Guidelines for Research on the Internet

1   Introduction

The following are guidelines rather than a code of practice to enable ethical research to remain flexible, be responsive to diverse contexts, and be adaptable to continually changing contexts.

2   Context

Depending on the research design, participants in research involving the internet can be identifiable or anonymous – they can explicitly consent to participate, or they can be invisibly observed without their knowledge. Thus, there are two key dimensions – the level of identifiability and the level of observation.

The different methodologies used in the case of internet research give rise to ethical issues over and above the University's standard research ethics guidelines. The University Code of Research Ethics accepts the following basic tenets as lying at the core of any research endeavour:

- ➢ Human dignity
- ➢ Autonomy
- ➢ Protection
- ➢ Safety
- ➢ Maximisation of benefits and minimization of harms

i.e., Respect for: persons; justice and beneficence.

These basic principles must of necessity continue to ground ethical enquiry, whether or not involving the internet, and are not to be 'trumped' by the use of the new technologies.

The scope and contexts of internet research have been dramatically expanded through the continuing global diffusion of the internet into nearly every country of the world, as facilitated through a growing array of devices (including game consoles, internet-enabled phones and other mobile devices.

It has been observed (AOIR 2012) that internet research encompasses inquiry that:

- Uses the internet to collect data or information, e.g., through online interviews, surveys, archiving, or automated means of data scraping;
- Studies how people use and access the internet, e.g., through collecting and observing activities or participating on social network sites, listservs, web sites, blogs, games, virtual worlds, or other online environments or contexts;

- Uses or engages in data processing, analysis, or storage of data sets, data banks and/or repositories available via the internet;
- Studies software, code and internet technologies; examines the design or structures of systems, interfaces, pages and elements;
- Employs visual and textual analysis, semiotic analysis, content analysis, or other methods of analysis to study the web and/or internet-facilitated images, writings and media forms;
- Studies large-scale production, use, and regulation by governments, industries, corporations and military forces.

Perhaps the most distinctive feature is the absence of physical contact between the researcher and the participant. This restricts the researcher's capacity to monitor, support or even terminate the study if adverse reactions become apparent. For example, experimental research online requires careful consideration of the design of stimulus materials to anticipate possible distress that might go undetected, and qualitative analysis of discussion boards requires an awareness that the author of quotations can be identified by using a search engine.

3 Principles

The University endorses the following AOIR Principles (2012) to be fundamental to an ethical approach to internet research:

- The greater the vulnerability of the community/author/participant, the greater the obligation of the researcher to protect the community/author/participant
- Because 'harm' is defined by context, ethical principles are more likely to be understood inductively, rather than applied universally – through the application of practical judgement attentive to the specific context
- Because all digital information at some point involves human persons, consideration of principles related to research on human subjects may be necessary, even if it is not immediately apparent how and where persons are involved in the research data.
- When making ethical decisions, researchers must balance the rights of subjects (as authors, as research participants, as people) with the social benefits of research and researchers' rights to conduct research. In different contexts the rights of subjects may outweigh the benefits of research.
- Ethical issues may arise and need to be addressed during all steps of the research process, from planning to research conduct, publication and dissemination.
- Ethical decision-making is a deliberative process, and researchers should consult as many people and resources as possible in this process, including fellow researchers, people participating in or familiar with contexts/sites being studied, research review boards, ethics guidelines, published scholarship and, where applicable, legal precedent.

3.1 Issues

Three major issues arise repeatedly in discussions about ethical practice in internet research:

- Human participants
- Private/public
- Data/persons

The internet complicates the fundamental research ethics question of personhood. Is an avatar a person? Is one's digital information an extension of the self? The essential question for research ethics committees to determine is: are we working with human participants or not? If information is collected directly from individuals, such as an email exchange, instant message, of an interview in a virtual world, the research context is likely to be recognised as one involving a person. If the connection between the object of the research and the person who produced it is indistinct, then there might be a tendency to define the research context as one which does not involve a person.

Example: a data set containing thousands of tweets or an aggregation of surfing behaviours collected from a bot is perhaps far removed from the person who engaged in these activities. In these scenarios it is possible to forget that there ever was a person somewhere in the process that could be directly or indirectly impacted by the research. Yet, there is considerable evidence that even 'anonymised' datasets that contain enough personal information can result in individuals being identifiable.

How might individuals be adequately protected when analysing such datasets in order to minimise harm? The issue can probably only be resolved on a case-by-case basis through research ethics committee reviews until a body of 'case law' emerges, based on the notion of reasonably foreseeable harm.

As a point of departure, the University, through the UREC and, by devolved powers, the CRECs, adopts the Recommendations from the Association of Internet Researchers Ethics Committee (Version 2.0) (2012), transposed as Appendix A of these Guidelines.

Appendix A

AoIR Guidelines (2012)

Triggers

Internet Specific Ethical Questions

The text below begins with common questions asked in the course of a research project. The bullet points represent more specific considerations that arise in internet-related contexts.

How is the context defined and conceptualized?

- Does the research definition of the context match the way owners, users, or members might define it? (Parameters such as 'culture,' 'person,' 'data set,' and 'public text' each carry different ethical expectations for researchers)
- Are there distinctions between local contextual norms for how a venue is conceptualized and jurisdictional frameworks (e.g., Terms of Service, other regulations)? For example, if the TOS defines the space as off limits for researchers but the individuals want to participate in public research of this space, what risk might exist for either the researcher or individuals involved?
- What are the ethical expectations users attach to the venue in which they are interacting, particularly around issues of privacy, both for individual participants as well as the community as a whole?

How is the context (venue/participants/data) being accessed?

- How are participants / authors situated in the context?
- How are participants/authors approached by the researcher?
- How is the researcher situated in the context?
- If access to an online context is publicly available, do members/participants/authors perceive the context to be public? What considerations might be necessary to accommodate 'perceived privacy' or the notion that individuals might care more about the appropriate flow of information as defining it as public or private?

Who is involved in the study?

- What are the ethical expectations of the community/participants/authors?
- What is the ethical stance of the researcher? (For example, a mismatch between the ethical stance of the researcher and the community/participant/author may create ethical complications).
- What are the ethical traditions of researchers' and/or author/participants' cultures or countries?
- If research data are housed in a repository for reuse, how might individuals or communities be affected later? For example, data collected for one purpose might be reused later for a different purpose but the researcher's relationship with the community from which the data came no longer exists. What possible risk or harm might result from reuse and publication of this information?

What is the primary object of study?

- What are the ethical expectations commonly associated with these types of data? (For example, working with aggregated, de-identified data carries different ethical expectations than working with interview data.)
- Does the object of analysis include persons or texts beyond the immediate parameters outlined by the study? What are the potential ethical consequences and how might these be addressed? (For example, collecting data from a blog often includes comments; collecting data from one social media stream reveals links to people or data outside the specific scope of the study.)
- If information collected in the course of a study can be linked back to an individual by means of internet search or other technology, what process will the researcher use to determine how that information will be treated? (For example, many challenges surround the responsible use of images and video.) To what extent might data be considered by participants to be personal and private, or public and freely available for analysis and republication? What other questions might arise as a result of the particular context from which these data were collected?

How are data being managed, stored, and represented?

- What method is being used to secure and manage potentially sensitive data?
- What unanticipated breaches might occur during or after the collection and storage of data or the production of reports? (For example, if an audience member recorded and posted sensitive material presented during an in-house research presentation, what harms might result?)
- If the researcher is required to deposit research data into a repository for future use by other researchers (or wishes to do so), what potential risks might arise? What steps should be taken to ensure adequate anonymity of data or to unlink these data from individuals?
- What are the potential ethical consequences of stripping data of personally identifiable information?
- How might removal of selected information from a dataset distort it such that it no longer represents what it was intended to represent?
- If future technologies (such as automated textual analysis or facial recognition software) make it impossible to strip personally identifiable information from data sets in repositories, what potential risks might arise for individuals? Can this be addressed by the original researcher? If so, how? How will this impact subsequent researchers and their data management?

How are texts/persons/data being studied?

- Does one's method of analysis require exact quoting and if so, what might be the ethical consequence of this in the immediate or long term? (For example, would quoting directly from a blog cause harm to the blogger and if so, could another method of representation be less risky?)
- What are the ethical expectations of the research community associated with a particular approach (e.g., ethnographic, survey, linguistic analysis)?
- Do one's disciplinary requirements for collecting, analyzing, or representing information clash with the specific needs of the context? If so, what are the potential ethical consequences?

How are findings presented?

- What immediate or future risk might occur by using exact-quoted material in published reports? (For example, while a participant might not think his or her information is sensitive now, this might change in five years. What protections might be put in place to anticipate changing perceptions?)
- Are individuals adequately protected in pre-publication reports, such as workshops, conferences, or informal meetings?
- Could materials be restricted because of copyright? (For example, many countries have strong restrictions on using screenshots or images taken from the web without permission. Certain sites have restrictions in their terms of service. Whereas there may be allowances for the scholarly use of copyrighted materials without permission, such as the U.S. doctrine of fair use, this is not a guarantee of protection against copyright infringement.)

What are the potential harms or risks associated with this study?

- What is the potential harm or risk for individuals, for online communities, for researchers, for research?
- Are risks being assessed throughout the study as well as in advance of the study? (Harm is only certain after it occurs. Thus, *a priori* assessments of risk might be useful but inadequate.)
- How are the concepts of 'vulnerability' and 'harm' being defined and operationalized in the study? How are risks to the community/author/participant being assessed?
- How is vulnerability determined in contexts where this categorization may not be apparent?
- Would a mismatch between researcher and community/participant/author definitions of 'harm' or 'vulnerability' create an ethical dilemma? If so, how would this be addressed?
- What harms--to life, to career, to reputation--may occur from the research (e.g., would the research "out" an individual who is not publicly out and perhaps cause them to lose their jobs? Would the research cause someone to face criminal or civil penalties?)?
- What possible privacy-related harms may occur? For example, might online groups disband or individuals cease to use an online support group or withdraw from blogging activities because of the presence of researchers; might individuals be upset that their perceived privacy has been violated; might individuals object to having their writing or speech anonymised, preferring to remain known and public in any published results?
- Who or what else could cause harm to the author/participant beyond the researcher?
- Are we acting in ways that minimizes risk?
- Does our research adequately protect the researcher as well as the community/author/participant?

What are potential benefits associated with this study?

- Who benefits from the study - do the potential participants? If not, what greater benefit justifies the potential risks?
- Is the research aiming at a good or desirable goal?
- Can we be sure the data collected from online sites, forums, communities, are "legitimate" and "valuable"?

How are we recognizing the autonomy of others and acknowledging that they are of equal worth to ourselves and should be treated so?

- Will informed consent be required from participants?
- If so, what procedures to obtain consent will be followed (e.g., print or digital signatures, virtual consent tokens, click boxes or waiver of documented consent)?
- Will consent be obtained just from individuals or from communities and online system administrators?
- In situations whereby consent is desired but written informed consent is impossible (or in regulatory criteria, impracticable) or potentially harmful, will procedures or requirements be modified?
- What harm might result from asking for consent, or through the process of asking for consent?
- What ethical concerns might arise if informed consent is not obtained?
- If a Research Ethics Committee deems no consent is required, will the researcher still seek subjects'/participants' consent in a non-regulatory manner?
- If informed consent is warranted, how will the researcher ensure that participants are truly informed?

What particular issues might arise around the issue of minors or vulnerable groups?

- Are minors being excluded from the study because of the difficulties of getting ethical permission to study them?
- In situations where identity, age, and ability of the participant is unknown or hidden, and harm cannot be determined as an *a priori* category based on known vulnerability of participant, how will harm be considered as an ethical concern and operationalized in the study?
- How are minors identified as 'minors' in contexts where demographic information is not required? What harm might result from asking (or not asking) for participants to reveal their age?
- How will parental or guardian consent be obtained in addition to assent where required by research regulations? What risks might arise in this particular consent process (for any or all parties, including the minor, the parents, and the researcher)?