Journal of Privacy and Confidentiality
Vol. 13 (2) 2023

NAHDAP-ICPSR

Submitted Nov 2022
Published Nov 2023

PROTECTING SENSITIVE DATA EARLY IN THE RESEARCH DATA LIFE CYCLE

SEBASTIAN KARCHER[†], SEFA SECEN[‡], AND NIC WEBER[◇]

[‡] Syracuse University

[‡] The Ohio State University

[◇] University of Washington, Seattle

Abstract. How do researchers in fieldwork-intensive disciplines protect sensitive data in the field? How do they assess their own practices? And how do they arrive at them? This article reports the results of a qualitative study with 36 semi-structured interviews with qualitative and multi-method researchers in political science and humanitarian aid/migration studies. We find that researchers frequently feel ill-prepared to handle the management of sensitive data in the field and that formal institutions provide little support. Instead, researchers use a patchwork of sources to devise strategies for protecting their informants and data. We argue that this practice carries substantial risks for the security of the data as well as their potential for later sharing and re-use. We conclude with some suggestions for effectively supporting data management in fieldwork-intensive research without unduly adding to the burden on researchers conducting it.

1. Introduction

The sharing of research data is slowly becoming the norm across disciplines (Goodey, et al., 2022). Research funders such as the members of UK Research and Innovation in the United Kingdom (UK Research and Innovation, 2022), the European Union (European Commission, 2016), the National Science Foundation (National Science Foundation, 2011), and the National Institutes of Health (National Institutes of Health, 2020) are requiring sharing of data produced with their funding. More and more journals are recommending or even requiring that data underlying empirical research be shared (Crosas, et al., 2018).

This increase in requirements and willingness to share data has been accompanied by an expansion of the infrastructure for sharing data (Kapiszewski and Karcher, 2020). The availability and range of data repositories and, in particular, the technologies offered by data repositories to protect sensitive data (Altman, et al., 2016) have increased significantly.

Key words and phrases: qualitative research, fieldwork, sensitive data, IRB, research ethics, data sharing.

Work funded through a grant by the Alfred P. Sloan Foundation and National Science Foundation grant 2116935.

JPrivacy Confidentiality

www.journalprivacyconfidentiality.org
DOI:10.29012/jpc.846

Absent from much of that discussion, however, are considerations about the security of sensitive data as they are collected by researchers. This is particularly important in fieldwork-intensive disciplines in the social sciences such as sociology, anthropology, human geography, and political science. The modalities and local circumstances of data management can be particularly fraught during fieldwork. Moreover, data collection tends to be significantly less regulated in the social sciences compared to, e.g., the health sciences, where many countries have specific legal frameworks governing the safeguarding of data, such as the Health Insurance Portability and Accountability Act (HIPAA) in the US. This is not to claim an absence of *scholarship* on the topic: a literature on data safety during fieldwork exists, ranging from concerns about digital trace data (Henry, et al., 2022) to the physical safety of data (Arias, 2014) and the safety of researchers themselves (Peter and Strazzari, 2017), an important topic that we mostly exclude from consideration in this paper.

Understanding how researchers protect sensitive data in the field is essential to advance both the security of such data and the ability of researchers to effectively share their data. Decisions that researchers make early on during fieldwork may have significant ramifications on shareability of data. In many cases, structural conditions may push researchers toward more conservative approaches to data security that inhibit the later sharing of information they collect. Investigating these choices and explaining why and how researchers make them can help generate not just better guidance for researchers but also better *structures* that encourage researchers to learn about and apply such strategies.

In this paper, we seek to advance a better understanding of researchers' views and practices regarding sensitive data during fieldwork, based on semi-structured interviews with 36 researchers in political science and humanitarian aid/migration studies. We show that researchers approach data security during fieldwork with significant uncertainty and, frequently, a degree of improvisation. The most effective researchers draw on a patchwork of sources to arrive at effective strategies for keeping sensitive information safe during fieldwork. Unsurprisingly, Institutional Review Boards (IRBs)[1] take on a central role in researchers' consideration of data security, but researchers' views on the role and benefits of IRBs vary significantly.

We begin by situating our research within the existing literature on protecting participants in qualitative research, and then describe data and data collection methods. The main section presents our findings about researchers' strategies for data security in the field. We discuss how researchers view their past and current practices, how they arrive at these practices, how training shapes their practices, and finally their views of IRBs.

2. Background

Fieldwork, or field research, is broadly defined to mean "leaving one's home institution in order to acquire data, information, or insights that significantly inform one's research." (Kapiszewski, et al., 2015, p. 1).

[1] We use the term IRB here to encompass any ethics review board, including those for scholars outside the US, known by different names such as "ethics review board" or "ethics council."

2.1. **Research in the Field.** Fieldwork is a common and well-documented approach to data collection in political science, sociology, anthropology, and many other social sciences (Okely, 1994). In each of these fields, researchers immerse themselves in intricate social and political environments to gather a variety of data including, but not limited to, personal accounts (interviews), demographics (surveys), and reactions to new or varied settings (experiments).

The success and safety of fieldwork depend critically on comprehensive planning that considers the unique challenges of a local context, which might include a complex web of political, social, economic, and cultural forces that will be encountered (Kovats-Bernat, 2002). Fieldwork in adversarial settings where a government or group may be suspicious of data collection may put researchers in the position of being surveilled, physically or emotionally intimidated, or censored (Demery and Pipkin, 2021). So while all research requires careful and purposeful design, researchers in hazardous field sites must spend considerable effort to plan research projects in order to protect their work, themselves, and their subjects/participants from harm. This includes devising secure research strategies, building local trust and support, understanding the agendas of political factions, and having emergency contingency plans (Wackenhut, 2018).

Such context-aware preparation should, ideally, include plans for how to manage, secure, and transfer digital data collected during fieldwork. Across disciplines and locations there are common challenges for field research collecting qualitative data. In Table 1 we briefly describe each of these challenges and cite relevant resources that describe these issues in depth. In Section 2.2, we describe how these preparations (managing, securing, and transferring data) map onto a "data life cycle model" and how participants in our study described their preparation and practices for managing data while in the field.

Table 1: Challenges in collecting data during field research

Challenge	Description	Citation
Personal Identifiable Information (PII)	Field research involves collecting contextual, qualitative data that can reveal identities and includes sensitive information about subjects.	Katz (2015)
Security	Securing data is challenging when collected away from home institutions, on digital devices, and across field sites.	Aldridge, et al. (2010)
Anonymity	Anonymizing data is labor intensive and often a post-hoc activity that leaves raw or original data susceptible to leaking PII.	Goodwin, et al. (2020)
Funding	Limited funding constrains the options field researchers have in adopting and deploying best practices in cybersecurity.	Kapiszewski, et al. (2015)
Training	Field researchers often lack data management and cybersecurity training.	Thurston and Pasternak (2019)
Managing Consent	Negotiating and managing consent in the field requires documentation and access that often include PII (e.g., participant names) that a research must secure.	Okely (1994)
Unplanned Data Collection	Personal data collection may precede protections when research is iterative, or improvisational as is often required in field settings.	Kapiszewski, et al. (2015)
Transferability	Digital data can be easily copied and transferred across devices. This leads to improved backups for field researchers, but also increases the potential for mismanaging or incorrectly sharing PII.	Demey and Pipkin (2021)

2.2. **Research Data Life Cycle.** A simplified model of the research data life cycle has three main phases: data collection, data deposit, and data retrieval for re-use. It has become a truism among advocates of shared data that sharing data (i.e., deposit and reuse) are greatly facilitated if they are taken into account during the collection process (Hedrick, 1988; Mannheimer, et al., 2019; Niu and Hedstrom, 2008). Decisions made early on in planning or data collection can significantly impact the ability of research to share data or to share them in a way that allows for effective reuse (Kirilova and Karcher, 2017).

Qualitative researchers, who interact with participants closely and frequently on highly personal questions, have long been discussing appropriate ways to ensure participant security throughout the life cycle, although we find some components of the literature to be surprisingly disconnected from each other. The most immediately relevant works for the interviews captured in the paper are those focused on data security during fieldwork. Much of this work consists of pragmatic how-tos, such as Aldridge, et al.'s (2010) list of 14 guidelines for digital security in the field, which includes both technical (strong passwords, encryption) and socio-technical (watch out for former employees) advice. Similarly, Arias (2014) provides a list of tools, such as threat assessment metrics, and recommendations focusing specifically on data security in highly violent and volatile settings. His article begins with a harrowing account of reporters in Brazil who were detained and tortured by a corrupt police gang they were investigating, as a motivating example for the importance of data security for researchers. Some of the strategies discussed in this literature are common sense, others, such as the use of a dedicated operating system such as the The Amnesic Incognito Live System (TAILS) (suggested, e.g., in Arias, 2014; Barratt and Maddox, 2016) may be reserved to technically sophisticated researchers operating in very high-risk contexts.

Beyond technology, qualitative researchers have long recognized the particular challenges in ensuring participant confidentiality, both given small populations and intimate settings, and because "the qualitative investigator is less in control of the research process than the laboratory scientist, and can only guide decision making. The inherent unpredictability of the research process undermines the spirit of informed consent and endangers the assurance of confidentiality." (Ramos, 1989, p. 58) Still, in earlier writing, the general importance of confidentiality as a relevant ethical category is largely unquestioned, even if tradeoffs are acknowledged.

Starting in the late 1990s, writers more directly question confidentiality as an appropriate, universal norm, arguing that it can be an obstacle to using qualitative research as a tool for societal change (Baez, 2002), and that many participants have strong preferences *against* confidentiality (Morse, 1998; see Kaiser, 2009 for a review of that literature). Finally, Jerolmack and Murphy (2017) argue against masking locations (and in many cases, identities) because masking can impact the quality and transparency of the research. Based on these views, there have been attempts to develop guidelines around confidentiality in qualitative research. For example, Kaiser (2009), in a widely cited paper, proposes to include various options around identity disclosure in informed consent conversations at the end of the research process. For those seeking to maintain strict confidentiality of participants, the US Census Bureau (Pascale, et al., 2020) provides a set of recommendations for reporting qualitative data while minimizing disclosure risk.

Closely related to critiques of traditional notions of confidentiality are critiques by qualitative researchers of the IRB process, which sets the institutional rules that researchers have to follow in order to protect participants. Just as simple rules such as keeping all participant information confidential may be ill-suited for the complex social relationships in which qualitative researchers operate, the one-shot, quasi-contractual relationship of the informed consent agreement at the heart of the IRB-prescribed ethics process may be insufficient at best, and counter-productive at worst, for long-term research projects that involve close, often personal interaction with participants (Guillemin and Gillam, 2004; Hammersley and Traianou, 2012; Shaw, 2008).

Finally, there is a somewhat more recent debate directly related to the *sharing* of qualitative data. While the debate is multi-faceted and extends far beyond the scope of this paper, ethics and participant protection loom large. Critics argue that sharing qualitative data may endanger vulnerable participants and violate their trust (e.g., Feldman and Shaw, 2019; Jacobs, et al., 2021). Others maintain that qualitative data may be ethically shared when handled carefully and when appropriate informed consent is given by participants (Bishop, 2009; Kirilova and Karcher, 2017). More recent empirical work shows that, in many circumstances including research on sensitive topics, many participants favor (responsible) data sharing (Campbell, et al., 2023; Kuula, 2011; Mozersky, et al., 2020; VandeVusse, et al., 2022).

This paper takes a meta-scientific approach: rather than focus on what researchers *should* do, we focus on what researchers *are* doing and, especially, how they *think* about what they are doing in the field to protect their participants and their data. The goal is to better understand how researchers see themselves in the complex web of recommendations, rules, and relationships laid out in the literatures on participant protection in qualitative research. By using purposeful sampling and a relatively small number of interviews, we are not aiming or claiming to present a complete picture of fieldwork practices among qualitative researchers. Rather, we provide an in-depth account of how qualitative researchers think about their use of data management strategies. This account might inform strategies to better prepare researchers for fieldwork as much as it might inform theoretical debates about the role of qualitative researchers in the field by providing a deeper evidentiary basis.

3. The Data

The data for this study are part of a larger, pre-registered qualitative study of researchers' views and practices regarding collection and sharing of sensitive research data (Weber, et al., 2021); the preregistration document contains much more detail on data collection procedures than we describe here. The original study covers four disciplines. For this paper, we focus on the two largest subsamples, qualitative and multi-method researchers in two fieldwork-intensive disciplines: political science and humanitarian aid/migration studies.

First, a key informant (Onwuegbuzie and Leech, 2015) was identified in each case study. The key informant suggested relevant researchers working with privacy-sensitive data that would be appropriate for participating in the study. Key informants have been active contributors to the research projects, and either formal or informal collaborators. Participants were recruited by email, and enrolled in the study after agreeing to participate and signing an informed consent agreement. For each enrolled participant, we generated a profile including relevant archival documents such as published papers, Curricula Vitae, professional affiliations, and a list of relevant journals in which the author published. During interviews, we explicitly asked participants how they would like to be identified in terms of disciplinary or specialty affiliation. For example, we observed that a researcher enrolled in the political science case study identified as an international relations scholar.

Second, snowball sampling (Naderifar, et al., 2017) was used to recruit additional participants, by asking participants for colleagues whom they feel are representative of their domain, or who may have important perspectives on the management of privacy-sensitive data in their discipline. Participants recruited through snowball sampling were treated the same as those recommended by key informants: a profile was developed, a disciplinary map was constructed, and an informed consent form was collected.

Third, within the migration studies case, purposeful sampling through bibliometric and scientometric studies was conducted to identify research specifically scoped to address relevant domains in our case study (similar to techniques used in Velden and Lagoze, 2013). To scope the dataset, we queried Web of Science filtered from 2017 through February 2021. From the total dataset of 2,909 author names and affiliations, we contacted a random sample of 145 (5%) in the first round of recruitment. This process was repeated two more times to increase the interview sample population size. In total, we contacted 435 researchers, scheduled 40 interviews, and completed 22. Four of those 22 participants did not wish to have their data analyzed and reported on in future research, and so are not included in the sample described here.

In order to systematically identify potential interview participants for political science, we reviewed recent issues of several journals that regularly publish qualitative research with human participants such as *Comparative Politics*, *World Politics*, *International Security*, and *Comparative Political Studies*. We focused particularly on research conducted in authoritarian or volatile contexts with vulnerable groups, and drew on interviews or ethnography as a primary research strategy. In addition, we also conducted Google searches and checked the databases of data repositories such as the Qualitative Data Repository (QDR) to identify potential interview participants. From these sources, we gathered a list of potential research participants and invited 70 individuals for an interview. Of the 70 individuals that we contacted and sent reminders to, 33 did not respond, 19 declined to do an interview, and 18 agreed to do an interview.

With these strategies, we interviewed a total of 36 respondents (18 from each discipline—see Table 2). All participants have experience collecting sensitive qualitative data, but vary in their epistemological and methodological approaches and views about data sharing. Moreover, the degree to which data collected were sensitive and the nature of the research setting varied significantly. All participants had completed a Ph.D. by the time of the interview. The majority were early career researchers (postdocs or assistant professors). Many of the interviews focused on data collected during dissertation research. We conducted semi-structured interviews, each lasting about one hour, over Zoom, with all respondents. Before each interview, we compiled a participant profile and interview checklist to be able to ask questions specific to the respondents' data and experience. Data were then coded in NVivo based on a pre-registered codebook.

Table 2: Participant demographics

Political Science		**Humanitarian Aid / Migration Studies**	
Subfield		***Subfield***	
Comparative Politics	16	Public Policy	7
International Relations	5	Migration & Resettlement	7
American Politics	1	Labor Migration	4
Law and Courts	2	Human Geography	1
Academic Rank		***Academic Rank***	
Junior Scholar	10	Junior Scholar	12
Senior Scholar	8	Senior Scholar	6
Research Method		***Research Method***	
In-depth interviews	17	In-depth interviews	14
Elite Interviews	2	Elite Interviews	0
Focus Group	1	Focus Group	2
Participant Observation	1	Participant Observation	10
Textual analysis	3	Textual analysis	2
Archival research	1	Archival research	4
Region		***Region***	
North America	14	North America	11
Europe	4	Europe	7
South America	0	South America	2
Gender		***Gender***	
Male	10	Male	3
Female	8	Female	15

Numbers in the table are counts of participants. The research method and subfield allow for multiple mentions per participant. Junior scholar refers to assistant professors and postdocs; senior scholar refers to associate and full professors.

4. Findings

We focus here on two questions and related follow-ups. We asked informants about their practices during fieldwork: to describe in detail their strategies for keeping data safe, how they developed these strategies, and if they consulted anyone or received any dedicated training. In our study design, we envisioned these questions would help us understand researchers' main concerns about their data and how they decide to trust specific technologies or institutions. In addition to some tentative answers to these questions, however, we organically observed a set of themes in which researchers expressed concerns about their strategies and training and, more broadly, the institutions offering them. We re-coded answers to these questions specifically with these topics in mind, and present the results below.

4.1. **"It was kind of make it up as you go along".** Perhaps the most striking finding was the number of researchers who described themselves as largely unprepared for managing their data in the field.

> *No, I didn't have a plan or strategy. I'm embarrassed to say I was really unprepared. I hadn't received good training in graduate school for fieldwork. I hadn't received much advice. (PS05)*

> *When the project started I was a graduate student and, you know, to be honest, it was kind of make it up as you go along. (PS06)*

> *My approach to that was sort of designed by the [major European research university] where I did my Ph.D. They didn't really provide a lot of training in this, I would say, I had, I think a 30-minute session on interviews, which is mainly about, you know, bring an extra pen if your pen breaks, which was pretty basic. (PS15)*

> *We are all really involved in our research in our classes, and we don't invest that amount of time to look for the right solutions or for the best solutions in terms of storage of information. (HR1)*

This is not to say that researchers do not care about data security. Especially respondents working in authoritarian, conflict, or post-conflict settings were often highly concerned and developed sophisticated strategies around protecting data.

Some combined standard technical solutions with a strong awareness of the physical security of data-holding devices:

> *Data security was probably one of the most important concerns for me while I was in the field. [. . .] So, for instance, I had my laptop under my pillow at all times. It was password protected and encrypted and there was no chance that anyone could and I was ready to destroy it at any moment if anything had happened. (PS04)*

One of the most impressive approaches to data security in the field was described by a researcher who combined technical solutions with a threat analysis (Who could try to compromise my data and how?), as well as socio-technical approaches—here, a trusted person outside of the field site holding the encryption key to sensitive data:

> *So, if my apartment was raided they would find a laptop with no information on it, and they would find an empty notebook and everything that was valuable was uploaded to Dropbox, double password protected and also encrypted. Oh, and the other thing is the encryption key necessary to decrypt the information was stored outside of [country]. So even if they tortured me or something: I literally had no way to decrypt the information on Dropbox. (PS07)*

Even in retrospect, however, researchers often wondered if the strategies they used were optimal (and, knowing our expertise on the topic, often phrased this as concern about being judged by us):

> *I am just a little nervous that maybe my practices haven't always been as good as they could be but …(PS10)*

4.2. **Learning about Protecting Data.** As described above, we repeatedly heard from participants that they improvised on data management and security. Moreover, not a single participant described strong training on such topics as part of their graduate training. This is true even for the institutions or grant programs that require a description of the handling of sensitive data, as this human rights scholar describes:

> *HR04: This whole ethics and data management protocols, so we have to fill forms and write reports on this, especially when—yeah, it's mainly for my institution, but the European Commission asks for that as well.*
>
> *[…]*
>
> *Q: And was this something that is like embedded in your training, while you are at university or how did you find out about making sure that you go through these workflows and processes in your research?*
>
> *HR04: When you have to do it. They asked you to do it, you just have to deal with it.*
>
> *Q: Okay.*
>
> *HR04: Nobody has ever taught me how to do that.*

As we have seen, nonetheless, some researchers came up with sophisticated and elaborate protocols to effectively protect data in the field. So, how did they learn about effective strategies? One important source of information was other researchers, especially senior researchers.

> *I mean, I think all the usual stuff that people say about field research applies in these contexts. I mean I talked to a lot of people who had done research there before I went. (PS21)*

Not in all cases was the experience of more experienced scholars similarly helpful: they may not have been fully aware of feasible strategies and available technologies:

> *I talked to other researchers, you know, the more senior people who already went through this process some professors, but not all professors do field work so it's not always like that they know what those risks are and many in the kind of the older generation, we would go to [research site], they would have someone else organize everything for them. (PS17)*

Perhaps surprisingly, dissertation advisors played a minor role in shaping the research data practices of most of our respondents, even when they were otherwise involved and helpful in the research process:

> *My dissertation advisor who I was speaking back and forth about this helped me through my first IRB proposal sort of putting it together and sort of what*

> *things were supposed to look like if I ran into problems in the field. Sometimes, I think that she had never even really sort of run into necessarily all that stuff that stresses you out. [...]. So, I think I was kind of on a limbo on that. (PS19)*

Only one respondent explicitly noted a dissertation committee member playing an important role in shaping the data management and security strategy. Several of the researchers conducting some of the most challenging fieldwork had turned to articles by researchers with prominent work in such circumstances (Lake and Parkinson, 2017; Parkinson and Wood, 2015).

Researchers also turned to local experts for guidance. For elite-oriented interviews, journalists may provide important information about both security threats and best practices and respondents' expectations. In other cases, community members themselves may have had important input on their expectations and local circumstances, especially for researchers conducting research outside the current mainstream of their disciplines:

> *[A] lot of you know how I develop my methods and my recruitment came from the communities I was working with because there was really kind of like I feel like a lot of you know at the time when I was learning about this stuff as a grad student, a lot of the classes and especially at [the Institute for Qualitative and Multi-Method Research] IQMR, I mean nobody was talking about qualitative research with marginalized populations like maybe, you know, talking about interviewing voters or political dissidents, but like not people kind of on the margins of politics. So, I learned the most from the communities I was working with. (PS06).*

4.3. **"It's not doing a lot for you:" The Role of the IRB.** The most cited source of input on practices related to sensitive data were the IRBs at our respondents' institutions, to which we turn now.

Almost all respondents talked about their IRB in the context of handling sensitive data, but the degree to which they found the IRB process helpful differed significantly. One group of researchers found the IRB largely unobtrusive but also of little help:

> *My IRB was the most hands-off. My entire project was given an exemption so I was extremely hands-off. (PS16)*

Other researchers found their IRB process useful as a starting point ("The IRB process was helpful to think about it at least on a basic level;" PS10) and in some cases (perhaps not coincidentally outside of the United States) praised their understanding and thoroughness:

> *But also, at the [university], the ethics board is extremely conscious that doing fieldwork in Indonesia is different from doing field work in the US or Canada. Their concerns are the same, but the way of, you know, mitigating risks are extremely different and require different strategies. (PS20)*

Other scholars view IRBs as a nuisance that makes qualitative research harder and less attractive to conduct:

> *[I]t is impractical. It never works, you never end up following the protocol, simply because these are just a set of practices that are not applicable to the real world most of the time. (PS11)*

> *I have to admit that I find the process [of getting IRB approval] a disincentive to doing interviews and, though I am happy that I am now doing more quantitative work that does not require endless applications to the IRB. IRB is one more step. It's one more headache. (PS12)*

A group of researchers echoes the longstanding concerns in the qualitative methods literature note above that IRBs are fundamentally at odds with the practice of qualitative research:

> *So, you know, the IRB forms ask you questions like, what are your treatments? And you're like, I don't know. I'm going to ask them questions. (PS06)*

Another researcher points out that IRBs often make rigid assumptions about what is "best" for participants that may contravene those participants' well-informed judgments. We are quoting from this interview at length because it demonstrates the level of local knowledge and nuance that goes into protecting participants and sensitive information.

> *So, in [my research site] it was interesting because often people preferred to have their interviews in a very public setting. So, it would be in their home compound. So, near their house or their farmland, but in view of the street, and that was because they wanted people walking by to know that I was there. That this Muslim with this white woman who had come through town that she was here that there wasn't anything shady or weird going on. [...] And so, but I really do those interviews to try to give the respondents as much space to kind of set that up as possible and, and that can be tricky within the confines of an IRB. So, you know things that I don't budge on are the consent procedure and things like I mean those are always rigid, but you know. I don't sometimes—the IRB says, 'well you know you're going to guarantee them a private anonymous space,' that's like, well, not if they don't want that, right? ()*

5. Discussion

Researchers are anxious about data management. None of the researchers we talked to—most with Ph.D. degrees from leading programs in their disciplines—felt that they had received adequate training in data management. Researchers who did feel they needed additional resources gathered these together from a patchwork of sources: more experienced researchers and, occasionally, their publications, local sources such as journalists, and respondents themselves. Even more notable than the resources listed by respondents are those that we did not encounter in our interviews. Very few participants who described their dissertation research attributed a significant role in informing their strategy to their advisor or committee. Equally, no respondent mentioned input from research data specialists, neither from an institution's research data services nor from a data repository, even though several talked about periods that saw a significant expansion of such services (Rachlin, 2022).

IRBs play a central role in shaping researchers' approaches to data, their management, and security. While some researchers did find the IRB process useful, many others considered it a mere formality with little application to their research process, and still others an active hindrance to good research practices. It is important to emphasize that this critique of IRBs is not due to a lack of attention to research ethics or data security. On the contrary, researchers who are critical of IRBs often find that they do not help promote ethical practices and may in some cases (as described by the researcher quoted at the end of Section 4.3) make ethical research harder. The skepticism of IRBs should not come as a surprise, given long-held criticisms of the IRB process by qualitative researchers noted in Section 4.3,

This patchwork approach to learning about data management has several downsides. The issue most clearly observable in our interviews is the dramatically varying levels of data protection measures used in the field, even where threat levels are comparable. This means that some data may be at risk, while others may be over-protected, leading to inefficiencies during data collection and potentially overly restrictive protocols impeding future use of data. Relying on networks of information, especially as passed on from senior scholars, has important benefits. Knowledge of the realities of fieldwork and, ideally, familiarity with a particular research context are major assets in crafting data management strategies. At the same time, widely different levels of access to experienced senior faculty and their advice by early-career researchers raise significant issues of equity. Moreover, senior faculty in the social sciences are not necessarily familiar with best practices in information security and may offer outdated advice. Finally, and most relevant to the possibilities of future use and data sharing, advice by more experienced researchers, who were rarely subject to data-sharing mandates, may be overly restrictive and prevent future sharing of data, e.g. due to lack of consent to data sharing by participants (Mozersky, et al., 2020 on consent for data sharing in qualitative research; VandeVusse, et al., 2022).

Information on good practices for data security in the field is not unavailable: there is an emerging literature, much of it written by other social scientists, that underlies strategies and good practices. There are also resources offered by institutions' research data service groups (most commonly through the library) as well as data repositories on data management and security. (Corti, et al., 2014 is a prominent and early example; similar resources have since been created by other repositories.) The problems are that researchers may not be aware of these resources, that they are very rarely referenced in graduate training, and that they are not part of any of the formal channels that scholars utilize to prepare for fieldwork. Moreover, sound strategies for data security in the field cannot follow a cookie-cutter approach: they must take into account the nature of possible threats, local infrastructure and customs, and, not least, researchers' comfort with given technologies.

It may be tempting, then, to advocate for an expansion of the role and expertise of IRBs—the only institution that reliably interacts with researchers prior to fieldwork and reviews detailed plans of their research. As we document above, however, there is a significant and justified skepticism among qualitative and field-work-intensive researchers about IRBs, and expanding their role would meet resistance and could well be counterproductive.

The results reported in this article are part of a larger research project into the security and privacy practices of researchers managing qualitative data (Weber, et al., 2021). While this paper, and the broader project, benefit from comparing data management strategies across disciplines there are important limitations to this work: We interviewed just 18 participants from each discipline. Wee draw upon a broad literature in data management and privacy in order to support our interpretations, but there are may be important perspectives missing from our report. Our sample also skews toward North American researchers, although these individuals do research around the world. Some of the observations we make are specific to the IRB structure of universities. They may not be applicable to researchers working outside the United States, or in settings where IRBs do not regulate data collection. Based on our sampling methods, which prioritized researchers with strong publication records, we believe that our participants are among the more sophisticated field researchers. The sample does, however, skew relatively young, with most researchers describing dissertation research.

6. Conclusion

The current state of data management in fieldwork-based research is concerning. This is not, we emphasize, because of any of the practices described by our respondents: the researchers we talked to cared deeply about the safety of their participants and any risks posed to them due to their data management practices. Yet, many view their practices with significant trepidation and lament the lack of formal training to help them do better. Instead, researchers draw on an incompatible patchwork of resources, many of them informal, to build strategies to manage and secure their data in the field.

As we have argued throughout this paper, this status quo has significant downsides. Given the lack of formalization, researchers may, despite best intentions, end up with either insufficient or unnecessarily cumbersome strategies for securing their data. More directly applicable to the re-use of sensitive data, most information resources they do access (such as senior faculty or IRBs) are biased towards restrictive views of data security and may steer them away from management practices that would allow later re-use of their data by other researchers—even when properly secured in a secure setting.

Mainstreaming better training for data management within social science graduate curricula would help to improve the status quo. However, few graduate programs have suitable faculty to teach such material, and many may favor more subject-specific content for the coursework.

Instead, we envision a model that promotes a consultative approach across and beyond the campus. IRBs, as a common point of contact, would mostly take a lightweight approach to reviewing applications and instead point to other resources that allow researchers to combine local/subject-level expertise with data management and information technology (IT) security knowledge. For example, a fieldwork-bound graduate student could consult with a data management specialist from the library or a domain repository alongside a faculty supervisor who is knowledgeable about the prospective field site. As part of that consultation, the student might access resources compiled by the university's IT department regarding available technologies (such as secure cloud storage and secure mobile data collection applications). Output from that consultation could be captured for the IRB application and satisfy requirements with minimal input from the IRB proper. Several universities are already moving in more collaborative directions to data security as we envision in this article—Indiana University's "SecureMyResearch" project, for example may serve as a model for some of these components (Center for Applied Cybersecurity Research, 2022). Data repositories can, where resources are available, be active participants in such consultations or inform them through intermediaries, such as data librarians with strong familiarity with the resources that they offer for securing sensitive data.

As more research data are shared in the wake of an ongoing shift in norms and regulations, more *sensitive* data will be shared. As we have argued, the ideal approach to securing sensitive data should include the whole life cycle, long before data ever enters, e.g., a secure enclave. This is particularly true for the types of data that have received less attention in discussions about securing sensitive data, such as the fieldwork-based sensitive data that we examined in this article.

Acknowledgement

We thank Sarah Nguyễn for exceptional research assistance on this project.

References

Aldridge, J., Medina, J., and Ralphs, R. (2010). The problem of proliferation: guidelines for improving the security of qualitative data in a digital age. *Research Ethics*, 6(1):3-9. https://doi.org/10.1177/174701611000600102

Altman, M., Wood, A., O'Brien, D., Vadhan, S., and Gasser, U. (2016). Towards a modern approach to privacy-aware government data releases. *Berkeley Journal of Technology Law*, (3):1967. https://doi.org/10.15779/Z38FG17

Arias, E. D. (2014). *Data security in highly violent settings* (7; DSD Working Papers on Research Security). Social Science Research Council. http://webarchive.ssrc.org/working-papers/DSD_ResearchSecurity_07_Arias.pdf

Baez, B. (2002). Confidentiality in qualitative research: reflections on secrets, power and agency. *Qualitative Research*, 2(1):35-58. https://doi.org/10.1177/1468794102002001638

Barratt, M. J., and Maddox, A. (2016). Active engagement with stigmatised communities through digital ethnography. *Qualitative Research*, 16(6):701-719. https://doi.org/10.1177/1468794116648766

Bishop, L. (2009). Ethical sharing and reuse of qualitative data. *Australian Journal of Social Issues*, 44(3):255-272. https://doi.org/10.1002/j.1839-4655.2009.tb00145.x

Campbell, R., Goodman-Williams, R., Javorka, M., Engleton, J., and Gregory, K. (2023). Understanding sexual assault survivors' perspectives on archiving qualitative data: implications for feminist approaches to Open Science. *Psychology of Women Quarterly*, 47(1):51-64. https://doi.org/10.1177/03616843221131546

Center for Applied Cybersecurity Research (2022). *SecureMyResearch*. Indiana University. https://cacr.iu.edu/projects/SecureMyResearch/index.html

Corti, L., Eynden, V. van den, Bishop, L., and Woollard, M. (2014). *Managing and Sharing Research Data: A Guide to Good Practice*. SAGE. ISBN: 9781526460264

Crosas, M., Gautier, J., Karcher, S., Kirilova, D., Otalora, G., and Schwartz, A. (2018). Data policies of highly-ranked social science journals. *SocArXiv*. https://doi.org/10.17605/OSF.IO/9H7AY

Demery, A.-J. C., Pipkin, M. A. (2021). Safe fieldwork strategies for at-risk individuals, their supervisors and institutions. *Nature Ecology and Evolution*, 5(1):Article 1. https://doi.org/10.1038/s41559-020-01328-5

European Commission (2016). *Guidelines on FAIR Data Management in Horizon 2020*. European Commission Directorate for General for Research and Innovation. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Feldman, S., and Shaw, L. (2019). The epistemological and ethical challenges of archiving and sharing qualitative data. *American Behavioral Scientist*, 63(6):Article 6. https://doi.org/10.1177/0002764218796084

Goodey, G., Hahnel, M., Zhou, Y., Jiang, L., Chandramouliswaran, I., Hafez, A., Paine, T., Gregurick, S., Simango, S., Palma Peña, J. M., Murray, H., Cannon, M., Grant, R., McKellar, K., and Day, L. (2022). *The State of Open Data 2022*. Digital Science. https://doi.org/10.6084/m9.figshare.21276984.v5

Goodwin, D., Mays, N., and Pope, C. (2020). Ethical issues in qualitative research. In *Qualitative Research in Health Care* (pp. 27-41). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119410867.ch3

Guillemin, M., and Gillam, L. (2004). Ethics, reflexivity, and "ethically important moments" in research. *Qualitative Inquiry*, 10(2):261-280. https://doi.org/10.1177/1077800403262360

Hammersley, M., and Traianou, A. (2012). *Ethics in Qualitative Research: Controversies and Contexts*. SAGE Publications Ltd. https://doi.org/10.4135/9781473957619

Hedrick, T. E. (1988). Justifications for the sharing of social science data. *Law and Human Behavior*, 12(2):163-171. https://doi.org/10.1007/BF01073124

Henry, C., Gohdes, A., and Dorff, C. (2022). Digital Footprints and Data-Security Risks for Political Scientists. *PS: Political Science & Politics*, 55(4):804-808. https://doi.org/10.1017/S1049096522000543

Jacobs, A. M., Büthe, T., Arjona, A., Arriola, L. R., Bellin, E., Bennett, A., Björkman, L., Bleich, E., Elkins, Z., Fairfield, T., Gaikwad, N., Greitens, S. C., Hawkesworth, M., Herrera, V., Herrera, Y. M., Johnson, K. S., Karakoç, E., Koivu, K., Kreuzer, M., … Yashar, D. J. (2021). The qualitative transparency deliberations: insights and implications. Cambridge University Press, *19*(1):32. https://doi.org/10.1017/S1537592720001164

Jerolmack, C., and Murphy, A. (2017). The ethical dilemmas and social scientific trade-offs of masking in ethnography. *Sociological Methods and Research*, 1-27. https://doi.org/10.1177/0049124117701483

Kaiser, K. (2009). Protecting respondent confidentiality in qualitative research. *Qualitative Health Research*, 19(11):1632-1641. https://doi.org/10.1177/1049732309350879

Kapiszewski, D., and Karcher, S. (2020). Making Research Data Accessible. In C. Elman, J. Mahoney, and J. Gerring (eds.), *The Production of Knowledge: Enhancing Progress in Social Science* (pp. 197-220). Cambridge University Press. https://doi.org/10.1017/9781108762519.008

Kapiszewski, D., McLean, L., and Read, B. (2015). *Field Research in Political Science*. Cambridge University Press. https://doi.org/10.1017/CBO9780511794551

Katz, J. (2015). A theory of qualitative methodology: the social system of analytic fieldwork. *Méthod(e)s: African Review of Social Sciences Methodology*, 1(1-2):131-146. https://doi.org/10.1080/23754745.2015.1017282

Kirilova, D., & Karcher, S. (2017). Rethinking data sharing and human participant protection in social science research: applications from the qualitative realm. *Data Science Journal*, 16. https://doi.org/10.5334/dsj-2017-043

Kovats-Bernat, J. C. (2002). Negotiating dangerous fields: pragmatic strategies for fieldwork amid violence and terror. *American Anthropologist*, 104(1):208-222. https://doi.org/10.1525/aa.2002.104.1.208

Kuula, A. (2011). Methodological and ethical dilemmas of archiving qualitative data. *IASSIST Quarterly*, 34(3-4):12. https://doi.org/10.29173/iq455

Lake, M., and Parkinson, S. E. (2017). The Ethics of Fieldwork Preparedness. *Political Violence at a Glance*. https://politicalviolenceataglance.org/2017/06/05/the-ethics-of-fieldwork-preparedness/

Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., and Wutich, A. (2019). Qualitative data sharing: data repositories and academic libraries as key partners in addressing challenges. *American Behavioral Scientist*, 63(5):643-664. https://doi.org/10.1177/0002764218784991

Morse, J. M. (1998). The contracted relationship: ensuring protection of anonymity and confidentiality. *Qualitative Health Research*, 8(3):301-303. https://doi.org/10.1177/104973239800800301

Mozersky, J., Parsons, M., Walsh, H., Baldwin, K., McIntosh, T., and DuBois, J. M. (2020). Research participant views regarding qualitative data sharing. *Ethics and Human Research*, 42(2)):13-27. https://doi.org/10.1002/eahr.500044

Naderifar, M., Goli, H., and Ghaljaie, F. (2017). Snowball sampling: a purposeful method of sampling in qualitative research. *Strides in Development of Medical Education*, 14(3). https://doi.org/10.5812/sdme.67670

National Science Foundation (2011). *Dissemination and Sharing of Research Results*. http://www.nsf.gov/bfa/dias/policy/dmp.jsp

National Institutes of Health (2020). *Final NIH Policy for Data Management and Sharing* (NOT-OD-21-013). https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

Niu, J., and Hedstrom, M. (2008). Documentation evaluation model for social science data. *Proceedings of the American Society for Information Science and Technology*, 45(1):11. https://doi.org/10.1002/meet.2008.1450450223

Okely, J. (1994). Thinking through fieldwork. In A. Bryman & R. G. Burgess (eds.), *Analyzing Qualitative Data*. Routledge.

Onwuegbuzie, A., and Leech, N. (2015). Sampling Designs in Qualitative Research: Making the Sampling Process More Public. *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2007.1636

Parkinson, S. E., and Wood, E. J. (2015). Transparency in intensive research on violence: ethical dilemmas and unforeseen consequences. *Qualitative and Multi-Method Research*, 13(1):22-27. https://doi.org/10.5281/zenodo.893081

Pascale, J., Willimack, D. K., Bates, N., Lineback, J. F., and Beatty, P. C. (2020). *Issue Paper on Disclosure Review for Information Products with Qualitative Research Findings* (Research and Methodology Directorate, Center for Behavioral Science Methods Research Report Series Survey Methodology #2020-01). U.S. Census Bureau. http://www.census.gov/content/dam/Census/library/working-papers/2020/adrm/rsm2020-01.pdf

Peter, M., and Strazzari, F. (2017). Securitisation of research: Fieldwork under new restrictions in Darfur and Mali. *Third World Quarterly*, 38(7):1531-1550. https://doi.org/10.1080/01436597.2016.1256766

Rachlin, D. J. (2022). Academic librarians and research data services: preparation and attitudes revisited. *Internet Reference Services Quarterly*, 26(4):1-13. https://doi.org/10.1080/10875301.2022.2072042

Ramos, M. C. (1989). Some ethical implications of qualitative research. *Research in Nursing & Health*, 12(1):57-63. https://doi.org/10.1002/nur.4770120109

Shaw, I. (2008). Ethics and the practice of qualitative research. *Qualitative Social Work*, 7(4):400-414. https://doi.org/10.1177/1473325008097137

Thurston, A. F., and Pasternak, B. (2019). *The Social Sciences And Fieldwork In China: Views From The Field*. Routledge. ISBN: 9780367311308

UK Research and Innovation. (2022). *Making your Research Data Open*. https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/

VandeVusse, A., Mueller, J., and Karcher, S. (2022). Qualitative data sharing: participant understanding, motivation, and consent. *Qualitative Health Research*, 32(1):182-191. https://doi.org/10.1177/10497323211054058

Velden, T., and Lagoze, C. (2013). The extraction of community structures from publication networks to support ethnographic observations of field differences in scientific communication. *Journal of the American Society for Information Science and Technology*, 64(12):2405-2427. https://doi.org/10.1002/asi.22929

Wackenhut, A. F. (2018). Ethical considerations and dilemmas before, during and after fieldwork in less-democratic contexts: some reflections from post-uprising Egypt. *The American Sociologist*, 49(2):242-257. https://doi.org/10.1007/s12108-017-9363-z

Weber, N., Karcher, S., Nguyễn, S., and Secen, S. (2021). *Privacy Encodings for Sensitive Data (PESD)—Case Study Pre-Registration.* https://doi.org/10.17605/OSF.IO/ABY9M