Privacy and Security Challenges in Doctoral Students' Human Subjects Research

Mary Anne Smart
University of California San Diego

Daniel Tan
University of California San Diego

Abstract

While universities focus on addressing the legal responsibilities for protecting the privacy of participants in research involving human subjects, the practical enactments of these responsibilities often fall to PhD student researchers. To explore the practical privacy and security-related challenges PhD student researchers face in human subjects research, we conduct an online survey at a research university as well as follow-up interviews for willing participants. We present both qualitative results from 18 students who participated in our study. We find that PhD student researchers struggle with a variety of issues including properly anonymizing participant data, securely sharing data with other researchers, and understanding compliance requirements.

1 Introduction

Researchers who work with human subjects have a responsibility to try to protect participants from harm [8]. One way that participants may be harmed is through violations of privacy. For example, depending on the research topic, participants' reputations may be damaged if their identities are revealed publicly. Therefore, it is important for researchers to think carefully about participant privacy and to follow best practices for securing participant data. Although researchers typically undergo various forms of training to prepare them for engaging in human subjects research, they may still face privacy and security-related challenges in their research.

Earlier work studying researchers at two public universities found that graduate students and faculty alike struggled with data management issues—including issues related to privacy and security [3]. If we want to help researchers better address the privacy and security-related challenges that they face, it would be helpful to first understand the range of specific challenges that researchers are facing. We focus specifically on PhD students, because as researchers-in-training, their relative dearth of experience likely creates additional challenges. We focus our investigation on the following research questions:

R1: What privacy and security-related challenges do PhD students face when conducting human subjects research?

R2: How do PhD students deal with privacy and security-related challenges they face in their research?

We conducted an online survey of PhD students in order to understand the privacy and security-related challenges PhD students face in their research with human subjects. We also conducted three follow-up interviews to add depth to the data gathered from our survey. We found that students face a wide array of challenges, including challenges related to sharing data with collaborators, challenges related to compliance with various regulations, and challenges related to anonymization of participant data. While many students know where to turn for help in dealing with such challenges, some do not. Our findings highlight important areas for future work.

2 Related Work

A 2011 study at Purdue University and the University of Illinois at Urbana-Champaign found that many faculty members shared concerns about their graduate students' competency in data management [3]. The study identified "data ethics"—a category defined to include privacy and security issues—as an area of data management where graduate students could benefit from further training. One particular data privacy issue mentioned by multiple faculty was data sharing; students were often uncertain about whether and how to share data with others safely. While this study examined data literacy needs with a focus on science and engineering disciplines,

USENIX Symposium on Usable Privacy and Security (SOUPS) 2022.
August 7–9, 2022, Boston, MA, United States.

we are interested in privacy and security challenges faced by PhD students from any discipline, as long as they are engaged in human subjects research.

One specific privacy issue that has received significant research attention is that of anonymization. Properly anonymizing data is much more difficult than it appears at first glance [9]. For example, it is possible to identify most individuals in the US with only three demographic fields: ZIP code, gender, and date of birth [12]. There have been several high-profile cases where individuals were re-identified from publicly-available "anonymized" datasets, including the Netflix Prize dataset [7]. A variety of methods—some heuristic and others cryptographic—exist for protecting quantitative data from this kind of re-identification [4, 13]. Qualitative information, on the other hand, poses its own distinct challenges. Qualitative researchers in particular may find that privacy requirements come into conflict with other goals of the research; for example, researchers aiming to "challenge [. . .] oppressive structures" with transformative research may find this goal at odds with the requirement to provide confidentiality [1].

3 Methodology

In order to understand the privacy and security-related challenges faced by PhD students in their research with human subjects, we conducted an online survey in addition to three follow-up interviews. Our Institutional Review Board determined that the protocol was exempt from full IRB review.

3.1 Survey Design

Although surveys are most often used in quantitative research, they can be useful for qualitative research as well [2]. For particularly sensitive topics, an anonymous survey may feel safer than a face-to-face interview. We expected that discussing privacy or security-related mistakes could be a sensitive topic for students. Furthermore, online surveys are generally more convenient for participants than interviews, since a survey can be taken whenever and wherever the participant pleases.

The structure of the survey is as follows. First, participants agree to the consent form and confirm their eligibility—namely that they are graduate students of at least 18 years of age who have conducted human subjects research. Next, participants are asked to rate their agreement with five statements related to self-efficacy, four of which are adapted from prior work [10]. Participants then answer a sequence of open-ended questions about challenges related to data security, challenges related to protecting participants' privacy, research participants' privacy concerns, new challenges raised by the COVID-19 pandemic, and conflicts between privacy concerns and other research goals. This section concludes by asking about any other concerns that the student would like to share.

In the next section, participants are asked to share their institutional affiliation. Participants from our institution are provided information about on-campus resources that offer assistance related to privacy and security. These participants are then asked to name any other resources that they have found helpful in addressing privacy and security-related issues. Participants from other institutions are also asked what resources have been helpful for them in addressing such issues. Next, all participants are asked to share their area of study, whether they are PhD students, whether they have advanced to candidacy (if relevant), and demographic information. Finally, participants are asked if they would like to participate in a follow-up interview. Those who wish to participate are rerouted to a separate survey to enter their contact information. The full survey instrument is included in Appendix B.

3.2 Interview Protocol

The follow-up, semi-structured interview was designed to give participants the opportunity to elaborate about their experiences and to probe participants further about the challenges described in their survey responses. We first ask the interviewee if they have anything in particular that they would like to share, in case there was some specific concern that led them to agree to the follow-up interview. Next, we ask general questions about their background and area of research. We continue by asking about training, sources of support, and methods and/or tools used to protect participants' privacy. Finally, we ask for further elaboration on some of the topics covered by the survey. The interviews were conducted over Zoom and recorded with participants' consent. The first author conducted the interviews, and the second transcribed them. In order to minimize privacy risks, the interview recordings were deleted soon after the transcriptions were completed.

3.3 Participants

Participants were recruited through flyers, social media posts, and emails to various mailing lists. Although we allowed any graduate student with relevant research experience to participate, we focused recruitment efforts on PhD students; in the end, only PhD students participated in the study.

Although 20 participants completed the online survey, two responses were omitted from our analysis due to issues with response quality and relevance. This left a set of responses from 18 PhD students, nine of whom had advanced to candidacy. All but three students were affiliated with our institution. They represented a range of disciplines, including biomedical sciences, cognitive science, computer engineering, computer science, economics, neurosciences, public health, psychology, and sociology. Information about participant demographics can be found in Appendix A. Three students participated in a follow-up interview. The median survey completion time was 11 minutes, and all interviews were less than 30 minutes. The students were not compensated for their participation.

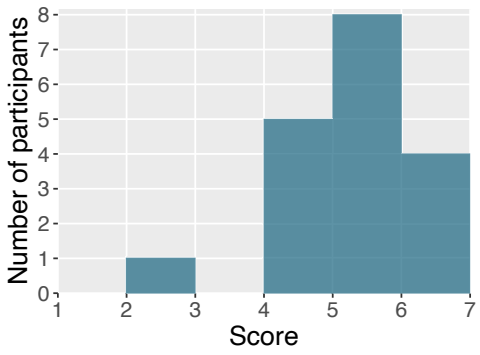

Figure 1: Participants' self-efficacy scores. The scores range from one (low self-efficacy) to seven (high self-efficacy) and represent averages over five questions.

3.4 Qualitative Analysis

We used open coding to analyze the data collected through the survey and interviews, taking an inductive approach. First, we read through all of the survey responses and interview transcripts to identify excerpts related to our research questions. Then, together, we developed a set of codes. Next, we went through the excerpts, tagging them with the relevant codes while iteratively updating the codebook and retagging excerpts as necessary. We resolved any disagreements or uncertainties through discussion.

4 Results

4.1 Privacy and Security Challenges

Although most students had high self-efficacy scores (Figure 1), the students nevertheless faced a wide array of privacy and security challenges in the course of their research. We outline these challenges below, beginning with the challenges that were mentioned most frequently and continuing with less frequently discussed challenges.

4.1.1 Data sharing

Data sharing was a pain point for many of the students in our study. Sometimes the difficulty was in obtaining data from a provider; one student stated that "data providers have too strict of limitations on data which make it difficult to analyze proprietary data." In many cases, difficulties arose when working with collaborators. For example, one student bemoaned the "amount of paperwork involved in [. . .] getting coauthors from other institutions access."

4.1.2 Anonymization

Several students described challenges with anonymizing participant data. In some cases, the data was inherently difficult to anonymize. For example, one student explained that "it can be a little hard to ensure that participant data is fully anonymized, since we have worked with participants from relatively small groups." For another student, it was the process of anonymization that caused challenges, since it created "months of extra work." In this case the challenge lay in the process rather than in the nature of the data itself.

4.1.3 Participant Recruitment

Privacy concerns sometimes created challenges for participant recruitment. For example, one researcher who worked with undocumented immigrants explained that many participants had privacy concerns related to their undocumented status. Challenges with recruiting immigrants for participation in research have also been observed in prior work [5, 6].

4.1.4 Confusion

Students expressed confusion about various aspects of dealing with privacy and security concerns and about issues of compliance. For example, one student was unsure about whether they were "allowed to collect certain information." Another student expressed uncertainty about the security of Zoom, explaining that: "since the IRB allows it, I accept that it is safe, but I don't know that for sure."

4.1.5 Technological Challenges

Students faced a variety of technological challenges related to privacy and security. For example, one student mentioned VPN connection issues that made it difficult to access the secure services they needed. A specific technology that raised some challenges was Zoom. For example, one student mentioned that a challenge resulting from the COVID-19 pandemic was storing Zoom recordings. Another issue that arose was the use of personal laptops. For example, one student expressed concern about "storage of information on personal laptop computers that may be lost, stolen, or compromised."

4.1.6 Compliance

Compliance with various privacy and security-related regulations was challenging for some students. For example, one student described "being in constant conversation with the Registrar" regarding challenges related to FERPA (Family Educational Rights and Privacy Act).

4.1.7 Physical Security

Some students experienced challenges related to physical security. For example, one student was not provided with any secure storage space and had to use a personal locking filing cabinet. Another challenge related to physical security was

that of selecting an appropriate location for meeting participants; one student mentioned that "it can be hard to find a suitable location to conduct research where the participant can feel that no one will recognize them."

4.1.8 Training Gaps

Gaps in training sometimes posed challenges for researchers. For example, one student described working with staff who lacked a research background. Part of this student's responsibilities involved training staff to understand regulations and best practices "from the research side." In other cases, students themselves had not received adequate training. For example, one student had "not been given any instruction" on data security and had to do their "own research on the topic."

4.1.9 Power dynamics

Data sharing issues were often exacerbated by the power dynamics at play. For example, one student struggled with senior researchers' casual approach to sharing data:

> *No matter how much I tried to avoid sharing files over WhatsApp, my team consists of much more senior researchers (boomers) in a context where privacy issues are of little concern generally (developing country).*

The context of the research could affect the behavior of collaborators; for example, some countries have fewer regulations related to human subjects research and data management. One student struggled to advocate for better privacy but noted that such struggle came with a cost:

> *I have found that conflict about privacy arises when senior researchers do not consider demographic information or phenotypic information to be potentially identifiable in part or in aggregate. As a graduate student, I have been allowed to use my best judgement and/or change the study designs to avoid such conflicts, but it did not come without a cost. When a student disagrees with a senior researcher, there is always tension and stress placed on the student.*

Finally, one student felt that department pressures to "tell stories [. . .] in an ethnographic way" led to privacy challenges. These students struggled to protect their participants while navigating the social hierarchies of academic workplaces.

4.2 Dealing with Challenges

The students in our study dealt with privacy and security challenges in a variety of ways. Some challenges could be addressed with technical solutions. For example, one survey respondent mentioned running "all analyses on a remote server" to avoid downloading sensitive data to personal devices. Students also mentioned taking precautions to protect participants' identities such deleting identifying information or assigning numeric codes to participants. In some cases, students then needed to explain these precautions in accessible language, in order to address participants' privacy concerns.

In other cases, students needed help addressing privacy and security-related challenges. Resources that students turned to for help included advisors, university websites, and the Registrar. Although some students knew where to seek help, others did not. For example, one student expressed frustration about not knowing where to find support:

> *At my old job, there was an IT department that was responsible for all of these things. Now, it's not clear who to turn to about a data infrastructure question.*

Even if resources exist to help with data management, leading students to these resources may require targeted outreach.

5 Future Work

Although survey responses from 18 students help us begin to map the terrain of privacy and security challenges that PhD students face in their research, it is likely that our findings in this preliminary study are incomplete. In particular, students at other institutions and students from departments that are not represented in our study may face additional challenges that we did not uncover in our investigation.

Additional limitations of this study point to areas for future work. One limitation is our sole reliance on students' own accounts. Students may not always be aware of all the privacy and security issues at play in their research and may unknowingly violate best practices. Future work could consider observing PhD students in their day-to-day activities and noting to what extent best practices are adopted. Another limitation is our sole focus on PhD students.

Usable security and privacy researchers may be well-positioned to form partnerships with campus libraries or IT departments that could address some of the challenges that PhD students face in their research. A number of the challenges identified in our study are familiar topics to the research community. For example, issues with file sharing systems have already been explored in prior work [11, 14].

Although our study focused on PhD students as researchers-in-training, our findings point to the need to consider faculty as well. While some students turned to faculty for help with privacy and security challenges, other students struggled with confronting faculty about problematic practices. Similar tensions around privacy and security may arise in other contexts as well. This too is an area worthy of further study.

Acknowledgments

We thank Professor Munyaka for teaching us about usable security and privacy and for providing guidance on this project.

References

[1] Benjamin Baez. Confidentiality in qualitative research: Reflections on secrets, power and agency. *Qualitative research*, 2(1):35–58, 2002. URL: https://doi.org/10.1177/1468794102002001638.

[2] Virginia Braun, Victoria Clarke, Elicia Boulton, Louise Davey, and Charlotte McEvoy. The online survey as a qualitative research tool. *International Journal of Social Research Methodology*, pages 1–14, 2020. URL: https://www.tandfonline.com/doi/full/10.1080/13645579.2020.1805550.

[3] Jacob Carlson, Michael Fosmire, CC Miller, and Megan Sapp Nelson. Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy*, 11(2):629–657, 2011. URL: https://muse.jhu.edu/article/428877.

[4] Cynthia Dwork. Differential privacy. In *International Colloquium on Automata, Languages, and Programming*, pages 1–12. Springer, 2006. URL: https://link.springer.com/chapter/10.1007/11787006_1.

[5] Tamy Guberek, Allison McDonald, Sylvia Simioni, Abraham H. Mhaidli, Kentaro Toyama, and Florian Schaub. *Keeping a Low Profile? Technology, Risk and Privacy among Undocumented Immigrants*, page 1–15. Association for Computing Machinery, New York, NY, USA, 2018. URL: https://doi.org/10.1145/3173574.3173688.

[6] Carina Katigbak, Meghan Foley, Lauren Robert, and M Katherine Hutchinson. Experiences and lessons learned in using community-based participatory research to recruit asian american immigrant research participants. *Journal of Nursing Scholarship*, 48(2):210–218, 2016. URL: https://doi.org/10.1111/jnu.12194.

[7] Arvind Narayanan and Vitaly Shmatikov. Robust deanonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy*, pages 111–125. IEEE, 2008. URL: https://ieeexplore.ieee.org/abstract/document/4531148.

[8] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. The Commission, 1978. URL: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html.

[9] Paul Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Rev.*, 57(6):1701–1777, 2009. URL: https://heinonline.org/HOL/P?h=hein.journals/uclalr57&i=1713.

[10] Hyeun-Suk Rhee, Cheongtag Kim, and Young U Ryu. Self-efficacy in information security: Its influence on end users' information security practice behavior. *Computers & security*, 28(8):816–826, 2009. URL: https://doi.org/10.1016/j.cose.2009.05.008.

[11] D. K. Smetters and Nathan Good. How users use access control. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, New York, NY, USA, 2009. Association for Computing Machinery. URL: https://doi.org/10.1145/1572532.1572552.

[12] Latanya Sweeney. Simple demographics often identify people uniquely. *Data Privacy Working Paper 3*, 2000. URL: https://privacytools.seas.harvard.edu/files/privacytools/files/paper1.pdf.

[13] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002. URL: https://www.worldscientific.com/doi/abs/10.1142/S0218488502001648.

[14] Tara Whalen, Diana Smetters, and Elizabeth F. Churchill. User experiences with sharing and access control. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, page 1517–1522, New York, NY, USA, 2006. Association for Computing Machinery. URL: https://doi.org/10.1145/1125451.1125729.

A Participant Demographics

Tables 1 and 2 show the breakdown of participant demographics. Since participants were allowed to select multiple identities, the counts may not add up to eighteen.

Table 1: Gender

Women	11
Men	4
Other	3

Table 2: Race & Ethnicity

Asian, Native Hawaiian, or Pacific Islander	8
Hispanic, Latino or Spanish	2
White	11
Other	3

B Survey Instrument

B.1 Self-efficacy Questions

Rate your agreement with the following statements. (Strongly agree - Strongly disagree)

- I feel confident understanding terms relating to information security.
- I feel confident using different programs to protect my information and information system.
- I feel confident learning advanced skills to protect my information and information system.
- I feel confident getting help for problems related to my information security.
- I feel confident following best practices in order to protect participant data.

B.2 Introduction to Open-ended Questions

Next you will be asked several open-ended questions. Do not worry about spelling or grammar in your responses to these questions. Please include as much detail in your responses as you are comfortable sharing. Remember that you can skip any question that you do not want to answer.

The questions will be related to privacy and security. For your convenience, definitions of these terms, taken from a UC Privacy and Information Security Steering Committee report, are provided below:

- Information privacy refers to the appropriate protection, use, and dissemination of information about individuals.
- Information security refers to the protection of information resources from unauthorized access, which could compromise their confidentiality, integrity, and availability.

B.3 Open-ended Questions

- What challenges have you faced in your research related to data security? How have you dealt with these challenges?
- What challenges have you faced in your research related to protecting participants' privacy? How have you dealt with these challenges?
- Have you ever had a research participant express concerns related to privacy or security? If so, please describe the experience. How did you address the concerns?
- Has the COVID-19 pandemic raised any new privacy or security-related challenges in your research? Please describe any challenges you have faced.
- Have concerns about participants' privacy ever come into conflict with other goals of your research? How did you deal with this conflict?
- Do you have any other concerns you would like to share?

B.4 Resources

- What is your institution?
- *If our institution:*
  - There are several resources on campus that can help you with privacy and/or security issues related to your research. UCSD Research IT Services can assist you "as you look for technology to support your research." The UCSD Privacy Office offers a variety of training programs that may be relevant to student researchers. The Research Data Curation Program can help student researchers with various aspects of data management. The UCSD Human Research Protections Program "exists to promote high quality, ethical research" and offers a variety of training opportunities.
  - Are there any other resources that have been particularly helpful for you in addressing privacy and security-related issues in your research?
- *If other institution:*
  - Are there any resources that have been particularly helpful for you in addressing privacy and security-related issues in your research?

B.5 Details and Demographics

Please remember that you are free to skip any questions that you do not wish to answer.

- What is your area of study? *72 choices*
  If other: You selected “Other" as your area of study. What is your area of study?
- What kind of student are you? *PhD student, Master’s student, Other: text entry*
- Have you advanced to candidacy? *Yes, No, Not applicable (not a PhD student)*
- What is your gender? Select all that apply: *Woman, Man, Non-binary, Prefer to self-describe: text entry*
- Which categories below best describe you? Select all that apply: *White; Hispanic, Latino or Spanish; Black or African American; Asian, Native Hawaiian, or Pacific Islander; American Indian or Alaska Native; Other: text entry*
- Are you willing to participate in a short follow-up interview? *Yes, No*