Considerations for protocol design concerning digital data

A lot of research currently makes use of digital data and apps in recruitment, data collection, and data analysis. These protocols deal with all of the issues that non-digital research deals with, but also have special considerations that are unique to them due to the amount of data, the unique properties of the information studied, and the complex perception of subjects when it comes to their online data. The following are examples of questions and concerns that one should consider when designing protocols involving digital data and apps. **Please note these are not requirements, just suggestions to improve the protocol's design.**

Population Perceptions

Users' perceptions of the technology are important components of digital research because different levels of technological and data privacy literacy exist in the public so users' perceptions can be drastically different depending on their level of understanding. The following questions will assist in developing an understanding of how potential subjects may perceive your study's technology and its impact on their data.

1. What is the study population's expectation of privacy when using the app or website? It is important to keep in mind, different populations have different expectations of privacy online and those expectations change depending on the context of each interaction.
    a. As an example of users' expectations, when a group of Twitter users was asked about whether researchers should use their tweets in research, only 35% said "yes." However, the 65% who said "no" did make qualifications about when it would be acceptable; when getting their permission first, when the tweets are part of a large group so their tweets are not singled-out, and depending on the value of the research.
2. How can user expectations of privacy be assessed?
    a. Examine the terms of use or service of the app or website to see if there is a description of the intended use of the content. For example, the agreement could say the content is not meant for the public domain or that any reproduction of the content must have written permission from the individual who created the content or the hosting entity who runs the platform.
        i. There is some risk in this method since it is well known that users do not read the terms of service and can be misinformed about the level of protection that the terms of service provide against researchers. When asked in the same Twitter study mentioned above, approximately 10% of participants said they thought researchers could not use their tweets without their permission for research because of Twitter's terms of service. In reality, Twitter's terms of service explicitly allow researchers to use them.
    b. Does the site have a membership requirement (i.e., log in) to see the content? If there is a log in, there is an expectation of privacy.
    c. Does the site have an explicit or implicit target membership (e.g., grieving parents, dating sites for specific populations/groups)? If so, this could create a sense that only individuals from that specific group will see the content, even if there are no restrictions.
3. What is the technological literacy of the study population? For researchers, this is important because it will help determine at what level of technical language the consent process should be discussed when reviewing it with the participants.

App functions

How an app functions is very important to the ethics of a protocol, just like any other tool used in the study. Any app used in a protocol must have a description of how it works included. The following things should be considered in order to make sure all possible issues are addressed. If the app has any medical function(s), please also review the FDA mobile app guidance.

1. Is the use of the app mandatory or optional?

   a. If the app is mandatory, how will that effect the potential study population? Will this requirement eliminate groups that the protocol wants to target because they do not have the ability to use the app for whatever reason?
   b. If it is optional what are the alternative means of participation?
2. Is there technical support for the app that participants know how to access?
3. Will participants need to provide their own devices or will the study provide them? Do you have a plan to retrieve any devices the study provides that will be communicated to participants?
4. Will participants need to pay to use the app? If so, will the study compensate them for the purchase?
5. How will use of the app impact subjects' data plans, if they have one?
6. Is the app custom-made? If so, has it undergone quality assurance testing for functionality, compatibility, performance, stability, and security?
7. Is there a license agreement, terms of use, privacy policy, or any other document(s) that subjects will have to agree to with a third party in order to use the app?
   a. How does the consent form compare to those documents? For example, does the language in the consent refer to any third party document that has exculpatory language? If so, there should be language in the consent form that indicates the exculpatory language does not apply in the context of the research.
   b. Will someone review the app's documents for future updates that could affect the protocol?
   c. Will the subject be encouraged to review the app documents in the consent process?
   d. Will subjects be notified if there are any changes to the app's documents that could affect their willingness to participate in the research?
8. Does the app create identifiable or linkable information? Keep in mind that re-identification is possible with information as trivial as zip codes and web search queries.
9. Will data be provided to any third parties, including the app developer?
10. Is the app using text messages to communicate with subjects? (If so, the protocol must explain that during the consent process and in the consent form due to federal law, which requires consent to receive text messages.)
11. Will patients' health care provider(s), if applicable, use the data collected?
12. Are the app developer and other entities that use the app as a source of information HIPAA compliant, if applicable?
13. Could the app gather data on persons other than the subject (e.g., an app gathering data on social contacts of the subject)? If so, does the protocol explain how this will be prevented and what will be done if it does happen?

Website functions

Protocols that use websites as a part of the research need to describe how the website functions in order to allow reviewers to better understand user expectations, data management, and the relationship the website will have with the research. The following questions are considerations to for researchers when designing protocols that use websites.

1. Can you see the information online without having to register a membership? This gives an indication of users' expectation of privacy. If you need to be a member to view the content, users could assume such content is not considered public information, no matter how easy it could be to create a membership account.
2. Does the website have a policy against research being done on the site?
3. Is there any indication that communication on the site is private, confidential, and/or selective in viewership?
4. How likely is it there might be unknown minors involved on the site? What, if anything, can be done to identify and/or filter them out of recruitment and data retrieval?

5. Does the website create identifiable or linkable information? Keep in mind, re-identification is possible with information as trivial as zip codes and web search queries. In addition, IP address is a new type of identifier with different countries declaring it private information.
6. Does the site have a comment section that could compromise subject confidentiality or privacy? If so, can this section be turned off or moderated to protect subjects?

Data protocols

Data management is even more important in protocols using websites and apps because of the potential amount of data to be gained and the number of subjects involved. The following questions can help researchers think through the potential problems and issues as they design their protocols.

1. Is the app or website third party? (If so, consult with legal when necessary.)
2. How do participant termination and withdrawal procedures work with the app or website?
    a. Will the app automatically delete from their device(s)?
    b. Can they still use the app or website after withdrawing?
    c. Are protections in place to stop gathering data from these withdrawn subjects?
3. Does the website have features that create an expectation of privacy that differs from what would ordinarily be expected? For example, is an online chatroom the same as conducting a conversation in a public place or does the login requirement imply that it is private?
4. Is the use of archived content (even in discussion venues open to non-members) considered secondary use of data?
    a. Users post digital content for a specific purpose (e.g., to provide or seek new information, to communicate with members of a specific community, and/or to possibly seek general public attention). It would be rare for users to expect their digital content will be used in an unknown person's research project. Thus, all content used for research purposes without the user's knowledge or expectation of such could be considered secondary data use.
    b. In many cases, users' content is archived indefinitely on the website to which it was originally posted. Users may or may not be aware that comments they made are archived for many years afterwards. In this respect, these archives of user content are also secondary data because the content is going to be used for a purpose other than originally intended. Being used in an unknown person's research study would qualify as such a purpose.
5. Where is the data stored?
    a. Mobile app data can be stored on the participant's device(s) or on a server. Either option has risks involved regarding breach of confidentiality. The device could be stolen or the transmission of the data from the device to the server could be intercepted. Encryption should be done in either case to protect participants' confidentiality. Devices should also be password protected.
    b. How long will the data be stored in each storage type/format that will be used (i.e., how long will the data remain on a subject's device, how long will it be on the server, and/or how long it will remain in your possession)?
    c. Is the data stored in multiple locations with different applicable laws? For example, European countries have regulations and standards for digital privacy that may conflict with local requirements and standards.
6. If you are using a third party server, does that third party also have access and right to use the data? Is this clearly stated in the protocol?
7. Does the protocol describe any password protection or data encryption for data obtained from digital sources?
8. Does the app gather any data not specified in the study protocol and, if so, does the protocol describe how the additional data will be dealt with/managed?

It is important to be aware that digital data coming from websites and apps can create unanticipated problems and/or violations of confidentiality and privacy. Some types of data can seem perfectly safe on an individual level but, when taken to the scale that is possible with digital data gathering methods, they can create problems. Even data that is completely de-identified can be dangerous. An example of completely de-identified data creating a breach in privacy involved an exercise-tracking app called Strava. The app released a map of their users' data that tracked their exercise routes to show the global impact of their app. The map had more than three trillion GPS data points that, inadvertently, created a map detailed enough to reveal the outlines of military bases in Afghanistan (identifiable to the level of individual buildings) which even Google Maps had greyed out for security purposes. For more information, hold down the control button and click here.

References

Buchanan, Elizabeth A., and Zimmer, Michael, "Internet Research Ethics", *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), Stanford University, 24 Aug. 2016, https://plato.stanford.edu/entries/ethics-internet-research/.

"Exemption Policy Re: Research Ethics Review for Projects Involving Digital Data Collection." Exemption Policy: Research Ethics Review of Projects Involving Digital Data Collection, Queen's University, 3 Oct. 2008, https://www.readkong.com/page/exemption-policy-re-research-ethics-review-for-projects-7465104.

Fiesler, Casey. "Participant Perceptions of Twitter." UCSD CORE Project Webinar.

"Guidance on the Use of Mobile Applications." Institutional Review Board, University of Kansas, June 2016, https://www.nursing.ku.edu/documents/ri/irb/Mobile-App-Use-Guidance.pdf.

Hern, Alex. "Fitness Tracking App Strava Gives Away Location of Secret US Army Bases." *The Guardian*, Guardian News and Media, 28 Jan. 2018, https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases.

"Institutional Review Board: Guidelines." Institutional Review Board: Guidelines, Internet Research, Bard College, www.bard.edu/irb/guidelines/.

"IRB Review of Mobile App Guidance." Office of Research, University of Pittsburgh, 10 Dec. 2014, https://www.irb.pitt.edu/sites/default/files/mobile_appreview_12_10_2014.pdf.

Peloquin, David. "Big Data and Confidentiality." AAHRPP 2018 Annual Conference. "Summiting New Heights in the Mile High City: Early Experiences Strategies and Solutions." 20 Apr. 2018, Denver, CO.

"Research Using Online Tools & Mobile Devices." *Research*, Indiana University, 14 May 2018, https://research.iu.edu/compliance/human-subjects/guidance/mobile.html.

Created 6-29-18
Revised 7-18-18, 2/15/2023 updated weblinks
J:\Master Outreach Documents\Survival Handbook\D - Guidance-Policy-Educational\D-132-Considerations for protocol design concerning digital data.docx