Ethical and Methodological Problems in Online Research

Margaret Emerton
Supervised by Dr Kate Bowles

This essay was written under the terms of a Summer Scholarship from the Institute of Social Change and Critical Inquiry and the Faculty of Arts, University of Wollongong, Australia.

Not for duplication without the permission of the author
Email: me21@uow.edu.au

March 2003
The Internet gives researchers access to an unprecedented –and unstable– archive of human activity and its traces (Jones, 1999; Lindlif and Shatzer, 1998). It differs from other media and communication technologies in a number of ways that affect research design. It reaches across cultural boundaries (Mitra and Cohen, 1999), national borders (Commercenet, 2002) and legal jurisdictions (Geist, 2001). It integrates multiple modalities of communication and types of content (Di Maggio, Hargittai, Russell-Neuman and Robinson, 2001; Morris and Ogan, 1996, in Mitra and Cohen, 1999), while facilitating play between them (Sosnoski, 1999). It provides technical utilities for efficiently collecting and processing data (Lindlif and Shatzer, 1998) but can obscure correlations by destabilising identity (Allen, 1996; Turkle, 1999).

These– and other– differences between online and off-line environments compel researchers to modify existing research methodologies and design new ones. The ethical conventions of some disciplines may become problematic when transplanted to online environments (Bruckman, 2002; Frankel and Siang, 1999), and resultant tensions between methodology, ethics and disciplinary discourse need to be resolved (Ess and Association of Internet Researchers Ethics Working Committee, 2002). Ethics and methodology are sometimes examined separately in the discussion that follows, but an underlying assumption is that regardless of how they are conceptualised, they are neither entirely separable from each other nor fixed. Every methodological decision contains –and is bounded by– ethical choices. They are contingent on each other and likely to co-vary throughout the processes of research design, implementation, evaluation and reporting.

This essay surveys literature on methodological and ethical issues relevant to online ethnographic research. It focuses on strategies for exploring reciprocal interactions and group discussions in online environments. These environments may be asynchronous (bulletin boards, mailing lists, Usenet, email exchanges) or synchronous (chatrooms, instant messaging, multi-user domains, groupware). The essay begins by examining how conceptualising online environments affects their characteristics as sites of inquiry. It reviews some key ethical traditions in humanities and social science research. Integrating these two sets of theorisations, it identifies areas where the methodological and ethical choices
available to online researchers may impede rigorous investigations of their subjects. It moves towards a consideration of specific problems that empirical research has identified as unique to online environments. Finally, it makes some conditional recommendations for conducting online research.

**Conceptualising Internet Environments**

**Community**

Ethnography, ‘writing about people’, requires in the first instance a group of people to write about. It assumes that they constitute a cultural entity, with shared understandings and common social processes that a researcher might articulate coherently (Burns, 1997; Kellehear, 1993; Lindlif and Shatzer, 1998). For ethnography to be a valid research paradigm online, sites of computer-mediated human interaction must be understood –at least in part– as forms of community.

This requirement is met by sites where participants share “common value systems, norms, rules and the sense of identity, commitment and association that also characterize [sic] various physical communities or other communities of interest” (Fernback, 1999, p.211). Lindlif finds the use of ‘community’ fitting for virtual spaces because they: “…often function as running conversations among many actors, and…it takes a measure of unselfish effort to sustain these conversations” (Lindlif and Shatzer, 1998, p.3). As in off-line communities, the degree of commonality –and commitment to sustaining conversation– varies enormously. So does the extent and tolerance of internal variation and of members’ simultaneous involvement with other communities. Online communities are more likely to be based on shared interests than shared demographic characteristics (Wellman and Gulia, 1999, in Di Maggio et al., 2001). While they are intentional (Lindlif and Shatzer, 1998) they are not necessarily voluntary: participation might be a compulsory component of some other activity. Community participants may or may not have pre-existing connections offline, or be members of analogous embodied communities. Their
only commonality may turn out to be their presence, or in the case of lurkers\(^1\), the trace of their potential presence. Care should be taken not to confuse or erase meaning by conflating models of community (Di Maggio et al., 2001; Kendall, 1999). A consequence of conceptualising online research sites as communities is that participants (and their textual traces) become human subjects: a casting with profound implications for research design that will be discussed further.

**Place**

Community has historically been identified with place (Fernback, 1999; Jones, 1999) and understandings of online environments as virtual or cyber-communities are supported by the ubiquity of spatial metaphors in Internet discourses (Bassett and O'Riordan, 2002; Cavanagh, 1999; Fernback, 1999; Nunes, 1995; Stone, 1995). Spatiality embedded in browser software greets the Internet user at the home page, helps them navigate the information superhighway: back, forward, home. Webpages use geographical metaphors: Geocities, site maps. Interactive sites use architectural ones: chat rooms, cyber lounges, virtual cafes, furnishing them with familiar objects: bulletin boards, visitors’ books.

Familiar ways of classifying space can be problematic in online research settings. Online sites “do not often have a tangible mooring” (Mitra and Cohen, 1999, p.197) and may move or disappear without warning. Borders may be difficult to define because of hyperlinks and other elements of software design. To limit data collection to available online spaces risks neglecting significant information found in off-line or private settings (Phillips, 1996 in Lindlif and Shatzer, 1998; Wynn and Katz, 1997). Depending on design, the validity of ethnographic research may be compromised by omitting community maintenance activities that occur outside their ‘home’ environment (Lindlif and Shatzer, 1998), or by misunderstandings caused by shifting contexts and codes (Paccagnella, 1997).

---

\(^1\) Lurkers have access to a group’s messages but rarely, or never, submit their own posts. As Cavanagh points out, lurkers are “at the very least tolerated in online environments” and “…often receive a warm welcome from communities when changing status” (Cavanagh, 1999).
It is difficult to sustain an understanding of public and private spaces as a binary opposition requiring differential treatment when (passwords and encryption notwithstanding) privacy in online environments is somewhat illusory (Hsia, 2000; Lindlif and Shatzer, 1998; Walther, 2002; Wynn and Katz, 1997). Internet technology relies on transmitting data through public infrastructures that facilitate monitoring and tracking information (Brunetti, 2002; Zuboff, 1989 in Di Maggio et al., 2001). Participants modify their behaviour accordingly, so that a degree of identity fabrication and guardedness about offline existence is normative in online environments. Whether and how this is significantly different from other research environments requires further investigation (Walther, 2002).

Aggregation of Internet services has changed the accessibility status of archived material. For example, messages consigned to Usenet news groups with the assumption that they were ephemeral (Mitra and Cohen, 1999) are readily accessible and subject to sophisticated full-text search capabilities since their absorption into Google’s Groups (Google, 2002; Mieszkowski, 2002).

**Site of Textual Production**

The adoption of community and spatial models is useful for exploring some aspects of Internet use, but there are dangers in assuming these are the only models available. The Internet is also a site of textual production and exchange; one where intertextual and intersubjective processes are foregrounded by the absence of concrete referents (Sandbothe, 2000). Systematic analysis of Internet texts and contexts offers to illuminate not only the texts and the textuality of the medium, but the users as readers, writers, producers and consumers of the texts (Mitra and Cohen, 1999). Hypertext, multimedia texts, interactivity and multilinear reading paths radicalise the processes of reading and authoring (Sandbothe, 2000).

People using the Internet to communicate with each other have unprecedented access to sharable information, communication tools and international reach (Mitra and Cohen, 1999). Their communications are mediated by computer technology, software design, representational politics (White, 2002), time-zone
differences and other aspects of their embodied context (Sussman, 2001). When ethnographers adopt community and place models uncritically they are less likely to evaluate the impact of the medium on texts and identities (White, 2002).

The Internet can be thought of as: “…a playground for amateur artists creating semi-published work” (Bruckman, 2002). To fail to consider the textuality of Internet postings and the contexts of their production risks conflating character with author, persona with person (Bassett and O'Riordan, 2002; White, 2002). The difference between these positions has profound implications for research design. The same action protects the rights of human subjects under one model and infringes the rights of human authors under the other. A decision to include anonymous quotes may trigger a breach of copyright (Bruckman, 2002). Conversely, attributing authorship could cause harm in a research subject’s embodied life (Boehlefeld, 1996).

**Liminoid Space**

Drawing on the work of the folklorist Arnold van Gennep and the anthropologist Victor Turner, some writers have investigated the liminality of Internet environments (Allen, 1996; MacWilliams, 2001; Shields, 2000; Strenski, 1995). Liminoid spaces are temporary performative spaces (Shields, 2000), the post-industrial manifestations of transitionary, often ritualised, social spaces where day-to-day realities, processes and expectations are suspended. For Turner, liminality allows the possibility of: “formulating a potentially unlimited series of alternative social arrangements” (Turner in Allen, 1996, p.183). Liminoid spaces are thresholds: they can be understood as cusps between embodied quotidian reality and imagined utopias, between the actual and the possible (Deleuze in Shields, 2000).

Applied to Internet interactions, this model alerts researchers to the need for circumspection when making connections between the participants’ online and offline worlds. They may operate under entirely different paradigms, and the boundaries might not be clear, consistent or mutually understood by participants or observers (Allen, 1996). The model illuminates empirical observations of Internet behaviour, particularly: identity experimentation (Turkle, 1999); the
degree of candidness and intimacy extended to mass audiences (Bassett and O'Riordan, 2002; Frankel and Siang, 1999; Strenski, 1995; Waskul and Douglass, 1996); the conflation of oral and written modes of address (Capurro and Pingel, 2002) and of autobiographical and fictional genres (Cavanagh, 1999; White, 2002).

### Ethical Traditions in Humanities and Social Science Research

Researchers exploring human activity in online environments nominally have access to a number of ethical traditions within which to design and evaluate their work. Institutional constraints on research design privilege certain positions that are not necessarily appropriate for online research: especially research that crosses cultural and disciplinary boundaries. This section sketches the orientations and concerns of ethical traditions in humanities and social science research to facilitate later discussion. It examines the regulation of research involving human subjects using the University of Wollongong as an example, and critiques human subjects protocols with reference to other ethical traditions.

#### Deontological Positions

Deontological positions proceed from sets of formally stated, codified rules. Act-deontological positions allow researchers to make situated judgments about appropriate courses of action, provided they operate within a framework of explicitly stated shared principles. Rule-deontological positions formally specify courses of action with the expectation that they will be universally applied (Thomas, 1996).

#### Teleological Positions

Teleological positions move in the opposite direction, evaluating courses of action according to the outcomes they produce. As a guiding principle, an action is ethical if “the consequences of that action are more favourable than unfavourable” (Fieser, 2001). Utilitarianism attempts to calculate “the greatest social good or the least social harm” as reference points for evaluating outcomes (Thomas, 1996, p.109). Act-utilitarianism evaluates actions on a case-by-case
basis. Rule-utilitarianism invokes generalised rules of conduct, evaluating the rules themselves against their outcomes (Fieser, 2001; Thomas, 1996).

These paradigms have dominated the discourse of research ethics in twentieth century Western cultures and have shaped the post-Nuremberg regulation of human subjects research (Capurro and Pingel, 2002; National Health and Medical Research Council NHMRC, 1999; Thomas, 1996). European and Australian researchers generally work from within deontological paradigms, while the United States uses utilitarian positions (Capurro and Pingel, 2002; Ess and Association of Internet Researchers Ethics Working Committee, 2002). This difference in orientation is reinforced by European laws that enshrine rights (for example: to privacy, data protection and consumer protection) that remain subject to negotiation and market forces in the United States (Ess and Association of Internet Researchers Ethics Working Committee, 2002). This becomes significant in cross-jurisdictional research, where researchers and their subjects operate under different paradigms with different sets of assumptions and expectations about their rights and obligations.

**Human Subjects protocols**

Human subjects protocols are intended to protect the rights and welfare of all participants and other parties who are affected, directly or indirectly, during research involving human subjects (NHMRC, 1999). Their development has been a multilateral project, gaining momentum after the judgment of the Nuremberg Tribunal delivered a set of standards covering medical experimentation (Capurro and Pingel, 2002). Ongoing concern at the abuse of power in research design, conduct and reporting across disciplines (Hale, 1998) has forced two international trends in the regulation of human subjects research: increasing codification, and the extension of coverage to include behavioural and social research (NHMRC, 1999).

In Australia, the National Health and Medical Research Council (NHMRC) has a statutory obligation under the *National Health and Medical Research Council Act, 1992* (Cwth) to issue guidelines and advice on medical research and ethical matters relating to health. Alongside this function, it:
… requires all institutions or organisations that receive NHMRC funding for research to establish a Human Research Ethics Committee (HREC) and to subject all research involving humans, whether relating to health or not and whether funded by the NHMRC or not, to ethical review by that committee (NHMRC, 1999, p.3).

The NHMRC’s dual function as adviser and major funding body introduces two problems. The ability to deny funding to an institution based on the compliance of projects that are auspiced by the institution but not directly funded by the NHMRC is extremely coercive for a non-regulatory body. The “National Statement on Ethical Conduct in Research Involving Humans” (the National Statement) describes itself as: “…a national reference point for ethical consideration relevant to all research involving humans” (NHMRC, 1999, p.1), but the funding policy forces a shift from reference point to compliance document. Secondly, evaluating non-health related research within a context dominated by medical discourse contributes to the pathologisation and universalisation of culturally-bound evaluative terms, such as: ‘well-being’, ‘welfare’ and ‘harm’.

The National Statement follows international precedents by enshrining three guiding principles of ethical research drawn from the (US) Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). These are: respect for persons; beneficence—the imperative to “maximise possible benefits and minimise possible harms” (NHMRC, 1999, p.4); and justice. The National Statement adds research integrity as a fourth. It augments these teleological principles with rule-deontological prescriptions that govern the operation of ethics committees and act-deontological guidelines for evaluating the proposals they review.

Under the National Statement, the University of Wollongong / Illawarra Area Health Service Human Research Ethics Committee (the Committee) is responsible for reviewing and approving human subjects research conducted within or by the University of Wollongong. The Committee interprets its
catchment as students or staff of the University, or anyone using the University’s facilities or information holdings (University of Wollongong / Illawarra Area Health Service Human Research Ethics Committee, 2003). The Committee’s documentation reinforces the priorities of the National Statement by emphasising: informed consent, privacy and confidentiality, risk assessment, the protection of vulnerable and dependent persons, deception, methodological appropriateness, cultural sensitivity, data storage and reporting as key sites of inspection (UoW HREC, 2003).

The imperative to engage with these matters at design, implementation and evaluation levels is difficult to dispute. What is open to contestation is whether the frameworks that inform human subjects research protocols can systematically deliver the just, beneficent and respectful research outcomes that they invoke.

**Critiques of Human Subjects Protocols**

Critics concerned with decolonising research methods argue that ethical research protocols:

…are framed in ways which contain the Western sense of the individual and individualised property – for example, the right of an individual to give his or her own knowledge, or the right to give informed consent. The social ‘good’ against which ethical standards are determined is based on the same beliefs about the individual and individualised property (Smith, 1999, p.118).

The National Statement raises this point, and acknowledges that: “cultural diversity in Australian society means that there may be a range of views on the relative weight of individual and collective values” (NHMRC, 1999, p.5). Despite this acknowledgement, key definitions and examples support the centrality of individual autonomy to the research process. Separate conditions apply to research that involves Aboriginal people and Torres Strait Islanders (UoW HREC, 2003), but this reinforces Smith’s point: that Western values are universalised and privileged while non-Western values are “othered”. Evaluating research outcomes for Indigenous communities using utilitarian models is
extremely problematic given the unresolvable conflicts of interest between the worldviews and values of colonised and colonising cultures. Designing ethical research requires historically situating the power relationships between researchers, the subject community and the cultures that have colonised it (Smith, 1999). Ethical and methodological protocols designed by Indigenous researchers and communities offer solutions that address power differentials, cultural differences and localised domains that might be usefully adapted to other research contexts (see for example: Martin, 2001; Rigney, 1997; Smith, 1999). Similarly, feminist researchers who analyse economies of power with reference to gender and sexualities have devised pluralist methodologies that historicise and consciously position the researcher and the research domain within the cultural contexts that produce them (Reinharz, 1992).

Postmodern ethicists reject the idea that ethical principles can be dehistoricised and universalised. Teleology must fail, along with all universal paradigms, because: “different subjective positions cannot be measured objectively against each other” (Yuthas and Dillard, 1999). While this position invokes ethical relativism, it does not necessarily lead to value-equivalence. It facilitates a shift of approach towards solving existing problems: using multiple, non-universalising strategies that privilege diversity, dialogue and localised contexts. Bauman argues that the codification of ethics has usurped the responsibility of individuals and non-institutional actors to make ethical decisions. He proposes an ethics of care based on: "being for the Other before one can be with the Other" (Bauman, 1993 in Yuthas and Dillard, 1999, p.4).

This return to decodified responsibility recognises that situated ethical dilemmas have been obscured by the overarching issues that dominate formal guidelines and steer the course of research (Yuthas and Dillard, 1999). Control of research and ethical agendas as an exercise of power has been an ongoing concern for feminist and Indigenous critics (Reinharz, 1992; Smith, 1999).
**Difficulties with the choices available to online researchers**

This section builds on the previous discussions to explore questions that arise from trying to reconcile the conceptualisations and ethical choices available to researchers in online environments. In particular, it questions whether the subjects of online research are necessarily human, and whether the default positioning of human subjects research protocols can compromise online research.

**Are the subjects of the research human?**

It may not be clear-cut whether a given piece of research ‘involves humans’, bringing it under the ambit of the National Statement. For example: does a piece of fictitious text, posted pseudonymously to a bulletin board and able to be accessed without registration, involve a human subject? Does a textual analysis of it require HREC review? The National Statement leaves the necessity for HREC review to be negotiated amongst stakeholders, citing the following possible risks of the research as reasons to require review:

- Physical, psychological, spiritual or emotional harm to participants
- The exploitation of cultural knowledge and/or property
- The potential for the use of personal or community-based information to infringe privacy, confidentiality or ownership rights that attach to that information
- The imposition of burden with little benefit (NHMRC, 1999, p.8).

There are additional factors for Internet researchers working in a global domain to consider. Compared to proximate, localised settings it may be very difficult to predict the degree of risk to participants without having already entered into direct dialogue with them. References to behaviours, political or religious affiliations and sexual preferences protected by Australian law may expose their authors to risks of social ostracism, economic and legal reprisal or violence should their online and off-line identities be correlated (Amnesty International,

---

2 Although as previously argued, NHMRC and Australian Research Council (ARC) funding policies, and their equivalents overseas (Bruckman, 2002), will affect the equity of stakeholders in these negotiations.
2003; Lee, 1996; Sussman, 2001). The rights attached to privacy and ownership—and participants’ understandings of them—vary significantly between cultures, jurisdictions and site genres.

The possible need to make direct contact with people affected by the research at some future time may lead the researcher to follow human subjects protocols as a precaution. Internationally, the nexus between institutional review and funding regimes is similar to that in Australia (Bruckman, 2002). In this context, exercising the precautionary principle leads to human subjects models becoming the default paradigm for online research design. Emerging Internet-specific ethical research guidelines generally start with the presumption that Internet research involves human subjects and should follow existing protocols (White, 2002).

Can the use of human subject models compromise online research?
Human research protocols provide necessary protections for research participants when the subject of the research is behaviour analogous to human behaviour in offline settings (Bassett and O'Riordan, 2002). They do not accommodate the degree to which online culture, sociality and content is mediated by technology and textual play (White, 2002). It is not the intent of this section to position ‘human subjects’ and ‘representations’ as binary opposites and reject the use of human subjects protocols for overseeing online research. Rather, it introduces tensions between human subjects and humanities models, and proposes that their resolution lies in research design and flexible attitudes to research oversight.

Humanities research has a long tradition of engaging critically with constructed texts3 without the consent of their authors or audiences. Research discourses under this paradigm—including, but not limited to: literary, art and music criticism, film and performance studies and some historiographical methodologies—prioritise the need to acknowledge texts as the intellectual property of their authors. Citing electronic sources is standard practice in

3 Throughout this discussion, unless otherwise stated, ‘text’ includes non-orthographic representations.
academic work, and the only concession to protecting the privacy of online authors is the recommendation in some style guides to withhold personal email addresses. These conventions reverse the human subjects models’ privileging of confidentiality. Researchers’ conflicting duties to authors: to seek consent or not; to name or not to name; and the methodological choices that are contingent on these decisions, are partly defined by publication status, and partly by where and how the research itself will be published.

While a human subject ultimately has agency for the production of text (Schirato and Yell, 1996), the relationship between author and text is problematic in humanities research. Models that propose text as a conduit to a unique and autonomous authorial mind best suit the human subjects model, but they have been contested throughout the twentieth century (Bassett and O'Riordan, 2002).

The ‘intentional fallacy’, a component of New Criticism models of literary theory, severs enduring connections between text, author and critic (researcher):

>The poem is not the critic’s own and not the author’s (it is detached from the author at birth and goes about the world beyond his power to intend about it or control it). The poem belongs to the public. It is embodied in language, the peculiar possession of the public, and it is about the human being, an object of public knowledge (Wimsatt and Beardsley, 1946, p.335).

If the moment of the text’s ‘birth’ is the point where the author loses control of it, the act of uploading it to a server –where it is immediately subject to the possibility of non-authorised change– consigns it to the public. Under this model, scrutinising online texts without their authors’ consent is as justifiable as using any other published text as research data.

Contemporary understandings of the relationship between author and text include the audience as an essential component in the production of meaning (Bassett and O'Riordan, 2002). Recursively, text is (partially) responsible for producing
the identity of its author. Textual production is always a politicised, historicised
and contingent act. The reader/researcher is implicated in the act of production,
and has a responsibility to: “…account for the dynamics of the relationships
between researched, researcher and results” (Bassett and O'Riordan, 2002). In
online environments, researchers also have a responsibility to account for the
‘entextualisation’ of these relationships. Any text under consideration is
produced by its intertexts and contexts as much as by human agency. This is
more intensely and immediately foregrounded online than in other media
(Sandbothe, 2000).

Bassett and O’Riordan propose readings that acknowledge the hybridity of
Internet texts as “…neither virtual selves nor objects completely distinct from
those who write them” (Bassett and O'Riordan, 2002). Drawing on Salazar’s
notion of ‘relations of production’, they advocate readings that consider relations
between content, form, history and researchers' response (Bassett and O'Riordan,
2002) with power as a central issue. Privileging the author over other agents of
meaning production, as required by human subjects models, materially affects
the discourse and outcome of the research.

Online, the conflation of ‘human cultural artefact’ with ‘human’ —of context-
specific consigned narratives with embodied selves— can lead to unjust results.
The possibility of inadvertently extending human rights to avatars and
pseudonymous representations is an absurdity that fundamentally threatens the
integrity and reputation of online research (Cavanagh, 1999; White, 2002). Conflation of character with author can disadvantage authors by failing to
acknowledge their creativity and intellectual property rights (Bruckman, 2002).
On a larger scale it can hinder the development of thinking about online
environments and humanness (White, 2002).

The blanket application of human research protocols may hinder research in
some cases. Solutions lie in carefully articulating research design, and ensuring
that ethics committees are familiar with distinctive problems that researchers
encounter in online environments. Rather than restricting the options for
approved research, ethical guidelines need to remain flexible enough to
accommodate changes in technologies and understandings of ‘human’ (Bruckman, 2002). Emerging negotiated guidelines for online research drawn up by practitioners are a useful reference point (see for example Ess and Association of Internet Researchers Ethics Working Committee, 2002).

**Problems identified in empirical research**

The sites available to online researchers vary enormously (Di Maggio et al., 2001; Kendall, 1999) and to produce valid, ethical research, it is desirable that the methodologies deployed match the characteristics of the site to the aims of the research. This section explores recurrent problems identified in published empirical research, and considers researchers’ solutions and suggestions.

**Publication and privacy status**

As noted previously, a number of choices depend on whether the research site is deemed a public place, and whether text posted there is considered published. The inherent insecurity of information shared through online technology is used to justify a default position that Internet content is in the public domain (Ess, 2002; Jones, 1994; Finn and Lavitt in King, 1996; Waskul and Douglass, 1996). The presumption is that participants should know that their data is vulnerable to interception, manipulation and republication. Conversely, it is invoked to confer provisional ‘private’ status to all sites to protect participants, on the basis that their perceptions of privacy are distorted by the medium (Boehlefeld, 1996; King, 1996). Clear delineation between public and private spaces online, while desirable (Frankel and Siang, 1999), is neither technically (Bassett and O'Riordan, 2002; Federal Privacy Commissioner, 2000; Sharf, 1999) nor conceptually (Waskul and Douglass, 1996) possible. Researchers have proposed several criteria to assist in determining the degree of privacy assumed by users of a site, and the research activities that are appropriate to those conditions.

The number of participants in a dialogue and the medium they choose gives some indication of their expectations of privacy (Waskul and Douglass, 1996). A direct email exchange between two people has a higher nominal privacy rating than the same exchange on a listserv (Waskul and Douglass, 1996). Although
the participants and the texts are the same, the potential audience and their implied invitation to participate shifts the status of the site and the researcher’s duty of care. Cavanagh suggests that tolerance of lurkers is an indicator that discussion is taking place in a public arena (Cavanagh, 1999).

King proposes sliding scales of group accessibility and perceived privacy, noting that participants’ perceptions of privacy sometimes conflict with accessibility (King, 1996). Unmoderated Usenet or web-based bulletin boards have the highest group accessibility; closed email groups with enforceable joining requirements and unpublished subscription addresses have the lowest. Perceived privacy depends on the subject matter of a group’s discussions, and the participants’ purposes in posting to it. Where publicity or the dissemination of ideas are desired, as in an academic discussion group or a dedicated ‘coming events’ mailing list, perceived privacy is low. It is highest in groups discussing sensitive or obscure material, for example therapeutic and support groups (King, 1996). Given that ‘sensitive’ is an ideological claim (Bassett and O'Riordan, 2002), this correlation should be regarded as tentative, and the preferences of participants about confidentiality, identification, and publication of their texts clarified. King’s analysis fails to take into account politicised identity projects that privilege open discussion and visibility. To err on the side of protecting participants by obscuring their identity or their concerns risks harming them by reinforcing cultural invisibility and stigmatisation (Bassett and O'Riordan, 2002).

Some sites post explicit privacy and publication rules that participants agree to as access conditions. For example, online writing workshops will sometimes specify that work uploaded to their site is draft-only, so as not to jeopardise deals with potential publishers seeking ‘previously unpublished work’. Posters in any forum can elect to copyright every post, consign them to the public domain as a political strategy (Bassett and O'Riordan, 2002), or attach customised intellectual property licenses to their work (Creative Commons, 1999). Bruckman’s conceptualisation of online texts as ‘semi-published’ work by amateur artists acknowledges that online publication is best understood as a continuum, and that traditional understandings of publication status need to be supplemented with other criteria when making decisions about research design (Bruckman, 2002).
Technical and rhetorical structures of a site—such as passwords, ‘members only’ areas, use of encryption software—can give clear indications of participants’ expectations about where is ‘public’ and where is ‘private’ (Allen, 1996; Boehljefeld, 1996). This is especially useful in heterogeneous sites that replicate some features—for example, chat rooms and bulletin boards—with different access or privacy conditions. In sites where clearly marked ‘public’ and ‘private’ spaces are available to users, their choice of forum indicates intentions and preferences about their audience and access to their texts (Allen, 1996). The Association of Internet Researchers’ Ethics Working Committee suggests that: “…the greater the acknowledged publicity of the venue, the less obligation there may be to protect individual privacy, confidentiality, right to informed consent, etc.” (Ess and Association of Internet Researchers Ethics Working Committee, 2002). Similarly, when a text is explicitly marked as published or in the public domain, obligations shift from protecting the author to protecting the author’s interests.

Informed consent
Many of the uncertainties posed by online environments can be resolved while seeking informed consent from potential research participants. Participants’ preferences about citing, quoting and contextualising themselves and their text can be ascertained during the consent negotiation process, minimising the perceived risk of harm. Some researchers advocate for this process to be ongoing and dialogic, allowing a tailored fit between methodology and site (Allen, 1996; Kendall, 1999; Sharf, 1999).

Negotiating informed consent in online environments presents unique problems. Determining who it is appropriate to ask is the first problem: the active participants in a given research setting are not immediately clear (Jones, 1994). Even when online communities have clearly delineated memberships, not all members are active online (Cavanagh, 1999; Lindlif and Shatzer, 1998). A considerable amount of community activity may occur ‘backchannel’: involving private emails, telephone calls and face-to-face meetings (Kendall, 1999; Lindlif and Shatzer, 1998). Asynchronous discussions may unfold over time,
accommodating changes in subject matter and key participants. Participants may enter into group discussions regularly, intermittently, or once-only, making open access groups in particular very difficult to define. Iterability of posted material means that the number of posts by a participant is not a reliable indicator of their involvement in group discussion. For example, with quoting and citation, a single post can influence a newsgroup thread over an extended period, without further activity by its author.

Having decided who the participants are, and where consent is required, the researcher confronts the logistical problem of contacting them. Traditional consent mechanisms—face-to-face discussion, obtaining a signed authority on paper—are cumbersome and likely to be unworkable in global research settings (Bassett and O'Riordan, 2002; Walther, 2002). Unless active email addresses are required for registration and are regularly updated, participants may be uncontactable via private channels (Bassett and O'Riordan, 2002; Bruckman, 2002). To negotiate consent in a public forum is potentially coercive to individual participants and disruptive to their community. King quotes a discussion group participant who regards being asked about research as an intrusion that inhibits her further participation in the group (Allen, 1996; King, 1996). Uncontactable participants are more likely when researchers include archived posts in their data or seek consent retrospectively. Participants may engage with a person or group using multiple identities (Ess, 2002). Two or more people may share an online identity (Allen, 1996; Dibble, 1998, quoted in White, 2002). There is no guarantee that the embodied research subject shares demographic classifications with the online identity participating in the research site (Frankel and Siang, 1999; Walther, 2002). This is particularly problematic if participants’ age or mental capacity makes them legally unable to give informed consent (Bruckman, 2002; Jones, 1994; Walther, 2002). The difficulty of obtaining informed consent from all participants in a research site increases with numbers, the casualness of their engagement with the site, and their location across political and legal domains.

Whether consent will be given—and maintained—depends to a large degree on how the researcher approaches the participants. Allen worked closely with her
informants’ community, always referring to her role as researcher and inviting questions and input. Aware that even in a dialogic context participants may withdraw their consent at any time, she proposes ‘over-recruiting’ (Allen, 1996). Some researchers run ‘laboratory sites’, setting up online facilities either as a point of initial contact, or as the research site itself (Lindlif and Shatzer, 1998). Consent to research is a condition of entry to these sites, while participants retain the right to withdraw consent at any time. This is not possible on an existing site, where screening by consent status disrupts and fundamentally shifts the dynamics of participation.

Waskul and Douglass suggest the use of ‘key informants’ as an entry strategy, and citing Schrum, recommend that negotiations should begin with the ‘owner’ or moderator of a group (Waskul and Douglass, 1996). Prior to collecting data, Allen met face to face with ‘the highest level system administrator’ of her site to discuss her research proposal (Allen, 1996). The AoIR draft recommendations propose that obtaining permission from the moderator / facilitator / list owner may be sufficient in sites with ambiguous group membership and high turnover rates (Ess and Association of Internet Researchers Ethics Working Committee, 2002).

Harm
Researchers, bound by an obligation to minimise harm (NHMRC, 1999), can harm their subjects directly, harm the research setting, or facilitate harm by third parties.

In the context of disembodied contact, where physical harm is not a risk, non-intentional direct harm by the researcher centres on two issues: breach of trust and expectations, and failure to monitor the consequences of research activities. If protocols for negotiating informed consent are carefully designed, they will result in a mutually understood set of conditions that apply during the research process. Failing to monitor the effects of research activities on the participants is a form of neglect, but the level of monitoring and intervention required is variable. The subject matter and genre of the site, and the degree of intrusiveness of the research are predictors of potential harm. Consideration of the site’s
characteristics will guide the design of feedback and debriefing procedures. For example, discounting idiomatic reactions, participants in an online support group where a high level of trust and respectful exchange is expected are more likely to be affected by research intervention than a group where guitar players swap chord tabulations. Maintaining contact with the research participants and responding to their concerns about the process and effects of the research reduces the risk of direct harm. Kendall proposes that all research on interactive online forums should include participant observation, where the researcher is embedded in and available to the site (Kendall, 1999). Allen questions the prioritisation of harm in the analysis of research ethics, citing her participants’ claims of having benefited from their involvement in constructing ‘ethnographies of the particular’ (Allen, 1996).

Harm by third parties requires an identity –embodied or virtual, singular or collective– to impact on. This in turn requires that the researcher has identified the research site or participants in the course of the research. Generally, quantitative research and analysis techniques, where neither the participants nor their communities are identified, are unlikely to cause harm (Walther, 2002). Concealing participants’ real names is not enough to protect their identities. Screen names tend to function like real names: they accrue reputation and social capital, often remaining constant across sites and containing clues to embodied lives (Bassett and O'Riordan, 2002; Bruckman, 2002; Frankel and Siang, 1999). Using full-text search engines, screen names and direct quotations used in publicly archived forums can be easily traced to source messages, revealing IP addresses that give pointers to geographical location (Bruckman, 2002; King, 1996; Walther, 2002). Researchers have addressed the matter by negotiating naming protocols with their participants (Allen, 1996), using codes (Boehljefeld, 1996), citing and quoting anonymously (Sharf, 1999), and assigning pseudonyms that suggest screen names (Kendall, 1999).

The types of harm that could be caused to research participants by third parties include some that are unique to online environments, such as interference with data and identity theft, including the creation of bogus obligations and liabilities (Cavanagh, 1999; Lee, 1996). Others – harassment, threats, defamation and
blackmail—also occur in offline settings, but take unique forms in the global, software-mediated environment of the Internet. Arguably, better social and technical defence mechanisms against non-physical harassment are available online than offline (Spertus, 1996). Online harassment carries the potential to extend into off-line domains, adding physical harm to the risk list (Working to Halt Online Abuse, 2002). Institutional harms, for example: intervention by law enforcement or medical authorities (Amnesty International, 2003; Ramble, Hermotimus and EverDawn, 2000), are mediated by the scale of the audience, evidentiary traces left in archives and software, and reporting requirements for intranet and ISP administrators.

It is questionable whether online research participants consistently experience a significantly greater risk of third-party harm than other Internet users, although the reporting of research in mass media can be a unpredictable factor. Risk assessments will be more accurate if they are made with specific reference to the aims of the research, the nature of its site and participants, and how data will be analysed and reported. Risk evaluation should consider the relationship between the participants’ cultures of origin and their online activities (Bassett and O'Riordan, 2002; Capurro and Pingel, 2002; Kendall, 1999; Sussman, 2001). Here a distinction needs to be drawn between the degree of risk the participants are prepared to accept for themselves and the degree of risk the researcher is prepared to expose them to, or accept on their behalf. Weighing up these matters requires attention not only to the participants’ preferences but also to the context of their production. Preferences may be informed by disinhibition within a liminoid site (Reid in Bruckman, 2002). Conversely, participants’ preferences may arise from overt political projects that can be harmed by over-protective reporting. For example, the interests of participants who are engaged in publicly asserting the visibility of marginalised sexualities or ethnicities are harmed by obscuring their identities and sites of engagement (Bassett and O'Riordan, 2002; White, 2002).
**Recommendations**

The following recommendations for designing online research are conditional. They are derived from a subset of the theoretical explorations and accounts of empirical research reviewed for this report. These texts (Allen, 1996; Bassett and O'Riordan, 2002; Bruckman, 2002; Lindlif and Shatzer, 1998; Mitra and Cohen, 1999; Sharf, 1999; White, 2002) were chosen because they promote situated, dialogic research that attends to the textuality of the medium, the politics of positioning the researcher, and the contextualisation of the site in relation to social institutions and local off-line conditions.

The imperative to limit, locate and contextualise online research sites is greater than that in analogous off-line environments. It is driven by the lack of certainty and fixity in online environments, the short history of online research, and the rate of change. Online sites can be tenuously anchored to the common reference points, material or historical, that locate similar research in offline domains. By way of illustration, a researcher planning to study a retirement village in rural NSW could infer a lot from geographical location and visual inspection. Immediately, provisional information about prevailing jurisdictions and the cultural context in which the residents are embedded is available. A study of an analogous online community, a ‘virtual retirement village’, can take very little for granted: all information relies on the participants’ testimony, whether directly solicited or gleaned from archives.

To compensate for the narrowed bandwidth of information available to online researchers, triangulation- using several different methods to gather and analyse data and contextualise the site- is recommended.

Preliminary surveys of potential sites are needed to assess the ‘fit’ of the site to the aims of the research and to design appropriate methodologies. Non-intrusive assessment methods should be deployed at this point to avoid multiple ethics committee applications and to avoid disturbance of sites that will not be included in the final research. Textual analysis software may be useful to determine whether a particular site lends itself to further investigation, while Web
‘mapping’ software can expedite the process of contextualising sites by genre, related content, degrees of separation between sites (Mitra and Cohen, 1999) and, if relevant, the nominal location of the participants.

Tailoring methodologies to fit the site, or selecting sites that are compatible with preferred methodologies requires attention to factors discussed in detail earlier. Specifically, researchers will need to consider the genre and subject matter of the site, perceptions of privacy, potential sources and degrees of harm, what legal, ethical and (sub)cultural traditions might inform the expectations of participants, and how the person/text relationship might be understood. Specific research methodologies and areas of inspection are contingent on the aims of the research, the characteristics of the site, and the resources available to the researcher.

Unless it is judged inappropriate to the site or the aims of the research, direct contact with participants is recommended as one of the sources of information. Contact is encouraged even where the primary data source relies on unobtrusive methods (for example, textual or discourse analysis). Where contact is made, dialogue that inscribes research as a partnership between researcher and researched is preferred. It is in the interests of research development that contact is ongoing and includes feedback about research design and the effects of publication (Allen, 1996). Key participants should be identified as early as possible, and their advice sought about site-specific culture and attitudes towards research. Consent requirements will need to be determined at an early stage, starting from a default position that consent from all participants is desirable but often not achievable. Provision should be made for design modifications arising from the denial, inaccessibility or withdrawal of consent.

It is recommended that a corpus of demographic data is collected in all research where there is direct contact with participants. Data that may be relevant to the participants’ discursive strategies- including, but not limited to: gender, age, geographical location, disability, employment status and language use- should be collected systematically. These categorisations will be useful when analysing
observations within the site and will help to position the research in relation to understandings of off-line environments.⁴

‘Laboratory sites’ are a tempting option for inexperienced researchers. Their use allows the researcher to reduce the number of variables, and to control some of the problematic aspects of conducting online research. For example, informed consent to the research can be made a condition of entry, and ambiguities about privacy can be partially resolved by having full control of site design. The use of WebCT at the University of Wollongong lends itself to constructing sites where online data can be contextualised by gathering verifiable information about the participants in local offline settings. Advantages are that existing sites are not disrupted, and close working relationships between researchers and researched are facilitated by proximity and cultural familiarity. Disadvantages are that participation will be implicitly shaped by the use of WebCT in formal educational settings, and explicitly regulated by University of Wollongong codes of conduct. While this latter point may not be significant for most projects, it is relevant to studies concerned with pseudonymity and transgressive behaviour.⁵ A compromise solution is the development of a ‘laboratory site’ for student participants hosted by an external server, perhaps with links back to UoW pages to facilitate access.

The diversity of potential research sites, aims and methodologies makes it inappropriate to make detailed design recommendations and apply them like a template. Rather, it is the task of online researchers to remain responsive to the demands of their unstable sites while (temporarily) situating them. This requires familiarity with the problems, dilemmas and successes identified in prior research and analysis. It also requires awareness that, like the medium itself, online research is changing rapidly: metamorphosing and shifting its emphases, connecting disparate applied research configurations into a cohesive whole.

⁴ Where there is a mix of avatars and legal identities present, the researcher will confront the problem of reconciling the two sets of data.

⁵ My proposed honours thesis is likely to engage at some point with ‘low-level’ transgressive behaviour, for example ‘flaming’ and ‘trolling’.
References


Creative Commons, 1999, "About us" Creative Commons. WWW page. Downloaded 6/2/03 from URL: http://creativecommons.org/learn/aboutus/


Hsia, I, 2000, "Every Step You Take: Online Privacy is an Illusion" Rutgers Law Record. Downloaded 28/12/2002 from URL: http://pegasus.rutgers.edu/~record/drafts/onlineprivacy.html


MacWilliams, M, 2001, "Virtual Pilgrimages on the Internet" WWW page. Downloaded 1/1/2003 from URL: http://it.stlawu.edu/~mmac/virtualpilgrimmacpaper.html


