Anonymising Research Data

Andrew Clark
University of Leeds

ESRC National Centre for Research Methods

NCRM Working Paper Series

7/06

Real Life Methods

Working Paper

Anonymising Research Data

Andrew Clark
University of Leeds
December 2006

Author contact details

Dr Andrew Clark
Real Life Methods
Leeds Social Sciences Institute
Beech Grove House
University of Leeds
Leeds
LS2 9JT

a.j.clark@leeds.ac.uk

Summary

This document outlines some thoughts and discussions we have been having about strategies of anonymisation of data to be collected through the ESRC / NCRM Real Life Methods Node *Connected Lives* project[1]. It is commonplace for social science research to adopt a policy of 'blanket anonymisation', whereby all names, places and other identifying features are disguised across a data set, including from interview transcripts, diaries and field notes. Here, I consider the practical and theoretical implications of such a strategy and suggest that anonymisation is not a process to be conducted – and assumed completed – at just one stage of the research process. Moreover, anonymisation strategies cannot be separated out from other methodological (such as issues around archiving or mixing methods) or indeed substantive issues (such as enabling deeper appreciation of the relationality of networks, or the ways in which space might be constructed). The implications of whatever anonymisation strategy researchers adopt on the future ability to appreciate the social and spatial processes behind networks, neighbourhoods and communities, need to be made clear throughout the research process. In summation, this document argues for a more reflexive, iterative approach to anonymisation and confidential that situates these, and other ethical concerns, in the context of the social process.

Keywords Anonymisation, Data, Ethics

---

[1] Though of course, any errors contained herein are mine.

1. Introduction

This document outlines some of the issues surrounding the anonymisation process in the *Connected Lives* strand of the ESRC's Real Life Methods Node[2]. Discussion begins with the rationale for anonymisation, outlines some practical and substantive issues concerning anonymising data and raises some concerns about how best to go about the practise of anonymisation. It ends with some suggestions for thinking through the challenges of anonymising 'real life' data. The discussion has implications for data analysis, user-engagement, research output and data archiving. This document purposefully avoids presents a prescriptive, and somewhat normative, outline of how best to go about developing an ethical anonymisation strategy. Rather, it calls for a more reflexive, iterative approach to ethical concerns (of which anonymisation and confidentiality are a part) that situates them more explicitly in the context of the research process. In this respect, much social research may be differentiated from legal and/or biomedical ethical discourses in that much social research requires ongoing, emergent ethical approach.

2. The ethics of anonymisation

> *'The anonymity and privacy of those who participate in the research process should be respected. Personal information concerning research participants should be kept confidential. In some cases it may be necessary to decide whether it is proper or appropriate even to record certain kinds of sensitive information.*
>
> *Where possible, threats to the confidentiality and anonymity of research data should be anticipated by researchers. The identities and research records of those participating in research should be kept confidential whether or not an explicit pledge of confidentiality has been given. Appropriate measures should be taken to store research data in a secure manner. Members should have regard to their obligations under the Data Protection Act. Where appropriate and practicable, methods for preserving the privacy of data should be used. These may include the removal of identifiers, the use of pseudonyms and other technical means for breaking the link between data and identifiable individuals such as 'broadbanding' or micro-aggregation. Members should also take care to prevent data being published or released in a form which would permit the actual or potential identification of research participants. Potential informants and research participants, especially those possessing a combination of attributes which make them readily identifiable, may need to be reminded that it can be difficult to disguise their identity without introducing an unacceptably large measure of distortion into the data.'*
> Statement of Ethical Practice, Social Research Online
> (www.socresonline.org.uk/info/ethguide.html)
>
> *'No matter how sensitive the information… ethical investigators protect the [participant's] right to privacy by guaranteeing anonymity or confidentiality. Obviously, information given anonymously secures the*

---

[2] More information about the *Connected Lives* project can be found in Appendix A.

> *privacy of individuals, but this safeguard is usually possible only in surveys using self-administered questionnaires without names attached… Most often the investigator can identify each individual's responses; therefore, the principal means of protecting research participants' privacy is to ensure confidentiality. The researcher can do this in a variety of ways: by removing names and other identifying information from the data as soon as possible, by not disclosing individuals' identities in any reports of the study, and by not divulging the information to persons or organizations requesting it without the research participant's permission'*
> (Singleton and Straits, 1999; 524).

It is common practice for researchers to protect the identity of those who participate in research. Although frequently considered in tandem (e.g. Christians, 2000; Homan, 1991; 140-150, Singleton and Straits, 1999; 524-525), it is important to recognise the distinction between anonymity and confidentiality. Anonymity is the process of not disclosing the identity of a research participant, or the author of a particular view or opinion. Confidentiality is the process of not disclosing to other parties opinions or information gathered in the research process. While this discussion is only concerned with anonymity, this is not to deny the link between the two. Singleton and Strait (1999) argue that complete anonymity in most social research is impossible to achieve, and, as I argue here, anonymity is perhaps best approached here as a characteristic of the relationship between the researcher and the research participants.

There are three broad reasons for anonymising research data. First, anonymisation aims to 'protect' or hide the identity of research participants. This is particularly important when sensitive, illegal, or confidential information may have been disclosed during the research process, or when information is disclosed which may cause the participant distress should other parties learn such information. Anonymisation is thus an ethical issue which must be considered throughout the research process.

Second, in addition to the anonymisation of individuals, there is often a requirement to disguise the identification of research locations. In part, this is to further protect participants from being identified through research locations, but also because there may be good cause to anonymise the research location. For instance, some localities have become synonymous with deprivation, reportedly 'anti-social behaviour', social tension and the like. Research monographs and papers, government policy documents and media reports can often contribute to the stigmatisation of particular people in particular places. Conducting and reporting on research about particular problems in particular locales has the potential to perpetuate stigmatising discourses about place. Consequently, while not necessarily preventing such perpetuation, ensuring that particular research places cannot be identified in research outputs will at least not contribute to these stigmatising processes (Clark, 2003).

Finally, beyond these ethical concerns, there are legal requirements to ensure the protection of personal information and participants' identities under the UK Data Protection Act (1998) which came into effect in March 2000 (Grinyer, 2002; Parry and Mauthner, 2004). Under the terms of the Act, regulations for obtaining, holding using or disclosing information about individuals have been tightened in order to maintain the anonymity and confidentiality of personal data about individuals collected during the research process. However, there are certain exemptions for personal data processed for research purposes. Under the

Act, research data may thus be 'processed for purposes other than for which they were originally obtained, they can be held indefinitely, and research subjects do not necessarily have the right to access these data' (Parry and Mauthner, 2004; p143), though whether such exemptions would apply without challenge is unclear.

I now consider the implications of anonymity in the research process, drawing where appropriate on recent experience from the *Connected Lives* project[3]. I argue that it is not adequate to assume that anonymisation at just one stage of the research process (say at the point of transcription) will be sufficient to either protect identities at all stages of research, or indeed, whether protecting the identities of participants at all stages is necessarily the best thing. In addition, there are practical and epistemological concerns in the anonymisation of spatial data which must be considered in the research process. The discussion ends with an outline of a proposed strategy for anonymising data in the *Connected Lives* project.

3. The practice of anonymisation

While there are strong ethical and legal justifications for anonymising research data, this process is fraught with practical difficulties. First is the issue of what, or who, to anonymise. Commonly a process of 'blanket anonymisation', whereby all people (including third parties) referred to in interview transcripts, field notes, diaries and other data forms, are anonymised at the earliest opportunity (usually, at the point of transcription). Usually, this is done by replacing real names with pseudonyms or relying on initials. Often places too undergo a similar process of anonymisation. Such a strategy can be summarised as an attempt to remove 'background data' from the opinions or information presented about particular individuals. Morse for example, is unequivocal that researchers protect the identities of participants thus;

> *'At the beginning of the study (when giving informed consent), the participants were promised anonymity for their participation. The researcher must check carefully that none of the quotations used [in publication] makes a speaker recognizable through some contextual reference. He or she must ensure that demographic data are presented in aggregates, so that identifiers (such as gender, age, and years of experience) are not linked (making individuals recognizable) and are not consistently associated with the same participant throughout the text, even if a code name is used. This prevents those who know all the participants in the setting from determining who participated in the study and who did not' (1998; 79–80).*

There are several concerns that make such blanket anonymisation of all people, including third parties, not so straightforward. First, is the way in which decisions are made about what sorts of information to anonymise and which to leave in original form. For instance names, age, gender, ethnicity, and location (or address), are often removed from research data, but this should not be an arbitrary decision. There is the potential to identify particular participants based on a combination of these features without having access to that individual's

---

[3] For the purpose of this discussion, I adopt a strategy of blanket anonymisation in spite of criticisms I make of such a strategy.

name. Yet while such information can be disguised or removed for publication, as I later argue, it is much more difficult to justify this in the case of data archiving. The second issue is the tendency to reduce such data to 'background information'. Yet such data is not just 'background' information, but also provides context for deeper, and fuller, understanding of the empirical data. Yet knowing when 'background data' might become 'context' depends on the purpose of the research (or on the research context itself). Take for example, the issue of anonymising age, gender, sexuality, or political or religious beliefs alongside people and place names. As I argue later, it may be that such characteristics are crucial for analysis, if not in the current project, than at a later date should it be archived.
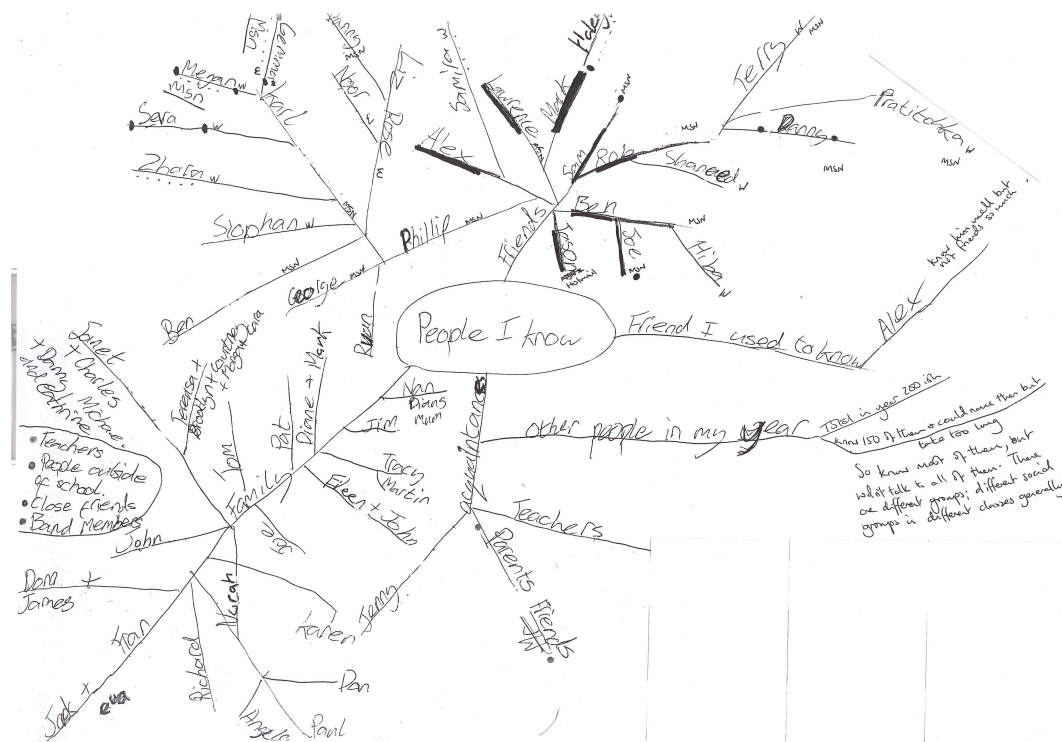
It is necessary to comment on the anonymisation of names and places. Perhaps of primary concern when anonymising names is the issue of pseudonym selection. It is well documented that names have social and cultural significance. Both personal (first) names and surnames imply particular ethnic, religious, class and age based connotations, which will inevitably be transferred to any pseudonyms. Thus through our anonymising process, we are in turn conforming to stereotyping practices and, potentially, inferring all sorts of connotative baggage onto research participants that may or may not be appropriate. For the *Connected Lives* project, it may be that names are themselves analytically significant. For example, friendships might initially be formed between young people who sit beside each other in school because they have alphabetically adjacent surnames. Parents may follow particular familial traditions in the naming of their children (such as following a particular theme: for instance I know of three siblings called Rose, Violet and Daisy) or follow family traditions and adopt ancestral names (such as sons named after fathers or grandfathers, or both). Should an anonymisation strategy take account of these issues?

Second is the issue of which individuals to anonymise. For example, should individuals speaking in a professional capacity (perhaps democratically elected ward councillors) be anonymised? What about individuals who represent a particular interest group but not an entire profession (such as residents' group members, local development workers, or GPs)? In some neighbourhoods (including the field site for the *Connected Lives* project), it may be that particular individuals are more well known than others, or are seen as 'key people' in particular neighbourhoods. The decision to anonymise such individuals is of ethical importance, for it implies that some individuals' rights to privacy are less important than others. Moreover, while a pseudonym may suffice for those unfamiliar with a locale, anyone familiar with it may still be able to identify the place, and people associated with it quite easily.

A third issue concerns the resources required to anonymise particularly large or complex, data sets. Again this is particularly relevant to the *Connected Lives* project which is creating participatory social networks. Such networks may contain many named individuals or groups of individuals (Figure 1). Should all these 'third party' names be anonymised? And if so, how? What about individuals mentioned in 'passing' in discussions during the construction of their networks? If all these individuals are to be incorporated into an anonymisation dataset, the length of time required to do this must not be underestimated. While other research methods, such as interviews or participant observation, might also reveal sensitive data about interrelationships and connectivity, I think it is particularly problematic in social network research. For example, permission to collect data was obtained from just one person in Figure 1 (the individual in

the centre of this ego-centric network) yet a whole range of data has been amassed about individual who have not expressed their permission to appear in the research data. Consequently, while this network may have been collected ethically with regard to the individual in the centre (the 'ego'), it is important to question whether this is an ethical act with regards the permission of the rest of the network. The information revealed in a network may be particularly sensitive for those individuals included within it and who could quite feasibly recognise themselves within it. Importantly, data that might not be seen as 'sensitive data' (such as a series of relationships between people) by one groups of individuals (say academics) at one particular time, might be considered particularly sensitive by other people, or at a future point in time. Consequently, and as argued later, while anonymising data may ensure the confidentiality of the data it is important to question the implications of this for network analysis (not least because such analysis may be rendered impossible because of the extent of anonymisation).

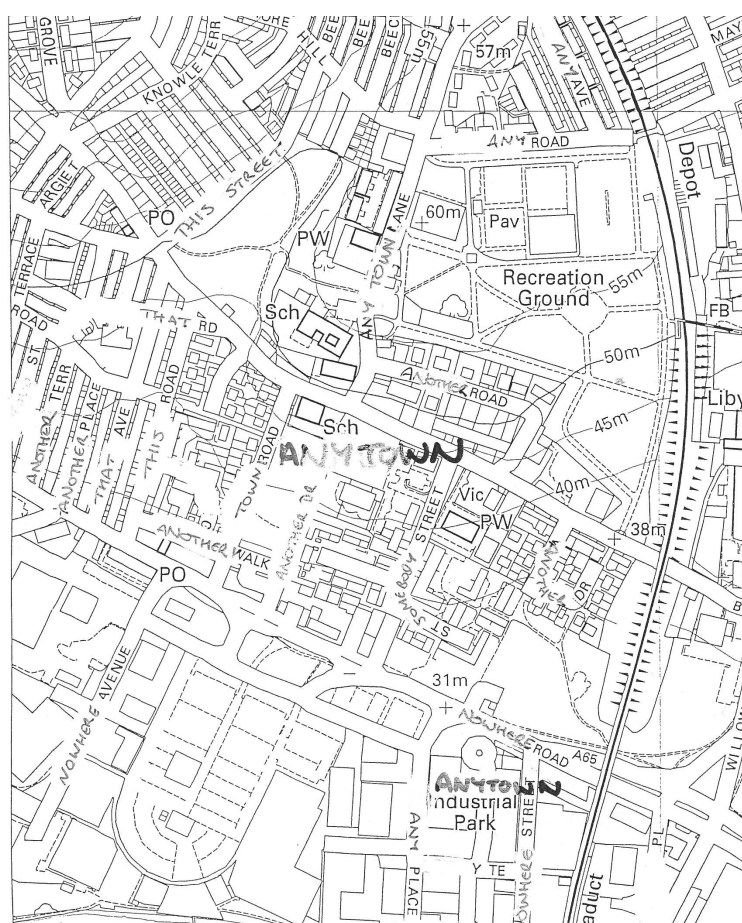**Figure 1: A trial participatory social network (surnames anonymised)**

Of course, it is not just an individual's name that defines his / her identity, but also ethnicity, politics, gender, sexuality, place of residence and so on. The UK Data Protection Act considers racial or ethnic origin, information on political affiliation, religious or other similar beliefs, trade union membership, information on mental or physical health, criminal convictions, and sexuality to be 'sensitive data' and thereby warranting particular protective attention. I do not want to discuss the implications of protecting such data, but rather suggest that 'context', be it biological, biographical, social, economic, or spatial, all contribute to identity construction. For this reason it is commonplace to anonymise the addresses and postcodes of research participants. Yet this too is not necessarily straightforward. Figure 2 is an attempt to anonymise all place names in a small section of the field location for the *Connected Lives* study. Hopefully, the complexity of the task of anonymising is self-evident, and it can

be noted that even here not all place names have been removed with some 'slipping through the net'. Yet does this ultimately matter? For anyone familiar with the location will surely recognise the street layout and position of particular landmarks.

The difficulties of anonymising place are as complex as those concerning the anonymity of people. For example, at what scale should the anonymisation occur? At street level? Or neighbourhood? Perhaps district or area? Or maybe even the city? Consequently, a decision to anonymise all place names is perhaps the least appropriate process of anonymisation, for reasons I outline below, for even if location identifiers are removed from a data set, places remain identifiable because they are constructed through stories and myths, gossip and historical events. Some areas for example, may become well known locally or nationally for particular events reported in the media such as civil disturbances, emergency evacuations, or links to terrorist networks. How can such stories be 'disguised'? Or, should they be excluded from the data set in some way? What if they form crucial elements of the everyday social practices in a given neighbourhood or street?

Figure 2: An attempt to anonymise 'place'

Even if all individuals and places are anonymised in research outputs, there is no guarantee that this will preserve participants' identities. Illustrative of this is the well cited case of Vidich and Bensman's (1958) research on the community of 'Springdale', a small town in upstate New York. Informants were promised that no individuals would be identified in printed reports. However, when their *Small*

Town in Mass Society was published, the people of the town could easily identify each character in spite of the use of pseudonyms. The participants were so outraged by this invasion of privacy that they featured a float in the annual Fourth of July parade with a large-scale copy of the jacket of the book. This was followed by residents riding masked in cars labelled with the fictitious names given them in the book, and then by a manure spreader, with the effigy of one of the authors (Vidich) bending over the manure (Singleton and Straits, 1999, Whyte, 1958).

It is not the researchers alone, however, who are responsible for preserving the anonymity of research participants, and there are instances when the scale and scope of anonymisation becomes relevant for participant recruitment. For example, when conducting neighbourhood based, or indeed any community-based researched, particular individuals may assume positions as local gatekeepers. Thus a researcher may gain access to particular participants either through direct introductions by such gatekeepers, or, through word of mouth. Either way, it is clear that the role of such individuals in the research process cannot remain anonymous, or even confidential, within the research team, but rather stretches across the research environment.
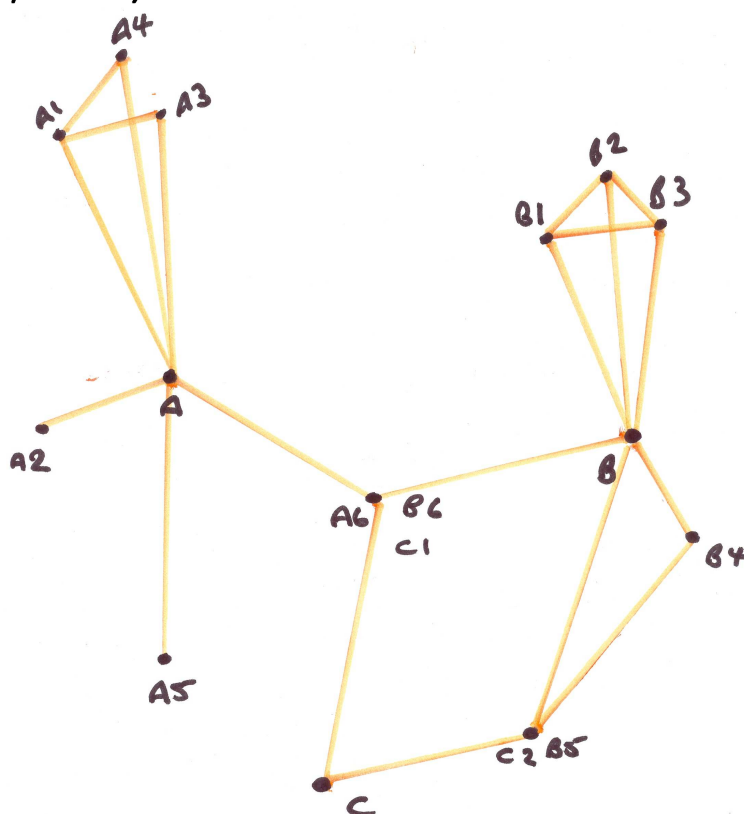
A fourth concern of anonymisation concerns the potential for user engagement in research. Much research, including *Connected Lives*, has an applied dimension, with the intention to at least contribute to the policy discourse, if not to policy decision-making directly. However, the anonymisation of places can have repercussions for the extent to which non-academics, such as those in national, and especially local, policy spheres, can engage with the 'real life' implications of the research. Yet openly declaring, say in a subsequent interview, presentation or dissemination document, which neighbourhoods or among which groups of individuals, a particular piece of research was conducted, makes any geographical anonymity quite pointless.

Fifth, there are methodological implications for anonymity in mixed methods research. Even if all personal and place names are anonymised in the data, there remains a strong possibility that these places and individuals will remain identifiable through the triangulation of different research methods. For example, one method *Connected Lives* are using is a 'neighbourhood walkaround'. These take the form of a tour of a neighbourhood as understood by a respondent. The resultant data include photographs, a walking transcript and a map (see Figures 3 and 4). The difficulties of anonymising such data are clear. First, while it is possible to erase the name of each street, neighbourhood and district, replacing these names with pseudonyms would be an enormous task. Second, even if all place names could be given pseudonyms, access to data collected through alternative methods, especially interview transcripts, but also photographs, makes such a process redundant. Consider for example, Figure 4, in which both the individuals, and the building, would need to be 'anonymised' in order to ensure confidentiality of research participants. I argue that anonymising place names, particularly in a multi-method research project, becomes a futile exercise because it is quite simple to identify such places through other data sources.

Finally, it is worth considering whether research participants want to remain anonymous (e.g. Grinyer, 2002). If an aim of some (particularly participatory) research is to provide voice to marginalised or disempowered individuals and groups. Yet if an individual chooses not to be anonymised in research outputs,

preferring say, to 'tell his / her story' up front, this raises important questions about who has ultimate control over the research data.

Figure 3: Annotated map completed by researcher with input from research participant

Figure 4: Photograph of a research participant embarking on a 'neighbourhood tour'

4. Anonymisation and the importance of context

I now turn to a substantive issue about anonymisation practices in social network research and consider the implications of anonymising data for subsequent analysis and the implications of this for understanding the nuances of real life.

An objective of the *Connected Lives* project is to understand how social networks operate over time and across space. One outcome of the project will be a number of participatory network diagrams (see Figure 1). I have already outlined that this will be a complicated and time consuming process, but there is also a more substantive issue at stake. If we assemble networks from a number of individuals in a given area, there is a strong probability that these will begin

to overlap or connect. This is not to infer that Milgram's 'six degrees of separation' thesis is necessarily adequate to explain connectivity (see Clark, 2006 for a review), but rather that the opportunity for interconnectivity is high. For example, of the participatory networks completed in the 'trial' stage of the research, at least two participatory networks overlap. Of these, one participant was known to the researcher from school some years previously, the other seemingly was not. During the latter interview, it unexpectedly emerged that the second participant had also attended the same school, at the same time. Consequently all three networks overlapped at a particular time in a particular place. This is represented diagrammatically in Figure 5. It can be seen that those individuals in different networks could be given different pseudonyms according to whichever network is being anonymised. Thus some individuals may receive two pseudonyms (e.g. C2 / B5) or even three (e.g. A6 / B6 / C1). As a result, anonymising data at the earliest stage possible (such as at transcription – see for example recommendations by Qualidata, 2003) may be detrimental to the analysis process[4]. Moreover, if two apparently randomly selected individuals happened to be connected, via a third party (in this case the researcher), it is highly likely that networks of individuals who reside in the same location will also connect. Hence it is vital that to keep accurate, detailed records of all individuals, and indeed places, in order to recognise interconnected networks.

Figure 5: Model of hypothetical interconnected social networks for three individuals (A, B and C)

However, there is a broader concern about anonymisation involving the eradication of context. Much has been made of the importance of context in the

---

[4] By analysis I mean all activity carried on out the empirical research data, including initial comprehension, identifying potential coding, and the exploration of early linkages between data from different sources or participants.

interpretation of social data. Yet it is this context (locations, times and dates, ethnicity, family names etc.) that are often removed or disguised through the anonymisation process. Parry and Mauthner (2004, 2005) specifically consider this with regard to anonymising data. They comment on the relationship between the researcher and primary research data is often considered as part of an interpretive and reflexive epistemology, and suggest that as a consequence archived material may be incompatible with any subsequent analysis beyond methodological exploration. They take issue with others such as Fielding and Fielding (2004; 697) who view the 'recovery of contextual features in secondary data analysis as a practical rather than an epistemological matter'. For Parry and Mauthner (2005; 339-340) this is an unsuitable 'practical solution' to an 'epistemological problem' (see also Mauthner *et al.*, 1998). This is not only because any 'recovery of context can only ever be partial' but also;

> *'even if it were possible to capture 100 percent of 'context', we do not feel its status, in respect of re-analysis, should be so elevated. On the contrary, we see it as different from, unequal to, and frankly a poor substitute for the original interpretative endeavor' (Parry and Mauthner, 2005; 340).*

While in agreement with Parry and Mauthner's (2005) argument, it is useful to consider further the relevance of 'context' in the *Connected Lives* project. While not wanting to enter into debate about what 'context' might or might not be, various 'lay definitions' of the term from various internet web-pages are listed in Table 1. A generic definition might consider it the circumstances and conditions (temporal, social, political, environmental, economic etc) that 'frame' or 'surround it'. However, rather than being mere 'background data' that can be 'added in' to research data, 'context' can be crucial for understanding real life.

The idea of 'context' has undergone a more thorough consideration in realist epistemology. Realist explanation is committed to 'ontological depth' in explanation. That is, since social events are interwoven between various layers of social reality, then so must be any account of them (Pawson, 1995). Rather than viewing 'context' as a single strata or layer of 'background data', it is crucial to understand how this context contributes (in realist terms, how it works as a mechanism) to the construction of the social world. With reference to Figure 6, Pawson explains that;
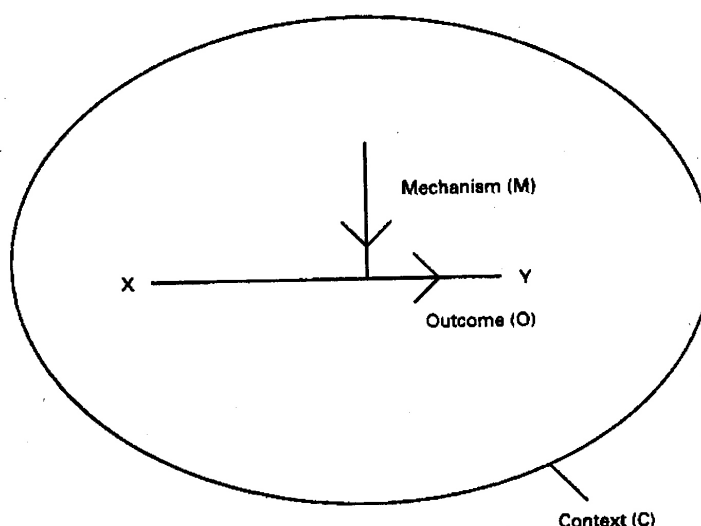
> *'[t]he basic task of sociological inquiry is to explain interesting, puzzling, socially significant outcome patterns (O) between events of happenings or social properties. Explanation takes the form of positing some underlying mechanism (M) which generates these outcomes and thus consists of propositions about how the interplay between agency and structure has constituted these outcomes. Explanatory closure requires that, within the same investigation, there is also an examination of how the workings of such mechanisms is contingent and conditional, and thus are only fired in particular historical or institutional contexts (C)' (1995; 301).*

Table 1: Lay definitions of 'context' in academic disciplines
(adapted from www.wikipedia.com. See http://en.wikipedia.org/wiki/Context; emphasis all mine)

Archaeology	The context (*physical location*) of a discovery can be of major significance. More precisely, an archaeological context is an event in time which has been preserved in the archaeological record. The cutting of a pit or ditch in the past is a context, whilst the material filling it will be another.
Communications and linguistics	Context is the *meaning of a message* (such as a sentence), its relationship to other parts of the message (such as a book), the environment in which the communication occurred, and any perceptions which may be associated with the communication. Thus, context is a *'frame'* through which one views a message.
Computer science	Context is the circumstances under which a device is being used, e.g. the current occupation of the user
Contemporary art	Context is often used to describe everything other than the content of the piece of work. For example the way in which a painting is hung within a gallery, the political situation at the time of viewing, the amount of wine consumed at the private view.
Psychology	Context refers to the *background stimuli* that accompany some kind of foreground event. For example, if a rat is foraging and is frightened by a cat, the place (and possibly time) of foraging is the context and the cat is the foreground event.

Figure 6 Basic elements of realist explanation (Pawson, 1995; 300)

Speculating on 'what works, for whom, and in what circumstances' (Pawson, 2005) then, requires paying attention to what others may dismiss as 'background' information, including locality and the social spaces of networks. If such data as that constructed in *Connected Lives* is anonymised and presented as a series of socially-networked pseudonyms and imagined geographies, there is a

danger of reifying this information, perhaps not at the first stage of analysis (within the *Connected Lives* Core Team), but in any subsequent research project that seeks to make sense of the data once archived (and out of the hands of the original research team). In such cases, there is a possibility of subsequent researchers falling into a trap of 'naïve realism' (Mauthner and Parry, 2004), in so far as either the reflexive and interpretive nature of the data, or its representation of 'real life' becomes just as 'real' as the experiences such data seeks to represent.

Thus the final issue worth exploring concerns the potential reification of anonymised research data. Here, it is useful to consider again the arguments put forward by Mauthner, Parry and Backett-Milburn about the importance of context in archived research (Mauthner *et al.,* 1998; Parry and Mauthner, 2004; Parry and Mauthner, 2005). Commenting on concerns about the re-use of archived qualitative data, they surmise that the usefulness of such data might be jeopardised if contextual information surrounding its production is not also provided. A major concern in the archiving of qualitative data has been raised concerning the preservation of confidentiality and the risk of exposing participants' identities, particularly given forceful arguments for supplying 'background knowledge about how, why, by and for whom the research was carried out' (Mauthner *et al.,* 1998; 734, cf Hammersley, 1997), along with detailed fieldnotes about the data. It has been argued that the provision of such 'contextual' information will complete archived datasets and enable the generation of new findings or theories (Hammersley, 1997). Mauthner *et al.,* (1998) claim that this ignores the reflexive and interpretive nature of many qualitative research paradigms. They argue that research data are not 'facts' lying round waiting to be found, but are social constructs 'created through the interaction of particular (either primary or secondary) researchers with particular respondents in particular locations at particular historical junctures' (Mauthner *et al.,* 1998; 735). Consequently, given the conditions under which data are produced are inescapable, its reinterpretation at a later point becomes somewhat problematic, even when missing 'background' (viz. contextual) data are 'put back in' by way of pseudonyms.

By way of illustration consider the role of spatial 'context' in anonymised data. It may be common to consider aspects such as time and location (viz. a form of space) to be contextual (viz. background) data that can, and should be anonymised in datasets. Yet as many have argued within (e.g. Harvey, 1989; Massey, 2005; Soja 1989) and beyond geography (e.g. Lefebrve, 1991), space is about much more than mere 'background'. Put simply, space and society are increasingly considered to be recursively constituted. Drawing on this, I have previously explored the relevance of spatial 'context' on the reproduction of social networks, contacts and ties (Clark, 2006). However, the relationship between the two is more complicated than a society-space dialectic. Understanding the productive power of space requires an appreciation of the nuances of how different social relationships are constituted in different times and spaces. Here, caution is required in anonymising so much 'context' – including place names and locations - that the capacity to understand the illuminary capacity of this space (as more than 'background context') may disappear into a fictive or imaginary geographic context. *Thus when anonymising contextual data, there are more important epistemological concerns than practical issues such as the complexity of the process.*

My concern here is not just about the potential for a highly complex strategy to anonymise people and place, nor about the potential to produce a somewhat fabricated geography. Rather it is about the way in which space may be essential to understanding processes which ultimately reproduce the structuring of society. It is somewhat straightforward to view spatial arrangements as a reflection of social divisions, and consequently the outcome of different social processes. Here, space might be seen as a medium that can 'fix' social processes long enough for them to be scrutinised. However, as critiques of such a 'spatial science school' approach to spatiality emphasise, the spatial cannot be set up as a measure of the social because this relies on the assumption that the two are quite distinct. Moreover, this also takes for granted many social categories (ethnicity, gender, class, sexuality etc.) that, far from being real, fixed, and mutually exclusive, are themselves representative of particular social constructions. Consequently, analysts have become more interested in the ways in which place and position actively contribute to the construction of society. Rather than asking why particular social outcomes might be located in particular places, some turned to questioning the difference that space makes. Fundamentally, this is a question about the ways in which 'context' can produce different social outcomes.

By way of illustration, it is useful to turn to s transcript of a walking interview conducted with a female participant - RP - as part of the Connected Lives research project. RP was invited to 'show us around' her version of the Connected Lives field site. It is widely accepted that 'place' is a personally felt construct, and inevitably the data produced was an individualised representation of RP's local geography. However, the method also reveals an intricate relational geography. The incident in the extract below concerns RP's attempts to explain the history of a series of disturbances in the area some years ago. The extract illustrates quite neatly how the field site – H---- P---- has import links with other parts of the city, including streets – Y---- R----, and pubs – the T----- . It is through appreciation of the relationships between these places that it becomes possible to understand some of the ways in which RP constructs her own view of 'community' in the area. For instance, it is perhaps less relevant whether this account is accurate or documents the 'real' cause of the disturbances, than the ways in which the incident is located 'outside' the area. For RP, the roots of the disturbances do not lie with anything to be found within the area (note how even the absence of marijuana is labelled as a problem about supplies from elsewhere). RP does not absolve 'blame' from the area's residents, but rather, attributes causes to relations between H---- P---- and Y---- R----. Understanding the relationality between places is thus crucial to enabling appreciation of how particularly communities are constructed. But such an understanding depends on as full an account of geographical 'context' as possible.

> RP: And it was a very bad time because there'd been the incident up at T----. Easy D----, you're alright?
> [man walks past]
> A: There'd been an incident up at?
> RP: Up Y---- R----. What had happened there was these guys that were kind of bullies. They must have beaten up two lads from down here. A gang, not of us lot, cos we call from here, H---- P----, a gang of men have gone from here, up to Y---- R----, started shooting at this pub, big fight's happened, so it caused a load of tension. So we're all living down here, we know what's happened and erm, so everybody's scared, cos before we know it we're seeing these different guys driving up and down the road and you're thinking [intake of breathe] is it them come from t' Y---- R---- coming down now to get their revenge back on whatever's happened up there, so everybody were on, we're all like that, hot weather like this, there's no weed about as I say, marijuana, so there'd been a drought on that which maybe sounds disgusting but erm the police decided to raid, one two, about three houses on here, so everybody's already scaredy, everybody's already fed up and then the police decide to do these raids. So it peed a lot of people off. So I remember the night cos I was sat on this wall. Group of lads, just lads here talking about we wanna do summat, we need to get back and start speaking, we're sick of how we've been treated by the police. How we're being treated. What can we do? So that's where the car was actually burnt out, up there, but as I say you're actually walking in a how they say no go area…
> *(Extract from walking interview with research participant)*

Finally, all the issues discussed here become more acute when the preparation of data for archiving is considered (see Bishop, 2005; Hammersely, 1997; Mauthner *et al.,* 1998; Parry and Mauthner, 2004; Parry and Mauthner, 2005). While this is a matter for another discussion, it is important to recognise that in preparing data for archiving, researchers may eliminate much 'context' which for them is seen as irrelevant, but which for others may be crucial for any subsequent re-analysis. Consider, for example, the strategy applied by Platt (1976: 7) in her study of professional sociologists:

> *'I have… gone to some lengths to make it difficult to identify individuals or projects in the text, though I have always been conscious of the danger of concealing relevant information by doing this. All names have been changed, and sometimes sexes; institutions, affiliations and research topics have been changed or described in general terms… If the same name appears in more than one place it does not refer to the same individual' (cited in Homan, 1991; 145)*

It is subsequently difficult to imagine what information was left uncorrupted in this study. While Platt may be acting with the best intentions it is questionable how useful this dataset would be for any future work conducted on it.

5. Ways forward

I now want begin to think about how some of the difficulties discussed here might be worked through. At the onset, it is important to recognise the ethical and legal obligations researchers have to research participants that their anonymity will be protected. However, it should also be recognised that participants may wish to be fully identifiable in research outputs and data archives (Grinyer, 2002). If anonymity is not possible, or (and perhaps more likely) if the potential for identification exists, then being open and honest with participants is of course the most ethical of all anonymisation strategies. If a strategy of anonymisation is pursued, it is vital that the personal identity of research participants remains hidden beyond the research team wherever feasibly possible. Of particular concern here have been the implications for anonymising spatial data. The possibility of identifying participants through particular locations, and the ethical implications of this, must be recognised. Yet ensuring complete anonymisation of place may be an impossible task, particularly if researchers are to fully appreciate the importance of context in the reproduction of social process, structures, and everyday life.

There may be a requirement to develop different levels of anonymity during the dissemination process. For example, user engagement and local dissemination may require a different strategy for the anonymisation of place than publication in academic journals. It may be difficult for those working at a local level to fully appreciate the implications or recommendations of research if the local context informing such recommendations is removed. Indeed, anonymising location specific identifiers from research outcomes may make the delivery of local-level policy or service delivery responses almost impossible. Yet the implications of this must also be considered; for revealing identifiers such as place-names at one scale of dissemination may make their anonymisation at other scales futile[5].

Third, it is necessary to leave a clear 'paper trail' of any anonymisation strategy (see also Mason, 2006). There must be careful consideration about where and how this is kept secure and who has access to it. While there is little point keeping the anonymity notes and codes in the same filling cabinet or hard drive as the data itself, the pressure on physical space imposed by any complicated strategy may make this difficult to achieve in practice.

Fourth, there needs to be more discussion about the process of anonymising archived data. In particular, and given the arguments considered above about the reflexive, interpretative epistemology of much qualitative data, there must be more discussion about the ethical implications of archiving complete data sets. For 'incomplete' datasets may be less valuable if archived with potentially crucial 'contextual' data omitted.

Fifth, it is relevant to consider the timing of anonymisation in the research process. It is frequently suggested that this must take place as close to the point of data collection as possible (e.g. ESDS Qualidata, 2004). For some projects (including *Connected Lives)* it may be more appropriate to instead anonymise transcripts and other data at as late a stage as possible. While this may make for a more time consuming process in terms of practicality (as has been shown) a

---

[5] For example, a close reading of the reference lists for many locality based research projects can identify that location through reference to other research outputs from a study.

blanket anonymity strategy will be far more time consuming where analysis is concerned. It is possible that some data is not used directly in research dissemination and which consequently does not enter the public realm. Notwithstanding the implications for secure data storage and future archiving, it may thus not be necessary to anonymise all the data.

Finally, it is clear that *anonymisation is an ongoing process of negotiation, reflection, and experimentation*. A strategy that seems appropriate at one stage of the research may not be by the end of it; likewise, what is appropriate for one research project, or indeed, thread of analysis, may not be appropriate for other projects or forms of analysis. It would appear that anonymisation must be an issue that the Real Life Methods Node should be approaching with an open mind as to how best to proceed, and learn, from the fieldwork process.

6. Conclusion

Anonymity is an ethical, practical and epistemological issue. Ultimately the suggestions proposed here may well remain a practical solution to an epistemological concern (Mauthner *et al.*, 2004). And I am certainly not advocating that we do not anonymise data - Horner (1998) for example has outlined some of the moral implications of the loss of privacy in research. But nor do I think a blanket process of anonymising all 'background data' is necessarily appropriate. Rather, I am calling for more reflexive scrutiny of whatever anonymisation strategy we choose to adopt. After all, an inappropriate strategy could result in inappropriate data analysis.

These issues have been discussed in the context of social network research, and here, I have also begun to consider the importance of connections within and between data sets. Specifically, I have highlighted the importance of understanding social networks through their social and spatial context, including the relationship between space and society in the reproduction of social structures and patterns, and the relationship within and between networks constructed by different participants. If networks are to be understood as being created in, and existing across, different sorts of context, then it is crucial that such context is made available for analysis without having been anonymised (and consequently 'decontextualised') first.

References

Bishop, L (2005) 'Protecting respondents and enabling data sharing: reply to Parry and Mauthner' *Sociology*. 39, 2, pp 333-336.

Christians, C (2000) 'Ethics and politics in qualitative research'. In N Denzin and Y Lincoln (eds.) *Handbook of Qualitative Research* 2nd edition. Thousand Oaks, Ca. Sage, pp 133 – 155.

Clark, A (2006) 'Understanding community: a review of networks, ties, and contacts'. NCRM Real Life Methods Node / University of Leeds. Unpublished Discussion Paper.

Clark, A (2003) 'Wish You Were Here? Experiences of Moving Through Stigmatised Neighbourhoods in Urban Scotland'. University of St Andrews. Unpublished PhD Thesis.

ESDS Qualidata (2004) Identifiers and Anonymisation: Dealing with Confidentiality. URL consulted 26th July 2006. www.esds.ac.uk/aandp/create/identguidelines.asp/

Fielding, N and Fielding, J (2000) 'Resistance and adaptation to criminal identity: Using secondary analysis to evaluate classic studies of crime and deviance' *Sociology*. 34, 4, pp 671 - 681

Grinyer, A (2002) 'The anonymity of research participants: assumptions, ethics and practicalities'. *Social Research Update*. 36. University of Surry. URL consulted 30th July 2006. www.soc.surrey.ac.uk/sru/SRU36.htm

Hammersley, M (1997) 'Qualitative data archiving: Some reflections on its prospects and problems'. *Sociology*. 31, pp 131-142

Harvey, D (1989) *The Condition of Postmodernity*. Oxford. Blackwell.

Homan, R (1991) *The Ethics of Social Research*. Harlow. Longman.

Horner, J (1998) 'Research, ethics and privacy: The limits of knowledge' *Public Health*. 112, pp 217-220

Lefebvre, H (1991) *The Production of Space*. [Trans. D Nicholson-Smith]. Oxford. Blackwell.

Mason, J (2006) 'Suggestions for anonymising our research records'. NCRM Real Life Methods Node / University of Manchester. Unpublished discussion paper.

Massey, D (2005) *For Space*. London. Sage.

Mauthner, N, Parry, O and Backett-Milburn, K (1998) 'The data are out there, or are they? Implications for archiving and revisiting data' *Sociology*. 32, 4, pp 733-745.

Morse, J (1998) 'Designing funded qualitative research'. In N Denzin and Y Lincoln (eds) *Strategies of Qualitative Inquiry*. Thousand Oaks, Ca. Sage, pp 56 – 85.

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.

Parry, O and Mauthner, N (2005) 'Back to basics: who re-uses qualitative data and why?' *Sociology*. 39, 2, pp 337-342.

Pawson, R (1995) 'Theorizing the interview' *The British Journal of Sociology*. 47, 2, pp 295 – 314.

Pawson, R (2005) *Evidence Based Policy: A Realist Perspective*. London. Sage.

Singleton Jr, R and Straits, B (1999) *Approaches to Social Research* 3rd edition. New York. Oxford University Press.

Soja, E (1989) *Postmodern Geographies: The Reassertion of Space in Critical Social Theory*. London. Verso.

Vidich, A and Bensman, J (1958) *Small Town in Mass Society*. Princeton, NJ. Princeton University Press.

Whyte, W (1958) Editorial. 'Freedom and responsibility in research: The 'Springdale' case' *Human Organization*. 17 (summer), pp 1 – 2.

Appendix A

The Connected Lives Strand of the NCRM Real Life Methods Node

The Connected Lives strand of the Real Life Methods Node of the NCRM is a research project based at the University of Leeds exploring the importance of networks, neighbourhoods and communities and the social support they can offer. It aims to understand more about the ways in which community networks are used and how they are built and maintained, particularly with increased mobility and the introduction of communication technologies.

The research aims to understand communities through an exploration of the interactions of social networks in the Hyde Park / Burley Road area of Leeds. It is exploring how different social groups (according, for example social class, gender, ethnicity) construct networks of friends, relatives, neighbours and service providers, and how these networks are maintained over time and across space.

A geographic area with a heterogeneous population was purposely selected in which to understand the social networks people use to maintain their wellbeing. The research aims to interrogate how and why the social networks discovered are perceived to be important to these groups, and are keen to understand the interaction between travel, communication technologies, and transport service provision and the creation and maintenance of these networks.

The research is methodologically innovative. It is qualitatively driven and includes the extensive use of visual methods. It will also blend quantitative approaches to explore the dissonance between the perceived communities of different social groups and the definitions of community used in policy making and academic discourses. The research will add to our understanding of the networks that constitute communities. It will contribute to ways of representing and understanding the connected, fluid, mobile and relational nature of these networks and will add to our understanding of the importance or networks in promoting people's wellbeing. It will also report on the opportunities and limitation of the multi-disciplinary methods used in understanding community networks.

Funding

This is one of four projects funded under the joint University of Leeds / University of Manchester *Real Life Methods Node* of the ESRC National Centre for Research Methods. The Node advances qualitatively-driven, creatively blended, mixed-method approaches to understanding real lives.

ESRC National Centre for

Research Methods