Security and Confidentiality Considerations in Web Surveys

Guidance to University-based Survey Research Centers

Prepared by:
The Survey Research Center, University of Michigan for the
Association of Academic Survey Research Organizations

December 2, 2009

General principles

Whatever their mode of data collection, researchers who work with human subjects have an ethical obligation to respect the autonomy of their participants and prospective participants, to maximize the benefits and minimize the risks to those participants, and to ensure that the benefits and burdens of their research are distributed equitably. To meet these obligations, researchers must respect the privacy of their participants. They must collect and hold identifying information only when the participants provide it voluntarily and must maintain the confidentiality of that information. In addition, researchers must protect the data they collect against loss, corruption, and exposure.

Do Web surveys pose special risks?

While the maintenance of privacy, security, and confidentiality is essential in all types of research, Web surveys raise novel concerns. Data collected over the Internet may be more vulnerable than other kinds of electronic data to system failures at the data collection phase, particularly interception during transit. As with all electronic data, data stored on servers are vulnerable to system failures, to adverse environmental conditions, and to unauthorized access when stored. Prospective respondents who are unfamiliar with computers or the Internet may be unaware of these risks and may require explicit information about Internet communications before they can consent to participate in a Web survey. Depending on the location of the researcher, the data collection and storage sites, and the respondents, Web surveys may be subject to a varied and changing array of privacy laws. Finally, the software and technology used to implement Web surveys may impose new constraints on respondents – for example, by requiring them to answer every question in a survey before moving on to the next item, or by requiring them to complete the survey in a single sitting.

An additional concern regarding the implementation of Web surveys is the widespread use of third-party vendors to collect and store survey data. While many vendors of Web

survey software are reputable and responsible, the researchers who employ them relinquish control over some aspects of their surveys. This document provides a list of questions researchers should ask themselves before conducting a Web or internet survey – and explains why they should ask. The questions call attention to issues of privacy, security, and confidentiality in Web surveys. Whenever possible, the explanations provide examples of best practices or strategies for implementing adequate protections.

Questions to ask before conducting a Web survey

What measures will be taken to prevent loss of data?

As with all data stored on computers, hard drives are vulnerable to damage from environmental conditions that would not pose a threat to paper questionnaires. For this reason, the server on which the data are stored should be housed in a room with environmental controls that maintain temperature, humidity, and air flow at safe levels. Other desirable security measures include redundant systems at the data collection and storage sites, to ensure continued operation in the event that a component fails, and frequent data backups.

Who will have access to the server on which the data is stored? What measures will be taken to protect the data against unauthorized access?

Data collected over the Internet should be stored in a secure location and should be accessible to a limited number of authorized individuals. It should also be protected by electronic security measures, such as firewalls, passwords or encryption. Individuals who have access to the server should make a commitment to protect the privacy of the survey respondents and to maintain the confidentiality of the data. Researchers should ask how and how often server traffic will be monitored.

If the survey is contracted out to an organization or vendor of Web survey software, the data may be accessible to employees of that organization. Researchers should also ask about the security practices and privacy policies of these organizations. For example, will employees of the organization sign pledges of confidentiality and maintain adequate privacy protections? Are firewalls in place? Do any outside parties have access to the servers?

What would be the consequences in the event of a security breach?

With any method of data collection and storage – and even with strong safeguards in place – it is possible that the data will be lost, corrupted, or accessed by unauthorized individuals. Researchers should ask what would happen if the data were compromised. Who would assume responsibility for the security breach? Who would contact the respondents to tell them about the problem? What remedies might be offered?

Can the data be encrypted during transmission?

In surveys conducted over the Internet, there is a risk that data could be intercepted by an unauthorized party during transmission from the respondent's computer to the researcher's (or organization's) computer. The use of encryption technology protects against this risk by transforming the data so that it will be meaningless to anyone who intercepts it. Survey data should always be encrypted when it contains identifiable information, such as phone numbers, e-mail addresses, credit card numbers or social security numbers or IP addresses. Data should also be encrypted when descriptive variable names are transmitted with responses that might be intelligible to an intruder.

Can the researcher avoid collecting identifying information about the respondents? If not, can the identifiers be stored separately from the survey responses?

Any data stored on servers connected to the Internet may be more vulnerable to unauthorized access than data stored on paper questionnaires in a physically secure location. For this reason, researchers should take special care to protect the confidentiality of electronic data. If identifying information is not needed for research purposes, researchers should avoid collecting it or should request that it be deleted before the data are stored.

If identifying information must be retained, researchers should ask whether the identifiers can be stored separately from the survey responses. One way to achieve this result may be to collect identifying information using a separate survey and to link identifiers to survey responses using random identification numbers.

When collecting data over the Internet, researchers should be aware that IP addresses are a form of potentially identifying information, especially if combined with other data about the respondent. IP addresses are numbers that identify the locations of machines connected to the Internet. Because they could be used to identify the computers from which survey data were submitted, IP addresses could potentially be used to identify respondents and to link respondents to their survey data. Researchers should take the same precautions with regard to IP addresses that they take with regard to other identifying information.

Will identifying information about the survey respondents be collected by the vendor? If so, how will it be stored and used?

Some survey vendors may routinely record and store identifying information and use it for their own purposes, or for the use of third parties. This practice may pose an ethical problem for some researchers. For example, researchers who adhere to the standards of the American Association for Public Opinion Research commit that they "shall not disclose or use the names of respondents for non-research purposes unless the respondents grant us permission to do so."

Researchers should ask what data, if any, will be collected about their respondents. Will names, phone numbers, e-mail addresses, credit card numbers, social security numbers, or other identifying information be collected? Will IP addresses be recorded? Will information about browsing preferences or habits be tracked? If such data are collected, researchers should ask how it will be stored and used. What measures will be taken to protect the security and confidentiality of the data collected? For what purposes will these data be used? Will data be made available to third parties and, if so, how will those parties store and use the data?

Can survey submissions be authenticated?

Authentication is the process by which researchers track and, in some cases, restrict access to their Web surveys. Authentication protects the integrity, quality and confidentiality of survey data by reducing the likelihood that data will be viewed or submitted by someone other than the intended respondent as well as limits the ability of respondents to make multiple submissions to the same survey.

In general, researchers can choose from a wide variety of authentication methods and the approach may well depend on the nature of the data collection. The least restrictive methods allow the researcher to track the source of survey submissions, but place few or no restrictions on access to the survey. For example, the researcher can record the IP addresses of the machines from which surveys are submitted and can examine the addresses for duplicates. Similarly, the researcher can install cookies on the machines respondents use to complete the survey and can use the cookies to identify machines that have already accessed the survey. These methods do not prevent problematic submissions in every instance, but they can be used when the researcher is not working from a list-based sample and they do not burden the respondents with login requirements.

A number of more restrictive authentication methods allow the researcher to limit participation in the survey to a defined group of respondents. Some Web survey software can generate a unique URL for each respondent and can use information embedded in the URLs to direct respondents to personal copies of the survey. Alternatively, the researcher can provide respondents with a URL to the survey Web site and can require them to enter an ID, PIN, password, or some combination of these credentials in order to access the survey. This manual method of authentication requires more effort from respondents than does automatic authentication through a unique URL, but it may also reassure respondents that their privacy will be protected. Finally, the researcher can vary the balance of security and ease of access by combining automatic and manual methods of authentication. For example, the researcher can require respondents to enter a short, simple credential in order to access the survey and can embed longer or more complex credentials in a unique URL.

When working with vendors, researchers should ask which authentication methods are available. They should think carefully about the kinds of information they will collect and should determine which, if any, of the available methods offer an appropriate balance of security and ease of access to the survey. When the survey does not cover sensitive topics and when it is not necessary to restrict access to the members of a list-based sample, authentication using IP addresses or cookies may be adequate. When the survey poses more significant risks or when it is important to control access, researchers should use more robust methods of authentication.

Is the survey subject to any local or national privacy laws?

Laws regulating research on the Internet vary from place to place and have changed quickly in recent years. When conducting a Web survey, researchers should determine which laws, if any, apply at their home location and at the data collection and storage sites – including the locations of the respondents and of any vendors or contractors. Researchers should be aware that privacy protections are stronger in some places than in others. For example, laws in the European Union are more restrictive regarding survey respondents than are laws in the United States. When research activities cross local or national boundaries, researchers may be required to respect the most stringent of the applicable laws.

How will the researcher ensure that consent is fully informed?

Prospective respondents who have limited experience with computers or the Internet may not appreciate the privacy risks involved in transmitting confidential information

electronically. To ensure that respondents have all of the information they need to make an informed decision about participation, the informed consent document for a Web survey should include a brief description of the general content of the survey and of any risks, even if they are similar to day-to-day computing risks.

Will the software and technology used to implement the survey afford respondents adequate control over the manner in which they complete the survey?

In some cases, the software and technology used to implement a Web survey may impose constraints on respondents that may violate principles of voluntariness and autonomy. One such constraint is a restriction on the ability of respondents to skip individual survey questions. Some Web survey software requires that respondents answer every question in a survey before moving on to the next question. When using a vendor, researchers should ask whether it is possible to allow respondents to skip questions they do not wish to answer. If utilizing their own software, programming should allow skipping questions.

A second constraint is the requirement that respondents complete the survey in a single session. Ideally, respondents should be able to return to a survey they have started and resume answering questions where they left off, although the importance of this constraint may depend on the content and length of the instrument. Obviously, surveys that require respondents to look up information or take a considerable time to complete, should allow for multiple sessions. If respondents will have the option to complete the survey in multiple sessions, researchers should ask whether they will be able to submit responses from multiple computers, or whether they will be required to complete each session at the same computer (as may be the case if respondents are tracked across sessions using IP addresses or cookies).

Web Survey Security and Confidentiality Checklist

What measures will be taken to prevent loss of data?	
Who will have access to the server on which the data is stored? What measures will be taken to protect the data against unauthorized access?	
What would be the consequences in the event of a security breach?	
Can the data be encrypted during transmission?	
Can the researcher avoid collecting identifying information about the respondents? If not, can the identifiers be stored separately from the survey responses?	
Will identifying information about the survey respondents be collected? If so, how will it be stored and used?	
Can survey submissions be authenticated?	
Is the survey subject to any local or national privacy laws?	
How will the researcher ensure that consent is fully informed?	
Will the software and technology used to implement the survey afford respondents adequate control over the manner in which they complete the survey?	

References

American Association for Public Opinion Research. "AAPOR code of professional ethics and practice." Retrieved October 21, 2009, from http://www.aapor.org/AAPOR_Code.htm.

Armstrong, Rebecca (2003). "Human subjects research and the Internet: Ethical dilemmas," Research Review, Office of Research Administration.

Benfield, Jacob A. and William J. Szlemko (2006). "Internet-based data collection: Promises and realities," *Journal of Research Practice*, 2(2), Article D1. Retrieved October 12, 2009, from http://jrp.icaap.org/index.php/jrp/article/view/30/5.

Berry, David M. (2004). "Internet research: Privacy, ethics and alienation: An open source approach," *Internet Research*, 14(4), 323-332.

Buchanan, Elizabeth A. and Erin E. Hvizdak (2009). "Online survey tools: Ethical and methodological concerns of human research ethics committees," *Journal of Empirical Research on Human Research Ethics*, 4(2), 37-48.

Couper, Mick P. (2008). *Designing Effective Web Surveys*: Cambridge University Press.

Ess, Charles (2007). "Internet research ethics," in Adam N. Joinson, Katelyn McKenna, and Tom Postmes eds. *The Oxford Handbook of Internet Psychology*: Oxford University Press, 487-502.

Eysenbach, Gunther (2005). "Evaluating the organizational impact of healthcare information systems," in James G. Anderson and Carolyn E. Aydin eds. *Using the Internet for Surveys and Research*: Springer New York, 2nd edition, 129-143.

Frankel, Mark S. and Sanyin Siang (1999). "Ethical and legal aspects of human subjects research on the Internet," Report of a Workshop, American Association for the Advancement of Science.

Im, Eun-Ok and Wonshik Chee (2002). "Issues in protection of human subjects in Internet research," *Nursing Research*, 51(4), 266-269.

Kraut, Robert, Judith Olson, Mahzarin Banaji, Amy Bruckman, Jeffrey Cohen, and Mick Couper (2004). "Psychological research online: Report of Board of Scientific Affairs Advisory Group on the Conduct of Research on the Internet," *American Psychologist*, 59(2), 105-117.

Naglieri, Jack A., Fritz Drasgow, Mark Schmit, Len Handler, Aurelio Prifitera, Amy Margolis, and Roberto Velasquez (2004). "Psychological testing on the Internet: New problems, old issues," *American Psychologist*, 59(3), 150-162.

Nosek, Brian A., Mahzarin R. Banaji, and Anthony G. Greenwald (2002). "E-research: Ethics, security, design, and control in psychological research on the Internet," *Journal of Social Issues*, 58(1), 161-176.

Walther, Joseph B. (2002). "Research ethics in Internet-enabled research: Human subjects issues and methodological myopia," *Ethics and Information Technology*, 4, 205-216.