Recommendations and Policies Relating to Online Data Collection and Online Data Storage

Xavier University Institutional Review Board

Two primary issues in conducting online research include privacy and the freedom from undue influence. Although these issues arise in any research, issues of privacy are particularly salient to users in modern online environments. The following represent the IRB's recommendations to researchers interested in conducting online research. Issues beyond those addressed here will be considered on a case-by-case basis.

I. Appropriate Recruitment of Participants

Recruitment of participants for online studies must observe the fundamental underlying principle of **respect for persons**. That is, recruitment must be done in a manner that demonstrates the researcher's respect for the individual whose participation is requested. A key concern in recruiting participants for online research studies is privacy. Although every study is different and consideration will be made on a case-by-case basis, the following general guidelines should be followed, with respect to privacy and the protection of participants in this type of research.

Avoid mass unsolicited emails. The sending of mass unsolicited emails (i.e., "spamming") is strongly discouraged, as it may be viewed as an invasion of participants' privacy. Sampling should be targeted and purposive. When unsolicited emails are sent, full disclosure of how the email addresses were obtained should be provided in the cover email sent to participants, and should be re-iterated on the informed consent document. The following language represents two examples of how this disclosure could be provided.

Example 1.
"Your email address was obtained from your university's website, and you are receiving this email because you are involved in the training of nurses and nurse practitioners."

Information accessible on a university's website, as in Example 1, is in the public domain. Because of concerns relating to privacy, however, it is still in the researcher's best interests to make full disclosure as to this element of their research. Note that, absent information about how the email list was generated, potential participants may infer that some professional organization of which they are members provided their contact information and may feel obligated to respond based on the ambiguity of the research situation. The IRB does not view recruitment based on ambiguity as appropriate.

Example 2.
"Your email address was obtained from your company's Director of Human Resources, Susan Day. She has reviewed the research protocol and agreed to allow me to distribute my study to you and your colleagues. However, she will never know whether you personally participated in the study, nor will she ever have access to any of the information you provide. All information from this study will be reported in aggregate form only, and your participation (or non-participation) will have no effect on your employment."

Note that the language in the second example includes several elements that will have to be included in the informed consent document as well. Redundancy of the recruitment email with the informed consent document is encouraged. Particularly when engaging in targeted recruiting that involves the support or endorsement of one or more organizations, it is important that participants understand precisely why they are being contacted, how this contact was initiated, and that despite the support of their organization *they are still free to choose not to participate.* Obtaining support from organizations for online organizations is addressed in Section III.

II. "Viral" (or snowball) sampling strategies

Online data collection often capitalizes on the tendency of individuals to forward information to friends, relatives, co-workers, and other acquaintances who may be interested in the topic. This form of dissemination is commonly referred to as "viral" or "snowball" distribution because of the tendency of information to spread from person to person based on level of contact.

The IRB does not prohibit viral sampling. However, one cautionary note is necessary, and several elements should be included in the consent form that would not otherwise be needed if this strategy is used.

Cautionary note*:* From a methodological perspective, viral sampling relies on extreme levels of self-selection into the study. If you utilize viral distribution of a study, you are virtually guaranteeing a non-random sample of participants. This is purely a methodological issue and one that the researcher should consider as a potential limitation to the study, but as with most methodological issues, it is not something that the IRB will view as cause to not approve the study in and of itself.

Assurance of sampling strategy appropriateness: A related issue that must be addressed to the IRB's satisfaction is whether the use of a viral sampling strategy will be appropriate to allow meaningful testing of the research question. It must be made clear that the sampling strategy will result in a sample that will not invalidate the results of the study. If a viral strategy would predictably result in the collection of data that are ultimately not usable, the research effort

reflects an inappropriate use of participants' time, and therefore a lack of **respect for persons**. As such, the researchers must demonstrate awareness of the potential weaknesses of viral sampling, and cogently argue why such a strategy is appropriate in their study.

***Consent form additions**:* It is recommended to include your informed consent document as the first element of the online study, even if elements of consent were documented in the initial email (or other) contact.

Disclosure of the sampling strategy and its meaning to participants are needed, to ensure that **no** support of the project by the employing organizations, schools, internet service providers, or other owners of email hosting software is implied. The following language, or a variation thereon, is strongly recommended for inclusion in your informed consent document if you utilize viral sampling.

"This study utilizes a "viral" sampling strategy in which participants are encouraged to forward the study link to others who might be interested. I understand that, if I received information about this study at any non-personal email address, there is no endorsement or awareness of this study by my employer, my school, or any other organization. My decision to participate or not participate in the study will have no bearing whatsoever on my employment, education, or any services provided by the owner of the domain at which this request for participation was received. My email address will never be associated with my responses, and my participation will not be reported to the owner of this email address."

Note that a number of the elements included in the example above are simply extensions of pieces of the informed consent document that the IRB already expects to see. The viral strategy, however, necessitates an explicit statement that there is no connection between the researcher and the owner of the email address/hosting domain, again to prevent the participant from inferring such a relationship exists and being unduly influenced by having received the request to participate at an "official" email address.

As a final note, relating to the informed consent document, researchers must provide an accurate estimate of the length of time necessary to complete the survey. This estimate should be included both in the consent documentation and in any preliminary communications (e.g., recruitment emails).

III. Supporting materials from organizations

As with all research involving data collection at specific organizational or institutional sites, online research conducted with the support or assistance of any organization or institution must provide documentation of this support. A letter from an authorized representative of the organization, on official letterhead, must be provided to the IRB prior to approval of the research study.

It is both important that any organization represented as supporting the research actually does support it (where "support" does not necessarily indicate anything beyond allowing the research to be conducted) and that participants not feel as though their employment or benefits are in any way conditional upon their completing the study. As such, the letter of support must be obtained from any organization which: (a) allows distribution of online study materials via company mailing lists; (b) distributes, or allows one of its employees to distribute, online study materials on behalf of the researcher; (c) posts links to the study on its website, intraweb, e-newsletter, Facebook/social networking site; or (d) otherwise makes employees aware of the study, or allows the researcher to do so.

The letter of support, in combination with assurance that participants can withdraw from the study at any time, serves to protect participants from undue influence or coercion.

IV. Online Privacy Issues

Collection of IP Addresses. Most online data collection tools have a setting which disables the tracking of IP (internet protocol) addresses and geo-locations. IP addresses are unique identifiers which, if enabled, allow users to be tracked, sometimes at the level of the individual computer. Unless a compelling reason exists to track IP addresses, *the IRB requires that you disable such tracking* to protect participant confidentiality/anonymity. Surveys designed to be anonymous or confidential should therefore not be tracked. However, when conducting longitudinal research it may be necessary to track participation. In such a case, collecting IP addresses may be one means of accomplishing this goal. Whatever method is to be used for tracking participation over time, it must be made clear to participants both (a) that their participation will be tracked, over time, and (b) how this will be accomplished. Although the IRB does not recommend tracking IP addresses for this purpose for technical reasons, such tracking does represent one possible option.

General Online Privacy Concerns. In order to further protect confidentiality, investigators may encourage participants to do some or all of the following:

(1) Because some employers may use tracking software, participants may want to complete the online study on a personal computer.
(2) Participants should not leave the study open if using a public computer or a computer others may have access to.
(3) Participants may wish to clear their browser cache and page history after completing the study.

V. Collection of identifying information

If the identities of participants must be tracked (for participation credit, entry into award raffles, etc.), a separate database must be constructed to contain participant names and contact information. At no time may participant names and responses to online surveys or other data collection formats be stored in the same database. When using online data collection tools such as Qualtrics, this means that after completing the initial survey, participants should be provided with a link to a second, separate, Qualtrics survey. This second, separate, survey should only contain those questions necessary to identify the individual in order to award appropriate credit (e.g., participant name and instructor) and should not contain any questions that will be analyzed in order to test the study's hypotheses.

When collecting protected health information or other information that poses a reasonable security risk to the participant, there are two options. If the researcher is going to use web-based tools to collect protected health information, then it is preferred that the web-based application should be 21CFR part 11 compliant. If you use desktop tools (e.g., Excel, Access) then it must be kept on a desktop in a secure location or in the department intranet. Under no circumstances should protected health information be accessible to anyone other than the principal investigator or specifically designated project personnel.

Information Regarding Online Data Storage

Researchers are responsible for protecting the privacy and confidentiality of their study-related data, whether it is in tangible (e.g.. interview transcripts) or electronic format. Regardless of format, all data must be securely stored so that only those who are authorized to do so can access the information. Storing data online is acceptable as long as measures are taken to ensure that its security is protected.

For additional information regarding storing data online at Xavier, please see Xavier's Cloud Storage Policy. Only Xavier approved online storage locations are permitted.

Additional Security Practices

Researchers can implement additional strategies to protect the security of data stored in any Xavier approved online service. These include:

1. Password-protect all devices (i.e.. computers, tablets, smart phones) that can be used to access the data (Carney et al., 2000).
2. As with hard copies of research materials, information containing identifiers should be stored in files that contain no other information (Carney et al., 2000). An additional level of security can be added by password-protecting the individual files (i.e.. a feature of Microsoft Word).
3. Avoid sharing files with research team members by email attachment or emailing a hyperlink to the files.