Methodological issues in qualitative data sharing and archiving

Briefing Paper 2

Qualitative data archiving and reuse: mapping the ethical terrain

Matthew Williams, Bella Dicks, Amanda Coffey and Bruce Mason

Abstract
This paper considers some of the key ethical debates and issues associated with preparing qualitative data for archiving and reuse. The paper particularly addresses the archiving and reuse of qualitative multi-media digital data, although many of the issues will be generic to qualitative data preparation more generally. The paper arises out of work being undertaken as part of a project concerned with exploring the methodological issues associated with qualitative data sharing and archiving.

Introduction
The sharing and re-analysis of qualitative data still remains underdeveloped within the wider UK social science community. In recent years there have been a number of initiatives, including the Economic and Social Research Council’s Qualitative Demonstrator Scheme (http://quads.esds.ac.uk/) and National Centre for E Social Science (http://www.ncess.ac.uk/) that have had as part of their remit, the exploitation of new technologies for storing and preparation complex qualitative data for re-use. Such initiatives have been grounded in; an understanding that the qualitative research community has not been consistent in making data available for subsequent analysis and; in a belief that there is considerable potential for the secondary analysis of qualitative data to widen the reach and impact of qualitative studies (beyond a specific or local case). However it has been recognised that qualitative data archiving (particularly with a view to sharing and reuse) presents a number of methodological, epistemological, practical and ethical dilemmas. This paper particularly addresses some of the key ethical considerations for qualitative data archiving, and focuses specifically on data collected in a range of media. Qualitative research practice is increasingly embracing digital and multimedia technologies for data capture and storage. Textual data now often sit alongside other qualitative data forms (such as visual images and sound files) within single research projects. There are, of course, ethical issues associated with the conduct of multimedia qualitative research, and these have been explored by commentators elsewhere (for example see Coffey et al 2006, Dicks et al 2005, Pink 2006, Rose 2007). However it also needs to be recognised that the emergence of new technologies, and the possibilities these offer for the wider access and sharing of data also generates significant ethical questions for qualitative researchers. The contemporary social science research landscape is rich with ethical debates, codes of conduct, ethical guidelines and regulations. The substance of these has emerged from a combination of (new and updated) legal requirements and discipline specific values about research conduct. Ensuring research designs are grounded in sound ethical principles, and gaining ethical approval from an appropriate ethics committee before research commences, has become increasingly complex (and of course necessary) as our codes of conduct are refined and redefined, and our awareness of the issues becomes increasingly astute. Guidance for ethical research design and conduct, and mechanisms for ethical approval are now well established parts of the research process. However there remains little debate or
development of ethical best practice in relation to the storage and reuse of qualitative data, especially data in multimedia form. This paper begins to map the ethical terrain in relation to these issues, particularly focussing on:

- Continuation of participant protection beyond original purpose
- Multi-media data integrity
- Restriction of access to archived data

**Continuation of participant protection beyond original purpose**

The protection of research participants during the research process is paramount in most of the ethical guidelines that underpin social scientific inquiry. Participant protection can refer to a range of issues including the maintenance of confidentiality (via practices such as anonymisation and stringent data access management); the physical, social and psychological well-being of participants; and the maintenance of the rights of participants. Maintaining these protections beyond the original fieldwork, analysis and publication is potentially problematic and raises a number of issues for future data sharing and reuse. Salient issues that need to be considered in relation to the continuation of participant protection beyond original purpose are clustered around issues of data ownership and informed consent and include various legal, as well as ethical considerations. Many of these issues are compounded when data is available in different media forms.

Data ownership and continuous informed consent

The archiving and reuse of qualitative data presents difficult consent problems for both potential re-users and for the originators of the research. While informed consent may have been achieved, perfectly properly, in the original research, participants are unable to acquiesce to the reuse (re-analysis) of ‘their data’ if the research questions and agendas of potential re-users are not made clear in the first instance. In debating the issue of informed consent and data reuse both ethics and law play key roles and impact on the practices of social science researchers. The Economic and Social Data Service (ESDS) outline some of these points (http://www.esds.ac.uk/aandp/create/ethical.asp), including the need to establish copyright and clarify ownership of datasets, complying with the data protection legislation, understanding the duty of confidentiality and managing informed consent before and after participation. These can usefully be expanded, particularly in relation to multi-media qualitative research and data.

**Legal protections and constraints**

Under the Copyright, Designs and Patents Act (1988) two separate copyrights are created when a recorded interview takes place:

- The words spoken
- The recording

The words spoken by a respondent in an interview remain the intellectual property of that respondent. If the interview is recorded by a researcher the recording remains the property of the institution in which the researcher is employed. Words transcribed from a recording, however, remain the copyright of the interviewee. Thus in order for the content of recordings to be textually reproduced in any analysis or subsequent publication the interviewer must ensure that the respondent assigns copyright over to the institution responsible for the recording. Normally this is assumed to be taken
care of through the usual processes of gaining informed consent, which should include provisions for the transference of copyright from the respondent to the interviewer’s institution. Legally speaking, without the transfer of copyright the researcher has no legal right to reproduce the words spoken by the interviewee. This in itself is worthy of note. On our experience the transference of copyright is rarely an explicit part of the process of negotiation around informed consent, taken at best as an implicit assumption.

Moral rights also have to be considered as part of copyright. These are best considered as the authors’ rights to be acknowledged as the creators of the work. Moral rights also protect authors’ works from being subjected to derogatory treatment. Such rights belong solely to the authors of the work. In the case of an audio-recorded interview the interviewee and researcher are co-authors and so both have moral rights over the work. Whereas ownership of copyright can be legally assigned moral rights cannot. However, moral rights can be relinquished meaning it is still possible for a researcher to seek a waiver from participants before or following an interview.

When considering the potential storage and reuse of qualitative data researchers need to be cognisant of these issues, and may wish to include special provisions when requesting the reassignment of copyright and the waiver of moral rights. When the transference and relinquishment of rights are discussed respondents need to be made aware that the words they speak, and any opinions expressed, become the property of the researcher’s institution and that the data may be used at the discretion of the institution, working within the usual constraints of defamation and other relevant laws.

The legal position with regard to still and moving images (such as photographs and video recordings) is quite distinct from copyright issues with sound recordings. With regards to photographs, copyright remains with the person taking the image (although not if the photograph was taken for 'private' purposes). A research participant who agrees to have their photograph taken has no legal rights over the subsequent use of their image. Copyright remains with the researcher’s institution of employment. Video recordings of consenting research participants would be treated under the law in the same way as films where copyright remains with the 'producer' or 'director'. Again research participants have no legal rights over their video image or any subsequent use (unless the use contravene defamation laws). However, in a video recording of an interview a case could be made in favour of the respondent retaining copyright over their words spoken. In light of this it is worth while requesting that the respondent assigns copyright over to the researcher’s institution. Unlike with audio recordings however, the producer/researcher retains all moral rights over the video recording. On the whole, contrary to audio data, the laws governing photographic and video formats favour the researcher and their employing institution over respondents’ rights. Of course, this only refers to the formal legal position. Qualitative research necessarily involves the formation of relationships with participants, and good (ethical) research practice would suggest that these legal positions should not be relied upon to the detriment of present and future research relationships. Where possible, the creation / production of still and moving images, and their potential future use needs to be part of ongoing negotiations with research participants.
The crafting of the Copyright, Designs and Patents Act privileged mainstream media and arts productions and their creators. This focus has resulted in the clear disparity in legal rights certain media afford research participants. As a consequence the details above are an extrapolation and should be taken as a general guide for researchers to avoid any legal ramifications. More recently privacy laws have also emerged that may afford the researched with more rights across a range of media. Article 8 of the European Convention on Human Rights, for example, (http://www.opsi.gov.uk/acts/acts1998/80042--d.htm) has clear implications for the rights of individuals within the UK. The article states that ‘Everyone has the right to respect for his private and family life, his home and his correspondence’ [sic]. There is also evidence that UK law will take privacy more seriously in the coming years, perhaps dramatically changing the rights of research participants and research practice. For example, the introduction of the Data Protection Act (1998) and the Freedom of Information Act (2000) both require new data management procedures of researchers and their institutions.

Legal issues and storing data for reuse
Data that are stored for the intention of reuse (such as within the ESDS Archives) remain the ownership of those who hold the copyright. In the case of audio interviews and transcripts the researcher’s employing institution retains copyright over the archived data (dependent upon the respondents assigning copyright and waiving moral rights). Photographic and video data also remain the ownership of the producer/researcher (taken to mean the HEI or ‘university’ in most cases of academic research). In legal terms, if all protections have been taken by the researcher, participants have no rights over their interviews or recordings once archived for later use. The Data Protection Act and The Freedom of Information Act do allow participants the right to access data while stored and for that data to be kept securely in the interests of anonymity and confidentiality. A duty of confidentiality also exists in law but is not easily definable due to its absence in statute. Often a duty of confidentiality arises from a verbal or written contract with the respondent, usually in the shape of an informed consent document. Where non-public or sensitive information is passed on in confidence (such as medical details) a duty of confidentiality can arise automatically without explicit written or verbal assurance. For data to be deposited in the ESDS data archive for example, researchers are asked to make it clear in an informed consent document that respondents details will be shared amongst registered users of the ESDS who are bound by an end user agreement to maintain such confidentiality (for further guidelines see http://www.esds.ac.uk/aandp/create/confidentiality.asp)

Ethical protections
Unlike the legal frameworks of copyright and ownership which tend to favour the ‘producers’ of work, there are ethical research conventions, especially so but not confined to qualitative research, that seek to empower research participants. Pink (2007) espouses the ‘collaborative method’ in generating research data, arguing that such a method assumes researcher and respondent are consciously working together to produce interviews, photographs and video footage. The moral ownership of the data is shared in this method and respondents are empowered to change interview transcripts and edit still photographs and video footage – thus producing a co-authored / co-produced final product. While this method is at odds with the absolute legal position in the UK, where joint ownership of photographs and video footage is
not easily established (or necessarily desirable for the researcher’s employing institution), efforts can always be made to ensure respondents are fully consulted as to the use and representation of the data. One might consider this as simply good (and ethical) research practice.

Respondents who consent to having their data published (whether interview extracts, photos or video footage) for an originating research project relinquish their legal rights to control the data in any future use by other researchers. While the legal position is clear, appropriate ethical obligations are still to be established in the area of informed consent beyond the original purpose of data collection. The temporality of consent, as something always in process, always in negotiation and always in a state of renewal is something that is recognised, and thus supported in the recent development of the UK Economic and Social Research Council’s research ethics framework, specifically in relation to ‘participatory research methods’:

In the case of participatory social sciences research, consent to participate is seen as an ongoing and open-ended process. Consent here is not simply resolved through the formal signing of a consent document at the start of research. Instead it is continually open to revision and questioning. Highly formalised or bureaucratic ways of securing consent should be avoided in favour of fostering relationships in which ongoing ethical regard for participants is to be sustained, even after the study itself as been completed. (ESRC 2005:24).

Even where it is possible for consent to be constantly renewed and negotiated at the various levels of participation, there will usually come a time, on ‘leaving the field’, when issues of representation in terms of future use of data cannot be continuously renewed (Renold et al. 2006). This is most apparent when the data have been deposited in an archive and reused by researchers for projects that may have significantly different research aims and objectives from the original research. It is quite possible that the theoretical position, topic and questions of a data re-user may be at odds with those of the data originator(s). In such cases the original research respondents’ informed consent, based on an understanding of the aims and purposes of the research, could be seen as null and void. While legally speaking respondents have very little leverage in preventing reuse, ethically it would be ethical and responsible for the owners of the data (the original researcher and their institution) to introduce safeguards to prevent respondents from any harm or anxiety, in relation to future data re-use. Best practice would be to consult respondents before any agreement is made to allow for the reuse of data in a secondary project by another researcher or institution, although this may not always be possible. Best practice would also suggest that informed consent should be sought from respondents for any secondary analysis or repurposing by other researchers (Corti and Backhouse 2005). The extent to which this is possible, over time and in relation to disparate or potentially hard to reach groups is of course highly debatable. Where it is expected that continuous contact with research respondents might be problematic (due to relocation, time lapsed, requests of privacy and so forth) best practice might be for researchers to make it clear in the initial informed consent document that the data generated may be used for other as yet indefinable research purposes. Data storage, archiving and future access could legitimately and ethically become part of initial and ongoing negotiations with research participants. Combined with assigned copyright
and moral rights waived, this kind of nuanced negotiation would enable researchers to assure themselves that they have appropriately attended to the rights of participants. However, it remains the case that the judgement of the researcher remains central in making decisions about the reuse of archived materials, as it is the original researchers who are (theoretically at least) best placed to consider the values, opinions and well being of their original research participants. This raises interesting questions about the repositories for the archiving of qualitative data, as well as access and permissions to re-use (something to which we return later in this paper).

Ethical Dilemmas in Creating Multi-media Recordings

As this paper has already indicated, photographic and video data have a distinct set of ethical concerns that set them apart from the more common research audio interview and subsequent written transcript. Pink (2006) acknowledges the power of the visual across cultures, and acknowledges many of the potential anxieties and pains associated with the medium. In research situations, where participants are often in a subordinate position to the researcher (which is certainly the case in law relating to multi-media), it is important that the sensitivities towards the visual are factored into the design of the study. In particular participants should be fully consulted before any photographic or filming work begins and any objections need to be taken with great seriousness.

The subsequent use of photographic and video materials also needs to be established with informants prior to data collection, while recognising that this may be difficult to anticipate. Such measures are especially important where the participants might be considered to be vulnerable (for example in the case of children). Particular attention must be paid to archiving and representation with multi-media data as less can be done to anonymise respondents and maintain confidentiality, while retaining the integrity of the data. While there have relatively few legal cases within the social sciences with regards to the collection and representation of visual data, journalists and broadcasters have been dealing with such concerns for decades. The Ofcom Broadcasting Code 2005 (http://www.ofcom.org.uk/tv/ifi/codes/bcode) clearly stipulates that due care must be taken over the physical and emotional welfare and the dignity of people under eighteen who take part or are otherwise involved in any programme. It further states that people under eighteen must not be caused unnecessary distress or anxiety by their involvement in programmes or by the broadcast of those programmes. Several high profile legal cases have emerged over the inappropriate use of library footage in broadcast television programmes heightening the media’s sensitivity to these issues. Similarly researchers working with photographic and video materials need to be acutely aware not only of the sensitivities of respondents to the visual, but also the potential legal ramifications that may result from ill-judged multi-media representations. Much of these potential pitfalls can be negated by maintaining close and trusting relationships with participants and adopting where possible Pink’s (2001) notion of the collaborative method in generating visual data. Issues remain however, as to the ways in which visual data are re-appropriated and used in the course of secondary analysis by other (non-originator) researchers.

Multi-media data integrity

Good qualitative research practice necessarily entails establishing a reciprocal relationship between the researcher and those researched, built upon trust and rapport,
preferably over time. Part of this trust building exercise would usually include transparency of the research process and collaborative approaches to data generation. The principle of interview respondents, for example, verifying interview transcripts and empowered to change the contents where they think it is appropriate to do so, is a practice that many qualitative researchers would adopt, providing it does not interfere with the researcher’s epistemological position. Such processes can help to ensure validity and reliability in the data collected and foster trust in those participating in the research. In such cases data can be seen as collaboratively generated. Such ethical practices, are however, more complicated when applied to data in multi-media forms. While copyright law fails to acknowledge the rights of participants within multi-media recordings, the social science community might judge such media as ethically challenging. Establishing such a reciprocal and empowering relationship is potentially more difficult in still photographs and visual recordings that are less easily manipulated (notwithstanding the impact that increasingly digital technologies might have on our capacities to manipulate the visual). For example, concealing elements of photographs and/or significantly editing video recordings might be seen as somewhat different from ‘tweaking’ interview transcripts to reflect the true meanings of participants. Similarly altering audio recordings feels more invasive that altering words on a page (see Pätzold 2005). Of course, one might want to argue that this unnecessarily privileges the visual and aural over the printed word, and that all might best be seen as created and produced (at best as the result of active collaborations between researcher and participant). Nevertheless preserving the integrity of multimedia qualitative data, however that is best interpreted, remains a significant issue for archiving and future data sharing.

Attempting to anonymise multi-media data recordings creates several methodological quagmires for data archiving, sharing and reuse. Blurring faces and editing video to conceal the identities of research participants reduces the potential for the data to be reused. For example, a photograph that has had faces blurred reduces the capacity for researchers to interpret what is going on in the setting. Facial expression, eye direction and physical characteristics may be key research foci for a range of researchers across disciplines. The extent to which anonymisation is possible, or indeed desirable with regard to visual data is a particularly methodological issue. Images of children (or others considered to be vulnerable), raise additional ethical concerns. One the one hand, there is an ethical unease, not to mention a considerable job of work, with ‘blurring’ faces or choosing not to include images in which faces are identifiable – and thus to an extent ‘taking away’ ‘identities’. On the other hand using visual images, particularly of children and young people, in the communication and representation of any research project is problematic. This is particularly so if data are to be digitally available, albeit access protected, in some form or another. Moreover informed consent to use visual images in the present does not preclude difficulties later – dated photographs and video footage can come back to haunt all of us in the way that transcripts and written texts perhaps do not. However it would be equally ethically problematic to exclude (for example) children’s voices, bodies, faces, identities, experiences and participation from innovative multi-media qualitative research (and thereby making such approaches exclusive rather than inclusive).

---

1 A discourse or conversation analyst may protest at such tinkering of transcripts given their theoretical positions on the spoken word.
In terms of future analyses, to retain maximum reuse potential when archived, photographs and video should be left intact in their raw formats. As noted, this does raise ethical issues that require careful consideration. These can be mitigated during the initial fieldwork phase. Taking on board Pink’s (2006) collaborative method (for example) leads to participants playing an active role in the production of photographic and video data, including input into the composition of photographs and direction of filming. The review of visual and aural data by participants is still possible, can protect their rights and afford them the power to advise on changes or request the data be erased and the interview, photo shoot or filming be redone. Thus, understanding data collection as active, created and collaborative can help to safeguard the rights of participants in relation to data archiving and reuse/repurposing.

Of course, there are instances where it may not be possible or desirable to ‘redo’ data recordings (for example, in unexpected situations in the field) and there may also be times when participants ask for editorial work to be undertaken, that may be difficult or undesirable to carry out. The rights of participants need to be sensitively balanced with the potential use and reuse of the data. The key is to ensure an open dialogue with research participants. It making these decisions it is beneficial to refer to the ESDS End User Licence Agreement which affords participants with further protections and has contractual force in law. In particular the licence, which must be agreed with the reuser before access, requires the preservation of confidentiality and anonymity if the data is not available in the public domain. Depositors of multi-media data can remain confident that while the raw archived recordings have not been subject to anonymisation procedures, any new data representation must be fully anonymised.

**Restriction of access to archived data**

Little attention has currently been paid to the access procedures for archived qualitative data. Given technological advances, particularly in relation to digitisation and web-based formats, it is possible but not necessarily desirable for open access to archived datasets. Given issues of consent and copyright, and moral obligations to research participants (and the originators of the research) it is important that appropriate systems are put in place. These systems need to be sensitive to sharing and reuse of data, as well as to the ethical landscape in which the data are located. Currently Qualidata (ESDS) has used Athens authentication to access archived datasets deposited with them. Athens is a service which was funded by the Joint Informations Systems Committee (JISC) to regulate access to resources over the UK’s education and research network (JANET). Any student or researcher wishing to access e-journals or other online resources required an Athens account from their university. This system has proved secure enough to comply with the Data Protection Act as it only allows members of academic institutions to access the protected data (in some cases data can be made available for commercial purposes on the agreement of the data registrar and/or data service provider). However as more data are archived, and if the perceived desire to reuse qualitative data is realised, then researchers and their institutions may want to put in place more locally based procedures, to ensure that archived data sets are securely managed and only accessible to appropriate communities. This of course raises a further set of issues, in relation to how far access should be restricted to other social scientists, others working within a similar paradigm, others willing to share their data, non-academic researchers and so forth. A balance is required between ensuring legitimate access while protecting participants.
and original researchers (where appropriate to do so). In all cases it seems important that research participants are made aware of the current level of security in place for the protection of data and that no other third parties, such as the media or private research establishments, can gain legitimate access without explicit permission. For additional security, researchers (and their host institution) that own the copyright to the deposited data might request that they be involved in the vetting procedure before a potential reuser is provided with access. The End User Licence Agreement of the ESDS also specifies that no reuser has permission to give access to the data collection to another who is not also a registered ESDS user, and that further permissions are required from data depositors for information to be passed onto other users or used for teaching purposes (particularly in relation to sensitive materials. Such measures would need to be put in place and extended if qualitative data archiving and sharing became more widespread and distributed.

Concluding remarks
This paper should not be seen as an exhaustive account of all of the possible ethical issues involved in preparing complex qualitative datasets for archiving, sharing and reuse. Nor does it claim to offer definitive to solutions to the ethical dilemmas and consideration it has identified. Rather the paper should be seen as a contribution to the wider social science project of critical engagement with the methodological, epistemological and in this case practical implications of working with qualitative data in multimedia form. There are intellectual, moral and technological drivers for ensuring that qualitative datasets are archived and made available for appropriate reuse and secondary analysis. There are real ethical issues to be acknowledged and sensitively worked with, if not resolved. This paper addresses some of these issues, and in doing so hopefully gives the qualitative research community and the institutions in which data will be archived and accessed, some useful ways forward.

References


ESRC (2005), Research Ethics Framework (REF), Swindon: ESRC

