Ethical Issues in the Documentary Data Analysis of Internet Posts and Archives

Judith Sixsmith and Craig D. Murray

*Qual Health Res* 2001 11: 423
DOI: 10.1177/104973201129119109

The online version of this article can be found at:
http://qhr.sagepub.com/content/11/3/423

Published by:

http://www.sagepublications.com

Additional services and information for *Qualitative Health Research* can be found at:

Email Alerts: http://qhr.sagepub.com/cgi/alerts

Subscriptions: http://qhr.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://qhr.sagepub.com/content/11/3/423.refs.html

>> Version of Record - May 1, 2001

What is This?
Pearls, Pith, and Provocation

Ethical Issues in the Documentary Data Analysis of Internet Posts and Archives

Judith Sixsmith
Craig D. Murray

The documentary analysis of email posts and archives for qualitative research has been outlined elsewhere. Although there is an increase in the number of studies being conducted on listserv and newsgroup material in health research, this has not always been accompanied by a careful, in-depth consideration of the concomitant ethical issues. Therefore, this article outlines the ethical considerations surrounding this form of research, including issues of accessing voices, consent, privacy, anonymity, interpretation, and ownership and authorship of research material.

The use of email methodologies within qualitative health research is currently being explored by social scientists, particularly the use of email interviewing (Murray, 1995; Murray & Sixsmith, 1998), focus groups (Murray, 1997), documentary analysis of email archives and posts (Nochi, 1998; Winzelberg, 1997), and hypertext questionnaires (Michalak, 1998). However, although an increasing number of studies employ these electronic methods, there is very little published material specifically on the ethical dilemmas that surround the use of computer-mediated research (King, 1996; Waskul & Douglass, 1996). This is particularly the case when considering ethical issues that surround the use of publicly available email posts and archives.1

Health-related email forums can provide valuable information through documentary analysis of naturally occurring discourse in posts and archives. Person-to-group email communication involves individuals sending and receiving messages to and from a variety of mailing lists, bulletin boards, and online newsgroups. People use such forums as these to discuss common topics of interest, exchange information, and form new social relationships. Email users compose textual messages at their computer terminals and, using a modem, transmit the text to a distribution list or electronic site where others may read the message and choose whether to reply or add to the message.

AUTHORS’ NOTE: We thank the editorial staff of Qualitative Health Research for their help in preparing this article for publication. We also thank the two anonymous reviewers for their supportive critiques of an earlier draft.
Once communication has occurred within the confines of such groups, it remains available for a variable period of time for other people to access at a future date (2 years or more for mailing lists). In the case of newsgroups, communications are usually stored for a much shorter time (days to weeks). To post to or access information from such lists, it is usually necessary to subscribe to a members list, although sometimes this information is also made available to anyone via the World Wide Web. By joining a mailing list, all posts are automatically sent to the researcher’s email address, almost as fast as the posts are made and archived stores of past posts can be accessed. Posts and archives can then be used as documents or texts in a similar manner to the way in which letters, diaries, or other textual materials might be used in research. This means that researchers can analyze the naturally occurring, unsolicited everyday talk (albeit conducted on-line) that Potter and Wetherell (1995) suggest can help social scientists to understand social phenomena.

The “nonreactive” nature of this documentary research source has been useful when researching sensitive health issues, such as studies on recovering addicts (King, 1994) and survivors of sexual abuse (Finn & Lavitt, 1994). Our own research on the experience of prosthesis use (Murray & Sixsmith, 1996) carried out documentary analysis of email posts and archives, which gave access to a very rich source of data. However, it became apparent that there were important ethical issues that the available literature on email methodologies did not fully address and indeed that formal ethical guidelines (such as the British Psychological Society’s [1993, 1995] published guidelines) did not consider. During our research, we found that such ethical guidelines did not cover the range of ethical dilemmas with which we were confronted. Rather, documentary research on email posts and archives has its own dynamics and generates specific issues that need to be discussed within the research community.

It is with this in mind that this article outlines key ethical dilemmas confronting researchers throughout the whole research process when conducting email documentary analysis. Such issues go beyond traditional notions of protection of participants into researcher obligations to access the experiences, thoughts, and feelings of people whose voices might not otherwise be heard and, similarly, to be sensitive to the interpretation and treatment of research data (see Josselson, 1996). Thus, this article discusses ethical issues of accessing voices, consent, privacy, anonymity, interpretation, and ownership and authorship of email research material in qualitative health research.

ACCESSING VOICES

It has been argued elsewhere (Moore, Sixsmith, & Knowles, 1996) that our ethical obligations as researchers go beyond the protection of participants. Even more important, they require the researcher to involve within the research process those people who, perhaps because of sensitive health issues, do not usually present their views for inclusion in research. It is important to hear about the experiences of people who may be unable or unwilling, because of health problems, to participate in research. For instance, Fleitas (1998) notes that children who are physically stigmatized because of their medical condition or treatments are very sensitive to the responses of others to their differences. It is then the researcher’s responsibility to
develop innovative means to enable such people to be heard. For Fleitas, email and Internet chat rooms were one resolution to this problem. Another possible method, and the one we discuss here, is using email documentary analysis. Such methods are fruitful ways of fulfilling our ethical obligations to help vulnerable people be heard.

CONSENT

One important cornerstone of ethical principles in social research is to ensure that any participants in the research have given their fully informed consent (American Psychological Association, 1992; British Psychological Society, 1993, 1995; British Sociological Association, 1993; Oral History Association, 1996). A notable exception to this principle is observational research, in which traditionally it has been accepted that behavior performed within the public domain may be observed and researched without consent (British Psychological Society, 1993). The reason for this exception is to ensure that natural behavior is observed in its context, uncontaminated by the researchers' aims and objectives.

It is in this sense that qualitative health research using email posts and archives is considered. The key question here is whether it is ethical to use email posts and archives without their authors’ consent, or should consent always be sought and obtained prior to use? This issue is highly contentious. Some researchers consider posts on the Internet to be in the public domain and therefore available without consent for research purposes. For instance, Garton (1997) suggests that researchers are “only participating in the electronic equivalent of hanging-out on street corners . . . where they would never think of wearing large signs identifying themselves as ‘Researcher’. “ Posts to email forums have been recorded and stored without consent in a number of studies (Finn & Lavitt, 1994; Reid, 1991). However, the practice of using email material for research purposes has been criticized within the social science research community (King, 1996). Indeed, the use of such material without the permission of its authors is potentially damaging to the research process, especially when group members discover their words have been used without their knowledge or consent. In these circumstances, participants to discussion forums can feel that their privacy has been invaded and may become distrustful of email forums and of the community of research scientists.

Some researchers have decided that consent is necessary. For example, Egdorf and Rahoi (1994) sought the permission of their computer-mediated communication (CMC) groups prior to conducting research on publicly available lists and archives. However, seeking such permission can create further ethical problems. Where researchers have sought informed consent within the context of the email forum, there has sometimes been an unforeseen impact on group processes. King (1996) cites one member of an email support group who in response to continual postings to the list from people wishing to conduct research, refused to “open up” on-line to be “dissected” (p. 122). The argument we wish to make here is that by seeking consent to use archive material, the researcher may change the dynamics of the group, an intervention that some would suggest is unacceptable.

Contributors’ objections to the use of email material in research might be expected to vary depending on the nature of the forum. For instance, Foster (1994),
with a more academically oriented group, found objections to such requests were about using up capacity on the network and wasting people’s time. Individual members of email discussion groups that are focused on more personal issues may well object to their discussions being the subject of research. Groups can and do operate under their own rules regarding the public availability of postings (Howard, 1993). It is perhaps for this reason that consent should be negotiated wherever possible. By doing this, email participants can decide for themselves what they reveal about their lives to the group forum.

In light of the ethical issues discussed above, it would be advantageous for researchers wishing to conduct documentary analysis of email posts and archives to consult the introductory notes or charters of electronic forums (Langford, 1996). Charters may openly request that research should not be carried out on the forum. If clear directives do not exist, it may be possible to contact the list moderator and gain permission to conduct research. However, researchers need to bear in mind that any permission gained may not necessarily be viewed as consent by all members of the group (see Reid, 1996).

In our own research, negotiated consent was achieved but was somewhat problematic. List moderators were made aware of our presence and purpose, and permission was sought to conduct the research. This permission was granted, and a notice was posted to the group at the outset of the research. However, any old subscribers who had left the group (but had left archived posts) or new subscribers to the group during the 6-month research period were not alerted to our presence unless they read through the group archives. In hindsight, it might have been preferable to take into account the fluid membership of on-line groups (King, 1996) and repeatedly advertise the presence of researchers at the site (Stone, 1995). This would allow participants to choose whether to be involved.

PRIVACY

Ethical guidelines for social researchers state that the privacy and anonymity of participants must be upheld in the research process (American Psychological Association, 1992; British Psychological Society, 1993, 1995; British Sociological Association, 1993). The concept of privacy within social research is itself contentious. Defining which behaviors and/or places are private or public can be difficult. Homan (1991) has suggested that some private behavior is observable in public places (e.g., beaches, railway stations). Thus, an understanding of privacy within social research should take into account its variability between settings and local cultural values.

In the context of posts to email discussion groups, the private or public nature of communications needs to be established. Two key issues can be identified here. The first regards the location within which email posts are written. Although participants are generally informed that their posts will be archived and publicly available, King (1996) argues that because email participants often post from their homes, a false sense of privacy may be engendered. This presents an ethical dilemma about whether such material should form the basis of research data.

Second, the notion of privacy can also hinge upon the nature of the email group. Herring (1996) argues that CMC can be seen as both published (and therefore public) and private material at different times and in different places. For instance, it is
possible to contend that postees perceive email interactions as public for the group but private to outsiders such as researchers. Gurak (1996) found that the use of excerpts from other people’s emails (i.e., not their own) within a particular forum was acceptable to contributors. However, when someone wanted to use those same words outside the forum, people felt uncomfortable and expressed concern. Thus, on some occasions, it can be argued that email posts made to a mailing list are intended for the limited circulation of subscribed, interested members (cf. Herring, 1996). This further demonstrates why the ethical issues surrounding analysis of email posts and archives vary from concerns with documentary analysis per se.

With email material, the conversational interlocutors are often still socially engaged with issues that have been discussed in previous posts and thus may be affected by research use of such material. In contrast, traditional texts viewed as documentary sources tend to be concerned with transpired events and sometimes people who are no longer living.

The sense of group privacy in CMC referred to above may account for increased self-disclosure in email communication. For instance, social presence theory (Short, Williams, & Christie, 1976) would suggest that the absence of visual, aural, and other information characteristics of face-to-face interaction would make email users less aware of their audience (see Kiesler, Siegel, & McGuire, 1984) and more likely to treat their posts as private communications.

Alternatively, when email posts are made to mailing lists, 10s, maybe 100s and 1,000s of people become the audience. Moreover, email posts and archives are available to anyone linked to the Internet. In this sense, they do constitute material that is firmly in the public domain. Indeed, anecdotal evidence suggests that people posting messages to discussion forums often consider their material, to some extent, to exist within the public domain: “I’ve always treated my contributions to email as public documents” (Burton, 1994, p. 108).

Clearly, researchers need to take into account the nature and purpose of the discussion forums and people’s understandings of what is public and what is private before committing them to research scrutiny.

Finally, Homan (1991) argues that special considerations are needed to safeguard privacy when data, which have been collected and stored under one set of circumstances, are made available to others (including researchers) who were not originally anticipated. This is the case with various forms of archive material, which were originally intended for limited circulation and rarely thought of as sources for research (Homan, 1991). In fact, one of the aims of the Data Protection Act (Her Majesty’s Stationary Office, 1984) is to prevent the use of data for purposes other than those for which it was originally collected (Lauer, 1996).

ANONYMITY

If archived posts are to be used in qualitative health research, then a number of precautions can be taken to safeguard an individual’s anonymity. All identifying information contained within an email message should be removed, which would ensure that personal anonymity is maintained. This would include names and pseudonyms used by participants in their emails as well as the names and locations of lists and newsgroups.
When transgressions of anonymity have occurred, CMC communities have been made vulnerable. Finn and Lavitt (1994) name the actual bulletin boards involved in their research on survivors of sexual abuse and supply the date and times of the posts. They argue that changing the names of the senders ensured anonymity. However, the actual identities of group members could easily be located using the information supplied by the researchers. The anonymity of the group and its members was thus compromised (King, 1996).

The problem with removing all information that could identify participants lies in the possible loss of research information that accompanies thick description (see Davis, 1991). Thick description includes describing personal features (e.g., age, nationality, occupation, and so forth) about a research participant, which are considered important to fully contextualize a researcher’s interpretations. Akeroyd (1991) points out that disguises and omissions may affect features critical for the analysis and create problems for validity, reliability, and replicability. In general, the researcher’s commitment to protection of privacy and anonymity should usually take precedence over problems of information loss, although this is an issue with which each researcher must deal (see Yow, 1984).

Also, the protection of anonymity can be more confidently achieved by avoiding verbatim quoting from archived posts. King (1996) suggests that this offers a degree of anonymity because it is much more difficult to link together the research issues identified with specific contributors to the discussion forums. However, for qualitative researchers, such an approach can be seen to impoverish the data because paraphrasing the content of email posts may undermine the researcher’s interpretive purchase.

**INTERPRETATION**

The analysis of data and the consequent interpretation of meaning are not without their own ethical implications. Some researchers, particularly those from the oral history tradition (e.g., Yow, 1984), refer specifically to the ethical importance of not misrepresenting the narrator’s meaning or changing the sense of his or her words. As Gilbert (1997) notes, CMC provides a space of social action that although discursive, has a phenomenological reality.

The issue here relates to the representation of the views of those people whose communications form the basis of research data. With both posts and stored archives, we have textual narratives of people’s lives. Researchers take such written material and reshape it for an academic audience (Borland, 1991). As with all research, the danger is that this may be a misappropriation of words and meaning.

In research using email, misrepresentations can occur when the researcher does not have available the totality of communications made within the discussion forum. A full record of email communication may not be available to the researcher for two reasons. First, the archive may be incomplete. Rheingold (1993) talks about the “shock” that occurred when a prolific writer deleted several years of posts and how this made “the fabric of recorded conversations, the entire history of the [group’s] discourse . . . look moth-eaten” (p. 36). For research purposes, not only are such archives rendered moth-eaten but also the whole integrity and meaning of the debate is lost.
Second, discussion group data may comprise only a selective portion of the total communication because some exchanges can take place in private email (person to person), that is, outside of the forum of the group discussion. Therefore, whereas archive posts appear inclusive of all discussion between members, this may not always be the case (Waskul & Douglass, 1996).

If misinterpretation of data is to be minimized, then tracts of related messages need to be considered holistically. Messages are often formed in part to preceding messages and discussion threads. Taking these other messages into account allows their discursive context to be considered and as such allows for a more grounded interpretation. This approach was taken by Howard (1993), who took the precaution of quoting whole series of messages so that readers could establish the context for the messages.

OWNERSHIP

The ownership of discourse is by no means a unique ethical problem to CMC research, but it does present some interesting challenges to conducting ethically sound research using email.

The notion of ownership is of particular interest when considering documentary analysis of email posts and archives. One question that requires thought is “To whom do the posts belong?” Do they belong to the poster (author), electronic group (community), or any observer (including researchers)? Straightforward answers to these questions are not forthcoming. In the previously discussed example, one email group contributor removed everything he had written over a 2-year period (Rheingold, 1993). In doing so, he asserted ownership of those messages while simultaneously destroying the integrity of the discussion and diminishing the communal identity of the group.

This situation is compounded when researchers appropriate such material for their own use. For instance, the communication that takes place between people with “gender identity disorder” might conceivably be used to help devise treatments and interventions. However, such an approach by health professionals may be perceived as oppressive by the very people whose discourse is used. Can researchers ethically use such email material in publications and reports when email communicators might not approve of the purposes of the research?

AUTHORSHIP

A final ethical issue concerns the authorship of email posts. Even at the publication stage of the research, power over the material rests primarily with the researcher (see Ribbens, 1989). However, there has been some concern that copyright regulations might be involved. Cavazos (1994) takes a legalistic copyright view, suggesting that emails are very similar to published work and as such, any quotation from them should include a credit to the source. This is also the view of Mailbase (1998), a service that runs electronic discussion lists for the UK higher education and research community. However, if credit is given, then we have already seen that this would compromise anonymity. Researchers such as Boehlefeld (1996) attempt to
reconcile these seemingly opposing views by insisting that permission is sought to use long quotes. Seeking permission empowers people to choose whether they want their words used within a research publication and whether they would like to be credited.

CONCLUSION

In this article, we have discussed some of the key ethical considerations pertinent to qualitative documentary analysis of email content. In particular, ethical issues of accessing voices, consent, privacy, anonymity, interpretation, and ownership and authorship of material have been problematized. As can be seen, these ethical considerations are not simply related to data collection but are located throughout the research writing and publication process.

An approach to research that more fully involves participants in the research program would alleviate some of the ethical difficulties discussed in this article (Banister, Burman, Parker, Taylor, & Tindall, 1994). For instance, as an alternative to rule-bound ethical codes of practice in electronic research, Allen (1996) advocates “creative ‘ethical’ work” (p. 186) in which researcher and researched take part in a dialogue. As such, ethics of research “should be situated, dialogic agreements that develop over time” (p. 186) between researchers and participants. Indeed, there is some anecdotal evidence that CMC participants would like to be in on the research.

One subscriber of a discussion group, upon hearing rumors of a study that was to be carried out using the list, wrote,

Unless these academo-dweebs get down and dirty with us, the study is bound to be bogus from the start. I’m highly unimpressed. They remind me of Masters and Johnson. All observation, no participation. (as cited in Sudweeks & Rafaeli, 1996, p. 123)

The ethical issues discussed in this article serve to alert potential researchers of email content to the often problematic nature of such research. One aim of this article has been to present advice, based on our own research experience, about best practice. However, it is a further aim of this article to stimulate researchers to engage actively with the ethics of electronic research and with the particular dilemmas that may face them. As such, it is intended that this article goes some way in facilitating the engagement of researchers with email material in an ethically responsible manner.

NOTE

1. We refer specifically here to email posts and archives. However, it is our contention that the arguments presented here pertain to a broad range of electronic communications, including listerv and bulletin board discussion groups and chat rooms.
REFERENCES


Judith Sixsmith, Ph.D., is a senior lecturer in psychology at the Manchester Metropolitan University. Her current research interests include the psychosocial and environmental impact of teleworking for individuals and work organizations, the therapeutic role of the landscape, and the relationship between social capital, health, and gender in socially deprived communities.

Craig D. Murray, B.A., is a lecturer in psychology at Liverpool Hope University College where he teaches health psychology and qualitative methodology. His research interests include the social and phenomenological body and psychology and technology. He is currently writing his doctoral thesis, titled *The Experience of Prosthesis Use: Phenomenology, Culture, and Method*.