what children learn from specific exhibits) - it is clear that this context still remains not only relevant but, to a degree, structuring. That is, the particular design of the centre and its exhibits has emerged from within the historical problematic underlying this context and is conditioned at least in part by it. Unless a re-user knows this context, and is alert to its traces in the dataset, important dimensions of the exhibits' framing – and hence of the possible ways in which children can interact with them – will risk remaining obscure.

Hence, we would suggest that an important section of the contextual information provided for re-users consists of *an overview of the social, cultural, political-economic and historical context that impinges upon the particular field of action*. In our view, this course of action is preferable than subscribing to the actually untenable notion that the data can somehow 'speak for themselves'. And yet, once again, we come up against the problem of information overload. In providing this level of contextual information, one is actually providing what is in effect one's analysis. That is, it is not possible to come to know and understand the full substantive context of one's study until one has finished analysing the data-records.

This again returns us to the dilemma of stripping data away from context. Can qualitative data stand alone, in isolation from the methodological and substantive context that informs and shapes their production? Or should they always remain situated within this context, i.e. within the analysis that has been conducted? Can we possibly provide ***all*** of the context that a re-user would require in order to make sense of what is going on in the data? Can we make sure our data records are ***so*** extensive that the re-user could potentially reach the same level of understanding of context as the originator was able to?

We believe this is an almost impossible undertaking. It effectively requires data to be deposited together with their analysis, with all kinds of links made between them, and archives are not necessarily best suited to this task (for example, they impose standardising requirements on deposited data-sets, and also have insufficient resources to create a bespoke presentation for each deposition). Data archives are about data, not analysis (they are not the same as libraries). In any case, would re-users be prepared to dedicate valuable study time to the preparation of material still presented as 'extra' or supplementary to the data records themselves? This is

why, in our work, we have found hypermedia resources so useful for drawing the material together and yet still allowing unfettered access to the data-records. In the next section, we take these issues a little further and reflect in particular, on a) the difficulties of providing context and b) the role of the hyperlink as a way of addressing some of these problems.

Discussion: the hyperlink as contextualising agent

In this section, we advance the argument that qualitative research can effectively be conducted and authored via digital hypermedia. This could be seen as a means of supplementing the ability of archives to communicate the full complexity of context as it has been discussed here. We focus on the potential of the hyperlink to allow originators of qualitative datasets to make connections between the records and their own analysis of context. We have repeatedly in this paper made the assertion that there really is no meaningful divide between context and text (text understood in the sense of achievable data-records). One of the difficulties is that it is very difficult to decide where context begins and text ends. Indeed, it is debatable what context actually means. For many qualitative sub-disciplines of social science (such as discourse analysis, critical psychology, cultural analysis, and many others), context is the 'trace of the social' which is precisely what needs to be studied. Perhaps we shall never be able to 'give a single, precise, technical definition of context, and eventually we may have to accept that such a definition may not be possible' (Goodwin and Duranti, 1992: 2).

One of the features of hypertext as institutionalised on the World Wide Web is the ability for web-designers to link items together in an associative way simply by adding in highlighted links to blocks of text or images. This capacity has become extremely familiar to internet users, and tends to highlight the networked nature of knowledge (where everything can be connected to everything else). Hyperlinks therefore have the capacity to make the distinction between text and context less of an issue, since each piece of data (record) can be linked into various other kinds of information. Links on a screen allow data to be embedded in a multiplicity of informational resources, including multimedia and graphic screen design. For example, in the illustration given above, about the Director of the science centre and his ambivalence about the term 'edutainment', links could be provided that connect his speech to a) other informants' records, showing the interconnections on this topic, information about the science centre's founding and history; b) information about the particular dynamics of the interview, and the interviewer's own role in eliciting the information, and c) information about the different effects the recording media have, in that the audio tape shows up particular qualities of the interview that are lost in the transcript; and so forth. In other words, details pertaining to methodological, substantive and media-related context can all be brought into dialogue with the relevant data-record. The hyperlink can highlight this interconnectivity, without sacrificing structure and argumentation.

There are, however, dangers in using the hyperlink for communicating scholarly research. For instance, there are many kinds of relationship that hyperlinks suggest between two items of information, but it is rare for web-writers to make the nature of these explicit. Indeed, this is (one of the) feature of the Web that makes it seem rather unsuited to the representation of scholarly argumentation, since ideas and information tend to be hitched together in a manner that is not explicit or which does not enable the connection to be interrogated. A number of writers have started to think about the implications of this and the need for using links in a motivated manner (e.g. Lemke 2002a). Barbules (1998), for example, warns that the invisibility of links can serve to naturalise relations between items, preventing the reader from seeing what kind of connection is being made. For example, one item linked to another may be read, inappropriately, as cause-and-effect, particularly where this is suggested by beliefs about the world or common-sense frameworks of understanding. We suggest in our own work (Dicks et. al. 2005) that part of the solution for scholars and researchers may be to make the functions of links explicit. In this sense, we could think of a scholarly repertoire of named links, perhaps, which could put into place a set of accepted rhetorical qualifiers such as 'for example', 'but also note this' or 'an alternative explanation would be'. These links could be used to direct the reader to contextual information pertaining to particular data records.

Conclusion

This paper has suggested that the problem involved in specifying context for re-use is that it implies a separation between context and data. For us, the hyperlink allows one way of bringing the connections sharply into view. Existing ESDS guidelines on providing contextual information for both quantitative and qualitative data suggest, perhaps, that contextual information is external to the data-set, and can be treated as 'informing' the dataset or needing to be 'factored in' to it. Against this, and citing Holstein and Gubrium, we argue: 'even as context is empirically documented, qualitative researchers should be careful not to reify it or "freeze" it into a static entity.' (2004: 309). We suggest that context needs to be seen more reflexively than this, as part and parcel of the dataset itself. Archives will always be necessary and standardisable information will always be useful to search across datasets. But the dilemmas of context will continue to haunt the possibilities and potentials of re-use. We suggest that the electronic publishing revolution that is already upon us (see the rapid increase in online journals), together with the ease of creating one's own web-pages for project-dissemination, will mean that researchers increasingly turn to digital resources in disseminating their work. If they wish to make their datasets public too (or rather, accessible to designated users, probably with password protection) they will need to think of ways of making sure that there is contextual information that can point to the subtleties and complexities of both their data and analysis. Online EHEs as in our own experiments are one solution, if a time-consuming one: making more of the data *and* context available for re-use. It is possible that the best way forward for archives such as ESDS Qualidata is to think about how their online repositories can interact with other digital resources, including web-sites that researchers and authors themselves might construct. This would provide value-added [noun missing] for archives [archivists?] themselves, who could then specialise in developing tools for standardising datasets for searching. Authors and researchers, meanwhile, would be able to ensure that the full datasets, contextual information and analysis are also available online to complement the deposited records. It could also be of benefit to researchers concerned to ensure that their data are preserved in their full complexity.

References

Banks, M. (2001). *Visual Methods In Social Research*, London: Sage.

Bauer, M.W. and Gaskell, G. (2000). *Qualitative Researching with text, image and sound: A Practical Handbook*. London: Sage.

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

Bishop, L. (2005a). 'Is Secondary Analysis best? A case study of reusing qualitative data', Paper presented at 'Re-using Qualitative Data: CRESC Methods Workshop', University of Manchester, 28th September 2005

Bishop, L. (2005b). 'The re-use of qualitative data: Issues from experience at Qualidata', paper presented at the British Sociological Association Conference 'Social Orders and Disorders', Harrogate, April 23.

Corti, L. and Thompson, P. (2004). 'Secondary analysis of archived data', in C. Seale, G. Gobo, J.F. Gubrium and D. Silverman (eds) *Qualitative Research Practice*, London: Sage.

Emmison, M. and Smith, P. (2000). 'Researching the Visual: Images, Objects, Contexts and Interactions in Social and Cultural Inquiry'. London: Sage

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

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

Holstein, J.A. and Gubrium, J.F. (2004). in C. Seale, G. Gobo, J.F. Gubrium and D. Silverman (eds) *Qualitative Research Practice*, London: Sage.

Holstein, J.A. and Gubrium, J.F. (2004). 'Context: working it up, down and across', in C. Seale, G. Gobo, J.F. Gubrium and D. Silverman (eds) *Qualitative Research Practice*, London: Sage

Mauthner, N., Parry, O. and Backett-Milburn, K. (1998). 'The data are out there, or are they? Implications for archiving and revisiting qualitative data'. *Sociology*, 32(4), 733-45

Moore, N. (2005). '(Re)Using Qualitative data? Thoughts from the CRESC Qualitative Research Laboratory'. Paper presented at 'Re-using Qualitative Data: CRESC Methods Workshop', University of Manchester, 28th September 2005

Thompson, Paul (2000). 'Re-using qualitative research data: a personal account' [48 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [Online Journal], *1*(3), Art. 27, http://www.qualitative-research.net/fqs-texte/3-00/3-00thompson-e.htm.

Notes

[1] We also note that data themselves should be thought of as representations, since they are the researchers' understandings of the phenomena under investigation. Hence, data records are representations of representations. Nevertheless, our point stands, in that the two types are not the same thing; nor are the records able to represent all the modes of meaning available in the field of action itself. Hence, it makes sense in our view to maintain this distinction, particularly as it points to crucial issues in data- re-use.

[2] In fact, writing employs other modes too in its graphical dimensions – such as print colour, directionality, size, etc.

[3] Of course, much depends on the camera style used: handheld cameras that get close to the action can also convey a sense of subjective positioning.

[4] We recognise that in suggesting written narrative's enhanced powers of conveying immersion in events, we are on contentious ground, since it must be recognised that there are many styles of camera-work, including both fixed tripods that distance the subject, and much more engaged, mobile camera styles that gets in close to actors (see for example, the celebrated 1950s/60s films of the French anthropologist Jean Rouch). Nevertheless, writing makes a particular subjective narrative explicit and relatively unambiguous; this is not the case with film.

[5] Marcus Banks's book (2001) is useful as a more detailed examination of the contextual factors that allow an image to be read - both one's own images and those generated by others.