THE UNIVERSITY OF ARIZONA
Research, Innovation & Impact

Internet and Social Media Research

Background

Internet and social media offer researchers the opportunity to extend the reach of their recruitment efforts and to integrate novel approaches to data collection through sustained interaction with subjects. The continuously evolving nature of social media, however, generates uncertainty for both researchers and Institutional Review Boards (IRBs) regarding the management of risk to human subjects. Issues of privacy, confidentiality, informed consent and potential risks to subjects arise when researchers and subjects interact virtually. For example, a subject participating in research-related social media may mistakenly divulge personally-identifiable information in a public forum that they assume to be private or monitored by investigators.

The UA IRB believes that internet and social media-based research protocols must address fundamentally the same risks (e.g., violation of privacy, legal risks, and psychosocial stress) and provide the same level of protection as the more traditional non-electronic methods of research involving human participants. All studies, including those using internet and social media technologies, must:

1. ensure that the procedures fulfill the principles of voluntary participation and informed consent;
2. maintain the confidentiality of information obtained from or about human participants; and
3. adequately address possible risks to participants including psychosocial stress and related risks.

In general, researchers utilize social media for two purposes: recruitment (both one-way ads and interactive recruitment) and as a component of research interventions or data collection methods.

What is the Difference Between Publicly Available and Private Data?

A common misconception of social media, and internet-based research in general, is that if it is on the web then the information is public. Data are available on the internet in a wide array of formats and degrees of identifiability. When collecting data online, a researcher should be aware of how participants typically use a given space, forum, or platform and what expectations they have for observations or analysis of their data. Did participants intend for a researcher to access their data or for it to become part of a research study? Should a researcher consider data private or publicly available? It can be difficult to decide what is “fair game,” and failure to adhere to regulatory definition of the difference between private and public can pose significant risks to participants or others in the research process.

The regulations imply a difference between data collected from public observation and private data collected from interaction or intervention with a research participant.

- Private Information consists of data accessed through special permission, password protection, or registration. Private information "includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public." Researchers should have a clear understanding of the expectations participants have regarding their privacy as they use an online space. A website or platform's Terms of Service often outline privacy expectations on behalf of users.
- "Public observation" occurs in spaces where a reasonable person does not have an expectation of privacy. This definition of "public observation" applies in online spaces when there is no assumption of privacy by those using, contributing to, or interacting in the online space. When publicly available sources are used, the data only remains publicly available if the researcher does not attempt to trace usernames or profiles back to identifiable information.
- Here are some examples of private and public internet sources:
    - Private: Facebook/Instagram/X (formally known as Twitter) and any other social medical platform (unless "unlocked" by user), listservs, chat rooms, anything made private by the user, when use violates the Terms of Service
    - Public: comments at news sites/blogs, forums with no registration required, crowdsourcing sites, X data analytics, comments on Federal Notices, freely available Federal and State databases

Terms of Service

Many social media sites have specific data collection and sharing policies. The researcher is obligated to maintain compliance with the Terms of Service for any resource they access for data collection. Researchers are also responsible for ensuring their research does not violate revisions or updates to Terms of Service during the conduct of research.

Recruitment

Computer- and internet-based procedures are increasingly being used by investigators for advertising and recruitment of study participants. The text of the recruitment script must follow the 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 by the IRB before use in the field.

The following are common methods used to recruit participants online:

- Posting flyers or advertisements on social media platforms. Researchers may place one-way or direct advertisements on Facebook or something similar. In such cases, a researcher must be careful to understand how potential participants may interact with that material in cases of research involving health information or potentially sensitive data. For example, it may be possible for participants to incidentally identify themselves to others as research participants by “liking” or commenting on a publicly visible social media post.

- Posting recruitment scripts in forums and social media spaces dedicated to specific causes or conditions. If a researcher is collecting data on a specific condition or interest, it may be possible to recruit participants directly from online spaces frequented by participants meeting those criteria. In these cases, researchers should seek permission to post information from site moderators.
- Initiating interactive recruitment on social media platforms. Researchers may initiate two-way communication through social media with potential participants. This type of recruitment requires interaction over time through messaging. Researchers and research centers often create social media accounts specifically for this type of recruitment.

All research has the potential to attract bad-faith actors, but the nature of social media makes it much easier for these individuals to find, share, and exploit opportunities for compensation. Because of this, you need to evaluate your risk and mitigate it by selecting appropriate social media communities, by designing your study to be able to screen out bad-faith actors, and developing a plan for how you might identify and remove bad-faith responders from data after data has been collected.

Publicly available, less moderated social media platforms or communities (such as X, Reddit, Instagram, etc.) may be more likely to attract bots or bad-faith actors, by virtue of being easier to access than a closed Facebook support group. Consider not listing compensation amounts on social media recruitment materials.

Similarly, consider if your social media recruitment efforts should prompt more rigorous screening procedures. You may need to screen out individuals who might not be your target audience (e.g., recruiting via a dyslexia subreddit does not guarantee you will get individuals with dyslexia), to screen out ineligible individuals who might be willing to submit bad-faith responses for compensation, and to screen out “bots'' who automatically complete hundreds of surveys to receive compensation. For example, consider if screening should involve contact with the study team or providing additional documentation to verify identity or eligibility.

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, authentication of subjects is important to the validity of the data. Examples of verification measures include:

- 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.).
- Collecting IP addresses to screen out repeat responders and ensure residency in the U.S. Ensure the informed consent states the collection of IP addresses will occur and develop a process to discard IP addresses as soon as possible.
- Using bot-targeting screening mechanisms (such as CAPCHA).
- Structuring the screening survey to separate out unusual, repetitive, or inconsistent responses.

Informed Consent

Investigators must follow the guidelines on Informed Consent and include required elements of informed consent regardless of the method of data collection. Performing a consent process online can pose interesting challenges, though the same regulations and policies for consenting participants in human subjects research apply regardless of the format or venue. If obtaining consent in an online forum or chatroom, a researcher must ensure that the consent process does not alter the flow of conversation or use of that space. Some methods of obtaining informed consent online include:

- Obtaining Written Documentation Consent:
    - In cases where the IRB determines that written documentation of consent is required, a researcher will share a consent document with participants, which they will then sign and return by mail, email, or a similar format. In such cases, researchers will return a fully executed consent form, which also includes the researcher's signature and date, to participants.
- Waiving Written Documentation of Consent:
    - 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 is generally acceptable. This is similar to a cover page or disclosure form providing the elements of informed consent to potential subjects. In order to utilize this consent procedure, the investigator must request a waiver of documented consent. See guidance on Informed Consent for further information.

The process of requesting consent should not disrupt normal group activity. Researchers need to be particularly sensitive to 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 application.

Social Identities

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. 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. 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.

Data Collection

Various data collection methods exist on the internet and in social media. Below are some examples of the type and method of data collection with description of some of the concerns.

1. <u>Existing data</u> - Data only accessible through special permission are generally not considered public. 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 and IRB approval sought. Researchers wishing to access data from the internet or social media that contain identifiers and are not publicly available must first obtain IRB review and approval.

2. <u>Observations</u> - 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, community support groups 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. It is important for researchers to remain authentic and true when posting. Avoid being deceptive about the research intent when interacting with participants on the internet.

   Research in spaces that are not public or that maintain an expectation of privacy must have permission from the site organizer. For example, if special permission is needed to join a listserv, Facebook page, or forum then permission from the organizers to be there and make observations for research is required.

3. 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.

4. 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. Use of Qualtrics, REDCap, SurveyMonkey, Mechanical Turk, and other online survey tools is generally permitted for most minimal risk studies employing online survey procedures. 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.

5. Interviews - Interviews may be conducted over the internet using email or chat technology such as Google Chat, AOL instant messenger, Yahoo! messenger, etc. When conversing with a research participant via online 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.

Inclusion of Minors

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.

Minors may be screened out by checking for internet monitoring software such as SafeSurf and RSACi rating or using Adult Check systems. This may be necessary if the study presents more than minimal risk to subjects or asks particularly sensitive questions. Investigators may 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, the informed consent document may simply ask participants to confirm that they are the appropriate age of majority.

COPAA

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.

Data Security

Investigators must consider additional data-security issues when conducting internet-based research. Review the University's Information Security policies and guidance.

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.

It is strongly recommended that 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.

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