

The SAGE Handbook of Online Research Methods

The Ethics of Internet Research

Contributors: Rebecca Eynon & Jenny Fry & Ralph Schroeder

Editors: Nigel Fielding & Raymond M. Lee & Grant Blank

Book Title: The SAGE Handbook of Online Research Methods

Chapter Title: "The Ethics of Internet Research"

Pub. Date: 2008

Access Date: October 17, 2013

Publishing Company: SAGE Publications, Ltd

City: London

Print ISBN: 9781412922937

Online ISBN: 9780857020055

DOI: <http://dx.doi.org/10.4135/9780857020055.n2>

Print pages: 22-42

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

<http://dx.doi.org/10.4135/9780857020055.n2>

The Ethics of Internet Research

Rebecca Eynon Jenny Fry Ralph Schroeder

[p. 22 ↓]

[p. 23 ↓]

Abstract

The ethics, and ethical governance, of online research have been much debated, and a number of professional organizations have promulgated guidelines for researchers considering conducting their research online. This chapter offers an overview of the current position, suggesting relevant considerations in respect of different kinds of project, and highlighting some of the challenges and dilemmas that online researchers face. The chapter is oriented to three principal approaches to gathering Internet-based data: use of online methods to gather data directly from individuals, analyzing online interaction within virtual environments, and large-scale analysis of online domains. Amongst issues covered are those relating to data protection and regulation, data intrusion, and issues raised by norms of privacy and to what extent information on the Internet can be regarded as intentionally 'public'. The discussion is grounded in the ethical regulatory framework developed in offline research, noting commonalities and differences.

This chapter will consider some of the main ethical issues that researchers are likely to encounter in Internet-related research. These issues have been discussed now for some time, and some guidelines for researchers have been established (Ess and the AoIR ethics working group, 2002). Yet there is still considerable debate about the ethics of Internet research – not least because the Internet is still in a formative phase and new phenomena continue to emerge. In this chapter, we cannot hope to cover all of the issues in research ethics; that would require a whole book by now. Instead, we will discuss some of the major issues that have been debated – as well as some that have

only recently come to the fore – and give some indication of how to go about addressing them.

Ethics and research methods are closely interrelated. One of the challenges to developing a coherent approach to ethical dilemmas in Internet research is that as the Internet evolves as a space for social interaction and information dissemination, the methods necessary to capture and document such activities are also emergent and novel. Consequently, consideration of ethical issues in a context-independent manner, divorced from matters of research design, methods, [p. 24 ↓] and conceptual frameworks, would be limited in scope and usefulness. In this chapter, therefore, we discuss novel ethical dilemmas for Internet researchers in the context of three predominant approaches to gathering Internet-based data: use of online methods to gather data directly from individuals, analyzing online interaction within virtual environments, and large-scale analysis of online domains.

The online methods used to gather data directly from individuals that are discussed in this chapter include surveys, interviews, and focus groups. In these cases researchers use online tools to ask participants for responses to particular questions or issues. The study of online interaction in virtual environments includes various research methods, such as participant observation and logging and visualizing the interaction between participants. Large-scale analysis of online domains is still quite novel and involves capture and analysis of digital traces that people leave online, such as patterns of their search behavior, text analysis of e-mail corpora, and hyperlinks. The techniques employed in large-scale analyses rely heavily on indirect observation, with the data being de-contextualized from its sources and the analysis often combined with powerful visualization tools.

NEW TECHNOLOGY, OLD AND NEW ETHICS

Ethical governance in traditional research settings

Before going into the specific issues relating to the three main approaches that we have identified, it is useful to take a step back and reflect on how ethical issues relating to Internet research might differ from research in traditional settings. Many ethical guidelines and regulations are well established in (offline) social research. Reviewing these provisions here is important given that, thus far, the governance of Internet research has been heavily influenced by them (Basset and O'Riordan, 2002).

Professional bodies such as the Australian Research Council, the Economic and Social Research Council (ESRC), and the Social Research Association (SRA) in the UK have been involved in the development of ethical guidelines. There are also committees set up that are responsible for governing research on an institutional level. The names used for these groups vary; common terms include ethical review committees/ethics committees (UK), Institutional Review Boards (US), or Human Subjects Review Boards (AU). Both these mechanisms for external research governance (e.g. beyond that of the individual researcher or research group) have historical roots in the 'human subjects research model'. Three ethical concepts are at the core of institutional and professional research governance based on the 'human subjects model': confidentiality, anonymity, and informed consent. These are derived from the basic human right to privacy, though these rights are interpreted differently in different jurisdictions (for example, the EU and the US; see Reidenberg, 2000).

The beginnings of formal guidelines for the human subjects model can be traced back to the Nuremberg Code in 1949 in response to 'research' conducted by Nazi doctors. Yet while there were initiatives highlighting ethical principles to researchers (e.g. the National Institute of Health Clinical Guidelines in 1961 and the Helsinki Declaration by

the World Medical Association, developed for the medical community in June 1969), it was not until the 1970s, after Henry Beecher's critique of the treatment of human subjects by researchers after World War II (Beecher, 1966), and the controversy of the Tuskegee syphilis study, that more formalized guidelines were published. In 1974 such guidelines were published by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.¹ The regulations stipulated that Institutional Review Boards (IRBs) were required to determine, in each research proposal reviewed, that the potential risks to subjects were outweighed by the benefits, the rights of subjects were protected [p. 25 ↓] adequately, and informed consent would be appropriately obtained (Reilly, 1998: 682683). In 1991 a uniform set of regulations, the Federal Policy for the Protection of Human Subjects, was adopted more widely; and today's IRBs are governed by these regulations. However, it is important to note that these regulations come from the medical sciences and are therefore not necessarily cognate with the social sciences. This human subjects research model is widely used in discussions of online research ethics, although the suitability of applying this model in some online contexts has been questioned (Basset and O'Riordan, 2002).

Institutional governance of research (the interrelationship between legal and ethical interventions) not only varies between institutions, but also from country to country. These range from close intervention, which in extreme cases can hinder the progress of research, to minimal guidance which relies on the self-policing of researchers – as is the situation in the Netherlands, for example. Differentiation in institutional/professional ethical rules and guidelines illustrates a tension between external (structural) governance and the freedom of self-regulation among individual researchers. Such institutional provisions do not necessarily exempt researchers from further ethical obligations and responsibilities.

Researchers have different relations with research participants and data provided by them, depending on the method and approaches they use. For example, it is not uncommon for ethnographers to develop a trust relationship with the people that inhabit the communities they observe, and they often come to perceive themselves as custodians of the data they gather. Ethical practices are also shaped by personal ethical frameworks, as well as researchers' cultural and professional ones. As Ess (2006) argues, any emphases on the rights of research participants must be considered further

alongside other important rights and values – including (deontological) emphases on the importance of knowledge developed through research, and (more utilitarian) emphases on research knowledge as contributing to public policy and debate, along with researchers' rights and interests in pursuing knowledge. This is a recurring issue in relation to the ethics of Internet research and closely related to considerations of 'harm' to research participants (see Ess, 2006 for a more detailed discussion of *Utilitarian* and *Deontological* frameworks).

As the above discussion of research governance illustrates, there is a blurring of the boundary between ethical and legal considerations and provisions. Ess makes a useful distinction (2002: 5) between institutional or legal requirements as against the ethical requirements that can go beyond these. In addition to the requirements set by Research Ethics Committees and professional bodies, as already mentioned, there are also laws regarding privacy and data protection that govern research in different countries. In Internet research, however, the institutional and legal context may be uncertain because research participants may be online in any geographical context. The global reach of the Internet may thus, as Ess (2006) suggests, entail that researchers take heed of contexts which go beyond their own jurisdictions. This also applies to considerations over and above these institutional and legal requirements, such as what we might do as individual researchers out of a sense of fairness. Here, as well, it is necessary to think 'globally,' as values such as privacy may be culturally specific and what is considered an appropriate balance between privacy and freedom of expression will vary between cultures (Fry, 2006).

New ethics for new settings?

Why should online research require separate or additional treatment? Indeed, this 'meta-issue' has itself been a major debate which runs through the various individual topics in Internet research ethics. Walther (2002), for example, has argued that many of the features of Internet research are similar to those found in other media or in existing offline research. Walther's arguments are directed against those, and Frankel and Siang (1999)[p. 26 ↓] in particular, who argue the opposite; namely, that new

rules are required for this novel setting because, to give just one example, people may misrepresent their identity online (to which Walther replies, among other points, that they can also do this offline).

Despite continuing disagreements in this debate, Ess argues (2002) that there has recently been a convergence on the view that research ethics for online settings are not special and can be derived from the ethics for offline settings. We shall encounter a number of instances below. At the same time, we shall also argue that in some cases there are special considerations that are needed for online research, such as the changed nature of disclosure and informed consent. The reason for the debate, however, deserves to be spelled out and has to do with the role of the Internet in society.

It is true to say that Internet technology is increasingly becoming 'domesticated' (Silverstone et al., 1992) or part of our everyday lives. In the early days of the Internet there were many concerns raised about, for example, identity fraud, or the authenticity of online relationships, and the like (see,

for example, Baym, 2006; Wallace, 1999).

Nowadays these – what can with hindsight be seen as 'moral panics' – have waned, partly because the technology has become more commonplace and many concerns have been addressed in various ways. Yet this should not lead us to overlook what is genuinely new with this technology and its implications for novel forms of identity, community, and interaction online. As Ess (2006) argues, there may be different requirements in research when dealing with online identities from the traditional human subjects model in research.

Sensitivity to context

Apart from the fact that online research must be sensitive to the offline context in which it is taking place, online research must also be sensitive to different online contexts, since the Internet is many things to many people. The AoIR guidelines place an emphasis on this context-dependence, which entails respecting people's values or

expectations in different settings. A few examples (in addition to those provided in the AoIR guidelines) will suffice:

Internet research ethics thus need to be tailored to different contexts. It may not be sufficient, for example, to stay within the strictures of copyright law (the institutional and legal requirements mentioned in the previous section) or to simply adopt the rule of 'fair use' as with offline publications (Walther, 2002; but see the discussion by Ess, 2002:3) in order to be fair in an ethical sense to research participants.

APPROACHES TO INTERNET RESEARCH

The Internet has emerged as a major data resource for social science research. Not only is it a lens through which to observe our subjects of research and how they construct their identities and communities online, but it can also be a tool for gathering and analyzing social science data on a large scale. The uptake of Internet research for quantitative studies in disciplines as diverse as information science, political science, and geography, together with its deployment in the emergent area of e-Social Science, means that the Internet is increasingly taking the form of a laboratory for the social sciences, in much the same way as astronomers might use a virtual observatory to understand the mysteries of the night sky. An issue that unites these quite different approaches is the question of what constitutes a private act on the Internet and how researchers might deal with the issue of 'privacy in public.' Nissenbaum (1998) has highlighted that there is a lack of guidance for social science researchers in terms of dealing with this issue. Consequently, confidentiality, anonymity, disclosure, and informed consent, concepts at the core of ethical governance in the social sciences, are cast into uncertainty when it comes to research online.

The following sections deal with the particular ethical issues that arise in using online methods to gather data directly from individuals, analyzing online interaction within virtual environments, and large-scale analysis of the online domain. These methods need not be seen as mutually exclusive. Indeed, a researcher studying an online discussion forum may wish to interview individual participants, either online or offline,

as part of a virtual ethnography. Conversely, they may wish to combine observation embedded within a particular setting with a large-scale mapping exercise of hyperlinks to understand the relative social position of a particular forum or person in cyberspace.

Use of online methods to gather data directly from individuals

Whilst the Internet might be considered to afford a rich resource for observation of research participants, covertly or otherwise, using traditional social science methods such as interviews and surveys in the online environment has recently become very popular. Online versions of these traditional social science methods raise slightly different ethical challenges to the face-to-face context (Mann, 2003). For a detailed discussion of how to conduct interviews and surveys see the chapter on interviews by O'Connor et al., and the Internet survey section in this Handbook. We next address the key ethical considerations of these online methods.

Benefits and risks of online research

It is primarily the investigators' responsibility to ensure, as far as they are able, that participants will not come to harm by taking part in any study. In the social sciences, psychological and physical harm to participants may be caused, for example, by research that evokes bad memories or reduces a person's sense of pride or dignity, or by cases where the anonymity of the participant is not maintained as originally agreed (Bier et al., 1996). Trying to ensure harm is not caused by the study is particularly challenging as there may well be unintended consequences of research unforeseen by the researcher. As Rees states, 'The ethical problems which will be encountered in a project cannot, or certainly cannot always, be foreseen and prepared for at the start, even if some topics and methods can be seen in advance to carry greater risks than others'(1991: 147).

Regardless of whether research is online or offline a balance must be struck between the potential and significance of harm to the participant and to the benefits of the

research to the individual and society more generally. Online research is not intrinsically more likely to be harmful than face-to-face methods, yet it does pose different challenges (Kraut et al., 2004). In online research it is more difficult to assess the risk of participants coming to harm, as fewer studies have been conducted that researchers can learn from; and it is harder to judge individuals' reactions to the research (e.g. if a person is getting distressed by an interview question or if a participant feels insulted or harassed by other group members in an online discussion) (Bier et al., 1996; Mann and Stewart, 2000). Strategies to try to address these issues include building a good rapport with participants, establishing 'netiquette' in group discussions (Mann and Stewart, 2000), and providing participants with an easy way to leave the study (Hewson et al., 2003; Nosek et al., 2002).

[p. 28 ↓] A second issue is the potential of harm to the researchers. Given the anonymity of the Internet, researchers can come across or receive distressing information of numerous kinds. What a researcher does with such information has ethical, and in some cases legal, implications for the researcher (see Stern, 2003). It is important for researchers to anticipate and assess these risks prior to beginning the study as far as is possible, to reduce the potential of harm to themselves and their research participants.

Ensuring confidentiality

Harm can also be caused due to breaches of confidentiality and anonymity caused by the misuse of storing or using the data (Fox et al., 2003). Researchers have a responsibility to ensure the confidentiality of data and the privacy of participants at all stages of the process: during all interactions with the participants and when the data is transmitted and stored (Nosek et al., 2002). In general, the extent to which a researcher should be concerned about confidentiality depends on the nature of the data being collected. If the data is not controversial, or if anonymity can be ensured, then this is less of a concern compared to controversial research topics or research where it is necessary to obtain personal information (Kraut et al., 2004). Given that the perceived anonymity of the Internet may encourage people to discuss topics or disclose more details than they would be willing to in face-to-face situations (Meho, 2006), researchers

need to ensure that participants' perceptions of anonymity are met, or if not, made explicit to the participant (see section on informed consent below).

In terms of securely transmitting data, potential solutions include: the use of encryption and secure socket layer (SSL) protocols, use of data labels that are meaningless to anyone but the researcher, and the separate transmission of personal data and experimental data (Nosek et al., 2002). In terms of data storage, the data needs to be protected from other people accessing it or tampering with it; this can be an issue in the networked systems commonly in place in universities (Fox et al., 2003). Password-protecting computer directories, saving personal data and experimental data separately (Kraut et al., 2004), encrypting the files so no one else can read them, or coding the data in a way that reduces the likelihood of people being able to trace the data to a specific individual (Pittenger, 2003), are all possible strategies.

The issue of ensuring confidentiality whilst interacting with the participants may arise at various points throughout the research. Participants may wish to contact the researchers up to and including the debriefing stage at the end of the study, yet directly emailing the researchers may compromise anonymity in a number of ways. Firstly, e-mail addresses are often identifiable as they can contain names, geographical location, and organizational affiliation. While people can make use of anonymous e-mail services to cover their identity, these are not 100 percent effective; and tend to promise 'best efforts' as opposed to true anonymity. Secondly, a copy of all e-mails is retained on the server of the sending account, any transmitting server and on the destination server and these copies are frequently retained on back up tapes for a number of years. These issues can be particularly problematic when certain activities are carried out online. For example, if verifiable names and addresses or signed agreements are required to fulfill informed consent procedures (see section below) and/or if participants are rewarded for the research in the form of prizes or payment and personal details are required for tax purposes. E-mail should be reduced to a minimum with offline methods or alternative web-based methods utilized where appropriate. For example, setting up a discussion thread on the research website or other appropriate site for participants to ask questions (Fox et al., 2003), and, when offering prizes for participating in the research (a technique that in itself raises ethical questions), maintaining anonymity by purchasing online gift certificates and then providing the certificate number to the participant (Kraut et al.,

2004).

[p. 29 ↓]

Informed consent

Individuals who choose to participate in any research project must do so on the basis of informed consent, where the individual understands what the goal of the research is and what they are agreeing to do, the potential risks (e.g. limits to confidentiality) and benefits of taking part (e.g. incentives), and have details of alternative options that may benefit them. Participants must have the option to ask anything they wish and understand that participation is voluntary and that they can withdraw at any time (Anderson, 1998). In practice, gaining truly informed consent is not straightforward in any context. The nature of informed consent changes throughout the research process and thus needs to be constantly renegotiated (e.g. Bier et al., 1996; Sin, 2005). Further, it is difficult to ascertain if informed consent is truly given by the participant; for example, problems and misunderstandings arise when potential participants do not read documents carefully or fail to ask for clarification from the researcher (see Varnhagen et al., 2005).

In face-to-face contexts it is potentially easier to ensure that the participant is fully informed about the study compared to online environments. For example, the researcher can discuss the research with the participant, assess whether the individual fully understands the implications of the research and evaluate whether they are freely entering into the study. Owing to the distance between the researcher and the participant in online settings, this is more difficult. It is harder to determine whether the participant truly understands what they are consenting to, and it may take more time to gain consent, as it may require more discussions via e-mail to ensure the participants fully understand the implications of participating. This e-mail exchange may put participants off clarifying or asking all the questions they wish about the research (Mann and Stewart, 2000). To try and ensure participants are truly informed in online settings, techniques of increasing the readability of the document can be used (e.g. reducing the amount of text, use of subheadings, and use of colour).

Using quizzes to check understanding can be another means; though this extra burden on the participants increases the risk of dropout (Varnhagen et al., 2005). Despite these challenges the advantage of online consent, as compared to face-to-face consent, is that participants are likely to feel less pressure to enter into and remain in the study and are therefore more likely to enter and participate in the research freely.

A second important issue is verifying the participant's ability to give informed consent (Kraut et al., 2004). Verifying the ability of an individual to give informed consent is harder in online environments, as it is more difficult to know whether or not the online sample includes 'vulnerable groups' (e.g. young people,² the elderly, or people with mental health issues), and because the extent to which individuals are able or competent to give informed consent varies widely and this is more difficult to judge online. To reduce the chances of a vulnerable group (e.g. young people) being part of a research project this can be addressed, to some extent, by the recruitment strategy utilized. For example, sending specific invitations to known adult participants to access a password-controlled site (Pittenger, 2003), or designing advertising materials that are unlikely to attract or interest young people when employing a more 'broad brush' strategy (Nosek et al., 2002) may help. Other options include asking for information that only adults would have, such as credit-card information, though such activities can increase dropout (Kraut et al., 2004). In practice, verifying identity is really an issue only in research involving controversial topics and/or where the study presents higher risks to potential participants (Pittenger, 2003). Indeed, whether one should try and obtain online consent for high-risk studies at all is open to question (Kraut et al., 2004).

The issues considered in this section have included protecting participants from harm, ensuring confidentiality, and informed consent. These are all areas that also need to be addressed in face-to-face environments; but the use of the Internet poses slightly different challenges to these more traditional contexts.³

[p. 30 ↓] It is a balancing act for researchers to ensure that participants are protected, but at the same time not placing unnecessary and excessive burdens on participants in terms of completing informed consent procedures, ensuring security, etc. (Kraut et al., 2004). While it is impossible to predict all eventualities in online research, it is useful to always pilot test instruments and consent forms, as what works in one context with

one group may not work and/or may well produce different ethical questions in another situation (Meho, 2006). The next section explores the ethical considerations when analyzing interaction in virtual environments.

Analyzing interaction in virtual environments

To address the many questions raised by online environments, we can take as an example online virtual worlds. These include social spaces where people, in the form of avatar representations of themselves, interact with each other online in the virtual setting for various purposes (see the chapter by Schroeder and Bailenson in this Handbook). These purposes include gaming, socializing, and collaborating. The best known are Massively Multiplayer Online Games (MMOGs, sometimes also known as Role-playing Games or MMORPGs) where, apart from socializing, people typically engage in elaborate rule-following interactions to achieve ever higher levels in the game. All of these online spaces include visual representations of people (again, avatars), most use text communication (and also sometimes voice), and they have sophisticated social conventions and online social institutions of their own (Schroeder, 2007). [Note: Online virtual environments also include purely text-based online environments such as MUDs and MOOs. The focus in this section is on graphical environments.]

Online and offline

For online virtual worlds, one question is: under what circumstances can – and should – the researcher contact the subject outside the virtual setting? In this case it is useful to ask why such a move from studying the subject online to offline is deemed to be necessary. There can be three main reasons for this. One reason is to validate the information that has been obtained online; the second is to embed the subject's online behavior in the context of their real-world social setting; and the third is to obtain more in-depth knowledge of the subject, such as their motivations or the significance they attach to certain events. As to the first of these, the question can then be raised of

why it should be necessary to think that face-to-face questioning should provide more valid answers (as Walther 2002 points out). In the second case, it may be awkward to engage with people offline if one has only previously dealt with them online (for example, Taylor, 2006: 1-19). This also applies to the third case, although it may be, for example, that the subject can be contacted without co-presence, by means such as the telephone or e-mail. One dilemma in all these cases is similar to contacting people offline who have previously been encountered or contacted by other means: what kind of burden may be imposed on the research subject? Some of the issues arising in this case have been dealt with in the previous section.

Online social settings

Virtual spaces in which people interact online as avatars can be treated as social worlds, regardless of whether they are primarily used for socializing, collaborating, or gaming. To appreciate this point, we can leave to one side for the moment the offline 'frame' of this interaction, such as the legal issues attendant upon the use of the software and other legal rules which govern use of the Internet, as well as the other offline consequences of online behavior such as 'addiction.' We can also leave to one side the question of whether these online environments constitute 'communities' or not – in online virtual environments people who regularly interact as avatars clearly see themselves as social groups just as people do in any other social setting with regular and bounded interaction. These social settings are perhaps most often akin to 'third places' [p. 31 ↓] (public parks, coffee shops, street corners, and the like): places that are neither public nor private but in between, as in Oldenburg's (1989) work.

Against this background, online social spaces clearly exemplify the imperative discussed earlier: to be sensitive to the values and aims of people in different online settings. Again, this sensitivity to context will involve treating different virtual worlds in different ways – for example, whether they are small private encounters or the movements of large-scale populations. Or again, there may be occasions or whole worlds in which people interacting online are behaving in a public way; as in a public meeting or in a virtual world that is open to all for, say, commercial or educational purposes. It may also be, however, that certain spaces within a virtual world, such as an online church (Schroeder et al., 1998), although formally public, include interactions that

should be treated as private – such as when personal details are revealed, or if a whole online world is expressly designed to provide a private forum for interaction among a group that would be difficult in an offline setting (or in another virtual setting).

Research ethics then requires treating online interactions in virtual worlds with the same sensitivity that other, offline social settings are treated. For participant observation or fieldwork there have been extensive debates in anthropology about the role of the observer – and these will provide some guidance. And in special cases, it may also be that online virtual worlds need to be treated as sensitive fieldwork, for example, where vulnerable groups such as children are involved (Lee, 1993).

The role of the observer

One issue that arises online is disclosing your identity as a researcher. There is a balance to be struck between revealing that you are a researcher and engaging in unobtrusive observation. There is also a difference here from real-world observation, since it is easy to hide completely – or lurk – in the online world. Note that in online virtual environments there is furthermore a technical possibility that is not available in offline research: namely, to attach an identifier tag to your avatar that identifies you as a researcher and provides details of the research project – but only when others *choose* to click on this information.

Clearly, it will often be good practice – beyond legal requirements – to identify yourself as a researcher in the case of ethnographic or participant observation in a virtual environment. There may be a tradeoff in this case between the advantages of covert observation which does not disturb the environment, and revealing one's identity as a researcher – which ensures transparency, but may also lead to changed behavior on the part of the subjects (for a particularly striking example, where the researcher became 'stalked' among other things, see HudsonSmith, 2002). Anecdotally there have been a number of cases when many researchers descended on an online environment and there was resentment against their presence. The well-established rule in anthropology – to leave the field so that future researchers are not disadvantaged – must be an important consideration.

Studies of online populations

Another set of issues revolves around gathering powerful data gained from the surveillance of online populations. Even if online virtual worlds are *prima facie* public spaces, it is nevertheless important to be sensitive to the social context – just as public spaces in the physical world need to be treated as such. Yet there is also a difference between the virtual and the physical: in online worlds, whole scenes or even worlds can be recorded and later reproduced for research purposes. In the physical world too, of course, people can be covertly recorded (as with closed-circuit television cameras), but in online worlds the possibilities of recording, reproducing, and analyzing interactions, especially covertly, are more powerful.

In quantitative, anonymized studies of online environments for gaming, collaborating, and socializing, just as in the offline world, population data for the most part do not raise ethical concerns. Note an important [p. 32 ↓] difference, however: namely, that in an online virtual world *all* the interactions between avatars can be captured. In other words, this is the equivalent of being constantly under surveillance, as in George Orwell's novel *1984* or in 'reality' television programmes, where participants are constantly under the gaze of the camera. This raises novel ethical issues, since people using these environments do not necessarily expect to have all their behavior recorded.

To give an example: Penumarthy and Borner (2006) analyzed where people moved and when they focused their attention in an online virtual world for education. This kind of recording of behavior is unlikely to be objectionable. If, however, they had counted the number of times that avatars had engaged in particularly unsavory behavior, even in a public place, users might reasonably object to this kind of surveillance. There is a fine line then between when data, for example, about a large online game is aggregated to reveal patterns about behavior without violating participants' sense that they are under surveillance – and the opposite. The same applies, of course, to the analysis of small-scale groups, which can be analyzed down to the granularity of the finest details of interaction (Schroeder et al., 2006).

However, online worlds also have novel possibilities for presenting research findings in anonymous ways. It is possible, for example, to blur the names of avatars which are

next to their text bubbles (see, for an example, Brown and Bell, 2006: 229). Or again, it is possible to record and reproduce the gestures of an avatar and yet anonymize their identity by changing parts of their appearance from which they could be identified. Researchers thus also have a range of choices that differ from those in the physical world.

The wider context is that even if the online world is formally a public space, researchers will nevertheless want to maintain the trust of those whose online behavior they are studying. This includes not disregarding the sense of privacy that, for example, people's avatar representations may have in the settings of particular online worlds – even if this may involve guessing what the intentions and values of the person ‘behind’ the avatar might be.

It is not so much, then, that the boundary between online and offline should be abandoned, as some have argued (Taylor, 2006: 153). Rather, here, as in relation to the other issues discussed in this section, researchers will need to weigh the same ethical considerations as they always do in dealing with human subjects, and to adapt them to the novel technological possibilities and constraints of online virtual worlds. They will continue to face the choice mentioned earlier: between Kantian duty-based or ‘deontological’ ethics, with their absolute respect for the individual's aims, as against the calculation of consequentialist or utilitarian ethics, which weighs the balance of harms and benefits. In the case of online worlds, can researchers violate the privacy of people's online behavior? Do the benefits of disseminating research results about online populations outweigh the harms of disclosure? Should researchers seek consent from those they observe and report on?

The uses and limits of virtual environments for experimental research

It can be mentioned, finally, that despite the various tricky ethical issues that have been discussed, online virtual worlds – as Schroeder and Bailenson point out in their contribution to the Handbook – offer many novel opportunities for social research that are not available in offline settings. They include, as we have seen, some possibilities

that are not available in online research that gathers data directly from individuals. Yet at the same time, there are limits to using virtual environments to do research that cannot be done in face-to-face or physical settings. These limits have recently been highlighted in a replay of the Milgram experiments in an immersive, Cave-type environment by Slater and colleagues (Slater et al., 2006).

What Slater and colleagues did was to investigate the responses of participants to inflicting pain on a virtual character (an avatar) and to see how far they would go in administering 'painful shocks' to this character. This was a reconstruction of the experiments conducted in the 1960s by Stanley Milgram, who was interested in the extent to which people obey others in authority in inflicting pain or suffering on others. Part of the background to these experiments was the Nazi atrocities during the Second World War. The experiments involved research subjects administering ever greater electric shocks to another person, strapped in a chair in an experimental setting, when the experimenter told them to do so. Milgram found that participants were willing to continue administering shocks even when the person strapped in the chair was screaming with pain. What they did not know, of course, was that this person was in fact an actor and no shocks were being given.

Milgram's experiments have become regarded as controversial in terms of research ethics in a number of respects: one was that subjects were being deceived. A second was the controversial nature of the findings, perhaps most of all because they revealed some unpleasant facts about human nature. Third, they were regarded as inflicting unnecessary mental anguish on the research participants. The experiment by Slater and colleagues avoided these shortfalls in several respects: one is that research participants were not deceived; the second that the experiment was not about obedience, but rather about 'presence' and the responses to virtual characters; and finally that the person to whom the shocks were administered was a virtual human and the experiment was carried out in a Cave-type virtual environment. Participants therefore knew that the shocks they were giving could not hurt the virtual human.

Slater and colleagues found similar results to Milgram in terms of the stress levels of participants, though with 'lesser intensity' of the responses – an important replication, since doing this experiment with 'real' recipients of pain is no longer acceptable (2006:

e39). Participants felt increasingly uncomfortable about the experiment and responded to the virtual character as if it were highly real.

In fact, some participants stopped administering the 'painful shocks' and approximately half the participants said afterwards that they had wanted to stop the experiment.

The results shed important light on presence, since subjects clearly thought that the virtual human's pain made them respond as if they were in the presence of a real person. These are important results for the community of virtual environments [p. 34 ↓] researchers who study presence, as well as for other researchers. Yet they also generated considerable debate about how far this type of research should go. No doubt virtual environments provide many excellent opportunities for doing experiments that, for various reasons, cannot be carried out in the physical world and in face-to-face interaction. But there is also a range here, some such situations are clearly acceptable: for example, Slater and colleagues have also conducted experiments about the fear of public speaking in front of a virtual audience (Slater and Steed, 2002: 164-8). This is a good illustration of beneficial research, since it may cause some participants considerable anxiety during the experiment but may also help them to overcome or alleviate this anxiety in real-world circumstances. Yet some such situations that one can think of are equally clearly unacceptable; for example, a research participant brutally killing a virtual human merely to see how far they will go.

Slater et al.'s 'virtual Milgram' thus points to some limits of this type of research: some extreme social situations should be studied, because the distress to participants is not great and the value of the experiment *is*. Some extreme social situations should *not* be studied, because the distress to participants is too great and is not justified by the benefits of the study. So even if virtual humans are not real, this does not mean that interacting with them cannot cause undue distress to the real participants interacting with them – even when they know that this cannot be the case. Put briefly: just because it's virtual does not mean that any type of research can be done with human participants, and even if the virtual humans are not real, that does not mean that 'anything goes.' Virtual environments are useful because certain experiments can be done that cannot be done in the physical world and with people interacting face to face. Others cannot be done because they will, for example, be too realistic. Where to draw the line will be an ongoing debate in the years to come.

Large-scale analysis of online domains

The trend towards exploiting the Internet as a social science laboratory has been intrinsically tied to the increasing sophistication and stability of Internet technologies, such as search engines and archives. The capacity of these tools to record traces of social interaction on a global scale – of how individuals and communities consume and contribute to the Internet through automated code, such as Internet cookies (enabling the unique identification of browsers and users' hypertext pathways) and server log files (recording search terms, date and times of requests) and their application within social science research – raises a unique set of ethical issues. Recent developments in Internet-based analytic tools and resources have increased the ability of social scientists to delve deeply into the structure of online social worlds. Projects such as VOSON (Virtual Observatory for the Study of Online Networks) are combining traces of social interaction as captured by Internet archives with visualization tools to render social structures visible in ways not previously possible. At the same time, developments in e-Social Science (some of which are described in this volume) are using the Internet to combine data and tools in novel ways, leading to new tools, techniques, and digital records that incorporate multiple forms of data from across a diverse range of data sources. As Thelwall and Stuart (2006) point out, the AoIR ethical guidelines focus mainly on issues related to observational research in the analysis of interaction in online environments, and do not address issues of automatic data collection and large-scale analysis of online domains. This section does not attempt to produce a set of guidelines for these emergent novel approaches; rather, it lays out the territory in terms of potential issues and the multiple factors that may impact on Internet research more broadly conceived than in terms of single projects.

Triangulation of datasets and third-party reuse

Advancements in the development of resources and tools available on the Internet make the triangulation and third-party reuse of data much more likely. While a

standalone dataset may preserve anonymity and privacy, new capabilities for aggregating and combining data could jeopardize such ethical integrity by enabling profiles of individuals to be constructed through triangulation. The following case-scenario taken from McKee and Porter (in press) illustrates the way in which ethical issues related to the direct gathering of data from individuals, the analysis of interaction in online virtual environments, and the large-scale analysis of online domains might converge in a single study, regardless of which of these is the primary method:

A researcher decides to conduct a critical discourse analysis of messages posted in discussion forums geared towards teenagers. Because the discussion forums are publicly available to anyone with access to the Internet, he decides that he does not need to seek permission from the individuals to research and use their online posts. He does, however, use pseudonyms when he refers to and quotes from their posts in his work. When he publishes his research in a print-based, peer-reviewed journal, he includes many direct quotations that, when entered into a search engine, could immediately provide the URL to individuals' posts and thus explicitly reveal their identity (McKee and Porter, in press).

The tracking capabilities built into the very infrastructure of the Internet itself, and tools being developed to exploit the gathering and aggregation of fine-grained data on a large scale, mean that the role of researcher as custodian and gatekeeper of personal data becomes radically altered. Tools that enable data to be easily reused by third parties and recontextualized in novel ways undermine the notion of 'context', for example, the [p. 35 ↓] norms, values, and beliefs of groups within online social settings (see earlier section), as a heuristic for developing ethical practices that are socially and culturally appropriate. Reuse and the emergent practice of data profiling by third parties reduces choice for both researcher and research participant in terms of how data is represented and how it travels through media and across actors. The researcher, therefore, may no longer be able to foresee all of the consequences and potential harm of their research, which has implications for 'informed consent' where it is deemed appropriate in large-scale studies. This distancing of data from its context of creation and creator serves to make the

issue of what constitutes a private act online, and whether the 'human subjects model' is the most appropriate one for Internet research, all the more pressing. Bassett and O'Riordan (2002) argue that choice about privacy embodied in the use of spatial metaphors in online social settings, such as 'going to a chat room,' has evoked the 'human subjects model' on the Internet, but that the invocation of such language does not necessarily indicate its appropriateness. Rather, they argue, a humanities model of research governance might be more appropriate, whereby texts produced in a chatroom are treated as creative works and thus protected under intellectual property rights legislation, such as Copyright, rather than as representing human beings that need protecting under social science ethical governance, such as informed consent.

Furthermore, the vast quantities of social science data being generated by the Internet are of significant commercial value. Consequently, social science data generated and used by academic researchers may travel beyond the professional boundaries of the social science disciplines and into the private sector, whose practices in relation to ethical considerations are by and large governed by legal jurisdiction, rather than ethical codes of practice. The recent Google case in the US is a case in point (Fry, 2006). Internet search-engine companies like Google, Yahoo! and MSN collect billions of 'data fossils' about how we use the Internet. These data are recorded for unspecified periods of time and unspecified uses. The data they hold is subject to legal jurisdiction and in August 2005 the US Government subpoenaed American Internet search-engine companies to provide lists of all URLs indexed in their search engines as of 31 July 2005, and all search term queries used between 1 June 2005 and 31 July 2005. The case came to media attention because whilst America Online, Microsoft, and Yahoo! are alleged to have complied with this request, Google resisted on the grounds that it was an undue infringement of their users' privacy. In the ensuing court case Google was required to submit to a less onerous request. Google was also caught up in another controversy when it entered into the Chinese Internet search-engine market. High-tech companies such as CISCO, Microsoft, and Yahoo! had already been operating in China and providing e-mail and blogging services. Internet search-engine companies operate under national legal jurisdictions and when the Chinese Government ordered Yahoo!

to hand over blog data, the files subsequently contributed to the sentencing of alleged cyberdissident Shi Tao to prison for ten years. As a result of this and related outcomes Google has decided not to provide e-mail or blogging services within China, on the grounds that they cannot provide data that they do not have.

These two cases illustrate the extent to which the Internet is being governed by commercial interests. This raises the question of the extent to which Internet researchers should be concerned with the collection and use of potentially harmful data – given that we cannot anticipate all the ways in which it might be reused and by whom. In terms of research excellence, social scientists have always been encouraged to consider only collecting sufficient data to satisfy the immediate objectives of their research, but with the Internet the capabilities for collecting and storing data are so vast that the practicality or desirability of maintaining such practices in the context of new technologies, methods, and techniques (such as webometrics) is brought into question.

[p. 36 ↓]

Transition from the private to public sphere online

Maintaining contextual integrity of data is closely related to personal perceptions of privacy. New contexts may necessitate different privacy protections. Status and interests in data may change over time. For example, data that may originate as academic could become a training tool and then be of commercial interest. The issue of privacy in public in relation to research participants has not been addressed in a systematic manner and more often than not surfaces as a result of highly publicized media events, rather than within the framework of developing ethical codes of practice for research. As Nissenbaum (1998) argues, practices of public surveillance fall outside the scope of predominant theoretical approaches to privacy, which have concerned themselves mainly with two aspects of privacy: (1) maintaining privacy against intrusion into the intimate, private realms of individuals, and (2) protecting the privacy of individuals against intrusion by agents of government (Nissenbaum, 1998). Nissenbaum argues that normative theories of privacy ought to be concerned with

privacy in public, that contemporary experience with information technology offers compelling reasons to expect that theory will provide a means of understanding the problem of privacy in public, as well as a means for adjudicating it.

Recent cases such as the release of the Enron and AOL e-mail databases illustrate Nissenbaum's point (*op. cit.*). Data gathered in the private sphere, such as company archives of e-mail interactions, are being added to this ever-growing corpus of publicly available data (for a different perspective on this debate see Janetzko's chapter in this Handbook). E-mail datasets are a rich source of data for Internet researchers and have been used to study a range of topics, from online humour to social network analysis. Typically, researchers gain access to the e-mail of research participants based upon trust and full disclosure about the nature of the research intentions and provisions for confidentiality. In 2003, however, this situation changed as Enron's Outlook e-mail database was seized by the Federal Energy Regulatory Commission (FERC), as part of an investigation of the 2000-2001 energy crises in the western United States, and released online for public scrutiny. The database contained nearly 1.6 million e-mails, tasks, and calendar entries written during the period 2000-2002 by 176 former Enron executives and employees from the power-trading operations.

This e-mail corpus was made public in an uncensored form for two weeks. It was not anonymized and is reported to have contained confidential information such as social security numbers and salary scales. The corpus was removed from the Internet following complaints from Enron's employees, and confidential information was removed using automated text extraction. Content of the database includes ninety-two percent of Enron's staff e-mails, and messages are identifiable according to senders and recipients' names. The database is hosted by the private company Lockheed Martin and was made available for third party reuse. Several academic institutions and private companies have purchased and modified the database, making it available in more user-friendly and searchable forms. For instance, users can conduct a search based on e-mail genre, for example, humour, sexist, breaching trade secrets, and sender's or recipient's name. At <http://www.cs.cmu.edu/~enron/> users are urged to 'Please be sensitive to the privacy of the people involved (and remember that many of these people were certainly not involved in any of the actions which precipitated the investigation).'

In terms of institutional governance, an academic study of social interaction based on this dataset would be exempt from ethical review, as it does not involve direct intervention with human subjects. The potential harm, however, arising from the unintended consequences of releasing such datasets could be great. Where does a lack of institutional or professional governance in the face of potential harm leave the social science researcher in terms of ethical responsibilities? One difficulty is [p. 37 ↓] that in the absence of governance, such as an ethical review committee, it is hard for researchers to assess the balance between potential benefit and harm. Is it the responsibility of the researcher reusing a publicly available dataset to contact persons named for informed consent, or do we abandon such research if it is not possible to acquire informed consent? Do we treat it as public information, on the grounds that it is available in the public domain, even though it may contain sensitive information? This brings into question the extent to which we as researchers can respect context and intentions online.

Moving beyond text-based data

Whilst shared commercial and academic interest in social science data is not new, as with market research for example, the capabilities for triangulation and reprocessing are. In the offline world ethical issues mainly relate to text-based data; in the online world image and video data are becoming more prevalent and therefore protecting anonymity and privacy is more challenging.

This raises two fundamental and related questions. How should data be treated that has been captured in offline public spaces? And where are the legal boundaries of informed consent? Legal protections are not necessarily catching up with ethical dilemmas, with many researchers at the technological cutting edge 'fudging it' based on preexisting codes of practice. It is necessary to deal with multiple settings, for example, domestic, educational, work, and public, with no core set of conventions. For example, what are the different ethical considerations between videoing children in the classroom and videoing them in public spaces? Some technological solutions are possible. In a classroom setting, for example, parental consent for one child, but not another, could be resolved by different anonymization levels when either representing or accessing the

data. However, even though there are technical solutions, such as pixeling out of faces, people can still be recognized by other physical features.

It is debatable as to whether or not legal statutes alone are sufficient to protect individuals from harm, for example, relying on data protection law to guide the boundaries of a study. Nevertheless, there is a school of thought in the application of novel technologies to social science that is on the side of pushing the boundaries until there is a legal intervention. This can be problematic, given that in many cases the technology and its capability for triangulating and reprocessing data is so novel that often legal intervention is lagging behind. Therefore, practice is often pushing the boundaries of ethical frameworks and legal interventions.

Overlaps in ethical and legal interventions

This chapter has highlighted that norms, ethical frameworks, and legal statutes are lagging behind not only technical developments related to the Internet, but also social developments in the visibility and penetration of Internet-generated data, such as the use of e-mails, search terms, and blogs as legal evidence. These socio-technical developments raise a series of questions about where the boundary is between private and public on the Internet, who has the possibility to delimit the boundaries, and what the ethical responsibilities and duties of social science researchers are amidst such ambiguity.

As Thelwall and Stuart (2006) point out, some techniques, such as web crawling, are inherently illegal in their mechanisms. Web crawling is illegal because crawlers make permanent copies of copyrighted material without the owner's permission. This has been an issue for the Internet Archive (<http://www.archive.org/index.php>) and from a legal perspective the issue has been negotiated by implementing an opt-out policy. There are also technical solutions, such as the robots.txt protocol (Thelwall and Stuart, 2006). These legal and technical solutions do not, however, address issues of privacy when it comes to the potential triangulation of datasets and third-party reuse. If we adopt a humanities approach to ethical governance on the Internet and see a text produced in a chatroom as an artistic work, rather [p. 38 ↓] than representing a 'human subject', then copyright becomes an issue (Bassett and O'Riordan, 2002). In this way

an ethical issue is leveraged into a legal issue for which there appears to be a solution through a set of conventions. Such leveraging is likely to become more and more prevalent in the ethical governance of Internet research. Copyright is also bound to become much more of an issue with the proliferation of user-generated content. This is an area in which institutions should recognize that researchers will need some training and support. The implications of such ethical 'leveraging' also mean that, increasingly, the boundary between legal and ethical issues will become blurred, and continuation of this distinction in the development of ethical codes of practice or guidelines may not be the most effective strategy.

CONCLUSIONS

One of the key challenges in devising a code of practice for Internet Research Ethics is in its global reach and the necessity to respect and incorporate diverse cultural practices, ethical governance, and legal frameworks. What's different about Internet-based research in contrast to research in the offline world is that the research object is no longer clearly delineated by national boundaries and protected by national research governance. The emergence of virtual methods such as virtual ethnography, text-mining, and webometrics across disciplines as dispersed as media and cultural studies, sociology, political science, and linguistics also brings an interdisciplinary focus to bear on the Internet as an object of study and challenges existing instruments of research governance that have traditionally been focused along disciplinary dimensions.

At the same time, the online world affords new modes of human interaction, and related ethical practices are shaped by the researchers' objectification of those being researched, for example, whether individuals participating in an online chatroom are perceived as research subject, research participant, artist (Bruckman, 2002), or author (Bassett and O'Riordan, 2002). There is also a potential convergence between research and commercial data on the Internet. Development of aggregator tools and services such as BlogPulse have led to the informatization of data, whereby data acquires additional value beyond the immediate research context. Consequently, the potential for third-party reuse is much greater than in the offline world. In the context of e-Social Science, data sharing and reuse are institutional imperatives with many funding bodies now mandating the submission of datasets to data archives and repositories upon the

completion of funding. There is, furthermore, an impetus within e-Social Science to make data publicly available through the Internet (using grids or online repositories). This would entail the development of practices and techniques to anonymize highly sensitive data, with some data being easier to anonymize than others. Progress has already been made in preserving confidentiality within quantitative Internet-based datasets, but qualitative data is much more challenging. In the UK the Qualidata initiative (<http://www.esds.ac.uk/qualidata/>) is looking into how these issues can be resolved.

The context of social interactions in online worlds is also important to bear in mind. If we take the position that traces of interaction on the Internet are public and should be treated as such, for example, participants have no rights to privacy considerations, how do we address the issue that online bodies and forms of expression have offline instantiations? To what extent do we need to protect these from harm? As tools for tracing social structures become more sophisticated, so too do our capabilities for triangulating data and getting a more holistic view of participants lives. So that, whereas participants may choose to draw a boundary between their online and offline worlds, and may in fact be online in order to escape the strictures of the offline world, the technologies currently being developed do not necessarily respect such boundaries. So the question for us as social scientists is to what lengths we should go to discover [p. 39 ↓] people's intentions. This, of course, means that we must disclose ourselves as researchers, which could alter the kind of results we were hoping to obtain. In the context of e-Social Science research, participants may be, but are not necessarily, already in the Internet domain. We cannot therefore simply assume that they have *chosen* to be online, or what their intentions are in being there. Again this raises the question of whether 'public in everyday life' is equivalent to 'public on the Internet.' All the while, the 'human subjects' research model remains in place and, as Bassett and O'Riordan (2002) have argued, what is required now is the trying and testing of different models of research governance.

The issues that we have raised in this chapter go beyond responsibilities towards a particular set of research participants and have implications for social, political, and ethical aspects of social science research. A significant proportion of the world will not be represented in online research and researchers need to ask whether this is ethical. Certain groups are likely to be under-represented and are therefore less likely to gain

benefits from participating. Such an emphasis on the interests of the information-rich may reinforce existing societal divisions (Mann, 2003). Researchers have an ethical responsibility to ensure that the research they carry out is of high quality, and that conclusions drawn from it can be inferred from the data collected (Pittenger, 2003). Finally, one obvious strategy to adopt under conditions with yet-to-emerge norms that have been sketched here is to be explicit about the ethical decisions that are made, so that others can learn from and debate the issues that arise when reporting findings.

NOTES

1 The report, 'Ethical Principles and Guidelines for the Protection of Human Subjects of Research' (commonly known as the Belmont Report) is available online at: <http://ohsr.od.nih.gov/guidelines/belmont.html> <accessed 12 Dec 2006>.

2 Acquiring informed consent for participation in research by children is subject to legal frameworks and regulations that differ from country to country. Gaining informed consent for those under 18 in offline and online research creates special problems for any researcher. See, for example, Wiles et al. (2005) for a discussion of the UK context.

3 For further discussion about some of the issues raised here in relation to ensuring confidentiality and informed consent see the 2007 report 'Dilemmas of Privacy and Surveillance: Challenges of Technological Change' by The Royal Academy of Engineering.

FURTHER READING

Schroeder (2007) has mapped the ethical and legal issues in relation to the study of shared virtual environments. Ess (2006) provides a valuable overview of ethical issues in social science uses of the Internet. An introduction to how the Internet is studied as a communication technology from a variety of disciplinary perspectives can be found in Walther, Gay and Hancock (2005). Varnhagen and colleagues (2005) provide an interesting discussion of informed online consent based on empirical work. A thought-provoking discussion about protecting privacy in public is provided by Nissenbaum

(1998). A useful overview of the security issues to consider when carrying out research online is given by Nosek and colleagues (2002); and Stern (2003) highlights the important legal and ethical issues that arise when encountering distressing information online.

<http://dx.doi.org/10.4135/9780857020055.n2>