Researching Online Forums

Ethics Case Study | 1

BRITISH
SOCIOLOGICAL
ASSOCIATION

Researching Online Forums

Lisa Sugiura
University of Portsmouth

Online forums are discussion groups where people converse about topics of mutual interest. Public forum data can be accessed with little difficulty or interaction with the group and do not require password access or user registration, with posts accessible in the same way as letters to a newspaper, or a conversation on a bus. It is not possible to see who is reading the conversations, but users who wish to comment identify themselves, often using a pseudonym. Private forums, on the other hand require registration and passwords to access. Text from forums can be gathered by a computer programme or by manual copy-and-paste functions.

Research involving online forums raises ethical issues relating to informed consent of human subjects, protection of privacy and anonymity of research subjects.

These issues are not independent, instead they should be addressed together to mitigate overarching ethical concerns. They will be addressed in turn, with suggestions for best practice ethical approaches. However, this advice should be tailored to individual research projects, and are not intended as a 'one size fits all' approach.

Research involving online forums raises ethical issues relating to informed consent of human subjects.

Obtaining Informed Consent

Data from online forums are readily accessible to anyone, and, if archived, are accessible to the public months or years after messages were posted (Frankel and Siang, 1999).

This type of research could be exempt from the informed consent requirement, if it is conducted in public (Liu, 1999). However, due to the lack of public awareness, some commentators/researchers have argued that messages within online communities should not be collected without the author providing prior permission (Marx, 1998; King, 1996). Wilson and Atkinson (2005) also question whether online ethnography might be a form of 'electronic eavesdropping'. An individual might post information on his or her public profiles to be shared with friends and peers; however, this does not mean that they have consented for this information to be collated, analysed and published, in effect turning them into research subjects (Eysenbach & Till, 2001).

Hudson and Bruckman (2004) found that while it might be widely considered ethically acceptable to capture and analyse interactions and conversations in a public square without consent, this model did not match the expectations of their participants in realtime chatrooms. Nevertheless, Eysenback and Till (2001) have contended that it is ethical to record activities in a public place without consent, provided individuals are not identifiable. Human subject research norms such as informed consent do not apply to material that is published. However, the nature of online content means that it is more complex to distinguish between published and non- published material (Bruckman, 2004:103).

Ethical approach – Informed consent is not legally required to access data from publicly available forums, as they are in the public domain (as with much of the web, the legal frameworks and case law have yet to be made to govern this aspect of digital technology), but this is not to say that consent should automatically be overlooked. There are pitfalls both from attempting to obtain informed consent and bypassing it.

The following discussion provides a consideration of both approaches and these could be applied to individual cases, to help determine the best course of action.

Obtaining informed consent in either public or private forums, may involve the researcher posting to communities or individually contacting users and providing them with participant information sheets and consent forms to sign. This would require the researcher to join the community as a user, revealing their true identity and the purpose of their study. In some situations it might be precarious for the researcher to reveal such personal information, for example if the topic was sensitive. Furthermore, disclosure might disrupt the 'naturalistic' research environment. There are also practical difficulties involved in procuring informed consent from all members of online communities, as not everyone may see posts, and some members may have left, leaving their contributions still visible. However, seeking such permission can also create further ethical problems. In other studies, researchers have sought informed consent and found similar unforeseen impact on group processes.

King (1996) cites one member of an email support group who, in response to continual posts to the list from people wishing to conduct research, refused to "open up" online to be "dissected" (122). Hewson (2003) also question whether contacting potential participants may be viewed as "spamming", itself an invasion of privacy (40). In contrast, the covert approach might enable research to be undertaken without risk or harm to the community, especially where a posted site policy notifies users of its public access, which is a point noted by Sveningsson (2004). It would be advantageous for researchers wishing to conduct analysis of posts and archives to consult the introductory notes or terms of electronic forums, this is a view supported by Langford (1996).

Terms may openly request that research should not be carried out on the forum. Where 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 (Reid, 1996).

Whether consent needs to be obtained from individual contributors or from communities and online system administrators is fraught with uncertainty. The issue of ownership/intellectual property of the data may be addressed in the terms and conditions, but the moderators cannot speak for the forum users. What is public and what is private is blurred on the web. It is not sufficient simply to rely on whether a site is public or not; privacy and confidentiality are further important considerations for researching online forums. These issues will now be discussed in more detail.

Terms may openly request that research should not be carried out on the forum.

Privacy and Confidentiality

In online environments that are publicly viewable, such as discussion groups, individuals' expectations may be different from in communications offline, or in private digital correspondence such as email (Smith, Dinev and Xu, 2011). It is not always possible to determine whether users are aware of the public status of their contributions from the contributions themselves, or whether interaction with the user is required.

Individual and cultural definitions and expectations of privacy are ambiguous, contested and changing. People may operate in public spaces but maintain strong perceptions or expectations of privacy. Frankel and Siang (1999) have suggested that people may be more open online due to a false or exaggerated expectation of privacy (Frankel and Siang, 1999: 6).

Other groups have attempted to clarify the boundaries of public data for research (Sveningsson, 2003; McKee and Porter, 2009). According to the ethical guidelines of the AoIR, public forums can be considered more public than, for example, conversations in a closed chatroom (Ess and AoIR, 2002: 5, 7). While Basset and O'Riordan (2002) state that the lacking of applicability of a private sphere implies that all discourse lies de facto in the public sphere. However, Bakadjieva and Feenberg (2001) offer a different perspective, suggesting that the type of research and corresponding forms of relationship between the researcher and the subject has an impact on whether or not a space should be considered public or private.

Though conversations may occur in public spaces, the content could be private. In such circumstances, people may accidentally disclose personal information that could identify them in the research. As noted in the 2002 version of the AOIR ethics guidelines, privacy is a concept that must include a consideration of expectations and consensus. When conducting research within such shifting terrains, when there is no consensus, or even assumption of consensus, the AOIR suggest that Nissenbaum's concept of contextual integrity (2011) is a valuable construct. The accessibility of online discussions may suggest that they are freely available in a public arena; however, some researchers question whether the availability of information on the web necessarily makes this information public. For example, Heath et al. (1999, cited in Grinyer, 2007: 2) suggest that research involving 'lurking' encroaches on privacy and creates an unequal power relationship.

Ethical approach

The following discussion applies to public forums. For collection purposes, merely treating forum data as public text used for documentary analysis is insufficient, as the thoughts and intentions of those who had produced the information should be considered. Examination of people's feelings about that situation – the ethic of reciprocity, or Golden Rule, where the researcher considers how they would feel if the roles were reversed was considered, in order to appreciate how those observed might respond to the research (Honderich, 1995; Rawls, 1958). This would impact on whether the environment is considered public or private; for example, if someone was talking in a public space, it would be reasonable to expect that their conversation could be heard and accessed by others. However, this is difficult online, as web spaces have ostensible boundaries. Content on websites can be accessed by anyone and is not necessarily meant for public consumption. Researchers can familiarise themselves with the place of study in order to ascertain whether it should be considered public from the perspective of those who occupy it. This requires continual reflection during the research process.

Individuals and their online privacy expectations should be respected. If an individual has posted information on a public website under a public "privacy" setting, they may be considered to have a very low or no expectation of privacy for the information they reveal; regardless, in such situations the researcher needs to be careful not to make undue assumptions. The discussion above has identified that establishing the privacy expectations of research subjects is a problematic issue and one that is intensified by the web, as is the possibility of intruding on private exchanges and risking personal information during online research. One way to protect privacy is anonymisation. Anonymising data is a process designed to protect research subjects and their personal information, and to satisfy legal requirements such as the Data Protection Act 1998. However, whether data can be appropriately or completely anonymised is also debatable in web research.

Anonymity

A central feature of research is to provide descriptions and explanations that are publicly available and accessible. One potentially harmful outcome of research, however, is the risk of disclosing an individual's identity, and it is the responsibility of the researcher to employ preventative measures such as anonymity (SRA, 2003: 389) where there may be negative effects from disclosure. Although complete anonymity may be difficult to ensure, it is advised to remove all identifying data prior to publication, and where an individual is identifiable, explicit consent is required before publication (Wiles, 2012). However, web research complicates attempts to ensure anonymity, as data can be easily put into a search engine and the initial source easily discovered.

Bruckman (2004) proposes guidelines that incorporate a "continuum of possibilities" in the level of disguise required for individuals' names when reporting research (Bruckman, 2004: 229). With respect to web data; steps should be taken to protect all the individuals participating in research by removing all names and any identifying information in the final thesis and in any stored data. URLs or "links" to the forum websites should not be provided, and other personal details should be disguised; however, quotes may be used to evidence any findings and ensure traceability. Bruckman (2004: 229) suggests adopting a "moderate disguise", whereby verbatim quotations may be used but names, pseudonyms and identifiable details changed. This approach was also adopted in Hookway's (2008) study of morality in everyday life, where he prioritised the protection of his participants' identity over providing credit to them as authors.

Some online discussions contain personal information. This is further complicated by the blurring of the private and public distinction. The ethical guidelines of the Association of Internet Researchers (AoIR) suggests a settingdependent approach to distinguishing between subjects and authors, distinguishing between "reasonably secure domains for private exchanges" such as chatrooms and "public webpages such as homepages, Web logs [i.e. blogs]" (Ess and AoIR, 2002: 7). Where the research context is placed on the public/private continuum, this has an impact on the need to anonymise data. If people are considered to be subjects, then they need to be afforded the protection of anonymity; however, if the information they have posted is considered to be published, then they should be credited as an author.

Acknowledging when anonymity should be used and when it is necessary to cite a web user by their name (or pseudonym) is problematic. There may be circumstances when some web users may not want to remain anonymous, for example writers of blogs (though these appear quite distinct from forum posts), and so it would be inappropriate to anonymise such individuals. This would be viewed as infringement of copyright and incur issues of intellectual property. If web users are treated as authors of public documents, then issues of ownership of material must be considered.

Web users may have chosen to deliberately publish in the public domain. Bassett and O'Riordan (2002: 244) argue that in such cases, rather than maintaining anonymity, researchers should acknowledge the user's authorship and cite their texts as they would more traditional media, but as Ess (2006) points out, this may compromise their anonymity. Removing all identifying data about the web user, site etc. prior to publication is one solution to the problem of anonymisation procedures. However, the use of verbatim quotes to substantiate findings can impair this, as the quotes can be traced back to the original website and potentially to the person who made them. This is a new challenge created by the web, and one that researchers should be mindful of, possibly making the checks to determine the risk of uncovering individual identities.

If protection cannot be ensured via anonymity, then perhaps such data should not be reported. Anonymity per se cannot be solely relied on to avoid the need for informed consent; along with the notions of privacy and confidentiality, it requires intense consideration specific to the research issue and setting, as well as to the individuals concerned.

Ethical approach

When quoting comments, anonymisation is fundamental, as negative consequences to participants could arise from disclosure that resulted in violation of privacy. Even though information may be readily available to anyone online, and could be found by anyone using the similar search terms, researchers should not bring any extra unnecessary attention to anything written in cyberspace by individuals, especially where it has been analysed in relation to specific research issues. Therefore, anything of an embarrassing or sensitive nature, such as information about personal illnesses for example, should be removed and not used within the analysis of the data.

Researchers who collect and analyse online forum data (whether it is from public or private forums) should take care to protect it from becoming identifiable to individuals. As such, conversations should not be copied verbatim into research publications, as those direct quotes can be searched and identities discovered. A small number of relevant conversations can be summarised without losing character in reports. The jury is still undecided over whether full quotations need permission, though the various principles of ethics that have been discussed would suggest that this is more likely the case.

> If web users are treated as authors of public documents, then issues of ownership of material must be considered.

References

Bakardjieva, M. & Feenberg, A. (2001). Respecting the Virtual Subject, or How to Navigate the Private/ Public Continuum. *Online Communities: Commerce, Community Action, and the Virtual University,* 195-214.

—

Bassett, E. H. & O'Riordan, K. (2002). Ethics of Internet Research: Contesting the human subjects research model. *Ethics and Information Technology,* 2: 233-47.

—

Bruckman, A. (2004). Opportunities and challenges in methodology and ethics. In: Johns, M. D., Chen, S. L. S. & Hall, G. J. (eds), *Online Social Research: Methods, Issues and Ethics.* New York: Peter Lang.

—

Ess, C. & the AoIR Ethics Working Committee (2002). *Ethical decisionmaking and Internet Research: Recommendations from the AoIR Ethics Working Committee.* Available at: http://aoir.org/reports/ethics.pdf

—

Ess, C. (2006). Ethical Pluralism and Global Information Ethics. *Ethics and Information Technology,* 8(4): 215-26.

—

Eysenbach, G. & Till, J. (2001). Ethical issues in qualitative research on internet communities. *British Medical Journal,* 323, 7321, 1103-5.

—

Frankel, M. & Siang, S. (1999). Ethical and Legal Aspects of Human Subjects Research on the Internet. *American Association for the Advancement of Science Workshop Report.* Available at: http://www.aaas.org/ spp/dspp/srfl/projects/intres. main.Html

—

Grinyer, A. (2007). The Ethics of Internet Usage in Health and Personal Narratives Research. *Social Research Update 49,* University of Surrey.

Hewson, C. (2003). Conducting Research on the Internet. *PsychologistLeicester,* 16(6): 290-3.

—

Honderich, T. (1995). *The Philosophers: Introducing Great Western Thinkers.*

—

Hookway, N. (2008). Entering the Blogosphere: Some Strategies for Using Blogs in Social Research. *Qualitative research,* 8(1): 91-113.

—

Hudson, J. M. & Bruckman, A. (2004). 'Go away': Participant Objections to Being Studied and the Ethics of Chatroom Research." *The Information Society,* 20(2): 127-39.

—

King, S. A. (1996). Researching Internet Communities: Proposed Ethical Guidelines for the Reporting of Results. *The Information Society,* 12(2): 119-28.

—

Langford, D. (1996). Ethics and the Internet: Appropriate Behavior in Electronic Communication. *Ethics & Behavior,* 6(2): 91-106.

—

Langer, R. & Beckman, S. C. (2005). Sensitive Research Topics: Netnography Revisited. *Qualitative Market Research: An International Journal,* 8(2): 189-203.

—

Liu, G. Z. (1999). Virtual Community Presence in Internet Relay Chatting. *Journal of Computer-Mediated Communication*, 5(1): 00.

—

Lomborg, S. (2013). Personal Internet Archives and Ethics. *Research Ethics,* 9(1): 20-31.

Marx, G. T. (1998). Ethics for the New Surveillance. *The Information Society,* 14(3): 171-85.

———

McKee, H. A. & Porter, J. E. (2009). *The Ethics of Internet Research: A Rhetorical, Case-based Process* (Vol. 59). New York: Peter Lang.

———

Miller, T., Birch, M., Mauthner, M. & Jessop, J. (eds) (2012). *Ethics in Qualitative Research.* London: Sage.

———

Nissenbaum, H. (2011). A Contextual Approach to Privacy Online. *Daedalus,* 140(4): 32-48.

———

Paccagnella, L. (1997). Getting the Seats of Your Pants Dirty: Strategies for Ethnographic Research on Virtual Communities. *Journal of Computer-Mediated Communication,* 3(1): 00.

———

Rawls, J. (1958). Justice as Fairness. *The Philosophical Review,* 164-94.

———

Reid, E. (1996). Informed Consent in the Study of Online Communities: A Reflection on the Effects of Computer-mediated Social Research. *The Information Society,* 12(2): 169-74.

———

Signorini, A., Segre, A. M. & Polgreen, P. M. (2011). The Use of Twitter to Track Levels of Disease Activity and Public Concern in the US During the Influenza A H1N1 Pandemic. *PloS one,* 6(5): e19467.

———

Smith, H. J., Dinev, T. & Xu, H. (2011). Information Privacy Research: An Interdisciplinary Review. *MIS Quarterly,* 35(4): 989-1016.

Snee, H. (2013). Making Ethical Decisions in an Online Context: Reflections on Using blogs to Explore Narratives of Experience. *Methodological Innovations Online,* 8(2): 52-67.

———

SRA (2003) *Social Research Association Ethical Guidelines.* Available from: http://www.thesra.org.uk/ethical.htm

———

Staniszewska, S., HerronMarx, S., & Mockford, C. (2008). Measuring the Impact of Patient and Public Involvement: The Need for an Evidence Base. *International Journal for Quality in Health Care,* 20(6): 373-4.

———

Sveningsson, M. (2004). Ethics in Internet Ethnography. In: Buchanan, E. A. (ed), *Virtual Research Ethics: Issues and Controversies.* Hershey PA: Information Science Publishing, 45-61.

———

Wiles, R. (2012). *What Are Qualitative Research Ethics?* London: Bloomsbury Academic.

———

Wilson, B. & Atkinson, M. (2005). Rave and Straightedge, the Virtual and the Real: Exploring Online and Offline Experiences in Canadian Youth Subcultures. *Youth & Society,* 36(3): 276-311.