

In addition to this 2012 report, AOIR has also produced a 2002 report to assist researchers in making decisions about internet research. Available at: <http://www.aoir.org/reports/ethics.pdf>

Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0)

AUTHORS

Annette Markham, Umea University, Sweden & Loyola University, Chicago
Elizabeth Buchanan, University of Wisconsin-Stout, USA

with feedback or contributions from the AoIR Ethics Committee and AoIR general membership, including: Maria Bakardjeiva, (Canada), Andrea Baker (USA), David Brake (UK), Charles Ess (Denmark), Radhika Gajjala (USA), Camilla Gronholm (Finland), Jeremy Hunsinger (Canada), Mark D. Johns (USA), Steve Jones (USA), Stine Lomborg (Denmark), Heidi McKee (USA), Jim Porter (USA), Soraj Hongladaram (Thailand), Janet Salmons (USA), Susannah Stern (USA), Eva Svedmark (Sweden), Leslie Tkach (Japan), Leslie Regan Shade (Canada), Michele White (USA), and Michael Zimmer (USA).

TABLE OF CONTENTS

INTRODUCTION AND BACKGROUND

- Purpose and Audience
- Limitations, or What this Document is Not
- What is Internet Research?

GENERAL ETHICAL GUIDELINES

- Key Ethical Principles
- Guidelines, not Codes

A PROCESS APPROACH TO ETHICAL DECISION MAKING IN INTERNET RESEARCH

- Major Tensions/Considerations
- Internet Specific Ethical Questions

CONCLUSION

ENDNOTES

WORKS CITED

APPENDICES

- Appendix 1: Heuristic Chart of Ethical Questions by Data Type and Venue
- Appendix 2: Regional Resources and/or FAQs (online-only compendium)
- Appendix 3: Bibliography of Internet Research Ethics Resources (online-only compendium)
- Appendix 4: Case studies (online-only compendium)

INTRODUCTION AND BACKGROUND

The work represented in this document emerges from theoretical, empirical, and field research conducted by members of the Association of Internet Researchers, including members of the AoIR Ethics Working Group.

The first version of the AoIR Ethical Decision-Making document was released in 2002, after two years of international and cross-disciplinary collaboration. The document and its guidelines emerged from a series of extensive dialogues among experienced researchers facing and resolving ethical issues in internet research, philosophers, and other members of AoIR's international, interdisciplinary community. The intention was to develop guidelines from the bottom up -- i.e., out of the day-to-day practices of researchers in a wide range of disciplines, countries and contexts, in contrast to a more usual top-down approach that tries to provide a universal set of norms, principles, practices, and regulations. This approach was crucial because the enterprise of internet research is expansive -- that is, globally informed -- but also situated in innumerable locales. The 2002 document has subsequently received much use, and has been cited and used in a wide range of publications by a diverse number of disciplines. The AoIR Guidelines document has also been used by research ethics boards (REBs) and institutional review boards (IRBs) when making decisions about internet research-based protocols.¹

Purpose and Audience

While the first AoIR document enjoyed extensive use, much has changed in the field of Internet Studies since 2002. The scope and contexts of internet research have been dramatically expanded through the continuing global diffusion of the internet into nearly every country in the world, as facilitated through a growing array of devices (including game consoles, internet-enabled phones and other mobile devices) and ever-increasing bandwidth; rapidly expanding suites of new communication applications; and the increasingly seamless interweaving of online and offline activities and experiences. Alongside these developments, the literature of internet research ethics has grown considerably, providing us with a far more extensive range of theoretical resources and practical examples to help recognize and guide ethical reflection.

The AoIR ethics working committee, with extensive feedback from AoIR membership including at workshops conducted annually at the AoIR internet research conferences, from 2005 through the present, has developed this 2012 version in an effort to recognize and respond to the array of changing technologies and ongoing developments that affect the ethics of internet research. We emphasize that no set of guidelines or rules is static; the fields of internet research are dynamic and heterogeneous. This dynamism is reflected in the fact that as of the time of this writing, no official guidance or “answers” regarding internet research ethics have been adopted at any national or international level.

This document aligns with and extends the first AoIR document. This document represents a series of considerations designed to support and inform those responsible for making decisions about the ethics of internet research. And, while primarily directed at researchers, it provides a resource for a wide audience of other stakeholders such as review boards, ethicists, and students, by providing a current discussion of important ethical issues and pertinent literature in the field.

2 Final Copy: Ethical Decision-Making and Internet Research: Recommendations from the AOIR Ethics Committee
Approved by the Ethics Working Committee, 08/2012. Endorsed by the AOIR Executive Committee, 09/2012.
Approved by the AOIR general membership, 12/2012.

Limitations, or What This Document is Not

In this document, we do not reiterate the many specific rules of governmental or institutional research policies, as these are well documented and widely available. We also do not include specific disciplinary best practices or codes of ethics, as this document seeks to function at the macro-level.²

Although we identify current internet technologies and contexts, we acknowledge that technologies themselves change rapidly. Therefore, this document is designed to emphasize processes for decision-making and questions that can be applied to ever-changing technological contexts.

At its most fundamental level, we recognize that ethical decision-making interweaves one's fundamental world view (ontology, epistemology, values, etc), one's academic and political environment (purposes), one's defining disciplinary assumptions, and one's methodological stances. Decision making occurs at many junctures in the cycle of inquiry, including research design, research conduct, and research production and dissemination. Because of the complexity of ethical decision making in individual cases, this document focuses on general principles, illustrating their application using some examples from the many possible ethical issues that may arise.

What is Internet Research?

The internet is a social phenomenon, a tool, and also a (field) site for research. Depending on the role the internet plays in the research project or how it is conceptualized by the researcher, different epistemological, logistical and ethical considerations will come into play. The term "Internet" originally described a network of computers that made possible the decentralized transmission of information. Now, the term serves as an umbrella for innumerable technologies, devices, capacities, uses, and social spaces. Within these technologies, many ethical and methodological issues arise and as such, internet research calls for new models of ethical evaluation and consideration.

Because the types of interaction and information transmission made possible by the internet vary so widely, researchers find it necessary to define the concept more narrowly within individual studies. This is complicated by the fact that studies of and on the internet cut across all academic disciplines.

This document uses the following working definitions:

Internet research encompasses inquiry that:

- (a) utilizes the internet to collect data or information, e.g., through online interviews, surveys, archiving, or automated means of data scraping;
- (b) 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;
- (c) utilizes or engages in data processing, analysis, or storage of datasets, databanks, and/or repositories available via the.
- (d) studies software, code, and internet technologies
- (e) examines the design or structures of systems, interfaces, pages, and elements

3 Final Copy: Ethical Decision-Making and Internet Research: Recommendations from the AOIR Ethics Committee
Approved by the Ethics Working Committee, 08/2012. Endorsed by the AOIR Executive Committee, 09/2012.
Approved by the AOIR general membership, 12/2012.

- (f) 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.
- (g) studies large scale production, use, and regulation of the internet by governments, industries, corporations, and military forces.

Internet research is not machine specific or dependent, and we recognize the impact of smart devices and increased mobility on internet activities. It is important to note that definitions of and experiences with these contexts vary widely. Technological convergence collapses many contexts and categories in evolving and sometimes surprising ways. The internet mediates everyday life in industrialized and developing cultures, whether or not we are actively accessing the web. Thus, internet research should be considered in its broadest sense.

GENERAL ETHICAL GUIDELINES

Principles of research ethics and ethical treatment of persons are codified in a number of policies and accepted documents, such as the [UN Declaration of Human Rights](#), [the Nuremberg Code](#), [the Declaration of Helsinki](#), and [the Belmont Report](#). At their core, the basic tenets shared by these policies include the fundamental rights of human dignity, autonomy, protection, safety, maximization of benefits and minimization of harms, or, in the most recent accepted phrasing, respect for persons, justice, and beneficence. While originally stemming from the biomedical contexts, these principles have been adapted beyond these early contexts and rise above disciplines and methodologies. We accept them as basic to any research endeavor. These basic principles provide a starting point for a range of ethical considerations that occur at various junctures of any specific research project. Different ethical issues become salient as the researcher develops research questions, seeks and gains access to individuals and/or information, manages and protects personally identifiable information, selects analytical tools, and represents the data through dissemination, in published reports, conference presentations, or other venues.³

Key guiding principles

We find the following principles 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 contextually, ethical principles are more likely to be understood inductively rather than applied universally. That is, rather than one-size-fits-all pronouncements, ethical decision-making is best approached through the application of practical judgment attentive to the specific context (what Aristotle identified as *phronesis*⁴).

Because all digital information at some point involves individual 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.

⁴ Final Copy: Ethical Decision-Making and Internet Research: Recommendations from the AOIR Ethics Committee Approved by the Ethics Working Committee, 08/2012. Endorsed by the AOIR Executive Committee, 09/2012.

Approved by the AOIR general membership, 12/2012.

This work is licensed under a Creative Commons Attribution 4.0 International License.

Ethical issues may arise and need to be addressed during all steps of the research process, from planning, 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 (within one's discipline but also in other disciplines), and, where applicable, legal precedent.⁵

Guidelines versus codes

Ethical conundrums are complex and rarely decided along binary lines. There is much grey area in ethical decision-making. More than one set of norms, values, principles and usual practices can be seen to legitimately apply to the issue(s) involved. It becomes difficult to make judgments as to which set(s) apply, especially when one set conflicts with another in some way. This forces the researcher to determine which is more relevant in a given context or at particular junctures during the course of the study. We emphasize that ethical concepts such as harm, vulnerability, respect for persons, and beneficence are not just regulatory hurdles to be jumped through at the beginning stages of research, but concepts that ground ethical inquiry. As such, they should be assessed and considered throughout each stage of the research. Multiple judgments are possible, and ambiguity and uncertainty are part of the process.

We advocate guidelines rather than a code of practice so that ethical research can remain flexible, be responsive to diverse contexts, and be adaptable to continually changing technologies. When one considers that ethical assessments are always operationalized via some sort of practice (method), and also contextualized institutionally and/or geographically, it becomes clearer that an adaptive, inductive approach can yield potentially more ethically legitimate outcomes than a simple adherence to a set of instantiated rules. The emphasis on a process approach highlights the researcher's responsibility for making such judgments and decisions within specific contexts and, more narrowly, within a specific research project. Hence, this 2012 version of the AoIR Ethical Decision Making document adheres to the same underlying principles as the first: a dialogic, case-based, inductive, and process approach to ethics.

We strongly urge that people who use this document in their work contribute a synopsis or case study of how they used it, whether it was in fact useful, and whether they might change anything about it. In this way, we can more fully appreciate the range of approaches to ethical research practice and continue to build a robust and diverse collection of best practices. This growing compendium of resources can be accessed at <http://ethics.aoir.org/>.

A PROCESS APPROACH TO ETHICAL DECISION MAKING IN INTERNET RESEARCH

At each juncture of a research project, different ethical issues become relevant. Decisions about gaining access to a site for data collection are different from decisions about how to interact with participants or whether or not informed consent is needed. Once these issues have been resolved, further ethical quandaries will arise as one makes decisions about how to store data, whether or not

5 Final Copy: Ethical Decision-Making and Internet Research: Recommendations from the AOIR Ethics Committee
Approved by the Ethics Working Committee, 08/2012. Endorsed by the AOIR Executive Committee, 09/2012.

Approved by the AOIR general membership, 12/2012.

This work is licensed under a Creative Commons Attribution 4.0 International License.

to use actual quotes in published reports, or how to represent one's findings for various audiences. This section is designed to help guide one's considerations through the process of ethical assessment. We begin by outlining three major issues that arise repeatedly in discussions about ethical practice in internet research: Human subjects, private/public, and data/persons. This is followed by a brief discussion of a fourth issue for researchers: The tension between top down regulatory models and case based, situated approaches. Finally in this section, we provide a detailed list of questions that can help raise awareness of ethical considerations throughout a research project. Note that the appendices provide further tools to think with, including a list of frequently asked questions and a heuristic chart that lays out an array of questions arising in studies of particular contexts/venues or with particularly types of information or data produced or retrieved from those contexts.

Major tensions/considerations

Operational definitions of ethical concepts and appropriate research conduct vary widely, within and across disciplines. This is not unique to internet research, but describes an inherent quality of any ethical discussion. Our research community has learned over the past two decades that internet research requires individuals and institutions alike to confront many irresolvable contradictions and tensions. Many competing interests must be negotiated by researchers, ethics review boards, and institutions.

Human Subjects: The concept of 'human subjects' is a persistent and contentious example of a dialectical tension between disciplinary/regulatory models and context-specific sensibilities. The term 'human subject' as applied to the consideration of research ethics emerged from early responses to harmful treatment of persons in medical experiments⁶. It persists as a guiding concept for ethical social research, yet has been long critiqued for being ill-suited for models of inquiry that follow non-biomedical procedures for interacting with people or don't interact directly with people at all, for example, studying published texts.⁷ Many regulatory bodies across the world⁸ have dealt with these issues by creating different levels of ethics board review based on the idea that some research might be exempt from review or require only limited review. The definition of 'human subject' has become, unfortunately, a litmus test for whether or not one needs to undergo ethical review before conducting research. In internet research, 'human subject' has never been a good fit for describing many internet-based research environments.⁹ Ongoing debates among our community of scholars illustrate a diverse, educated range of standpoints on the answers to the question of what constitutes a 'human subject.' We agree with other regulatory bodies that the term no longer enjoys the relatively straightforward definitional status it once did.¹⁰ As a community of scholars, we maintain the stance that when considered outside a regulatory framework, the concept of 'human subject' may not be as relevant as other terms such as harm, vulnerability, personally identifiable information, and so forth. We encourage researchers to continue vigorous and critical discussion of the concept of 'human subject,' both as it might be further specified in internet-related research or as it might be supplanted by terms that more appropriately define the boundaries for what constitutes inquiry that might be ethically challenging.

Public/Private: Individual and cultural definitions and expectations of privacy are ambiguous, contested, and changing. People may operate in public spaces but maintain strong perceptions or expectations of privacy.¹¹ Or, they may acknowledge that the substance of their communication is public, but that the specific context in which it appears implies restrictions on how that information is -- or ought to be -- used by other parties.¹² Data aggregators or search tools make information accessible to a wider public than what might have been originally intended. In mediated contexts,

6 Final Copy: Ethical Decision-Making and Internet Research: Recommendations from the AOIR Ethics Committee
Approved by the Ethics Working Committee, 08/2012. Endorsed by the AOIR Executive Committee, 09/2012.

Approved by the AOIR general membership, 12/2012.

This work is licensed under a Creative Commons Attribution 4.0 International License.

as Nissenbaum points out, “what people care most about is not simply restricting the flow of information but ensuring that it flows appropriately” (2010, p. 2). As noted in the 2002 version of these AOIR ethics guidelines, privacy is a concept that must include a consideration of expectations and consensus. Social, academic, or regulatory delineations of public and private as a clearly recognizable binary no longer holds in everyday practice. When conducting research within such shifting terrains, when there is no consensus, or even assumption of consensus, Nissenbaum’s concept of contextual integrity (2010) is a valuable construct.

Data(Text)/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? In the U.S. regulatory system, the primary question has generally been: Are we working with human subjects or not? If information is collected directly from individuals, such as an email exchange, instant message, or an interview in a virtual world, we are likely to naturally define the research scenario as one that involves a person. If the connection between the object of research and the person who produced it is indistinct, there may be a tendency to define the research scenario as one that does not involve any persons. This may oversimplify the situation--the question of whether one is dealing with a human subject is different from the question about whether information is linked to individuals: Can we assume a person is wholly removed from large data pools? For example, a data set containing thousands of tweets or an aggregation of surfing behaviors collected from a bot is perhaps far removed from the persons who engaged in these activities. In these scenarios, it is possible to forget that there was ever 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. Scholars and technologists continue to wrestle with how to adequately protect individuals when analysing such datasets (Sweeny, 2009; Narayanan & Shmatikov, 2008). These are important considerations because they link to the fundamental ethical principle of minimizing harm. Does the connection between one’s online data and his or her physical person enable psychological, economic, or physical, harm? One way of evaluating the extent to which these ethical dilemmas may be hidden is to focus on the way that procedures for data collection or analysis extract data from lived experience.

These three issues represent ongoing tensions for internet research. Although we as researchers might like straightforward answers to questions such as “Will capturing a person’s Tweets cause them harm?” or “Is a blog a public or private space?” and we as authors of this document might like to supply answers, the uniqueness and almost endless range of specific situations defy attempts to universalize experience or define in advance what might constitute harmful research practice. We take the position that internet research involves a number of dialectical tensions that are best addressed and resolved at the stages they arise in the course of a research study. In saying so, we reiterate the value of a casuistic or case-based approach.

Top-down versus bottom-up approaches to ethics: This case based approach is complicated by the fact that most of us will grapple with other sets of tensions beyond the specifics of the research context we are studying: ethical versus legal considerations, regulatory-driven versus context-specific models for ethical approaches, and top-down versus bottom-up decision making. These tensions should be acknowledged and considered, even if there are no easy resolutions.

For example, when acknowledging the tension between regulation-driven and context-specific considerations, it may be useful to consider that while regulations or disciplinary traditions are intended to encourage ethical practice, they can also inadvertently function to restrict the researcher

by making universal declarations and *a priori* determinations about what constitutes harm and what will cause harm. This is often done by identifying particular categories of persons who are vulnerable, or by implying that by gaining informed consent, harm may somehow be mitigated. We agree that in certain extreme situations, there will be broad consensus about whether certain actions were ethical or not. Yet we also agree as a community of researchers that in most social situations, the issues and ethics are more fuzzy.

One must also balance the need to predict what ethical consequences might occur with the need to be open to how ethical issues emerge during the research process. In the United States, for example, most scholars are required to gain approval of institutional review boards (IRB) prior to conducting research. One is asked to make many decisions about possible ethical issues in advance of one's study. This can be daunting. Foreseeing harms or risks associated with a research project is always difficult, given that individuals and groups are unique, situations change, research directions can shift, and continually-evolving technologies may alter the ethical situation. On the other hand, researchers working without the direct approval of ethics review boards in countries where these are not a part of the research process have different responsibilities and may have additional challenges to face.

The challenge for researchers remains: how to balance contextual requirements with disciplinary, institutional, legal, cultural, or other constraints.

We offer the following detailed sets of questions to prompt reflection about ethical decision making within the specific confines of one's study. These are intended, along with the appendices, as tools to engage with prior to, during, and after the research process.

Internet Specific Ethical Questions

The list 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?¹³ (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?¹⁴
- 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?¹⁵
- 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’¹⁶ or the notion that individuals might care more about the appropriate flow of information as defining it as public or private¹⁷?

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).¹⁸
- What are the ethical traditions of researchers’ and/or author/participants’ cultures or countries?
- If research data is 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?¹⁹

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 re-publication?
- What other questions might arise as a result of the particular context from which this data was 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?²⁰ (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 this 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?²¹)
- 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?²²)
- Are individuals adequately protected in pre-publication reports, such as workshops, conferences, or informal meetings?²³
- 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 inadequate²⁴).
- 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 LGBTQ 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;²⁵ Might individuals be upset that their perceived privacy has been violated;²⁶ 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, fora, communities, is “legitimate” and “valuable”?²⁷

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)²⁸
- 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?²⁹
- What ethical concerns might arise if informed consent is not obtained?
- If an ethics board 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 persons?

- 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)?

CONCLUSION

This document provides a basic overview of internet research considerations. It advocates a process approach to ethics, which emphasizes the importance of addressing and resolving ethical issues as they arise in each stage of the project. We also advocate a casuistic, or case based approach, which attends to the specific needs of each case. Rather than prescribing a set of approved practices, we have deliberately chosen to suggest a characteristic range of questions that should be asked by internet researchers as well as those responsible for oversight of such research.

Guidelines such as these should not be considered complete or final. More than two decades of research practice provide multiple examples of ethical complications, unique solutions, and best practices. To continue to build, rather than reinvent the wheel, we believe it is vital for researchers to maintain a strong awareness of previous scholarship. To help in this endeavor, we provide a diverse bibliography in Appendix 3. We also recognize that as internet researchers encounter new venues, contexts, inter/actions, and tools, additional issues, questions, and responses will inevitably arise, either as variations of those mentioned in this document or as distinctively new. We encourage researchers to contribute synopses of their ethical challenges or specific case studies to our web-based compendium of cases. [website link here]. Our intent is that this document will remain useful in those new contexts in so far as it suggests starting points and possible analogies for raising new questions and developing new responses.

At the same time, we anticipate that as this document comes into use, it may help researchers further identify its limits as they encounter new contexts and ethical issues that emerge in the ongoing evolution of both the internet and our multifaceted efforts to research the communicative engagements these technologies make possible. We intend this document to serve both as a first way of addressing contemporary ethical challenges and as an occasion for inviting further critical evaluation and ongoing dialogical development of internet research ethics.

In sum, we hope this document will prove helpful to researchers, ethicists, and others concerned with the important ethical challenges of internet research, and the Association of Internet Researchers remains grateful for further critical comment and examples that help to carry our dialogues onward from this point.

ENDNOTES

¹ See Buchanan and Ess (2009), who found that U.S. based IRBs referred most regularly to the Code of

² For example, the Forum for Ethical Review Committees in Asia and the Western Pacific (FERCAP); U.S. Copyright Law, Norway's National Committee for Research Ethics in the Sciences and the Humanities (NESH); the EU Data Privacy Protection Act represent specific guidelines; each country and/or region maintains specific and changing guidelines. See also Privereal, which examines "the implementation of the Data Protection Directive 95/46/EC" in relation to research ethics and law (<http://privereal.group.shef.ac.uk/content/rec/countries.php>). We refer to some of these in Appendix 3 (the wiki-based bibliography located at <http://ethics.aoir.org>), but our listing is not all-inclusive. Readers should consult their professional and discipline-specific codes of conduct or practice, as well as extant regulation, to ensure they are acting in accordance with their disciplinary and jurisdictional bounds.

³ As one example, copyright laws vary by country and in the digital age, complexities in copyright laws have grown. Copyright is one particular issue that bridges legal and ethical realms, and is a matter of consideration in internet research. In some disciplines, for example, linguistics, copyright laws are more significant concerns than human subjects regulations. For more information on copyright law by country, see http://en.wikipedia.org/wiki/Category:Copyright_law_by_country

⁴ For further discussion of phronesis as applied to internet research ethics and to digital media ethics, see for example Ess (2009) and McKee and Porter (2009).

⁵ For example, in addition to general ethical principles, researchers must be aware of legal requirements in the countries implicated in the research. This can create important and potentially unresolvable tensions that should be acknowledged. While guidelines and regulations are intended to encourage ethical practice, they can also inadvertently function to restrict the researcher by making universal declarations and a priori determinations about what constitutes harm and what will cause harm. This is often done by identifying particular categories of persons who are vulnerable, or by implying that by gaining informed consent, harm may somehow be mitigated. We agree that in certain extreme situations, there will be broad consensus about whether certain actions are ethical or not. Yet we also agree as a community of researchers that in most social situations, the issues and ethics are more fuzzy. It can be difficult if not impossible to predict beforehand what might cause immediate or eventual harm, whether or not someone is vulnerable, or even whether or not we can call something a ‘human subject.’ Hence the need for deliberation.

⁶ E.g., the Tuskegee syphilis experiments:

http://en.wikipedia.org/wiki/Tuskegee_syphilis_experiment

⁷ Manipulation and close study of information generated by social media networks certainly constitutes a different research environment than sticking a needle into a volunteering person in a medical laboratory. On the other hand, entire communities have felt harm from use of their DNA data more than a decade after it was collected and anonymously aggregated. Interacting with a person in a face-to-face interview seems to raise different considerations than interviewing a person in an anonymous chatroom. Yet, anonymous members of online communities have felt harm after researchers published reports about what they perceived to be private community activities. Blogs are often considered public, published texts. On the other hand, users have described their blogs as a part of their identity, not to be treated as simply publicly accessible data.

⁸ For a fairly comprehensive listing of these, see

<http://www.hhs.gov/ohrp/international/intlcompilation/intlcompilation.html>

⁹ Should authors be considered human subjects? Should studies based on observation of people in public be scrutinized with the same criteria as studies involving drug testing?

¹⁰ The AoIR discussion list and archives (<http://listserv.aoir.org/listinfo.cgi/air-l-aoir.org>) provide many examples of researchers seeking advice from a world-wide community.

¹¹ This is widely discussed in ethics and internet research literature. A particularly keen analysis appears in Sveningsson (2003), and is modified/expanded somewhat in McKee & Porter (2009). The ethical parameters for collecting information in online public spaces are ambiguous and contested. Hudson and Bruckman (2004) found that while it might be widely considered ethically acceptable to capture and analyze interactions and conversations in a public square without consent, this model did not match the expectations of their participants in real time chatrooms, who felt strongly that “one may not ethically record an otherwise ephemeral medium without consent from participants” (p.118). Bassett & O’Riordan (2002) note that this sort of mismatch can occur in many ways: In the same online space or in the same online group, some users may believe their words are important public documents and they should receive credit as authors while others might believe their contributions are private, deserving anonymity. Likewise, the service provider, operator of an online forum, or terms of service may state that the content is public, yet individuals

perceive the space to be private. Cases by authors such as Bromseth (2002), Gajjala (2004), and Ikonomidis Svedmark & Nyberg (2009) illustrate the difficulties these issues can pose for the researcher.

¹² For example, youth interviewed by Boyd and Marwick (2011, p.6) felt angry that their Facebook information was used by teachers and administrators in a school-wide public presentation to illustrate the dangers of posting private information in public spaces. They knew the information was relatively public. It was the manner in which it was used that violated their expectations of privacy.

¹³ See for example Hudson & Bruckman (2004); Gajjala (2004); Ikonomidis Svedmark & Nyberg (2009); Baker (2009).

¹⁴ See for example Chee, Taylor, & de Castell (2012).

¹⁵ See for example Markham (2009).

¹⁶ For further discussion of this, see for example Sveningsson (2003), Sveningsson-Elm (2009) and McKee & Porter (2009).

¹⁷ For further discussion, see Nissenbaum (2010).

¹⁸ See Ayers (2003).

¹⁹ A case involving blood and DNA samples taken from the Havasupai Indian tribe in Arizona provides an illustration of the complexities of informed consent, misuse of data/samples, and community harms resultant from biomedical research. See <http://www.whoownsyourbody.org/havasupai.html>

²⁰ See for example Sweeny (2003); Zimmer (2010).

²¹ See for example Ikonomidis Svedmark & Nyberg (2009).

²² Markham (2012) makes the case for revising or refashioning data to resonate with the situation without using direct quotes to protect privacy.

²³ CF Footnote 18.

²⁴ Traditional notions of harm or risk come from the medical or biomedical models of research, which can often predict with more accuracy how a particular medical or psychological procedure can affect a research participant. Regulatory language around the issue of risk and harm may be predicated on such medical models, which are often different from social research environments, where potential risks or harms may not be clearly foreseeable. Likewise, in inductive or qualitative research frameworks, procedures for interaction with contexts or persons are added or modified in situ, so that the risks associated with particular methods may only be known at later stages of the research. Risk and harm in less formalized research frameworks can arise in many stages of the project, from data collection to analysis to presentation of research.

²⁵ See for example Bromseth (2002); Kraut et al., (2004).

²⁶ CF Footnote 8

²⁷ For example, Karpf (2012) notes: “Spambots, commercial incentives, proprietary firewalls, and noisy indicators all create serious challenges for the researcher.”

²⁸ Although in some cases, traditional informed consent procedures may be warranted, in other cases, these may not be the most ethically sensitive approach. As recommended by the 2010 AOIR Ethics workshop: “Informed consent may happen at different points throughout a study, Sometimes it may be more ethical to get informed consent at the end when you want to present a specific case study or quote an individual or focus on a particular element. Therefore, informed consent should be always an inductive process” (Buchanan, Markham, & Ess, 2010).

²⁹ See for example Stern (2003); Ikonomidis Svedmark (2010).

WORKS CITED

- Ayers, M. (2003). Fact or fiction: Notes of a man interviewing women online. In Buchanan, E.A. (Ed.). *Readings in Virtual Research Ethics: Issues and Controversies* (pp. 262-273). Hershey: Idea Group.
- Baker, A. (2009). Small Talk. *International Journal of Internet Research Ethics* 2, 1, 1-5.
- Bassett, E. H., & O’Riordan, K. (2002). Ethics of internet research: Contesting the human subjects research model. *Ethics and Information Technology*, 4 (3), 233-249. Available: http://www.nyu.edu/projects/nissenbaum/ethics_bas_full.html
- Bromseth, J. C. H. (2002). Public places – public activities? Methodological approaches and ethical dilemmas in research on computer-mediated communication contexts. In A. Morrison (Ed.), *Researching ICTs in Context* (pp. 33-61). Inter/Media Report 3/2002. Oslo: University of Oslo.
- boyd, d., & Marwick, A. (2011). How Teens Understand Privacy. Unpublished Manuscript. Available online: <http://www.danah.org/papers/2011/SocialPrivacyPLSC-Draft.pdf>
- Buchanan, E., & Ess. C. (2009). Internet research ethics and the Institutional Review Board: Current practices and issues. *Computers and Society*, 39 (3) 43-49.
- Buchanan, E., Markham, A., & Ess. C. (2010). Ethics and internet research commons: Building a sustainable future. Association of Internet Researchers 11th Annual Conference Workshop. Gottenburg, Sweden.
- Chee, F. M., Taylor, N. T., & de Castell, S. (2012). Re-mediating research ethics: End-user license agreements in online games. *Bulletin of Science, Technology & Society*, 32(6), 497-506.
- Ess, C. (2009). *Digital Media Ethics*. Cambridge: Polity.
- Gajjala, R. (2004). *Cyber Selves: Feminist Ethnographies of South Asian Women*. Walnut Creek: Altamire Press.
- Hudson, J. M., & Bruckman, A. (2004). Go away: Participant objections to being studied and the ethics of chatroom research. *Information Society*, 20(2), 127-139.
- Ikonomidis Svedmark, E. (2010). Med nätet som fält: urvalstankar, känslostormar och etikproblem. *Kulturella Perspektiv* 2, 11-21.
- Ikonomidis Svedmark, E., & Nyberg, A. (2009). Om det privata i publika och digitala rum. In *Se mig: Unga om sex och internet* (pp. 354-383). Stockholm: Davidsons Tryckeri AB.
- Karpf, D. (2012). Social science research methods in internet time. *Information, Communication, & Society*, 15(5), 639-661.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Cooper, M. (2004). Psychological research online: Report of board of scientific affairs' advisory group on the conduct of research on the internet. *American Psychologist*, 59(4), 1-13.

Markham, A. (2012). Fabrication as ethical practice: Qualitative inquiry in ambiguous internet contexts. *Information, Communication, and Society*, 5(3), 334-353.

Markham, A. (2009). How can qualitative researchers produce work that is meaningful across time, space, and culture? In Markham, A. N., & Baym, N. K. (Eds.). *Internet inquiry: Conversations about method* (pp. 131-155). Thousand Oaks, CA: Sage.

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

Narayanan, A., & Shmatikov, V. (2008). De-anonymizing social networks. *IEEE Symposium on Security & Privacy*. Oakland, CA. Available: http://www.cs.utexas.edu/~shmat/shmat_oak09.pdf

Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets (How to break anonymity of the Netflix prize dataset.) *IEEE Symposium on Security & Privacy*. Oakland, CA. Available: <http://arxiv.org/pdf/cs/0610105v2>

National Science Foundation. (2008).

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford: Stanford University Press.

Stern, S. (2003). Encountering distressing information in online research: A consideration of legal and ethical responsibilities. *New Media and Society* 5, 249-266.

Sveningsson, M. (2003). Ethics in internet ethnography. In Buchanan, E. (Ed.). *Readings in Virtual Research Ethics: Issues and Controversies* (pp. 45-61). Hershey: Idea Group.

Sveningsson-Elm, M. (2009). How do various notions of privacy influence decision making in qualitative internet research. In Markham, A. & Baym, N. (Eds.). *Internet Inquiry: Conversation about method* (pp. 69-87). Thousand Oaks, CA: Sage.

Sweeney, L. (2003). Navigating computer science research through waves of privacy concerns: Discussions among computer scientists at Carnegie Mellon University. In Tech Report, CMU CS 03-165, CMU-ISRI-03-102. Pittsburgh.

Zimmer, M. (2010). But the data is already public: On the ethics of research in Facebook. *Ethics & Information Technology*, 12(4), 313-325.

APPENDIX I: Charting Ethical Questions By Data And Type

This figure is intended to provide a reference guide for researchers and research reviewers to recognize where a particular research project might fit in the internet research continuum, in what specific type of venue a project exists, what types of information/data may be derived from these contexts, and what common ethical questions have been asked within these contexts.

This graphic is best viewed as a conceptual heuristic, as the categories overlap and merge. The first column lists the types of information that might be collected. These data types are not tied to certain internet spaces or genres, such as blogs, email, or the web, but cut across all contexts. The second column lists specific types of contexts. Again, because categories continue to converge, we have organized this column by major interaction type, rather than specific type of interface. The third column lists certain key ethical questions that might be raised. Notably, there are varying ways to conceptualize or define each of the contexts mentioned and varying ways to perceive one's research in or of these contexts. Two people can be utilizing the same technological context and yet have very different perspectives on it. These distinctions are also blurred as technologies converge and become less obvious as a part of everyday life.

Researchers and review boards can refer to this chart as a guide or checklist of considerations when writing or reviewing research proposals, developing methodological approaches for a particular study, designing specific protocols, and so forth. Ideally, a researcher should be able to answer these questions and justify those responses within the context of his or her research question, discipline, specific internet venue or context, and region or jurisdiction.

Note: This is shown as a small graphic on the following page. A higher resolution PDF is available at: <http://ethics.aoir.org/reports/2012aoirgraphic.pdf>

This chart provides a useful starting point for internet researchers to consider ethics. Convergence of technologies and capacities continues to break down the strict boundaries between these categories. Not intended to provide answers, this graphic promotes consideration of a range of issues and questions that may become relevant in the course of any internet related research.

Types of Data collected	Types of Venues/Contexts	Commonly asked questions about ethical practice
<p>Interactions, behaviors, transactions</p> <ul style="list-style-type: none"> • Hyperlinks • Comments or Recommendations • File or Information Sharing (file or snippet) • Forwarding /Replying • Interpersonal Interactions, conversations • Networks (e.g., maps visualizing communication flow or strength of relations between persons) 	<p>Direct communication (formal or informal interviews via real-time or asynchronous text, audio, or visual)</p> <p>Special Interest Forums (email- or web-based conversations and archives, e.g., threaded discussion forums, chatrooms)</p>	<p>How is protection of autonomy of participant/author achieved through informed consent or protection of vulnerable persons? How can researcher ensure that author/participant understands and agrees that content or interaction may be used for research purposes? Is the communication archived or easily searchable and retrievable? Is the data subject to open data laws or regulations? How long does the third party provider or ISP preserve the data and where? Could privacy be achieved through anonymization of email content and/or header information?</p> <p>How do terms of service (TOS) articulate privacy of content and/or how it is shared with 3rd parties? Regardless of TOS, what are community or individual norms and/or expectations for privacy? Does the author/subject consider personal network of connections sensitive information? Is the data easily searchable and retrievable? If the content of a subject's communication were to become known beyond the confines of the venue being studied – would harm likely result? Is the conversation thread or forum perceived as public or private by the author(s)/subject(s)? How is profile, location, or other personally identifying information used or stored by researcher? Is the data easily searchable and retrievable? How is informed consent or protection of privacy achieved? How are vulnerable persons identified and protected? If non-active archives are used, how is vulnerability or harm defined and how are potential or actual subjects protected?</p>
<p>Production, presentation, performance</p> <ul style="list-style-type: none"> • Texts (e.g., authored texts, naturally occurring discourse, interview transcripts) • Images (presented or produced by user or captured by researcher) • Video (presented or produced by user or captured by researcher) • Audio (presented or produced by user or captured by researcher) • User motions and movements (any and all activities produced or presented by user and/or captured by researcher) • Configurations or personalization of devices 	<p>Social Networking (e.g., LinkedIn, google+, Facebook, Twitter, Tumblr, Flickr, FourSquare)</p> <p>Personal spaces/ blogs (e.g., homepages, blogs, youtube, and all forms of multimedia presentation)</p>	<p>How do the terms of service articulate privacy of content and/or how it is shared with 3rd parties? Does the author/participant consider personal network of connections sensitive information? How is profile or location information used or stored by researcher? Does author/participant understand and agree to interaction that may be used for research purposes? Does research purpose and design balance possible conflicts between participant and researcher perceptions of public/private and sensitive/nonsensitive? Does the dissemination of findings protect confidentiality? Is the data easily searchable and retrievable? If the content of a subject's communication was ever linked to the person, would harm likely result?</p> <p>Could analysis, publication, redistribution, or dissemination of content harm the subject in any way? If the content of a subject's communication were to become known beyond the confines of the venue being studied would harm likely result? Does the author/participant consider personal network of connections sensitive information? Does author/participant consider the presentation of information or venue to be private or public? Do the terms of service conflict with ethical principles? Is the author/subject a minor?</p>
<p>Locations and movements</p> <ul style="list-style-type: none"> • Physical locations (GPS) • Physical movements • Surfing behaviors 	<p>Avatar-based social spaces, virtual worlds, and online gaming spaces (e.g., Second Life, SIMs, MUDS/MOOS, MPORPG)</p>	<p>Should these virtual worlds be considered "public"? What constitutes "privacy" in such places? Should avatars be considered as persons and afforded the same protections as human subjects? Will the process of requesting consent itself cause harm? How and when should consent be sought? What requires consent? To what extent do users perceive their interactions and communication to be private in these spaces? How do Terms of Service specify researcher presence, anonymity of users, and privacy/confidentiality? To what extent and in what ways could research activities interfere with or compromise a user's play or outcomes in the game? How should researchers juggle their own multiple roles? Could data be used to identify a user's physical location and other sensitive demographic information?</p>
<p>Archived information</p> <ul style="list-style-type: none"> • Demographic information • Bookmark collections • Discussion archives • Data banks • Transaction logs • Clickstream data • Trace data 	<p>Commercial Web Services (e.g., Google, AOL, Yahoo, Bing, MSN, SurveyMonkey, Cloud Storage)</p> <p>Databanks/ Repositories</p>	<p>What are the participant/author's expectations of privacy? Is the data easily searchable and retrievable? Is the data subject to open data laws or regulations? Does the service's privacy policy contradict ethical principles? What measures safeguard data at the site of data collection? How long will the data be stored on the servers? Does this contradict the time frame indicated by the researcher or institutional policies? What happens to the data after the researcher completes work on the service? How are the data destroyed? How will cross-border data be handled if IP addresses are considered by one country to fall under privacy regulations? Where is the data stored? How long will the data exist in the repository? What consent or permission might be needed for subsequent data use? Does the remixing/mashing of data enable identification of individual or group identities or enable any additional risks to participants? In the case of shared data, what conditions were placed on data use by the original researcher, if any? Regardless of conditions, what ethical responsibilities may require consideration by later users? What mechanisms are in place to ensure appropriate data provenance and ownership? How will images/audio be effectively anonymized?</p>

Created by Annette Markham as Appendix 1, AOIR Guidelines: Ethical Decision Making and Internet Research Ethics: 2012. Original document located at <http://ethics.aoir.org/>

APPENDIX 2:

Regional resources and/or FAQs

can be found online at : <http://ethics.aoir.org/>

APPENDIX 3:

Bibliography of resources, categorized by topic area

can be found online at: <http://ethics.aoir.org/>

APPENDIX 4:

Case Studies

can be found online at: <http://ethics.aoir.org/>