INTERNET-BASED RESEARCH

This guidance document is intended for researchers conducting recruitment, informed consent, and/or data collection procedures over the Internet. Should you need additional assistance, please contact OPHS at 510-642-7461 or ophs@berkeley.edu.

Table of Contents:

A. Introduction
B. Specific Guidance
1. Recruitment
2. Informed Consent
3. Data Collection
4. Data Security

C. Glossary of Terms

A. Introduction

Recent studies have shown that Internet-based research is becoming more and more commonplace. Popular because it is a quick way to gain access to a large number of respondents without expending too many resources, Internet-based research is, increasingly, the research method of choice for surveys. Observations, interventions, and analysis of existing data are also commonly used methods of Internet-based research. While this may be good news for investigators, Internet-based research brings up difficult issues concerning human subject protections in the application of federal regulations 45 CFR 46 and the ethical principles of the Belmont Report. Topics such as privacy, confidentiality, recruitment, and informed consent become complicated when research is conducted online. This guidance intends to shed light on a variety of issues surrounding Internet-based research and to help investigators design studies that are in line with 45 CFR 46.

B. Specific Guidance

1. Recruitment:

a. Internet-based procedures for advertising and recruiting potential participants must follow the Committee for the Protection of Human Subjects (CPHS) guidelines for recruitment that apply to any traditional media, such as flyers and newspaper ads. Examples of Internet-based recruitment methods include emails, online advertising, and chatroom postings. For non-exempt research, these texts must be submitted for review and approval before use in the field.

b. The proper identification and qualification of subjects is a challenge in Internet-based research. Without face-to-face or voice-to-voice interaction, it is difficult for investigators to be sure that participants are not misrepresenting themselves. In

certain situations, investigators should discuss measures taken to authenticate subjects. These situations may include studies for which authentication of subjects is important to the validity of the data or that consist of particularly sensitive topics. Examples of such measures include:

i. Providing each study participant (in person or by U.S. Postal Service mail) with a Personal Identification Number (PIN) to be used for authentication in subsequent computer- and Internet-based data collection. In this example, the PIN used must not be one that could be used by others to identify the individual (e.g. social security number, phone number, birth date, etc.).

ii. Minors may be screened out by checking for Internet monitoring software such as SafeSurf and RSACi rating or using Adult Check systems. This can be necessary if the study presents more than minimal risk to subjects or asks particularly sensitive questions. Or investigators might want to increase the validity of their study by screening out minors if their research is focused on adult subjects. On the other hand, in most studies involving no greater than minimal risk, it is sufficient for the informed consent document to simply ask participants to confirm that they are the appropriate age of majority.

2. Informed Consent:

a. Investigators conducting non-exempt research must follow the CPHS Informed Consent Guidelines and include required elements of informed consent when generating consent documents. When online surveys are employed, the CPHS consent template for online surveys may be adapted. Investigators conducting research that qualifies for exempt review should refer to the CPHS guidelines on exempt research for information regarding informed consent.

b. In general, investigators conducting Internet-based research with minors must obtain both child assent and parent permission. Researchers may request a waiver of parent permission provided the study fits the appropriate criteria.

c. CPHS generally accepts the use of "I agree" or "I do not agree" buttons (or other electronic methods for indicating affirmative consent) on online pages in lieu of signatures. For surveys sent to and returned by participants through email, investigators should include a consent document and inform participants that submitting the completed survey indicates their consent. This would constitute unsigned consent. In order to utilize this consent procedure, the investigator must request a waiver of documented consent. See CPHS Informed Consent Guidelines for further information.

d. If the CPHS determines that documented consent is required, the consent form may be mailed or emailed to the participant who can then print and sign the form and return it to investigators via email, postal mail, or fax. Note that electronic signatures do not satisfy the requirement for documented consent.

e. The process of requesting consent should not disrupt normal group activity. Researchers need to be particularly sensitive of this when entering online

communities and chatrooms as the process of requesting consent is often perceived as disruptive. If seeking informed consent will harm the validity of a study or make the research impracticable, it may be possible to obtain a waiver of consent provided the study meets the appropriate criteria. When requesting a waiver of informed consent, issues regarding deception or incomplete disclosure may need to be addressed in the researcher's eProtocol application. Please see the CPHS Guidelines on Deception and Incomplete Disclosure for further instruction.

f. Personas, or avatars, are social identities that Internet users establish in online communities and websites. These personas allow individuals to reveal varying levels of personal information and also allow them to navigate the virtual world as a particular character or alter-ego. Names of Internet personas (characters or avatars) or real names may be used in reports and publications only with consent from the participating individual (see Data Security below for more information). In these situations, specific language concerning the release of identifiable information must be included in the informed consent document and specific consent must be sought from subjects for this release. If research participants give consent to be identified, data must still be secured properly to avoid any misuse by a third party.

g. Collecting data over the Internet can increase potential risks to confidentiality because of the frequent involvement of third party sites and the risk of third party interception when transmitting data across a network. For example, when using a third party website to administer surveys, the website might store collected data on backups or server logs beyond the timeframe of the research project. In addition, third party sites may have their own security measures that do not match those of the investigators' (see Additional Considerations below for more information). Participants should be informed of these potential risks in the informed consent document. For example:

   i. "Although every reasonable effort has been taken, confidentiality during actual Internet communication procedures cannot be guaranteed."

   ii. "Your confidentiality will be kept to the degree permitted by the technology being used. No guarantees can be made regarding the interception of data sent via the Internet by any third parties" (Pennsylvania State University).

   iii. "Data may exist on backups or server logs beyond the timeframe of this research project."

3. Data Collection:

Internet-based data collection methods can range from the use of existing data and observations to interventions and survey/interview procedures.

a. *Existing Data:* Research utilizing data that are both existing and public is not considered human subjects research and does not require CPHS review (http://cphs.berkeley.edu/secondarydata.pdf). Data only accessible through special permission are generally not considered public. However, if steps are required to access data (e.g., registration/login, payment, etc.), but access is not restricted beyond

these steps (e.g., anyone who creates a username and password can access the data), the data may qualify as publicly available. When determining whether or not data are public, the investigator must decide if there exists an expectation of privacy. If it is determined that the data were not intended for public use, even if the data are technically available to the public, the data should be considered private. For example, data available on WikiLeaks are technically public but include information about individuals who did not authorize the release of such data (http://en.wikipedia.org/wiki/WikiLeaks). Researchers wishing to access data that contain identifiers and are not publicly available must first obtain CPHS review and approval.

b. *Observations:* When online research procedures are employed, the investigator must be sensitive to the definition of "public behavior." Despite navigating in a public space, an individual may have an expectation of privacy, and investigators need to be sensitive to that expectation. For example, an investigator wishes to collect data from discussions posted in an online community support group for substance abusers. The online community is technically public, in that anyone can view the discussions and join the group, but some group participants are there to provide personal experiences and support regarding substance abuse and may believe that all discussions and personally identifiable information will remain private.

   Research in spaces that are not public or that maintain an expectation of privacy must be reviewed at the non-exempt level. In order to make this determination, it is important that investigators be familiar with the online space in which they intend to conduct research. Not only do investigators need to have an insider's viewpoint in order to know whether or not participants have an expectation of privacy, but investigators will often be met with hostility if they are not sensitive to the online community's expectations. Participants of an online community may see the presence of a researcher as intrusive. If an investigator has prior experience in an online community and is already known to its participants, the researcher may have a better chance of being welcomed into the space.

c. *Chatrooms:* When navigating in a chatroom, it is important that those present are able to let the researcher know if they are not comfortable with the researcher's presence and that the researcher respects these wishes. Because access to chatrooms can prove difficult for investigators and chatroom participants are not always eager to have a researcher in their midst, one suggested technique is for investigators to create their own chatrooms just for research purposes. Investigators can greet individuals joining the chatroom with a message informing them about the study and asking them for their informed consent. This is a good way to be sure that all participants are fully aware of the research and have consented to participate.

d. *Surveys*: Survey research is one of the most common forms of Internet-based research. Researchers are advised to format survey instruments in a way that will allow participants to refuse to answer specific questions. For example, the list of responses can include an option such as, "Decline to answer." In addition, participants must always be given the option to withdraw from a study, even while in the middle of a survey.

Use of Qualtrics, Mechanical Turk, and other online survey tools is generally permitted for most minimal risk studies employing online survey procedures. Investigators should indicate within the protocol where the survey will be hosted (i.e., on the survey platform or if there will be a link to an external survey site). The CPHS recommends that investigators use Berkeley Qualtrics to collect survey data.

Investigators should review confidentiality measures and data security policies for the given online survey tool and make sure that they are described in the protocol. If security measures are not in line with what CPHS requires, use of the given survey company may not be approved. Research participants also need to be informed of data security measures.

For more information please see:
CPHS Guidelines on Mechanical Turk for Online Research.

e. *Interviews:* Conducting interviews online allows researchers to gather information from respondents who would be difficult to contact otherwise, such as a very geographically dispersed population. Interviews may be conducted over the Internet using cross-platform communication technology such as Google Chat, WhatsApp, Skype, etc. When using such an application, researchers should state in the protocol which method of communication they will use. For example, if participants will be voice-dialed, researchers need to affirm that the video function will be disabled. When conversing with a research participant via chat, investigators should take into account the inability to read visual and auditory cues, which can lead to possible misinterpretation of both questions and responses. Voice intonation and facial expressions are often used to convey meaning. Thus, investigators may need to ask clarifying questions in order to accurately interpret responses, and provide additional information in order to be sure that participants understand the questions.

f. *COPPA:* Operators of commercial websites and online services directed towards children under 13 years of age that collect personal information from these children must comply with the Children's Online Privacy Protection Act (COPPA). The goal of COPPA is to protect children's privacy and safety online, in recognition of the easy access that children often have to the web. COPPA requires website operators to post a privacy policy on their website and create a mechanism by which parents can control what information is collected from their children and how such information may be used.

For more information please see:
http://www.ftc.gov/privacy/privacyinitiatives/childrens.html.

g. *Additional Considerations:*

i. When recording material that would otherwise be temporarily posted online, consideration should be given to whether the act of recording this information potentially creates risks for subjects. For example, information is, at times, posted on the Internet by a third party without the consent of the involved individuals. If a study is likely to record illegal or socially undesirable activity, the investigator

should judge whether or not recording this information would create risk for the subjects and, if so, reconsider using or retaining the data.

ii Investigators should make sure to review any applicable Terms of Service (TOS). TOS outlines the rules a person or organization must observe in order to use a service. Internet service providers (ISPs) and all websites that store personal data for a user have TOS, in particular, social networking sites, online auctions and financial transaction sites.

4. Data Security

Investigators must consider additional data-security issues when conducting Internet-based research. (See the CPHS website for additional data security guidance and policy requirements.)

a. Researchers must take special care to treat online identities (personas or avatars) and their corresponding character names just like real ones. People care about the reputation of their personas and these aliases can usually be traced back to real-world names.

b. Even when it is not the intention of the researcher to collect identifiable information, Internet protocol (IP) addresses are potentially identifiable; thus, if IP addresses will be collected, proper confidentiality measures must be in place in order to protect the subject's identity. These measures include password protection and encryption.

c. All identifiable or coded data transmitted over the Internet must be encrypted. This helps ensure that any data intercepted during transmission cannot be decoded and that individual responses cannot be traced back to an individual respondent. It is important to note that encryption standards vary from country to country, and there are legal restrictions regarding the export of certain encryption software outside US boundaries. It is the investigator's responsibility to research possible restrictions and plan data security measures accordingly.

d. The level of security should be appropriate to the risk. For most research, standard security measures like encryption and secure socket layer (SSL) will suffice. However, research involving particularly sensitive topics may require additional protections, such as housing data on a professionally managed server.

C. Glossary of Terms (PRIM&R, 2010)

Blog: A website used as a journal; can be personal or professional in nature.

Chatroom: An online locations where individuals can come together to have text-based chat discussions that occur in real time.

Cloud computing: Distant storage or data management servers typically owned and operated by a third party.

Confidentiality: Pertains to the treatment of information that an individual has disclosed in a relationship of trust, and with the expectation that it will not be divulged without permission to others in ways that are inconsistent with the understanding of the original disclosure.

Cookie: A text file placed on user's computer by a website or web server. Often used to keep track of individuals as they navigate a site, and more broadly, the web.

Encryption software: A piece of software that is used to obfuscate information to all of those who do not have the means to decrypt the information.

Internet Protocol (IP) address: A numeric address assigned to every computer that connects to a network, or more commonly, the Internet.

IRC: Internet Relay Chat, a protocol used for hosting and participating in chatrooms.

Lurking: A behavior specific to online communities, wherein an individual remains silent, observes, and does not participate in the community.

Online persona: An online character or avatar used by an individual.

Online survey: Any tool used to collect responses to survey questions via the Internet.

Privacy: Control over the extent, timing, and circumstances of sharing oneself (physically, behaviorally, or intellectually) with others.

Publicly available: The general public can obtain the data and they are readily available to anyone (without special permission/application) regardless of occupation, purpose, or affiliation.

Secure website: A site that conforms to best current security practices, such as use of the Secure HyperText Transfer Protocol (https), application of relevant patches, and sound auditing and certificate management.

Virtual community: A group of individuals networked together through association with a virtual environment, homepage, or other Internet medium (e.g., SecondLife).

D. References

Bruckman, A. (April, 2002) *Ethical Guidelines for Research Online*. Georgia Institute of Technology.

Buchanan, E. (2009). *Online Survey Tools: Ethical and Methodological Concerns of Human Research Ethics Committees*. Journal of Empirical Research on Human Research Ethics, pp. 37-48.

Buchanan, E, Gallant, D., Miller, M. (December, 2010) *Navigating Research Regulations and Research Ethics in the Internet Age.* PRIM&R San Diego Conference.

Buchanan, E., Odwazny, L. (March, 2012) *Ethical Internet Research: Informed Consent Regulations and Realities*. PRIM&R webinar.

Guidance for Computer and Internet-Based Research Involving Human Participants. Retrieved from http://irb.uconn.edu/internet_research.html

UCLA Guidance on Research Involving the Internet. Retrieved from http://ora.research.ucla.edu/OHRPP/Documents/Policy/8/Internet_Research.pdf