

From data archive to ethical labyrinth

Annamaria Carusi¹, Marina Jirotko²

¹Oxford Computing Laboratory

²Oxford e-Research Centre

Email address of corresponding author: annamaria.carusi@comlab.ox.ac.uk

Abstract. Researchers in the social science are increasingly encouraged or obliged to deposit data in digital archives for greater transparency of research or for secondary use by other researchers. However digital archives raise many ethical challenges at the institutional, disciplinary and personal level, and researchers can find themselves caught between conflicting requirements. This paper considers the ethical challenges of qualitative data in particular showing what specific ethical challenges qualitative researchers face. There is generally a lack of policy or guidelines as to how to deal with digital data, or else there are conflicting requirements set by funding and academic institutions and by the law. In the face of this, researchers themselves need to be aware of the ethical and legal dimensions of their data, so that they are in the best position to enter into negotiations concerning whether and how it is archived. The options for archiving are outlined, and an interdisciplinary approach is recommended.

Introduction

With the rise of digital archives in the research cycle of social science, even research that may not initially construe itself as e-social science acquires that status. It is the digital archive that will normalise e-social science, as it makes digital material more readily accessible, and at the same time necessitates the tools and technologies associated with e-social science research for analysis and interpretation of large quantities of digital data. Several funding bodies have made research funding conditional on the deposit of data in a digital repository, and increasingly qualitative as well as quantitative data is being included in this policy (Parry & Mauthner (2004); Corti (2000)).

The motivations for requiring deposit of material within an archive are scientific and pragmatic. The justifications put forward are that the publication of data will contribute to a more rigorous scientific approach by facilitating the transparency of research, allowing for the testing and checking of results and interpretations by other researchers. Equally important is the pragmatic reason to avoid the repetition of expensive research by allowing research that has already been carried out to be exploited as fully as possible. This is achieved by secondary use of data

Across the sciences digital publication is increasingly pushing against the boundaries of paper-based models of publication (Borgman 2007). One of the challenges to this model is in the separation that has traditionally existed between publication and data, and there is increasingly a drive to publish data alongside research papers or archiving it to make it more available (Heely et al 2004). The digital archive is fast becoming a new publication medium that may eventually change the parameters of scholarly research articles, and indeed of scholarship in general. Despite predictions concerning such changes, there is currently still much resistance by researchers to depositing their data with digital archives. It is claimed that one reason for this is that it is not currently part of the publication culture of social science, and particularly not of *qualitative* social science (Corti 2000). The publication of data makes a profound difference to the nature of data collection, organisation, interpretation and use. Gathering and storing data for ones' own use or with an eye to other possible addressees besides oneself are two very different activities¹. Digital media by their nature – for example, their greater publicity from inception – are already exerting pressure on the publication culture of all forms of representation, even the lowly family snapshot, as is witnessed by the rise of sites such as Facebook and Flickr'r.

Ethical aspects

It is not, however, a matter of only attitudes to publication and data or representations changing. In the case of social science research which is conducted with research subjects in all kinds of mostly off-line situations, there are ethical implications to be considered for the research subjects and sometimes for the researchers too (Corti (2000); Parry (2004); Corti, Day & Backhouse (2000); Dickson-Smith et al (2007)). Even though e-social science and Internet research in general has a broad spectrum of ethical dimensions, in this paper, we are considering only those ethical considerations that come to the fore with the archiving of data in a digital format on an Internet-enabled platform of some sort. A characteristic of these data is that they are not necessarily gathered or collected on the Internet, and that the social behaviour or phenomena studied do not necessarily occur on the Internet. Typically, data will be gathered in interviews, focus groups, and/or video and audio supported observations. Increasingly the data will have digital form (textual transcriptions of interviews in digital documents, video, audio); the real shift comes when the digital data is Internet-enabled in a digital archive. Ethical issues arise with both quantitative and qualitative data – for example, issues of consent, anonymity and privacy occur in all forms of data, and certain risks attendant on data on the Internet affect quantitative as well as qualitative data – for example, the risks attendant on aggregation (see for example Tavani). However, there are also specific risks attendant on qualitative research. Some of these specific risks are:

- *the context and method of collection makes the archiving and re-use of qualitative data methodologically more difficult:* whereas quantitative data is arrived at through abstraction from a context, qualitative data is highly contextualised. Interpretation of qualitative data requires an understanding of the

¹ In this respect, it may be useful to consider the theory of dialogism that sees the nature and meaning of any utterance or discourse dependent on its addressees rather than only on its sender (even when the addressee is the sender, as in diary notes). For example Bakhtin 1930/1981.

context in which the data were collected (See Van Den Berg (2006) Gillies & Edwards (2006) and Kelder (2006));

- *the relationship between research subjects and researcher which is necessitated for obtaining qualitative research:* there are of course as many different kinds of relationships between research subjects and researchers as there are kinds of research, and experimental research rests on a fundamentally different relationship to ethnographic research. However, common to all these forms of research is an element of trust, which is often built up in a close inter-personal relationship (Dixon-Swift et al, 2007);
- *the difficulty of anonymisation which is both effective and does not render the data methodologically useless:* anonymisation needs to be both of primary and secondary subjects (eg, not only interviewees but also the people that they mention). However a thorough anonymisation of all affected subjects can render even textual data difficult to understand. In the case of visual data, anonymising facial features can remove aspects which are crucial to research (see below for an example);
- *the nature of possible harms to subjects beyond anonymisation – especially in view of the nature of digital data;* in this respect it is interesting to note how the Economic and Social Research Council, a major funder of social science research in the UK, defines risk:

Risk is often defined by reference to the potential physical or psychological harm, discomfort or stress to human participants that a research project might generate. This is especially pertinent in the context of health-related research. But, in addition, social science raises a wider range of risks that needs to be considered by RECs [Research Ethics committees] These include risk to a subject's personal social standing, privacy, personal values and beliefs, their links to family and the wider community, and their position within occupational settings, as well as the adverse effects of revealing information that relates to illegal, sexual or deviant behaviour. (ESRC Research Framework, p21)

Even if data are anonymised, the re-use of data for purposes which go against research subjects' personal values and beliefs, or which affect a group of which they are a member – that is, family or community – are considered to be risks. Digital data compounds this risk because it is inherently susceptible to being copied, manipulated, and de- or re-contextualised (see Carusi forthcoming). This affects quantitative as well as qualitative data, but in view of the greater dependence of qualitative data on context, can be greater in the case of qualitative data – and may certainly be *experienced* by research subjects as being a greater risk. These risks go hand in hand with methodological issues.

- *the vulnerabilities of the researcher:* as mentioned, qualitative research often requires a close inter-personal relationship between research subjects and researchers, in the course of which researchers make themselves vulnerable in many different ways (Dixon-Swift et al 2007). Thus 'data' concerning researchers are embedded in qualitative data too, without the possibility of even superficial anonymisation.

Ethical indeterminacy

Indicating that there may be ethical implications of archiving data and addressing those implications are two different things. Guidelines, policy and common practice lag behind the reality of creating archives and depositing data. As in many other Internet-enabled technologies, the creation of data archives is creating a vacuum not only with respect to articulated policy, but also with respect to researchers' own moral intuitions and judgements. Existing practice can provide some guidance, but this is limited precisely because the technologies are pushing us beyond existing practice and is often challenging its moral grounds.

This is most obvious in the case of informed consent, which is a fundamental ethical requirement of the majority of ethical procedures (see below for a discussion). However, it is unclear what constitutes informed consent in the case of archived data. The ESRC and Qualidata both recommend that research subjects should be informed that their data will be accessed by other researchers once it has been suitably anonymised, yet they do not indicate whether research subjects should be informed that it will be archived in an Internet-enabled repository. Yet there are many research subjects who may have had or heard of unfortunate experiences with even photographs attached to emails and for whom this information would not be insignificant.

Several of the e-social science projects which form part of the UK NCeSS programme continued the ethics policy that they had used beforehand, and which was usually either that of their institution or of a related discipline: for example, psychology has a longstanding ethics framework in place and many other forms of social research adopt the consent forms used by psychology, in the absence of alternatives. This was the strategy used by the 'HeadTalk' project, based in the Applied Linguistics Department at the University of Nottingham. This project uses qualitative textual, video and audio data in that it is attempting to push beyond the boundaries of corpus linguistics by providing the means to study embodied language, that is, language use with the gestures that accompany it. The data for this research are not only textual transcripts of language, but also audio and verbal digital recordings. Like other corpus studies, the project aims ultimately to publish a resource which is available to other corpus linguists. Unlike other corpus studies, this resource will include visual and audio data. Apart from ethical procedures adopted from psychology, legal conditions for anonymisation and copyright which are required by publishers are also standardly followed. However these legal conditions relate to textual language, not images. The project followed all ethical and legal requirements, and at all times respected the wishes of research subjects.

The research subjects for this research are generally colleagues or students who are well-versed in corpus linguistics and who therefore could be considered to be well-informed of the general nature and purpose of the research. In addition, the data gathered are not highly sensitive or personal as they are for the purposes of linguistic and not behavioural or social study. However, the data subjects were not used to the visual nature of the research, or to the mode of visual analysis of which it consists: for example, the identification of gestures such as nods or hand movements which can, on the visual plane alone, have a strange effect

of making the subject represented appear like a mannequin - a good example of a reification through research (see Strathern (1999 and 2002). In addition, the images are used on posters and presentations relating to the project. This can prove to be embarrassing for research subjects, and in fact in an early study resulted in a subject withdrawing from the study. This is in line with research subjects' rights as set out in the UK Data Protection Act (1998).

The salient unexpected results of this specific case were the publication of images where text had set the standard for previous research; and a particular manipulation and use of images which could be experienced as disconcerting for research subjects. There is no doubt that these are fairly trivial harms, in particular in view of the non-sensitive nature of the data. However, these do fall under the risks listed by Research Ethics Framework of the ESRC – which requires the offering of data for archiving purposes. A colleague who finds his/her image on posters and presentations in contexts where s/he is likely to be recognised – and where there is the added possibility of an unusual manipulation of the image – could experience an unsettling exposure in her occupational setting. Anonymisation is impossible because it would result in the direction of the gaze being lost, which in turn would prevent the tool for head nod recognition not being able to work. In addition, it must be borne in mind that participation in research is voluntary and it is up to research subjects to define for themselves what they may find distressing. This is a major difference between data in e-social science research and data in Internet Research, where the research subjects are already on the Internet, and so it can be assumed that they do have some understanding of Internet discourse and representation, and have consented to by participation.

It is not difficult to imagine a possible extension, with other harmful or distressing consequences, in an archived data set. The form of publication in the HeadTalk project does not allow for the downloading and copying of images (except by taking a snapshot of the screen), but it is not automatically the case that other archives will prevent or prohibit this. In fact, our own investigation of Qualidata showed that much data was easily copied.

HeadTalk and similar projects raise questions concerning the form that informed consent should take for this type of data to be published in an archive or other form. Should research subjects be shown examples of archived data, and should their attention be drawn to the possible uses to which archived data can be put? In order to meet the conditions that informed consent should be an ongoing process, and that research subjects can withdraw from research at any point, should research subjects automatically be given a registration to any archive in which data concerning them is stored so that they can see for themselves how it is archived should they want to, or perhaps see to what uses it is being put should they want to know?

Related to issues of informed consent it is also unclear what constitutes a protection of privacy and confidentiality. Researchers may find themselves in a position of having to reconcile several different sets of ethical requirements. Some of these are listed below:

I. The funding body may require that data be made available for archiving. In the UK, the ESRC requires that data be offered for archiving in Qualidata. Whether it is in fact archived and in what form with what access is subject to negotiation – a point to which we return.

II. The research institution – university or other – has its own set of ethical requirements, which are often motivated by a concern to avoid litigation and so are especially cautious. The University of Oxford research ethics guidelines state that a reason for obtaining ethical clearance from the institution is that funding bodies often require it. However, the requirements of this same institution also contain the following conditions:

- a) State in the information sheet that the material will be seen only by members of the research team, and will not be shown to anyone else without the explicit written consent of the participant. The participant should be offered the opportunity to review the material, if desired, before giving consent.
- b) The relevant recordings should be kept in a locked filing cabinet
- c) Researchers should be sensitive to the possibility of recordings being ‘lost’ after being archived, and only discovered years later after the researcher who collected the data has disappeared. The researcher should make a plan for the storage and ultimate disposal of the material. Any material that is archived must be labelled as confidential, with the name and contact details of the researcher attached. (<http://www.admin.ox.ac.uk/curec/oxonly/protocols/datastorage.doc>)

III. Legal requirements can differ from both of the above. For example, privacy and confidentiality are regulated by the Data Protection Act (1998) in the UK. The 8 principles of this act are as follows:

1 Personal data shall be processed fairly and lawfully and, in particular, shall not be processed unless—

(a) at least one of the conditions in Schedule 2 is met, and

(b) in the case of sensitive personal data, at least one of the conditions in Schedule 3 is also met².

2 Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.

3 Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed.

4 Personal data shall be accurate and, where necessary, kept up to date.

5 Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.

6 Personal data shall be processed in accordance with the rights of research subjects under this Act.

7 Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.

8 Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of research subjects in relation to the processing of personal data.

(http://www.opsi.gov.uk/acts/acts1998/ukpga_19980029_en_9#sch1)

IV The European Union Directive on Data Protection similarly “make[s] archiving research data an illegal activity without the data collector or person storing data having obtained

² These schedules relate to conditions under which data may be used with informed consent.

prior permission to store and allow access to data.” (Corti, Day and Backhouse (2002)), and even though there are exemptions for research and scientific purposes, these are particularly difficult to apply to qualitative data because of its nature (Parry & Mauthner (2004)). At any rate, the legal situation is such that researchers need to be aware of how to interpret the law, as well as how to interpret the data in the light of possible concerns of their research subjects.

V There are different disciplinary sensitivities over data. e-Research projects often involve inter-disciplinary work across a number of different disciplines including computer science, that are not used to dealing with data about human subjects. While following protocols ‘to the letter’, there can be tensions where researchers do not necessarily appreciate the sensitivity of others. There may be deep differences between psychology and social sciences too, with experimental psychology and cultural anthropology potentially demonstrating the greatest degree of difference. However researchers within both these disciplines expressed grave concerns concerning the archiving process and the consent process. For example, experimental psychologists view it as yet another hurdle to be overcome; and where the research is dependent on naturalistic settings, the dimension of archiving is yet another artificial intrusion. Anthropologists often have serious problems with informed consent, as their subjects often do not share the culture of informed consent for educational or other reasons; in addition, they may deal with political and extremely personal topics. Some of the problems associated with Internet publication of data have also been encountered with the digital collections of museums (Carlson and Anderson forthcoming, Srinivasan 2006) Lastly humanities researchers tend to be more concerned with authorship than with confidentiality (Bassett & O’Riordan (2002), White (2002)).

VI Even within disciplines, there are different sensitivities often relating to methodological practices. For example, detailed ethnographic or ethnomethodological social science requires a close relationship with research subjects. Trust is an integral part of this relationship, and researchers may be particularly aware of the need to respect the privacy and other values of research subjects. For this reason, they could be particularly reluctant to share this data with others in an open access system. This has been evident even in the development of technologies for collaborative distributed analysis and interpretation of video data for ethnographic or ethnomethodological studies. For example, developers of MiMeG were particularly careful to ensure that data is always under the control of the researcher who collected it, even in the context of a collaborative interpretation. In the first version of the software, video remained on the hard disk of the researcher, and was not ever transferred to the hard disk of other researchers or any other server. As this resulted in problems with time lag, the second version now allows for data to be transferred, but only with the explicit consent of the researcher and only with trusted others. This kind of collaborative interpretation is predicated upon a pre-existing trusted collaboration with other researchers. The situation is different with highly constructed experimental data. However, our research with the ESDS Qualidata has also shown that the type of data that is preferred for archiving is precisely that which is not constructed for one narrow research question but which contains much residual material that could be exploited by other

researchers³ (interview with John Southall). This is precisely the kind of data that is produced by rich ethnographic and ethnomethodological data collection.

In view of these different sets of requirements, guidelines and sensitivities, researchers can easily find themselves caught between conflicting or contradictory pressures. The way forward involves negotiating a path between existing practices, cautious prediction regarding the possible consequences of new technologies, and hypothetical reasoning regarding what is morally and legally acceptable. Above all, the existence of these conflicts, tensions and indeterminacies means that researchers have to be proactive, and capable of positioning and protecting data as they deem appropriate. There are not absolute constraints or requirements. Rather there is a broad scope for interpretation of data in the light of duties and responsibilities to research subjects, institutions (research institution and funding body) and their disciplines (other researchers and their peers).

Options for negotiation

Currently in the UK, researchers who are funded by the major funding council are constrained by the requirement that they must at least make data available for archiving. However, beyond this constraint there is enormous scope for negotiation about whether data actually are archived and how they are archived (Corti, Day & Backhouse, 2000). In this paper, we highlight some of the most important points for proactive intervention of researchers in this process.

Two aspects of researchers' attitudes to data are common throughout these points: that is a proprietary attitude and an attitude of moral responsibility towards research subjects, expressed in the way in which data are handled (ranging over the way they are gathered, interpreted, made available to others, and stored – see Goguen 1997). Moral responsibility can be expressed both in the relationship between researchers and subject(s) as it unfolds, and in the way in which researchers handle data about subjects subsequent to the research. The willingness of subjects to participate in research and their comprehension of the research are at the core of Western ethical attitudes towards research, and are based upon notions of autonomy and upon an individualistic outlook of the relevant ethical aspects research. There are however also broader ethical issues, having a social, cultural and political aspect. For example, anthropologists dealing with subjects who have never used a computer let alone understand what an online archive is may feel that they are deciding for others and possibly representing them in negotiating whether and how to archive. This is not a new phenomenon in anthropology, as for centuries anthropological collections were displayed in museums without the permission or even knowledge of the subjects whom they represented. This, however, is a good example of a paternalism, which came to be recognised as having had serious cultural consequences for which museums currently struggle to make amends. Paternalism – and a form of deciding for others – can also occur in the way in which researchers negotiate on behalf of research subjects in line with researchers' own understanding of what constitutes protecting research subjects. For example, because of exclusions for purposes of research in the Data Protection Act, and

³ Interview with employees at Qualidata.

vagueness about what such purposes (or ‘same purposes’) consist in, researchers in fact have a great deal of leeway in interpreting whether archiving material in digital form is in accordance with the provisions of this Act. Their interpretation will not necessarily coincide with that of research subjects, and a form of paternalism can ensue, resulting either in the under or overprotection of research subjects. With an eye on these kinds of historical precedents and ethical consequences, there is good reason to abstain from paternalism in any form regarding the dissemination or publication of data and to take the attitude that without some form of participation from research subjects, either there ought not to be digital, Internet-enabled archiving of social science data in an open access forum, or it must be made clear how the decision has been made. This is an extremely complex topic where we need to steer a careful path between a patronising and paternalistic attitude towards research subjects, and an entirely ‘laissez faire’ attitude (see Nespor 2000, Srinivasan 2006). Throughout the following points, the over-arching principle has been that of avoiding deciding for, or representing research subjects where possible, and putting in the forefront the relationship to research subjects and their control of data in which they are represented.

- 1) Informed consent options: what constitutes genuinely informed consent with respect to digitalised Internet-enabled data? Advice concerning informed consent from legal and institutional bodies includes the point that apart from being voluntary and non-coerced, research subjects must be given relevant information regarding possible risks and harms, informed consent must be ongoing – in particular at any points where the research changes, and that there must be an opt-out clause. That is, it must be possible for the subject to withdraw from the study at any point where s/he no longer feels comfortable with it. Much of this becomes exceedingly difficult when consent is being requested for possible secondary uses, which, by the nature of the case, are unknown to the researchers gathering the data. However, it is difficult to articulate all the risks that may be attendant on digital Internet-enabled data without scaremongering. With subjects who have knowledge and understanding of the Internet, one option to be considered is that they be shown (or have access to) examples of archived data, which is relevantly similar to the data that are being gathered about them (see Luff, Heath & Jirotko 2002) Research subjects could be allowed to see and possibly to edit the data in which they are represented (see for example the CUREC guidelines for video and audio data). It is also possible to use the functionalities of digital archives to enhance the ongoing consent process and research subjects’ choice of opting out of the research. Research subjects who are interested could follow their data through different uses that are made, and opt out at any point that they feel uncomfortable. For all this to be possible, research subjects should be registered participants on data archiving sites, if they are interested in being so. Responsibilities for data do not all lie with researchers; instead the very process of informed consent which is a form of respect for the autonomy of individuals, by the same token puts a certain amount of responsibility on them too.

Apart from being given relevant information regarding the research, what it is for, and how data will be used, informed consent also needs to cover anonymisation and access.

- 2) To anonymise or not to anonymise: Anonymisation is often seen as a kind of general panacea for many of the possible ethical harms associated with social science research. Ethical considerations have become focused on privacy and confidentiality, and this is generally also mirrored in a concern for individuals rather than for groups and the broader social, ethical and political issues associated with social science research. Researchers are repeatedly told that they must anonymise particularly in the case of ‘sensitive’ data, with the range of data that could be construed as sensitive being left vague. This vagueness means that it is up to researchers together with their research subjects to arrive at a judgement regarding the sensitivity of the data, and whether the data ought to be anonymised due to that sensitivity. However, research subjects do not always want to be anonymised, sometimes precisely because the research involves so-called ‘sensitive’ data. Some research about sexual preferences, household arrangements or political beliefs may be seen by participants as having a political dimension, with which they want to be identified (see Nespor 2000).
- 3) Access, tagging, copying: there are several options regarding access to data. Access could be granted simply upon registration of a user; or there could be a requirement that permission must be requested. Permission could be requested from the archive host institution, from the researcher’s institution, from the researcher alone or in conjunction with the data subject if this was part of the original consent given (for example, so that s/he can exercise their choice to opt in or out). Access is also related to other activities that can be undertaken by secondary researchers: for example, tagging for data mining purposes, downloading, copying, etc. Decisions must be made about all of these possible functionalities and to whom they are permitted⁴. The new potentials for access to research data raises issues concerning the distinction between the private and the public, and many concerns about anonymity and privacy have to do with the ambiguity of the boundaries between private and public spaces. The Internet itself is challenging the boundary between these spaces in ways that force us to reconsider in which spaces we demand privacy and is challenging social, philosophical and legal conceptions of privacy. Nissenbaum proposes that Internet challenges to privacy ought not to curtail our demand for privacy, but rather to extend it to some public spaces (see Nissenbaum 1998 and Nissenbaum 2004). Considering the possibilities of data mining and aggregation that archives may be exposed to, this is not an insignificant question for social science research archives.

Conclusion

Digital archives place ethical demands on social science researchers. If nothing else, the technologies for archiving, analysing and interpreting research data brings to the fore

⁴ See for example the ‘Data Enclave Project’ based at the University of Chicago <http://www.norc.org/projects/Data+Enclave+Project.htm>

ethical processes and procedures that have become merely procedural and automatic, and forces a reconsideration of what has been taken for granted. The archives are an opportunity for reflection on the ethical presuppositions of research, and a degree of disciplinary self-examination with regards to ethics. The need to reflect upon the possible impact of archiving on one's own disciplinary interpretive and ethical procedures invites comparison and contrast with other disciplines. For example, anonymity is not considered to be as important in humanities research as it is in social science research, in contrast with authorship which is more important to humanities (White (2002)). A debate does not mean adopting the attitude of other disciplines to anonymity and other issues, but rather a reconsideration of whether and why it is important. Similarly Parry & Mauthner (2004) provide an illuminating comparison of the attitudes of oral history and social science to data ownership and copyright and the implications that this has for data storage. Apart from understanding what existing practices regarding research data are, placing data in its historical context can also be illuminating (when and why did specific practices emerge, and are they still appropriate?), as is fully addressing its philosophical aspects. Parker (2007) argues that it is necessary to go beyond multidisciplinary collaboration in research situations towards ethnography/ethics as a new discipline where combines investigation of moral practices (existing, actualised practices that emerge during ethnography) and ethical challenges, such as questions regarding why (or even whether), negotiation is an ethically better route than other routes, what can be negotiated about and what are its limit or guiding notions – for example, respect, recognition, dignity and so on (Parker 2007). While in this paper we have stressed the ways in which informed consent can be safeguarded in the transition of data from the offline into the online world of the digital archive, it is also clear that in the process of negotiation that the researchers, subjects and archivists will embark upon, the kind of deliberation required will mean that nothing about what constitutes an ethical handling of data can be taken for granted in advance. This is not least because of the challenges of the digital world to long-held ethical principles as embodied in current ethical guidelines and requirements. Digital archiving thus challenges all researchers to question the ethical dimension of their data practices.

Acknowledgements

We would like to thank our interviewees at the DRESS and MiMeG nodes of NCESS, as well as at Qualidata.

Research for this article was supported by the ESRC Grant No RES-149-25-1022.

References

Bakhtin, M. M. [1930s] (1981) *The Dialogic Imagination: Four Essays*. Ed. Michael Holquist. Trans. Caryl Emerson and Michael Holquist. Austin and London: University of Texas Press.

Bassett, E. and O'Riordan, K. (2002) 'Ethics of Internet research: Contesting the human subjects research model,' *Ethics and Information Technology* vol 4, no 3, 2002, pp. 233-247.

Borgman, C.L. (2007) *Scholarship in the Digital Age: Information, Infrastructure and the Internet*. MIT Press.

Carlson, S and Anderson, B. forthcoming 'What are data; The many kinds of data and their implications for data re-use,' in (Ed Jankowski, N.) *eScience: Transformation of Scholarship through Cyberinfrastructures*. Peter Lang.

Carusi, A., (forthcoming) [3] Carusi, A. (forthcoming) 'Ethics in e-science: Data as representation,' *Ethics and Information Technology*.

Corti, L., Witzel, A., Bishop, L., (eds) (2005), 'Special Issue: Secondary Analysis of Qualitative Data', *FQS*, Vol 6, No 1, 2005.

Corti, L. (2000) 'Progress and Problems of Preserving and Providing Access to Qualitative Data for Social Research—The International Picture of an Emerging Culture', [58 paragraphs], *FQS* Vol 1, No. 3 – December 2000.

Corti, L., Day, A. & Backhouse, G. (2000) 'Confidentiality and Informed Consent: Issues for Consideration in the Preservation of and Provision of Access to Qualitative Data Archives', [46 paragraphs], *FQS* Vol 1, No. 3, 2000.

Dickson-Swift, V., James, E.L., Kippen, S. & Liamputtong, P., 'Doing sensitive research: what challenges do qualitative researchers face?' *Qualitative Research* 2007, vol 7, pp. 327-353.

Ess, C. Introduction. (2002) *Ethics and Information Technology* vol 4, no 3, pp 177-188.

Ess, C. & AoIR ethics working committee. (2002) *Ethical decision-making and Internet research: Recommendations from the aoir ethics working committee. Approved by AoIR, November 27, 2002.* www.aoir.org/reports/ethics.pdf.

ESRC, *Research Ethics Framework*

http://www.esrc.ac.uk/ESRCInfoCentre/Images/ESRC_Re_Ethics_Frame_tcm6-11291.pdf [accessed 20/09/2007].

Gillies, V. & Edwards, R. (2005) Secondary Analysis in Exploring Family and Social Change: Addressing the Issue of Context [30 paragraphs], *FQS* vol 6, no. 1, 2005, Article 44

Goguen, J. (1997). Toward a social, ethical theory of information. Geoffrey C. Bowker, William Turner, Les Gasser (eds) *Social Science, Technical Systems and Co-operative Work: Beyond the Great Divide*. Mahweh, N.J., Lawrence Erlbaum, 1997, pp. 27-56.

Heery, Rachel, Duke, Monica, Day, Michael, Lyon, Liz, Hursthouse, Michael B., Frey, Jeremy G., Coles, Simon J., Gutteridge, Christopher J. and Carr, Leslie A. (2004) Integrating research data into the publication workflow: the eBank UK experience. In, *PV-2004: Ensuring the Long-Term Preservation and Adding Value to the Scientific and Technical Data, ESA/ESRIN, Frascati, Italy, 5 - 7 Oct 2004*. Frascati, Italy, European Space Agency, 8pp.

Luff, P., Heath, C., & Jirotko, M (2002) 'Notes Towards an Applied Ethnography. Working Paper'.

Kelder, J. (2005) 'Using Someone Else's Data: Problems, Pragmatics and Provisions' [56 paragraphs], *FQS* vol 6, no. 1, 2005, Article 39.

Nespor, J. (2000) 'Anonymity and Place in Qualitative Inquiry,' *Qualitative Inquiry*, vol 6, no 6, 2000, pp. 546-569.

Nissenbaum, H. (1998) 'Protecting Privacy in an Information Age: the Problem of Privacy in Public,' *Law and Philosophy* vol 17, 1998, pp.559-596.

Nissenbaum, H. (2004) 'Privacy as Contextual Integrity,' *Washington Law Review* vol 79, 2004, pp. 119-157.

Parker, M. (2007) 'Ethnography/ethics,' *Social Science and Medicine*, in press.

Parry, O. & Mauthner, N.S. (2004) 'Whose Data are They Anyway?: Practical, Legal and Ethical Issues in Archiving Qualitative Research Data,' *Sociology* vol 38, 2004, pp.139-152.

Strathern, M. (1999) 'Pre-figured Features,' in *Property, substance and effect: Anthropological Essays on Persons and Things*, London & New Brunswick, Athlone Press, pp.29-43.

Strathern, M. (2004) 'The whole person and its artefacts,' *Annual Review of Anthropology* vol 33, 2004, pp. 1-19.

Srinivasan, R. (2006) 'Where Information Society and Community Voice Intersect,' *The Information Society*, vol 22, 2006, pp. 355-365.

Thomson, D., Bzdel, L., Golden-Biddle, K., Reay, T. & Estabrooks, C.A. (2005) 'Central Questions of Anonymization: A Case Study of Secondary Use of Qualitative Data,' [33 paragraphs] *FQS* vol 6, no. 1, 2005, Art.29.

Van den Berg, H. (2005) 'Reanalyzing Qualitative Interviews from Different Angles: The Risk of Decontextualization and Other Problems of Sharing Qualitative Data,' [48 paragraphs] *FQS* vol 6, no. 1, 2005, Article 30.

White, M. (2002) Representations or people?, *Ethics and Information Technology* vol 4, no 3, 2002, pp 249-266.
