Internet Research

Content author:

- Lorna Hicks
  Duke University

There are 5 sections in this module and a short quiz at the end. You will need about 20 minutes to complete this module.

Introduction

The Internet, with an estimated 500 million users worldwide, has much to offer researchers, both as a research tool and the object of study:

- It allows researchers to conduct survey research without the expense associated with mailed surveys.
- It facilitates the process of recruiting groups of subjects with characteristics or experiences not found in local communities.
- It provides a way to pre-screen participants.
- It provides a huge pool of potential subjects.
- It provides the opportunity to observe communications among individuals and to code and analyze an extensive range of phenomena, such as the use of online support groups or the decision-making behaviors of online consumers.
- The Internet itself may be a subject of study, including linguistic and discourse-based analysis of computer-mediated communication and the Internet as a cultural entity.

However, researchers must deal with problems unique to the Internet environment. There are issues that do not have ready solutions, such as the difficulty of confirming the adult status of online subjects and the "personhood" of pseudonymous identities.

Researchers also need to consider the problem of mischievous conduct. It is possible that individuals will intentionally seek to disrupt or damage a study. "The Internet, like any other society, is plagued with the kind of jerks who enjoy the electronic equivalent of writing on other people's walls with spray paint, tearing their mailboxes off, or just sitting in the street blowing their horns." (1)
Investigators must become conversant with new technology, particularly when sensitive information must be safeguarded. If an investigator plans to employ a private company to conduct an on-line survey, the investigator must be able to determine that sufficient protections for the research subjects are in place.

Although this module will not offer solutions for problems inherent in Internet research, it will raise some issues for discussion and offer guidelines for protecting subjects and data in the new environment.

**Module Contents:**

1. Observing online communications
2. Designing Internet research: the consent process
3. Designing Internet research: Privacy issues
4. Assessing risk
5. Technical issues
1.0 Observing Online Communications

One of the most controversial issues regarding Internet research involves the observation of online communications. Part of the attraction of the Internet is that researchers can get verbatim transcripts of interactions without announcing their presence. For example, a researcher could study process and content differences between moderated and un-moderated support groups by observing open groups of the two types.

There is currently no consensus in the research community about whether online communications in open forums constitute private or public behavior. Conclusions about whether they are public or private behavior will affect if and how the regulations are applied.

One view is that the act of posting to an open site, accessible to millions, constitutes public behavior and may be observed and recorded without consent. According to this view, if no identifiers are recorded, such observations may not even meet the definition of research with human subjects. An opposing view is that, in spite of the accessibility of their communications, people participating in some of these groups make certain assumptions about privacy, and that investigators should honor those assumptions. If one subscribes to this second view, either consent would be required or it would have to be waived in accordance with the regulations.
2.0 Designing Internet Research: The Consent Process

The principle of respect for persons requires researchers to give prospective participants adequate information for making the decision to participate in a study. Some of the responsibilities of the researcher are discussed below.

2.1 Ensuring Comprehension

The Internet can be used in many ways to recruit subjects with similar characteristics, thus making it possible for researchers to design consent processes with some confidence that they will be understood. For example, a researcher may recruit subjects via e-mail using a list of people within known parameters, such as undergraduates at the researcher's institution. And topic-specific sites may allow researchers to post recruitment messages for studies related to the topic, again with some prior knowledge about the kind of individuals who will visit that site.

However, organizations such as university psychology departments and national professional associations post web sites for general use. When anyone who has access to the Internet is a potential subject, it may be difficult to ensure comprehension of the consent information.

Methods of enhancing comprehension include:

- Incorporating short questionnaires within the consent process to assess the potential subject's understanding of the information presented in the online consent document and crafting appropriate responses such as routing the subject to additional material.
- Designing interactive consent processes that are tailored to potential subjects' primary language.

Not all research conducted online is free of risk. If subjects are invited to participate in a study that involves risk, comprehension becomes critical.

2.2 Documenting Consent

The regulations for protecting research subjects make a distinction between 1) providing subjects the information they need to decide whether to participate in
a study, and 2) documenting that the information was provided and that the subject agreed to participate. Documentation generally consists of a signed form that investigators keep in their possession.

The process for obtaining consent online usually includes a written statement of the basic elements of consent followed by a statement such as, "Clicking below indicates that I have read the description of the study and I agree to participate." This process does not produce a signed consent form for researchers to store as part of their research records.

The requirement to document consent may be waived in accordance with criteria provided in the regulations. The criterion for a waiver of the required to document informed consent often employed in Internet research is: "When study participation presents minimal risk of harm to the subject and the research involves no procedures requiring consent outside the context of participation in a research study."

When studies in the social and behavioral sciences involve no more than minimal risk to subjects, choosing to participate after reading the consent information provided and clicking to the next screen may be thought of as the virtual equivalent of deciding to participate in a telephone survey, a procedure for which the requirement for documentation of consent is often waived.
2.3 Allowing Withdrawal from a Study

The principle of respect for persons requires that individuals be allowed to withdraw from participation in research without negative consequences. Persons who agree to complete paper surveys and who then wish to withdraw from the study may simply discard the instruments. But an online survey may permanently record responses as they are entered, thus raising the issue of what will be done with the data if a subject wishes to withdraw.

Online surveys are sometimes designed so that one cannot proceed without answering every question. This design compromises the principle that consent is an ongoing process and may be of particular concern when a survey is about a sensitive issue.

Thus, online survey instruments must explain at the outset what options are available, if any, for retrieving and discarding responses, and for some studies it may be appropriate to provide a "no response" option for questions likely to be viewed by subjects as sensitive or intrusive.

2.4 Identifying Underage Participants

People under the age of 18 cannot legally consent to participate in a research study. Unfortunately, researchers recruiting from the general population via the Internet cannot know whether respondents are 10 or 100. Admonishing subjects that they must be 18 years of age to participate does not guarantee compliance.

At present there are no reliable methods for determining the age of Internet users. At one time, the possession of a credit card was a determiner of adult status, but this is no longer true.

One possible response to the problem of identifying minors online is to limit general recruitment to research for which parental permission can be waived in accordance with the Common Rule.

2.5 Using Deception

The Internet offers researchers compelling opportunities to study behavior unmediated by the presence of an observer. Observation of public behavior is often approved by Institutional Review Boards (IRBs) through the exemption process. However, the Internet provides unique opportunities for observational research in private settings. For example, an investigator can join a closed group with relative ease to observe interactions among the members, concealing his or her identity.
3.0 Designing Internet Research: Privacy Issues

Doing research on the Internet creates challenges not present when doing research in traditional settings. Among these challenges are respect for privacy and maintenance of participant anonymity.

3.1 Observing Internet Communications

As noted above, there is not yet a consensus about whether communications in open forums, such as support groups, constitute private behavior. Is observing such behavior equivalent to accessing public information such as a quotation in, say, the Miami Herald?

It is incontestable, however, that the contrast between investigators' assumptions and those of group members is leading to some unanticipated consequences. For example, one web site, rather than closing the group, posted what amounts to a "Researchers Keep Out" sign. A researcher who was observing the group interactions and recruiting subjects felt obligated to stop his study mid-stream. In another case, members of an open support group contacted an IRB to complain strenuously about a researcher making contact with the group, based on the assumption that she had been "lurking" prior to making contact. In both of these cases, the group members felt that observation violated their privacy.

3.2 Being Aware of Identifiers

IRBs and investigators must contend with applying the ethical principle of respect for persons in an environment where the meanings of identity and privacy are shifting. Identifiers in the Internet environment include more than just standard demographics.

Protecting subjects' privacy, as well as the confidentiality of their responses, requires considering ways people identify themselves when using the Internet.

Should individuals' Internet identities, often very different from their "real" identities, be afforded the same protections as their identities off the Internet? Consider a case where a researcher wants to mention someone by (online pseudonymous) name in an article. Would using that name violate the privacy of the pseudonymous identity?
Now, consider a case where a researcher wants to use a quote obtained during a web-based communication. Merely supplying a pseudonym for the author might not be sufficient to ensure anonymity. Someone could, for example, enter the quote into a search engine and possibly find out the identity of the author.

4.0 Assessing Risk

The greatest risk to subjects participating in social and behavioral sciences research is the inadvertent disclosure of private identifiable information that could damage their reputations, employability, or insurability, or subject them to criminal liability.

Many online research studies ask subjects to provide sensitive information. For example, subjects might be asked about their use of illegal drugs, sexual practices, suicidal thoughts, or the use of violence to solve interpersonal conflicts.

Researchers must be aware of the vulnerability of the data at every stage of the research process. Data are vulnerable during collection, transmission, and storage. Protecting data will be covered in more detail in the fifth section of this module.

Potential subjects may or may not be savvy enough to evaluate the confidentiality procedures used in a study. So it falls upon researchers and IRBs to explain risks to subjects, and to make sure that subjects and data are protected. Researchers may also need to caution subjects about Internet use practices that may put them at risk, such as failing to close their browsers or using public kiosks that record "trails" of web use on individual computers.

Potential participants may also not be savvy about securing their own computers. For example, computers can be infected with keystroke loggers. Keystroke loggers can capture screen shots, including e-mail messages, monitor Internet activity, and send data collected before encryption and after decryption, to a predetermined e-mail address or web service. They can be attached to a computer manually, for example, by parents wishing to monitor a child's Internet activity, or implanted by unknown parties via viruses or
worms. Thus, even the most careful security procedures employed by a researcher can be voided if the subjects' computers are not secure. Consent forms should make it clear to participants, when appropriate, that they are responsible for the security of their own computers.

5.0 Technical Issues

Using the Internet to conduct research involves activating a web of communications among computers:

1. Communication between the researcher and the subject.
2. Communication between the subject and the web server.
3. Communication between the web server and the researcher.

Each communication carries the risk of a breach of confidentiality. Investigators and their IRBs need to have the expertise to evaluate the systems put into place to protect both subjects and research data.

Using a commercial service means that researchers do not have to do any web programming and do not have to maintain a web site. Someone else will assign passwords, merge data from different instruments, and download the data into the statistical package of the researcher's choice. Firms offer automatic invitation and tracking systems, including personalized e-mail invitations. They will even archive data for future use by the researcher. However, there may be some conflicts between the firm's practices and the need to protect subjects. It is imperative that researchers have an expert in Internet security review contracts with outside firms.

Case Study. The discussion about technical issues will be based on the following case study:

A researcher is studying writing as a means of recovery from childhood sexual assault. She sends a recruitment message to a random sample of female students and employees through the University's e-mail system. Potential
subjects reply to her via e-mail if they are interested in being in the study. She then sends each one a unique password to the website where consent information and the instruments reside.

The web server is managed by a commercial firm that created HTML versions of the surveys for the researcher. The firm will maintain the data and download it to the researcher's computer in a format of her choice.

The firm will assign each respondent a unique ID number to correlate the responses to the instruments, and will maintain a key linking e-mail addresses with ID numbers. The researcher plans to call some subjects for interviews (those who have consented to be contacted). Therefore she will need access to the key that identifies the subjects' e-mail addresses.

There are significant issues in this study with regard to safeguarding communications that contain private information, as well as to protecting data. The following sections examine these issues in more detail.

5.1 Communication Between the Researcher and the Subject

In the case study, the researcher is relying on e-mail for recruitment and data collection. So she needs to keep the following points in mind:

- E-mail is not secure. E-mail programs maintained by employers often are not private. A survey conducted by the American Management Association found that 52% of companies monitored e-mail use. Personal computers, and the information contained in them, might be available to users other than the research subject, including institutional auditors.
- The recruitment message should very clearly state that the researcher used publicly available sources and that she has no information at all about whether any members of her random sample had been assaulted as children. That way, if a friend, family member, or employer reads the e-mail, it will be clear that the message did not target assault survivors.
- Access to the password could allow unauthorized persons to access the site with mischief in mind, and thus violate subject confidentiality. Ideally, the password should work only once. If a subject wishes to amend her responses, she should have to contact the researcher and ask for a new password.
- Inadequate passwords can easily be cracked. All passwords should
meet security standards to protect the database from intruders.
5.2 Communication between the Subject and the Web Server

Sensitive data must be protected as they move along communication pathways between computers. When a subject enters sensitive data in the blank spaces of a web-based questionnaire, it is analogous to a shopper providing a credit card number when shopping online. Online merchants use a Secure Socket Layer (SSL) protocol that allows secure communications across the Internet. Basically, an SSL protocol encrypts what the user inputs and decrypts it at the web server. Sensitive research information needs the same protection.

In the case study, the researcher needs to ensure that the firm she hires will use SSL, and then explain that technology to subjects, perhaps using the credit card analogy. (If data transfer will not be secure, subjects need to be warned: "Transfer of information across the Internet is not secure and could be observed by a third party.")

5.3 Communication between the Web Server and the Researcher

Sensitive identifiable data need to be protected in transit by using either an SSL protocol or a secure file transfer protocol. If the data are not protected, they may be observed.

The data being transferred in this case study will not have identifiers. Individual records will have unique ID numbers generated by the firm. However, the transfer of the key from the firm to the investigator becomes a security issue. How will the firm get it to her? Will the firm send it in an encrypted e-mail file or keep it in a database and provide her a password? Is there a research-related need for the firm to keep the key?

5.4 Protection of Data

All databases storing sensitive and identifiable information must be protected, regardless of whether they are created and maintained by commercial firms or by individuals. Encrypting the databases probably provides the most security. The researcher's agreement with the firm she has hired should include specific provisions about how, and for how long, it will store the data. The agreement should also clarify that the researcher owns the data, and should describe any circumstances under which the firm will be permitted to use de-identified data.