Out of Sight, Out of Mind? Exploring Data Protection Practices for Personal Data in Usable Security & Privacy Studies

Florin Martius
martius@cs.uni-bonn.de
University of Bonn
Bonn, Germany

Luisa Jansen
luisa.jansen@unibe.ch
University of Bern
Bern, Switzerland

Lukas Struck
struckl@uni-bonn.de
University of Bonn
Bonn, Germany

Arthi Arumugam
arumugam@uni-bonn.de
University of Bonn
Bonn, Germany

Lisa Geierhaas
geierhaa@cs.uni-bonn.de
University of Bonn
Bonn, Germany

Anna-Marie Ortloff
ortloff@cs.uni-bonn.de
University of Bonn
Bonn, Germany

Matthew Smith
smith@cs.uni-bonn.de
University of Bonn
Fraunhofer FKIE
Bonn, Germany

Christian Tiefenau
tiefenau@cs.uni-bonn.de
University of Bonn
Bonn, Germany

Abstract

Adherence to data protection measures such as pseudonymization or anonymization is critical in human subjects research because it has a direct impact on the confidentiality of participants' sensitive information, trust in research practices, and compliance with ethical and legal standards. Regulations such as the General Data Protection Regulation (GDPR) and guarantees made by researchers in informed consent forms mandate strict protocols for data security. However, compliance with these is not always straightforward. To gain qualitative insights into data protection practices in the field of Usable Security and Privacy (USP), we conducted interviews with 22 practitioners (five professors, eight researchers, nine data protection officers) and one focus group with five researchers. Overall, our results show a high awareness of ethical and legal responsibilities but highlight many practical and procedural issues. Based on these, we make concrete recommendations on how to improve the protection of personal data in research.

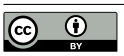
CCS Concepts

• **Social and professional topics** → **Professional topics**; • **Security and privacy** → **Privacy protections**; • **Human-centered computing**;

Keywords

Human Subjects Research, Personal Data Handling, Ethics, Usable Security and Privacy

ACM Reference Format:
Florin Martius, Luisa Jansen, Lukas Struck, Arthi Arumugam, Lisa Geierhaas, Anna-Marie Ortloff, Matthew Smith, and Christian Tiefenau. 2025. Out of Sight, Out of Mind? Exploring Data Protection Practices for Personal Data in Usable Security & Privacy Studies. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3706598.3713654



CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713654

1 Introduction

When conducting human-centered studies in the field of Human Computer Interaction (HCI), researchers must adhere to legal and ethical standards. Regulations such as the General Data Protection Regulation (GDPR) in Europe [63] or similar acts in the U.S., e.g., the California Consumer Privacy Act (CCPA), Virginia Consumer Data Protection Act (VCPDA), Connecticut Data Privacy Act (CTDPA), Utah Consumer Privacy Act (UCPA), or Nevada Privacy Law and Senate Bill 220 (SB220) [9, 10, 35, 44, 45], mandate strict protocols for data security and data protection. To comply with ethical standards, researchers need to maintain participants' privacy to avoid harm to the participants. Data protection measures, such as anonymization and pseudonymization, help mitigate risks like identity theft or misuse of sensitive information, aligning with the obligation to do no harm. Additionally, breaches of personal data not only harm participants directly but also erode trust in researchers and potentionally discouraging future participation in research.

It is worth noting that breaches of confidentiality can also have legal consequences. Consequently, it is important that researchers protect the personal data of their participants both for ethical and legal reasons.

The Association for Computing Machinery (ACM) guidelines, which must be adhered to when submitting to conferences like CHI, also emphasize the importance of preserving the privacy of research participants [1]. They reflect a broader ethical commitment within the research community to ensure that all research activities are conducted with the utmost respect for participant autonomy and confidentiality.

However, in practice, adhering to ethical and legal standards can be complex since researchers may face various issues (e.g.,

legal ambiguities, lack of processes, and academic pressures) when handling personal data. When reviewing our own practices, we saw room for improvement and identified several issues we needed to fix - some of which were straightforward and others more complex. As researchers in the field of Usable Security and Privacy (USP), we felt we could be doing better. Conversations with colleagues painted a similar picture in other USP groups.

To explore the handling of personal data within the research community, we conducted an exploratory, open-ended study. Our aim was to understand the general measures in place to protect the participants' data, while also considering broader challenges as raised by participants. Rather than assessing compliance with specific legal frameworks, we sought to obtain an overview of current practices and identify common issues faced by researchers in our field. To get an overview of current data protection practices in our field, we defined the following research questions:

RQ 1: How is personal data currently handled in USP research?
RQ 2: What challenges do USP researchers face when working with personal data regarding data protection?
RQ 3: What processes are used to ensure proper data protection?

To answer these questions, we conducted 13 interviews with primary and supervising researchers from the domain of USP, followed by 7 interviews with 9 Data Protection Officers (DPOs) of research institutions that have published in the field of USP. We chose to study the USP community because most members possess an understanding of security and privacy technologies, providing both the awareness and capability to implement protective mechanisms. In addition, we wanted to gather insight into usability issues concerning data processing, for which this community is well suited.

Our study highlights that (USP) researchers are highly aware of the importance for privacy-conscious data handling. They prioritize protecting participants' data, complying with laws, upholding ethics, and maintaining participants' trust. However, our results also show that managing personal data is often challenging for researchers, sometimes leading to unintentional violations of data protection regulations. We identified two main issues: First, there is uncertainty about how to handle personal data during processing. Researchers struggle with decisions like what qualifies as identifiable information and whether certain data should be deleted. Second, even when aware of proper procedures, researchers often face difficulties in following them due to external factors like time constraints or pressure to publish, which can hinder the implementation of proper data protection practices despite their good intentions.

Our results suggest that holistic processes that involve multiple stakeholders and accompany the entire research project help to ensure compliance. We thus argue for a more systematic approach to data handling in USP research.

To this end, we provide actionable recommendations for all relevant stakeholders, including researchers and organizations, aimed at strengthening data protection processes in research and minimizing the risk of harm to participants.

2 Related Work

There are two closely related terms in the realm of data protection: *Personally Identifiable Information (PII)* and *personal data*. While both refer to information that can identify individuals, they differ in context and scope.

PII refers to any data that can directly identify an individual, such as a name, social security number, or email address, and is widely used in the United States [18]. The broader term *personal data*, on the other hand, includes any information related to an identifiable person, either directly or indirectly, which can encompass not only *PII* but also pseudonymized or aggregated data that could still be traced back to an individual [63]. This term is more used in Europe, particularly under data protection laws like the GDPR [63].

In this paper, we primarily use the term *personal data*, as it covers a wider range of information, including *PII*.

2.1 Research Ethics

Research ethics emphasize participant rights and well-being. Safeguarding data protection and participants' control over how their personal data is used – referred to as informational self-determination – are central to ethical research guidelines such as the ACM's policy on research with human subjects [1]. Institutional Review Boards (IRB), introduced after WWII, have since become a standard requirement for research involving human participants [26].

In recent years, traditional techniques to ensure participants' informational self-determination have been critiqued. A growing body of literature has shown the ineffectiveness of traditional informed consent mechanisms [33, 62]: Consent forms tend to be overly complex, mirroring the usability issues seen in privacy notices, which leads to participants consenting without fully understanding the risks involved [52, 53, 58]. Also, anonymization, which is frequently used as a protective measure, becomes increasingly insufficient as advances in technology make it easier to re-identify individuals from de-identified datasets [40]. Apart from these techniques, IRBs have been critiqued for their bureaucratic processes, which can delay research without necessarily preventing ethical issues [7]. These problems stress the responsibility put on individual researchers to handle the personal data they collect with great care, going above mere compliance with regulations [37, 66].

In the HCI community, the last years have seen a noticeable shift toward improved research ethics, openness, and transparency: A comparative study of CHI papers between 2017 and 2022 revealed that practices related to acquiring IRB approval, reporting consent collection, and being transparent about participant compensation have significantly improved [56]. However, despite these developments, each of these ethical practices was still only observed in around half of the CHI 2022 papers, indicating room for improvement.

2.2 Legal Aspects

Data protection regulations like the CCPA in the U.S., the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada, and the GDPR in Europe have placed legal obligations on researchers handling personal data [6, 9, 63]. These laws seek to protect individuals by regulating the collection, processing, and sharing of personal information.

Among these regulations, the GDPR is recognized as particularly stringent and far-reaching. Enacted in 2018, it applies not only to entities within the European Union (EU) but also to any organization that processes the personal data of individuals located in the EU [21, 63]. A unique aspect of the GDPR is its broad definition of personal data, which includes any information that can directly or indirectly identify an individual. Pseudonymized data also remains protected under the GDPR if it can potentially be traced back to an individual [63]. However, even under GDPR, the provisions of the overarching EU regulation allow for member states to specify some aspects within their national laws [63]. As a result, navigating the different legal frameworks across jurisdictions can be a significant challenge for researchers [59]: Which laws are applicable does not only depend on the location of the researcher's organization but also other factors (e.g., the country of residence of participants [21, 63]) and the type of data [27]. The extensive body of work by legal scholars explaining data protection regulations for scientists showcases the complexity of the current laws [29, 64]. Legal experts have also tried to estimate the impact of GDPR on researchers and their workflows [13, 22], but have not yet presented empirical assessments of the actual effects.

The few existing investigations into actual compliance of researchers when handling participants' data in behavioral sciences primarily stem from open-science-focused investigations. For that reason, they are centered around data-sharing practices and touch on issues of data protection only superficially. In these interview and survey studies, the researchers under investigation report hesitations to share their datasets due to concerns about compliance with legal obligations [3, 32, 67], showcasing the contrary interests of open science and data protection. Hussey [28] also reports the ethical concerns of authors (e.g., the protection of participants' data) as a reason not to share data. Zillich et al. find that regulations of data protection not only create a feeling of uncertainty but pose a bureaucratic burden on scientists [69], highlighting the need for clearer frameworks that balance data protection with scientific openness. Hallinan et al. assessed the compliance with the GDPR by researchers in terms of informed consent forms [25]. The authors found that the 101 assessed psychological informed consent forms deviate significantly from the GDPR. While the responsible researchers seemed to be well-intended when creating informed consent forms, they lacked the resources to fulfill their legal obligations [25]. Going beyond informed consent forms, actual data protection practices by researchers handling personal data in any empirical discipline are yet to be investigated in detail.

2.3 Participants' Trust

Participants' trust is essential for research, particularly in studies involving personal data. Without trust, people might refrain from participating in research or could be less likely to provide honest responses, which can undermine the validity of research findings [23, 41]. Researchers, generally, are aware that a trusting relationship with participants is a necessity for their work and understand it as part of their professional ethics [24].

Krause et al. found in an analysis of poll data that trust in science is high, especially for controversial topics [34]. While individuals are skeptical of companies' privacy-related communications, they generally believe information on data handling provided by researchers [55] and expect research to be performed in accordance with ethical standards [16]. This trust seems to be prominent in participants' perceptions of risks associated with data handling: Even though some participants do not understand the information researchers provide, they are unconcerned due to their belief in scientific standards [31]. From the perspective of participants interviewed by Guillemin et al., research institutions guarantee compliance with ethical standards [23]. This reliance on researchers' practices implies a special responsibility for the personal data of human research subjects.

2.4 Summary

Previous work on data handling practices in research across all empirical sciences has examined the start of the research process: informed consent [25, 33, 62] and a possible outcome: data sharing practices [3, 32, 67], as well as highlighted challenges due to bureaucratic obstacles [7, 69] and legal complexity [29, 64], but has not evaluated how researchers studying humans manage personal data throughout the entire research process.

In our work, we fill this gap through an exploratory investigation of the challenges USP researchers perceive throughout the complete research process from study planning to publication and beyond, and the support mechanisms in place to aid them. Through our additional interviews with DPOs we specifically focused on institutional processes to aid the researchers.

3 Methodology

To investigate the data handling process in USP human subjects research, we followed a process of theoretical sampling and continuous analysis, taken from constructive grounded theory [8]. We first conducted and analyzed interviews with five supervising researchers (i.e., professors) and eight primary researchers (e.g.., graduate students) about their practices regarding the handling of personal data during research. In addition, we conducted one focus group with five primary researchers to identify uncertainties in classifying what identifying information is. During the focus group, further problems in handling personal data were discussed. In order to investigate the data handling process from a different angle, we conducted seven interviews with nine DPOs of research institutes. The purpose of these interviews was to provide both a legal and an institutional perspective to assess how data handling processes should be properly implemented. Additionally, we wanted to find out how institutions support their researchers in complying with data protection regulations. Figure 1 shows a visual representation of our process. Our study materials, i.e., the informed consent forms and demographic questionnaires, can be found in the supplemental material. The guidelines for all interviews are in Appendix A.

3.1 Ethics

We were very conscious of the fact that our study leads to participants sharing information on ethical or legal mishaps that could cause trouble if linked to a participant. To minimize this risk, we collected very little personal data and fully anonymized our data

before submission. We weighed this risk against the potential benefit to the research community and future study participants. In our view, the risk-benefit ratio was acceptable.

The study received approval from our affiliated IRB. Prior to participation, all participants were informed about the study's purpose and their rights as participants. Participation was entirely voluntary, and participants were free to withdraw from the study at any time without facing any consequences. We also encouraged participants to ask questions both before and after the study to ensure clarity and understanding. All data from the surveys and interviews were collected, stored, and managed in full compliance with the GDPR.

Despite our work highlighting aspects of data security that should be improved, we want to explicitly state that all participants were active in maintaining the security of participant's personal data, and we saw no signs of gross negligence. Where issues arose, we do not blame the researchers; instead, we think these are clear indications of the need for better processes, support infrastructures, and tools.

3.2 Positionality

Since background, experience, and prior knowledge can influence the data collection and analysis process [19, 46, 57], we communicate these in the following. We are eight researchers working with empirical methods in USP. R1 and R3-R8 belong to a university that is subject to the GDPR, while R2 belongs to a university that is subject to the Swiss Federal Act on Data Protection (FADP). R1-R4 are PhD students. R5 and R6 are student assistants. R7 is a postdoc. R8 is a professor. All but R5 have experience with qualitative fieldwork and analysis. In the context of data responsibilities in this work, R1 is the primary researcher, R2-R6 are supporting researchers, and R7 and R8 are supervising researchers. Neither university has centralized data management processes. The IRBs require statements about data handling but not a formal data management plan. In R7's and R8's group, the primary researchers are in charge of study data and the technological measures to ensure security and data protection, as well as keeping track of who has copies of the data. Supporting researchers are in charge of their data and the technological measures to ensure security and data protection. The final responsibility of ensuring compliance and long-term archiving lies with R8, supported by R7. Before embarking on this project, data minimization during collection, access control, and encryption were mandated. What was considered identifying data and the associated risks were discussed on a per-study basis. However, no formal processes were in place - in particular, there were no processes to ensure timely deletion of data no longer needed. This was done by the primary or supervising researchers in an ad-hoc manner which was not ideal. This was a key motivation for embarking on this work. Since the lack of processes is also a key result we draw from our analysis, we want to highlight that this is something we suspected and consequently want to inform the reader of this potential confirmation bias.

3.3 Recruitment of USP Researchers

We recruited the participants for our researcher interviews in person at a USP community event. This event had an international audience, with approximately 60 attendees, the majority of whom were affiliated with institutions in Europe or the United States. In order to participate in our study, researchers had to have conducted at least one study on a USP topic in the past. We made sure to recruit PhD students and post-docs as well as professors to ensure that we had both the perspectives of people who are in the role of a primary researcher as well as supervising researchers. To maintain data integrity, we ensured that participants in the same role were not affiliated with the same research department. Given the exploratory nature of our study, which does not test compliance with specific regulatory frameworks, we included a diverse sample of researchers from both the EU and the US to capture a wide range of perspectives. While we acknowledge the differences between data protection laws in these jurisdictions (e.g., the GDPR in the EU versus state and federal frameworks in the US), our focus was on identifying general practices and principles that researchers use to protect sensitive information across regulatory contexts.

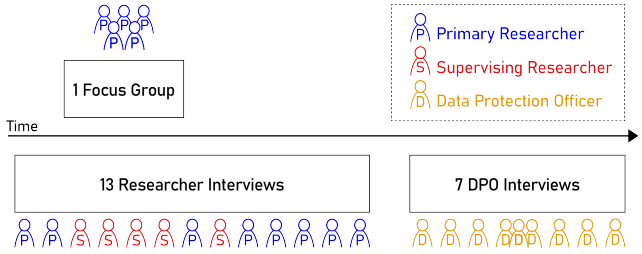


Figure 1: Data collection process for interviews and focus group.

We refer to researchers who actively collect and handle human subjects' data as "primary researchers." This is often but not necessarily the first author of a project. Importantly, they need to feel ownership and responsibility for the data. "Supporting researchers" are colleagues who also work with the data but in a supporting role without the same feeling of ownership (often co-authors). Finally, we call those who oversee a study and manage the primary and supporting researchers "supervising researchers" (often professors or post-docs). However, these roles are not tied to the positions. We refer to all of these groups together as "researchers." In the interviews, the professors in the study sample spoke from the perspective of their supervising role, while the PhD students and post-docs spoke about their experience in the primary researcher's role.

Researchers participating in our interview study did not receive monetary compensation but got a small toy present as a sign of our gratitude.

3.4 Recruitment of Data Protection Officers

For the interviews with the DPOs, we specifically targeted individuals associated with institutions in Europe and the United States that had demonstrated expertise in conducting numerous studies in USP, to reflect the research area of our researcher participants. To identify appropriate participants, we reviewed the program of the most recent Symposium on Usable Privacy and Security (SOUPS) conference and selected institutions known to the authors for their

strong presence in this area. Using this information, we invited DPOs from these institutions to participate in the study, reaching out to them directly via email. We progressively expanded the pool of invited institutions until reaching data saturation. A total of 25 emails were sent, and DPOs from seven institutions agreed to participate in the study. As an incentive, we offered to provide a preprint of the results of our study and an additional compensation of €30 or $30 based on the location.

3.5 Demographics

We decided to collect only the demographic data we needed to contextualize the results obtained from our sample. The demographic information for the three participant groups (researchers, focus group, and DPOs) can be found in Table 1. The researchers group consisted of participants from five countries, with the majority coming from Germany and the USA, and their professional positions included PhD students, post-docs, and professors. The focus group was composed entirely of PhD students, primarily from Germany. The participants in the focus group and the interview group did not overlap. Six interviews were conducted with DPOs from Europe and one with a DPO from the U.S. One of these interviews involved three DPOs working at the same institution, with one of them serving as the primary DPO. All DPOs were fully qualified lawyers.

Table 1: Demographics of study participants.

Demographics	Researcher n=13	Focus Group n=5	DPO n=9
Country			
Germany	5	4	5
USA	5		1
Netherlands	1		2
Belgium			1
Switzerland	1	1	
Luxembourg	1		
Position			
Professor	5		
Post-Doc	2		
PhD Student	6	5	

3.6 Data Collection

3.6.1 Researcher Interviews. We approached participants by clearly explaining the purpose of our study. Although standard interview practices generally advise against sharing personal opinions or attitudes about the research topic beforehand, we chose to acknowledge that we saw room for improvement in our own research data handling practices. This intentional deviation from common practices aimed to foster trust and encourage openness with the participants. By being transparent about issues we had identified in our work, we reassured them that our goal was not to assess their compliance with data protection regulations but rather to understand the challenges and potential issues in order to help the community.

If a researcher indicated their willingness to participate, they first completed a short questionnaire. In addition to questions about their demographics and experience as a researcher, the questionnaire included an informed consent form. Participants were assigned a pseudonym to separate research data from contextual demographic data.

As a warm-up question, participants were invited to discuss a previously conducted empirical study. Following our semi-structured interview guideline, we asked open questions about what personal data the researchers collect in their studies, how they process and store these data, and when the data are deleted. We explicitly asked about processes in their organization that guide them in handling research data. We asked follow-up questions, especially about practices that could cause issues or that seemed challenging for the researcher.

All researcher interviews were conducted in person, either in German or English, depending on the participant's preference.

All of the interviewees consented to be audio-recorded; one of the two interviewers, however, took notes as a backup. On average, the interviews lasted 11.2 minutes, ranging from 4.2 minutes to 17.3 minutes.

3.6.2 Researcher Focus Group. In addition to the researcher interviews, we conducted a focus group with five participants to delve deeper into challenges around handling personal data. This focus group took place after four one-on-one interviews because we recognized that many participants were uncertain about what data types and combinations could be considered identifiers. Given that this is a complex topic, we felt that enabling a discussion would be valuable. After obtaining consent from all participants and explaining the study's goals, we opened the session by asking what data researchers consider to be identifiers. This prompted a rich dialogue among participants.

3.6.3 Data Protection Officer Interviews. To investigate the regulations and processes that govern data collection from the institutional view, we conducted seven interviews with nine data protection officers.

Six of the interviews were conducted online, while we had one joint meeting with three DPOs of the same institute that was in-person. Four of the interviews were conducted in English and three in German. Prior to each interview, the DPOs were asked to fill out a questionnaire about their professional role, work experience, and education.

We based our DPO interview questions on themes and open questions that emerged from our researcher interviews. For the first four interviews, we started with a warm-up question about publicly available personal data in research. We continued the interviews with questions about the correct process of handling personal data in the research process, focusing specifically on informed consent, the role of personal data, responsibility, retention periods of data, the end of an empirical study, and guidelines available for researchers. We asked follow-up questions where applicable.

According to our process of theoretical sampling and continuous analysis, after four interviews, we visualized the process described by DPOs and researchers alike to provide a focus point for the remaining DPOs when answering our questions; see Figure 2. We also changed our warm-up question into a more precise question

about the definition of personal data. This was done since the previously used warm-up question led to very broad answers that would serve no further purpose.

Like the previous interviews, two researchers conducted the interview. Again, all participants agreed to be recorded, while we had one interviewer taking notes as a fallback. On average, an interview lasted approximately 32 minutes, ranging from 20 minutes to 51 minutes. We ended data collection when we reached saturation for our research questions.

3.7 Data Analysis

For both interviewed groups, the audio was transcribed and anonymized before further processing. Due to the limited time window of the conference where we recruited researchers, we did not start formal analysis until after all these interviews were conducted. However, the data-collecting researchers discussed their findings and notes at the end of each day. Data analysis was continuous and ongoing during the DPO interviews to enable us to incorporate new insights into our data collection process.

MaxQDA[1] was used for qualitative analysis, following an iterative procedure. We followed a hybrid coding approach, starting with deductive categories we wanted to analyze and adding subcategories as they emerged inductively from the data.

Starting with the researcher interviews, R1 developed a codebook based on the first five interviews. The codebook was then discussed and restructured by R1 and R2 until they reached full agreement. Using these codes and creating new ones as they emerged, both coders analyzed all researcher interviews independently. Subsequently, all codes were discussed, and discrepancies were resolved until a full consensus was reached. Both coders jointly reorganized the resulting codebook to allow for a better overview.

R1 and R2 then analyzed the interviews with the DPOs. Using the same codebook and again adding new codes where necessary, each analyzed all interviews. Then, the coders again discussed any discrepancies until they found a consensus. The final codebook is available in the supplemental material.

3.8 Limitations

Sample: Our total sample size consisted of 13 primary researchers, five professors, and nine DPOs. While this provides a diverse range of perspectives and experiences, the relatively small size of the sample limits our ability to conduct meaningful quantitative analysis or make claims that our findings are broadly generalizable across wider populations. However, it is important to note that the purpose of this study was primarily qualitative and exploratory in nature. In line with established qualitative research methodologies, we did reach theoretical saturation [8] for our analysis. The researchers we interviewed are primarily from Europe and the US. We acknowledge that conceptualizations of terms and specific contents of data protection laws vary between these regions and individual countries (but also depend on other factors). Nevertheless, we combined both samples in this study to provide a holistic overview of USP researchers' data protection behaviors. Consequently, this exploratory study adopts a broader perspective on data protection concepts to accommodate these variations.

[1]https://www.maxqda.com

Generalizability: We only recruited in the field of USP. These researchers might handle research data differently from others in HCI. As an example, researchers in USP, given their familiarity with privacy issues, may overestimate their own abilities in data protection. We are planning follow-up research in further domains, including quantitative surveys, to get a broader view. Still, the DPOs we interviewed are not only associated with USP researchers but deal with all researchers working with personal data, indicating that findings about processes and organizational challenges could concern researchers regardless of disciplines.

Researcher bias: As USP researchers, we are part of the population we studied. As insiders, we might have the advantage of an especially trusting relationship with the participants: Knowing that we face similar challenges, participants might have been more comfortable opening up to us. This is particularly important due to the sensitive research topic and the potential need to disclose unethical behaviors. However, being insiders also carries the risk of bias, potentially blinding us to patterns or behaviors ingrained in our field.

Confirmation bias: We started this project because we found our own data handling processes had room for improvement, which may have introduced a bias toward confirmation of our prior beliefs. We explicitly discussed these potential biases during the coding and analysis process. Still, we would like to encourage researchers from other fields to replicate our findings from an outsider's perspective.

Social desirability: Given the sensitive nature of our research topic, we cannot rule out that participants may have responded in a socially desirable manner. They could have presented their handling of participant data as being more rule-compliant than it actually is. Before each interview, we told participants that we see room for improvement in our data handling processes to counter this bias as much as possible. We hope that by applying this framing, the general social desirability bias (i.e., adhering to ethical and legal standards when dealing with personal data) and the concrete situational social desirability bias (i.e., confirming our beliefs that problematic data handling processes occur in research) counteracted each other.

Group dynamics: The inclusion of both individual interviews and a focus group may have introduced variability in data due to group dynamics. In group settings, such as our focus group and the three-person interview, dominant personalities may influence the conversation, potentially suppressing diverse viewpoints and leading to conformity that might not reflect individual perspectives.

Focus: Our study focuses on data protection practices within the research community, particularly measures such as anonymization, pseudonymization, data retention, and access control. However, data protection extends beyond these security-focused measures and also encompasses participants' rights, such as access to their data, control over its usage, and lawful processing. While we also touch on privacy-related concepts, such as data minimization, privacy extends further to include aspects like informed consent, purpose limitation, and broader ethical considerations in research data usage. While our research provides insights into governance mechanisms for securing participant data, a comprehensive examination of the full scope of both data protection and privacy would require further investigation.

4 Results

We greatly appreciate the openness of our study participants in discussing challenges when protecting research data. To protect them and minimize the risk of identification, we assign pseudonyms to primary researchers and DPOs, where 'P' is for the primary researcher and 'D' is for the DPO interviews. For the group interview of DPOs, we further distinguish between interviewees using pseudonyms D4A, D4B, and D4C. However, we do not assign individual pseudonyms to the supervising researchers to further protect their anonymity, as the number of professors in the field of USP is comparatively small.

4.1 General Attitude

Although we did not specifically ask, almost all researchers we interviewed explicitly mentioned the importance of rule-compliant and ethical data handling in their work, underscoring their motivation to protect their participants' data. However, we saw uncertainty when it came to the practical implementation of these principles in their day-to-day research practices.

From the perspective of the DPOs, regulations, particularly the GDPR, act as an "administrative burden" for researchers, as D3 notes. This sentiment was echoed by D5, who, while acknowledging being "a fan" of GDPR, also recognized that its bureaucracy can sometimes impede research. In this context, D5 described data protection in research as "a question of motivation." They also highlighted the tension between regulatory compliance and academic freedom, noting that adhering to data protection regulations is "not always easy for researchers" [D5].

In some cases, we observed that the act of participating in our study led to introspection among the participants. For instance, P7 described the interviews as "a good opportunity to reflect on some of the things we do." Similarly, the interviews motivated several members of the DPO group to consider providing more informational resources and guidance on data protection to researchers.

4.2 Current Practices

To investigate how personal data is currently handled in USP research (**RQ1**), this subsection highlights findings about current data handling practices. Through interviews with researchers, we identified four main stages in the research process that involve personal data: study design, data collection, data analysis, and publication. In addition to these chronological phases, we identified data storage as a cross-stage area critical to the handling of personal data. In the following sections, we present primarily the findings from researcher interviews and, where appropriate, interweave the perspectives of the DPOs to provide an alignment of viewpoints. An overview of these sectors and their corresponding data handling processes is provided in Figure 2.

4.2.1 Study Planning. The study planning phase entails defining what data will be collected, how it will be processed, and securing approval from an IRB. Researchers begin the study planning phase by making decisions about what types of data will be collected and how the data will be managed throughout the study. In addition to the research observations themselves, personal information is often collected for logistical reasons, such as maintaining contact with participants or sending compensation. The type of data collected may also depend on the institution's chosen compensation method. Some institutions require bank transfers, which necessitates the collection of sensitive personal information such as full name, address, and banking information. Others choose to minimize the collection of personal information by using alternatives such as online vouchers or PayPal, which only require an email address.

A central part of the data management process is the creation of informed consent forms that explain the purpose of the study, the types of data that will be collected, and the associated benefits and risks to participants. All participants reported that the informed consent form included a section detailing the personal information that would be collected. However, researchers indicated that they prefer not to specify an exact timeline for data anonymization in these forms. Instead, they retain all data until the end of the study, leaving the timeframe for deletion vague. This flexibility allows researchers to retain data for possible future analysis or follow-up studies rather than committing to a fixed deletion schedule.

After determining the types of data to be collected, researchers often need to obtain approval from their IRB and, in one case, create a data management plan (DMP).

4.2.2 Data Collection. Once the study has been planned and the necessary data are identified, the next phase involves collecting this data from participants. In this phase, we observed that participants differentiated between research observations and contextual data, such as demographics or contact information. According to several participants, avoiding the collection of personal data was another measure, when applicable. This includes avoiding identifiers such as names or employers in interviews or collecting demographical data that is not relevant to the research. A further action was to refrain from the collection of exact data and to collect binned data instead. An example was P8, who does not ask for the specific age but only for age bins. As another option, the same researcher enables participants to opt out when a survey asks for personal data, although this complicates analysis.

P4 and P8 solved this challenge by not collecting the data from the participants themselves but outsourcing this task to a third-party data collection service. This service collects the necessary data from the participant and only releases anonymous data that is not linked to any participant.

4.2.3 Data Analysis. After the data collection, the data needs to be analyzed. Respondents who had conducted interviews before mentioned that they transcribed the interview records and deleted the recording after the analysis. The recordings are kept for the duration of the interview analysis in order to correct any transcription errors. Also, as a data protection measure, some participants mentioned that they do not analyze personal data when it is not necessary. For instance, P8 explained that IP addresses are collected but deleted immediately after data collection.

4.2.4 Publication. When publishing research results, the participants we interviewed take special care not to publish any identifiable information about their participants. This was ensured by publishing data in either anonymous or aggregated form so that no connection to the individual person could be made. If there was any doubt as to whether participants could be identified through a

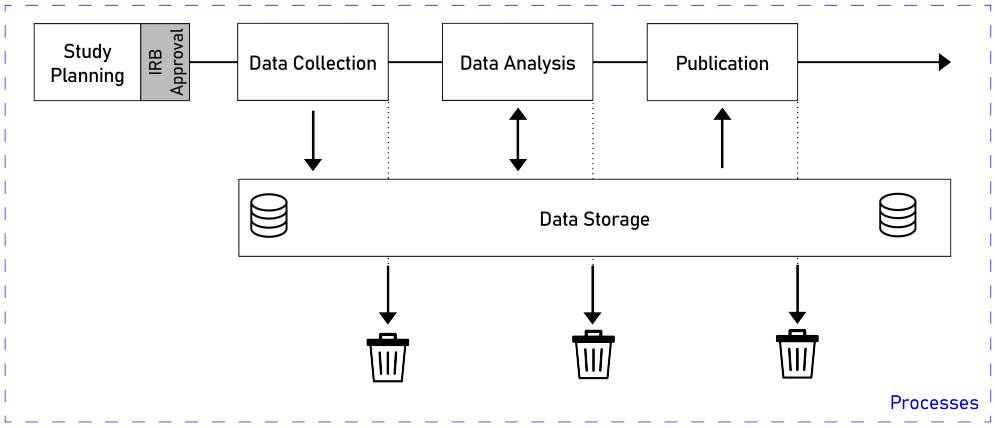


Figure 2: Simplified diagram of identified stages.

combination of data points within a small population, the data were not published, despite the goal of sharing datasets. Researchers also mentioned withholding interview transcripts, because statements could potentially identify participants, especially if the studied population is small.

4.2.5 Storage. Data storage emerged as a recurring consideration throughout the research process, particularly when handling personal data. We discuss two key data protection measures related to storage. The first, **access control**, involves the implementation of technical measures to ensure that only authorized individuals can access identifiable data, thereby preserving confidentiality. The second is **pseudonymization and anonymization**. Pseudonymization involves separating research data from contextual information, such as demographics, while still allowing for re-identification through a linking table. In contrast, anonymization completely removes identifying information, ensuring participants can no longer be linked to their data. At the end of the research process, data is **archived** with only the observations retained, without any identifying information.

Access Control. Access control was identified as an important factor in ensuring the confidentiality of personal data during the research process. However, our findings revealed considerable variability in how this was implemented in different contexts. In some cases, access was tightly restricted; for example, researchers P2 and P3 reported that they were the sole administrators of the research data and restricted access to only themselves. In contrast, other practices were less restrictive. One supervising researcher noted that access was granted to the entire research group, regardless of individual involvement in the specific project.

Participants also highlighted the importance of using privacy-compliant services for storing data. Many reported using their institutions' cloud solutions to ensure compliance with data protection regulations. However, lapses in practice were noted. One participant stated that despite having access to a privacy-compliant institutional cloud service, they had sometimes inadvertently stored personal data in a non-privacy-compliant cloud spreadsheet.

Many researchers further stated that the data would be encrypted to protect it from unauthorized access.

Pseudonymization and Anonymization. Pseudonymization was recommended by the DPOs we interviewed as the preferred approach when anonymization is not yet feasible. Researcher participants mentioned collecting and storing data separately, using a link between the datasets. As an example, one supervising researcher explained using three separate databases for contact information, demographics, and actual survey content. Other participants reported storing data separately without a linking table. This serves the purpose of knowing who took part in the study (e.g., for compensation purposes) without needing to link the data back to individual participants. To minimize risk to participants, DPOs advised that identifying data should be deleted once it is no longer needed. We found that many researchers anonymized their data, especially in interviews and when data has been collected unintentionally.

Archiving. According to the DPOs, after the research project is completed and the findings are published, the data should be archived. Proper archiving ensures accountability and may also be essential for conducting follow-up studies. In most cases, anonymized data is sufficient to meet the requirements for accountability.

These anonymization practices varied considerably across participants. While some researchers reported deleting personal data that was no longer needed after publication, others stated that data they should have deleted may still be stored. One supervising researcher reported that all raw data, including personal data, is archived in encrypted form after publication. Despite the intention to delete all data that could identify an individual participant, researchers were not certain that this was actually done in their prior studies. As an example, P7 remarked: "I honestly think, there could still be some pieces of information in some places, yeah." This deletion sometimes only occurs when researchers happen to think of it, often triggered at random when they come across completed projects

and question whether the data is still needed. P2 described that the data processing outlined in their informed consent information was adhered to during the project's runtime, but once a project was completed, the data management was largely forgotten: "Once it is completed, it's out of sight, out of mind." P7 noted that the fast-paced nature of graduate research contributes to this problem, as students often move quickly to the next project without ensuring that all identifiable data from previous projects has been properly deleted.

Sometimes, researchers do not think of data copies, as P2 mentioned that "they are probably still exactly where they were on my hard drive, on my backup hard drive." This can also apply to external data collection services since one professor noted that data was not deleted from the survey platform used: "We didn't delete any projects. So [the survey platform] still has everything."

4.3 Challenges

When we investigated the data handling processes of the researchers, we identified several challenges researchers have to deal with (**RQ2**).

4.3.1 Study Planning Phase. One of the most fundamental challenges is the lack of a common understanding of what constitutes personal data. While all participants agreed that data types such as email addresses, names, and IP addresses qualify as personal data, there was uncertainty about combinations of seemingly anonymous data points that could still identify individuals, especially within small or specific populations, e.g., age in a sample where certain ages occur seldomly.

This uncertainty was recognized by the DPOs, who acknowledged that determining what qualifies as personal data can be tricky and is context-dependent. D1 explained that "it depends on the individual case", which was echoed by several DPOs. This variability complicates the research planning phase, as researchers must navigate these uncertainties while attempting to comply with data protection regulations.

In many cases, personal data only affects demographics, which is an accompanying data point and can often be stored separately. However, in some cases, personal data is part of the observation data itself, such as when a regression analysis is performed with demographic co-variates. This makes techniques such as storing demographic data separately complex and leads to personal data being stored long-term as needed for accountability reasons.

Compliance with regulations such as GDPR was considered an administrative burden by multiple DPOs, particularly when it comes to multi-institution collaborations. For instance, determining responsibility for data control requires legal agreements between universities, which can slow down research progress and introduce additional complexity to the data management process. Some researchers in the U.S. mentioned that they explicitly do not do research with participants from the EU, or do this only when they have a European partner that handles the regulation parts.

4.3.2 Study Execution Phase. In some cases, researchers unintentionally collect personal data due to technical limitations. For example, one supervising researcher mentioned that survey platforms like Qualtrics collect IP addresses by default, even when researchers do not intend to collect this information. Researchers must actively disable these settings, a process that requires awareness of the platform's default behaviors. Similarly, in qualitative research, participants may mention identifiable information such as names or employers during an interview, complicating efforts to preserve anonymity.

A recurring challenge is the uncertainty about when to anonymize or delete personal data. While DPOs emphasized that personal data should be anonymized or deleted as soon as it is no longer needed, researchers had different views on the appropriate timing (i.e., after data collection, after analysis, or after publication). Many researchers have, therefore, been vague in their informed consent forms, indicating that data would be kept as long as needed without specifying clear criteria for deletion. This was done to avoid having to delete data that is still needed.

In anticipation of potential requests from reviewers for additional analysis, researchers reported keeping data even after the analysis was completed and collecting more data than intentionally intended. In case reviewers require more contextual data than collected, researchers may need to recontact study participants. This is impossible if the data has already been anonymized and difficult even if it has not because it increases the burden on participants and may result in low response rates. For example, one focus group participant mentioned that their response rate was only about 50% when they tried to collect follow-up data.

4.3.3 Study Completion Phase. The distribution of data across multiple locations during the research process poses challenges when it comes to the deletion of data copies. Personal data is often stored on collection platforms (e.g., survey platforms), local hard drives, institutional cloud storage, backup systems, potentially communication platforms (e.g., email, Slack, etc.), or being physically printed. Participants reported that these distributed data copies are difficult to track, particularly when it comes to backups or external data collection services.

Finally, with pressure to publish and time constraints causing researchers to quickly move on to the next project, data may not be deleted at all. This problem is compounded by the long delay – sometimes months or years – between the completion of a study and the acceptance of the paper for publication. By that time, the primary researchers are often already working on their next project or may have moved on to another institution.

4.4 Formal Processes in Data Handling

We investigated formal processes that research institutions implemented to ensure proper data handling, addressing **RQ3**.

4.4.1 Data Management Plan. In an effort to address the challenges associated with data handling, some institutions have developed formal processes to guide researchers. These processes varied greatly across institutions.

Several institutions encourage researchers to develop a DMP during the study planning phase. This document specifies what data will be collected, how it will be managed, and what safeguards will be in place to ensure that personal data is handled securely. In some cases, only a simplified form of a DMP is necessary for obtaining

IRB approval, which incorporates data protection considerations early in the study design.

However, the treatment of DMPs varies significantly across institutions. We found one institution requiring the DMP to be formally submitted, while others only appeal to researchers to create it for their own use. This variation can lead to differences in accountability and how closely data handling practices are monitored. Some institutions had no requirements for the study planning phase concerning data management.

From the perspective of DPOs, decisions on data collection and processing are closely linked to risk management. DPOs emphasized that the sensitivity of the data collected determines the level of care and security measures required. While anonymous data is generally considered risk-free and can be handled flexibly, highly sensitive data – such as medical or financial information – requires more stringent protocols to mitigate the risk of data breaches. DPOs emphasized the need for researchers to ensure that sensitive data is handled securely, including the implementation of strict access controls, encryption, and clear policies for anonymization or deletion when the data is no longer needed. DPOs explicitly recommended to carry out a data protection impact assessment as a tool to formalize these considerations.

4.4.2 Guidance and Training. While some institutions provide mandatory data handling training during onboarding or before researchers are allowed to work with personal data, the majority of participants indicated that they had not received such training. One supervising researcher expressed support for the idea, stating, "As I am having this conversation with you now, I think to myself that [data management training] might not be so bad."

Some institutions provide internal guidelines or checklists that researchers can use to ensure they follow proper data handling procedures. Several DPOs noted that institutional privacy policies exist but are not specifically tailored to research needs. To complicate things further, one participant (D4B) mentioned that it is difficult to find the guidelines in the internal information portal. The guidelines and checklists mentioned across all interviews differ substantially in several dimensions. In terms of abstractness, they range from one overarching document on data protection guidelines for the whole university to specific templates for the creation of informed consent forms or DMPs. They vary in scope, including some very specific guidelines about data protection when reusing data or guidelines on information security. In the level of privacy expertise required, they span from accessible tools like an internal Wiki with explanations to advanced resources, like a checklist for data protection in research projects that demands substantial legal understanding and thorough motivation of researchers. All DPOs mentioned offering advice on the handling of personal data. Typically, however, this service is only used for studies involving highly sensitive data or longitudinal research where data handling changes over time.

In practice, data handling is largely based on trust: Institutions trust their researchers to comply with data protection rules and laws. Meanwhile, supervising researchers trust primary and supporting researchers to handle the personal data of participants responsibly. In many cases, this trust-based model does not include monitoring processes verifying that personal data is appropriately anonymized or deleted.

Only a few institutions have established formal procedures that vary in depth. One supervising researcher mentioned that the IRB protocol includes a closeout form which they have to complete at the end of the study to verify compliance. One other supervising researcher and P8 both stated that they have checklists, but only for their own use since no one verifies compliance with the checklist.

Many institutes only conduct in-depth investigations in cases where incidents are suspected. Some implement random monitoring of data handling practices. For certain institutions, DPOs reported conducting annual reviews of relevant data protection documents for large longitudinal studies spanning over several years.

We also want to highlight one university having an exceptionally well-defined process for data handling and the corresponding responsibilities within a three-level hierarchy in addition to the researchers themselves. The lowest level consists of data stewards and their supervisors. The data stewards are employed at each faculty and assist the researchers in carrying out data protection impact assessments. Formally, the data stewards' supervisors are responsible in case of errors; they review every data protection impact assessment. On the second level, a privacy team assigned to the information services department supports all data stewards and manages policies and procedures. This team also keeps track of all data protection impact assessments and reviews them annually. The third level consists of the DPO of the research institution: The DPO is independent and can monitor all processes, both in an ad-hoc and a planned manner. The monitoring of specific data protection measures is done by the most suitable party: For example, data stewards monitor data retention practices, while the security team of the research institution audits security-related measures.

Two DPOs mentioned that they plan to implement processes to further assist researchers in handling personal data. These plans include providing more research-specific material or a data coordinator to oversee the handling of these data.

One of the most significant points of ambiguity relates to who is ultimately responsible for data handling. We found a difference in perceptions of responsibility between primary and supervising researchers: While most of the primary researchers saw themselves responsible for the proper handling of personal data, the supervising researchers had a mixed view. Two supervising researchers acknowledged that based on the IRB form, they are effectively responsible for the process while stating that this is not always the case in practice. One supervising researcher attributes all responsibility to the primary researcher, while another one takes the responsibility upon themself. This supervising researcher made sure that the data handling of their students complied with internal regulations: They explained that they regularly check in with their students about data deletion along the research process, especially during the analysis stage. Even though these check-ins happen spontaneously, the supervisor ensures the confidentiality of participants' data at the latest during an IRB check-out phase implemented in their research institution's processes. In almost all cases, the responsibilities were not clarified at the beginning of the study.

The DPOs had differing views on the person responsible for data protection compliance. While they all agreed that the DPO is not

the responsible person, there were arguments in favor of both the primary and the supervising researcher having this role. Multiple DPOs mentioned that the primary researcher is in charge because they are the ones who actually handle the data. In contrast, other DPOs saw the supervising researcher being responsible because they direct the research and "just send [primary researchers] out to get the data," as D4A stated. This is especially problematic since most scientific staff who act as primary researchers only have temporary contracts and are no longer employees of the university at the end of most retention periods. Nevertheless, all DPOs agreed that it is necessary to determine in advance who is responsible for data handling. According to D2, diffusion of responsibility may also lead to data not being deleted because everyone involved assumes that any other party will do it.

In the interviews, the researchers did not link their own data handling processes with the formal processes of their institution. However, we observed that the "out of mind, out of sight" problem did not occur for participants whose institutions' processes required formal verification of adequate archiving at the end of the process.

5 Discussion

Our study reveals a high level of awareness among USP researchers for the need to handle personal data in a privacy-conscious manner. The researchers clearly stated that they believe the topic to be an important one and presented the many efforts they invest to protect participants' personal data. They genuinely intended to comply with legislation, respect ethical values, and honor the trust of participants.

Nevertheless, researchers' success in complying with legal and ethical mandates varied: While some interviewees reported adherence to strictly regulated processes, others acknowledged that they had deviated from best practices to some extent in the past. Due to access control measures, these issues are unlikely to have caused any harm, although they potentially could.

In the publication phase, errors can be severe: In this phase, noncompliance can result in the publishing of participants' personal data. Consequently, researchers reported that they were very careful about publishing data points because they understood the potential impact of errors. In addition, errors would be public, providing an extra incentive for compliance at this stage, especially when publishing data as encouraged by best practices in Open Science [15, 43]. Ethically, researchers may have to balance increased research integrity through openness and data protection through restriction of access. This can be challenging, for instance, if the complexity of proper anonymization processes discourages researchers from sharing [3, 32, 67] or serves as a convenient justification to avoid complying with Open Science requirements [28]. Nevertheless, in line with the opinion of the interviewed DPOs, small violations at other stages before publication could also accumulate and increase the likelihood of a data breach and, therefore, need to be corrected.

Our research highlighted significant challenges that impede the straightforward implementation of good data-handling practices. In the following, we discuss three key areas where improvements could greatly ease the process for researchers: legal frameworks, research ecosystem, and organizational factors.

5.1 Legal Ecosystem

Although the GDPR and similar laws have provided a foundational framework for data protection, the complexity and variability of legal obligations remain a challenge for researchers, supporting results of prior work [69]. Our findings indicate that legal uncertainties at a practical level persist, making it difficult for researchers to fully comply with data protection laws.

HCI researchers and legal experts have highlighted the need for more uniform and usable laws for researchers [5, 14, 30]. We support this view, as we observed that even DPOs sometimes struggled to apply the law in specific research contexts. This shows that navigating legal issues is understandably daunting for individual researchers, especially those without legal training.

Designing more usable legal frameworks is a challenging endeavor, as it requires balancing the interests of diverse stakeholders. Such considerations extend well beyond the scope of this article.

5.2 Research Ecosystem

Our findings indicate that the demands of the academic environment – particularly time constraints and the pressure to publish – also contribute to researchers' difficulties in fulfilling their good intentions considering data protection. In the rush to move on to the next project, it seems easy to overlook completing data anonymization or deletion tasks from previous studies. This "out of sight, out of mind" issue can lead to unnecessary data storage, even when researchers are committed to following best practices and are willing to delete data.

However, as stated by some DPOs the importance of research data to academic careers cannot be overstated. The interviewed researchers often collected more data than strictly necessary for the immediate project, anticipating that it may be useful for future work. This tendency was sometimes reinforced by peer reviewers, who may request additional demographic or classification data to contextualize research findings. While additional data may improve the understanding of results, it conflicts with the principle of data minimization[49].

Additionally, the academic system encourages frequent movement of researchers between institutions, which further complicates data management [2, 42]. When staff who were originally responsible for handling data leave, it becomes unclear who should take responsibility for ensuring that data are deleted or anonymized in accordance with policies written in the informed consent. This challenge underscores the need for long-term, institutional oversight of data management, as well as clear processes that transcend individual projects or researchers.

Finally, USP researchers specifically might overestimate their data protection abilities due to their familiarity with privacy issues. Moreover, unlike end users who can learn from security incidents discussed by peers and widely published in the news [48, 50, 51], the lack of similar documented breaches in academia might lead these researchers to underestimate the risks in their work.

5.3 Institution's Ecosystem

The challenges posed by regulatory and academic ecosystems highlight the need to improve organizational processes to support ethical data management. While most IRBs require some statements

about data management during the study planning phase, only few institutions had centralized processes to support researchers throughout the project. However, we found that having processes in place to assist researchers was very valuable. Without these, our study suggests there is a high risk of personal data being retained longer than necessary.

A significant institutional issue is the unclear allocation of responsibility for data management within research teams. As primary researchers – often graduate students – move on to the next project or job in their careers, there are no policies in place to ensure that the research project is completed in terms of proper archiving. This includes both anonymization before archiving and deletion of archives after the retention period.

Supervising researchers play a critical role in overseeing their students' data protection practices. They should be actively involved in the creation and implementation of DMPs and ensure that personal information is deleted or anonymized when it is no longer needed.

At the institutional level, DPOs and other responsible authorities e.g., data stewards, can help establish clear, formal processes for tracking data management throughout a project's lifecycle. This could include periodic audits or spot checks to ensure compliance with data retention and deletion policies, especially for high-risk studies. While we recognize that some institutions have already implemented such processes, many others have not, leaving gaps in data protection. Guidance from legal experts and DPOs will be critical in helping institutions develop these processes in a way that is both comprehensive and practical.

6 Recommendations

Our investigation into data protection practices within the USP field has revealed areas that could benefit from enhanced data protection measures. During our research for this paper, we took deliberate steps to safeguard participant data by collecting only essential information, anonymizing or deleting raw data when no longer needed, and crafting a privacy policy that clearly outlined our data handling procedures. Still, despite these efforts, a DPO criticized our policy for lacking precision and clarity, underscoring how challenging it can be to meet high privacy standards. Standardized templates or best-practice frameworks would have been helpful for us in navigating these complexities.

Although the potential risk to participants due to breaches of confidentiality in HCI studies generally appears low, we encourage the CHI community to re-evaluate their procedures, especially for studies involving vulnerable populations [20, 61, 65] or with sensitive data [4, 36, 47, 61]. To support the community in addressing these challenges, we offer our recommendations for researchers and their institutions. The first two apply to all stakeholders, while the next five address specific roles.

6.1 Process

Based on our findings, we highly recommend a structured data management process throughout the research lifecycle. The DMP is an important starting point during the planning phase, but continuous oversight is vital as the research progresses. Figure 3 summarizes our recommended process for data handling.

After each research phase, a brief clean-up phase should be appended to assess whether personal data is still necessary for the research purpose. When possible, pseudonymization or anonymization measures should be applied. We appreciate that the research process can be dynamic and objectives may shift; it can be a completely valid result of a clean-up phase that all data is retained, but it should be done consciously.

A final, thorough clean-up should take place after publication, carefully considering which data needs to be archived. Any data no longer needed should be fully removed from all storage devices, including external services, backups, and any devices used by supporting researchers who may have had access to personal data. If personal data needs to be stored long-term to satisfy scientific best practices, we recommend the use of encryption and/or offline storage [60].

6.2 Data Management Plan

DMPs are a promising tool to oversee the handling of research data, and numerous resources exist to guide researchers in their creation and execution, especially provided by funding agencies [12, 39]. A notable example is the DMP Online[2] [12], an online tool for creating and maintaining DMPs of various pre-defined schemes. It is suitable for smaller projects due to its straightforward design which allows for collaboration with colleagues. The default DMP template provided by the Digital Curation Centre [11] prompts researchers to address ethical aspects of data management in addition to other core elements. We would like to explicitly emphasize the following aspects, derived from the underlying questions that the researchers we interviewed could have benefited from addressing prior to conducting their studies:

- What data is going to be collected?
  - For what purpose?
  - What risks are associated with the data?
  - What data points could identify an individual?
    - When can identifying data be removed?
- What data should be archived after the study?
- What data protection measures are used?
- Who are the primary, supporting and supervising researchers?
- Who has access to what data?
  - overseeing the DMP
  - the technical protection measures
  - the in-project clean-up phases
  - the post-project clean-up
  - the archiving of the data

Future research could evaluate these proposed elements in terms of exhaustiveness and usability when planning the management and protection of participants' data.

Researchers may benefit from institutional assistance to write and maintain a DMP: Experienced data stewards at research institutions can play a key role by combining legal expertise, Open Science methods, and case-specific knowledge to guide researchers in balancing their ethical responsibilities [54, 68]. Digital tools could help with maintaining a DMP over time, e.g., by unifying data storage or sending automatic data deletion reminders.

[2] https://dmponline.dcc.ac.uk/

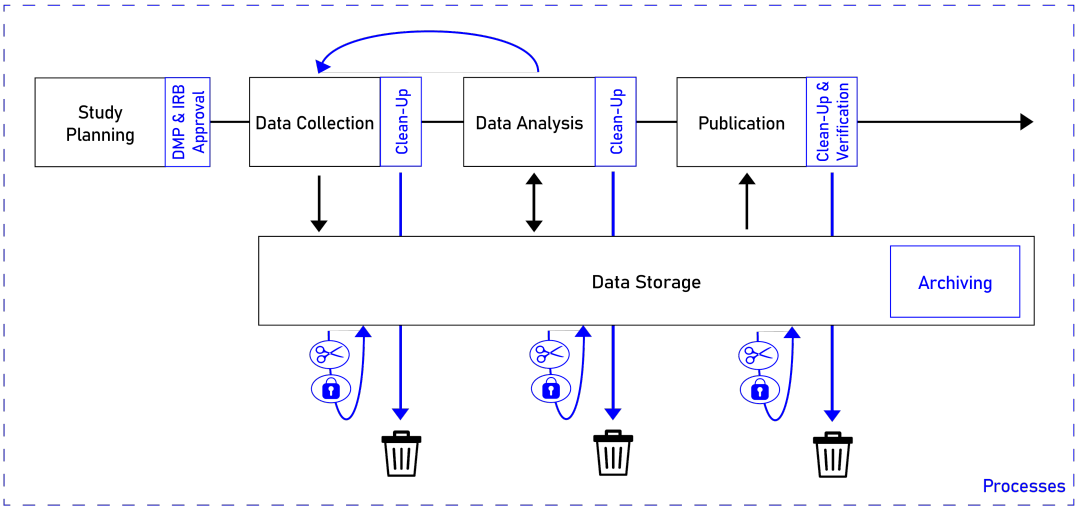


Figure 3: A diagram of an example research process with data protection measures. The scissors, lock, and bin symbols represent the separation, encryption, and deletion of data, respectively.

6.3 Primary Researchers

We recommend a study design process that includes a detailed DMP as described above. Having such a plan will make it easier to implement sensible protection measures and decide when to remove data. In case of uncertainties when creating the DMP, primary researchers are recommended to reach out to supervising researchers and organizational support personnel such as DPOs early.

Researchers should also keep track of data copies made during the project to ensure that all personal data can be deleted or anonymized when it is no longer needed.

6.4 Supervising Researchers

For supervising researchers, such as professors, we recommend increased oversight of their students' data protection practices. Supervisors should ensure that a DMP is created early in the research process and that it is adhered to throughout the project. Ideally, supervising researchers should have one or more DMP templates tailored to their research which could be developed in exchange with the DPOs. This can mean that studies on the usability of encryption interfaces will have different DMP requirements than studies about the data protection behaviors of abuse victims. When publishing data, researchers may follow practical recommendations for ethical data sharing as proposed by Meyer [38]. After conclusion of a project, supervisors should ensure that the practices described in the DMP were implemented and all data was anonymized – including copies and backups.

6.5 Organizations

We recommend that organizations develop clear policies and procedures to support all researchers in the responsible management of personal data. This could include providing standardized flowcharts, DMP templates, or checklists for researchers to follow, as well as institutional support in the form of consultations or training sessions with DPOs or data stewards. While regular audits of ongoing research can be helpful [17], this should be balanced to avoid overburdening researchers, particularly in low-risk studies, which are more common in USP than, for instance, medical research. Institutions should also facilitate the sharing of best practices among research groups to ensure that common data management challenges are systematically addressed.

6.6 Policymakers

While the introduction of the GDPR and similar legislation was an important step, more needs to be done to ensure that the legal provisions are clear and applicable in specific research contexts. Legal experts have called for more accessible guidance [5, 30], and we support this view. Lawmakers should strive to simplify the legal landscape for researchers by harmonizing regulations across jurisdictions where possible while maintaining the flexibility needed for case-specific interpretations.

6.7 Reviewers

Finally, we encourage peer reviewers to consider whether certain demographic or personal data are necessary to contextualize research findings. Researchers often collect or retain additional personal data in response to reviewer requests, which may conflict with the principle of data minimization. Since there can be good reasons for this, we do not want to discourage this entirely but do recommend the same level of data protection consideration as during the IRB process.

7 Conclusion

The protection of research data is not only necessary for compliance with regulations but also to fulfill ethical responsibilities. The trust in researchers to protect the personal data of participants is one of the cornerstones of human subjects research. Motivated by the issues we identified in our own data management processes, we conducted the first qualitative examination of USP researchers'

personal data handling processes to get a better understanding of what issues they face and how they are addressed. We conducted interviews and a focus group with 18 USP researchers and nine data protection officers. We found that researchers were motivated to protect their participants' data, but faced many practical issues which could hamper their efforts and lead to suboptimal results. Based on our analysis, we make recommendations for the different actors involved to aid them with the task of research data protection.

7.1 Future Work

We focused our investigation on the USP domain, which could be a best-case environment. Therefore, it would be interesting to look at other areas of HCI and also other disciplines working with personal data that do not have a security and privacy background to investigate how they compare to USP. Since our study was qualitative, a quantitative follow-up study would be valuable to assess the prevalence of common issues and best practices. While our work primarily focused on the confidentiality of participant data, privacy and data protection extend beyond these aspects to include participants' rights, lawful data processing, and regulatory compliance. Future research could explore these dimensions in greater depth. Another avenue for future work is to evaluate the impact and usablity of interventions such as a DMP or standardized templates and to develop tools supporting the data management process.

Acknowledgments

We would like to thank our study participants, researchers and DPOs, for their valuable Insights. Also, we thank the Werner Siemens-Stiftung (WSS) for supporting this project, and our anonymous reviewers for their help and feedback.

References

[1] ACM 2021. ACM Publications Policy on Research Involving Human Participants and Subjects. https://www.acm.org/publications/policies/research-involving-human-participants-and-subjects

[2] Amrei Bahr, Christine Blume, Kristin Eichhorn, and Sebastian Kubon. 2021. With #IchBinHanna, German academia protests against a law that forces researchers out. *Nature Human Behaviour* 5, 9 (Aug. 2021), 1114–1115. https://doi.org/10.1038/s41562-021-01178-6

[3] Siviwe Bangani and Mathew Moyo. 2019. Data Sharing Practices Among Researchers at South African Universities. *Data Science Journal* 18, 1 (July 2019), 28. https://doi.org/10.5334/dsj-2019-028

[4] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M Vardoulakis. 2020. A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[5] Ute Bernhardt, Ingo Ruhmann, and Thilo Weichert. 2018. *Die Forschungsklauseln im neuen Datenschutzrecht.* Technical Report. Netzwerk Datenschutz. 27 pages.

[6] Legislative Services Branch. 2000. Consolidated Federal Laws of Canada, Personal Information Protection and Electronic Documents Act. https://laws-lois.justice.gc.ca/eng/acts/p-8.6/index.html

[7] Barry Brown, Alexandra Weilenmann, Donald McMillan, and Airi Lampinen. 2016. Five Provocations for Ethical HCI Research. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 852–863. https://doi.org/10.1145/2858036.2858313

[8] Kathy Charmaz. 2014. *Constructing grounded theory* (2. ed ed.). SAGE, Los Angeles, Calif.

[9] Cal. Civil Code. 2018. 1.81.5. California Consumer Privacy Act of 2018 [1798.100 - 1798.199.100]. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5

[10] Utah Code. 2022. Chapter 61 Utah Consumer Privacy Act. https://le.utah.gov/xcode/Title13/Chapter61/13-61.html

[11] DMP DCC [n. d.]. Checklist for a Data Management Plan. https://www.dcc.ac.uk/DMPs/Checklist

[12] Martin Donnelly. 2012. Data management plans and planning. In *Managing Research Data* (1 ed.), Graham Pryor (Ed.). Facet, 83–104. https://doi.org/10.29085/9781856048910.006

[13] Edward S. Dove. 2018. The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era. *Journal of Law, Medicine & Ethics* 46, 4 (Dec. 2018), 1013–1030. https://doi.org/10.1177/1073110518822003

[14] Rossana Ducato. 2020. Data protection, scientific research, and the role of information. *Computer Law & Security Review* 37 (July 2020), 105412. https://doi.org/10.1016/j.clsr.2020.105412

[15] Florian Echtler and Maximilian Häußler. 2018. Open Source, Open Science, and the Replication Crisis in HCI. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–8. https://doi.org/10.1145/3170427.3188395

[16] EU 2013. *Responsible Research and Innovation (RRI), Science and Technology*. Technical Report 1096 / SP401. European Commission. 39 pages. https://europa.eu/eurobarometer/surveys/detail/1096

[17] José Fernandes, Carolina Machado, and Luís Amaral. 2022. Identifying critical success factors for the General Data Protection Regulation implementation in higher education institutions. *Digital Policy, Regulation and Governance* 24, 4 (Sept. 2022), 355–379. https://doi.org/10.1108/DPRG-03-2021-0041

[18] Hildegard Ferraiolo, Ramaswamy Chandramouli, Nabil Ghadiali, Jason Mohler, and Scott Shorter. 2015. *Guidelines for the Authorization of Personal Identity Verification Card Issuers (PCI) and Derived PIV Credential Issuers (DPCI).* Technical Report NIST SP 800-79-2. National Institute of Standards and Technology. NIST SP 800–79–2 pages. https://doi.org/10.6028/NIST.SP.800-79-2

[19] Nollaig Frost, Sevasti Melissa Nolas, Belinda Brooks-Gordon, Cigdem Esin, Amanda Holt, Leila Mehdizadeh, and Pnina Shinebourne. 2010. Pluralism in qualitative research: the impact of different researchers and qualitative approaches on the analysis of qualitative data. *Qualitative Research* 10, 4 (Aug. 2010), 441–460. https://doi.org/10.1177/1468794110366802

[20] Aakash Gautam, Deborah Tatar, and Steve Harrison. 2020. Crafting, Communality, and Computing: Building on Existing Strengths to Support a Vulnerable Population. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[21] GDPR.EU 2018. Does the GDPR apply to companies outside of the EU? https://gdpr.eu/companies-outside-of-europe/

[22] Travis Greene, Galit Shmueli, Soumya Ray, and Jan Fell. 2019. Adjusting to the GDPR: The Impact on Data Scientists and Behavioral Researchers. *Big Data* 7, 3 (Sept. 2019), 140–162. https://doi.org/10.1089/big.2018.0176 Publisher: Mary Ann Liebert, Inc., publishers.

[23] Marilys Guillemin, Emma Barnard, Anton Allen, Paul Stewart, Hannah Walker, Doreen Rosenthal, and Lynn Gillam. 2018. Do Research Participants Trust Researchers or Their Institution? *Journal of Empirical Research on Human Research Ethics* 13, 3 (July 2018), 285–294. https://doi.org/10.1177/1556264618763253

[24] Marilys Guillemin, Lynn Gillam, Emma Barnard, Paul Stewart, Hannah Walker, and Doreen Rosenthal. 2016. "Doing Trust": How Researchers Conceptualize and Enact Trust in Their Research Practice. *Journal of Empirical Research on Human Research Ethics* 11, 4 (Oct. 2016), 370–381. https://doi.org/10.1177/1556264616668975

[25] Dara Hallinan, Franziska Boehm, Annika Külpmann, and Malte Elson. 2023. (Un)informed Consent in Psychological Research: An Empirical Study on Consent in Psychological Research and the GDPR. *Journal of Open Access to Law* 11 (2023), 1–28.

[26] Carol A. Heimer and JuLeigh Petty. 2010. Bureaucratic Ethics: IRBs and the Legal Regulation of Human Subjects Research. *Annual Review of Law and Social Science* 6, 1 (Dec. 2010), 601–626. https://doi.org/10.1146/annurev.lawsocsci.093008.131454

[27] HIPAA 1996. Health Insurance Portability and Accountability Act. , 169 pages. https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf

[28] Ian Hussey. 2023. Data is not available upon request. https://doi.org/10.31234/osf.io/jbu9r

[29] Claudia Irti. 2022. Personal Data, Non-Personal Data, Anonymised Data, Pseudonymised Data, De-Identified Data. In *Privacy and Data Protection in Software Services*, Roberto Senigaglia, Claudia Irti, and Alessandro Bernes (Eds.). Springer, Singapore, 49–57. https://doi.org/10.1007/978-981-16-3049-1_5

[30] Timo Jakobi and Maximilian Von Grafenstein. 2023. What HCI Can Do for (Data Protection) Law—Beyond Design. In *Human Factors in Privacy Research*, Nina Gerber, Alina Stöver, and Karola Marky (Eds.). Springer International Publishing, Cham, 115–136. https://doi.org/10.1007/978-3-031-28643-8_6

[31] T.J. Kasperbauer, Colin Halverson, Abby Garcia, and Peter H. Schwartz. 2022. Biobank Participants' Attitudes Toward Data Sharing and Privacy: The Role of Trust in Reducing Perceived Risks. *Journal of Empirical Research on Human Research Ethics* 17, 1-2 (Feb. 2022), 167–176. https://doi.org/10.1177/15562646211055282

[32] Mary Anne Kennan and Lina Markauskaite. 2015. Research Data Management Practices: A Snapshot in Time. *International Journal of Digital Curation* 10, 2 (July 2015), 69–95. https://doi.org/10.2218/ijdc.v10i2.329

[33] Mohammed I.U. Khan, Lawrence Mbuagbaw, Matthew Holek, Faris Bdair, Zoha H. Durrani, Katie Mellor, Saskia Eddy, Sandra M. Eldridge, Claire L. Chan, Michael J. Campbell, Christine M. Bond, Sally Hopewell, Gillian A. Lancaster, and Lehana Thabane. 2021. Transparency of Informed Consent in Pilot and Feasibility Studies is Inadequate: A Single-Center Quality Assurance Study. *Pilot and Feasibility Studies* 7, 1 (Dec. 2021), 96. https://doi.org/10.1186/s40814-021-00828-w

[34] Nicole M Krause, Dominique Brossard, Dietram A Scheufele, Michael A Xenos, and Keith Franke. 2019. The polls—trends: Americans' trust in science and scientists. *Public Opinion Quarterly* 83 (Sept. 2019), nfz041. https://doi.org/10.1093/poq/nfz041

[35] Nevada Legislature. 2019. Senate Bill No. 220 - AN ACT relating to Internet privacy; prohibiting an operator of an Internet website or online service which collects certain information from consumers in this State from making any sale of certain information about a consumer if so directed by the consumer; and providing other matters properly relating thereto. https://www.leg.state.nv.us/App/NELIS/REL/80th2019/Bill/6365/Text

[36] Yuan Liang, Hsuan Wei Fan, Zhujun Fang, Leiying Miao, Wen Li, Xuan Zhang, Weibin Sun, Kun Wang, Lei He, and Xiang'Anthony' Chen. 2020. OralCam: Enabling Self-Examination and Awareness of Oral Health Using a Smartphone Camera. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–13.

[37] Konstantinos Mersinas, Maria Bada, and Steven Furnell. 2025. Cybersecurity behavior change: A conceptualization of ethical principles for behavioral interventions. *Computers & Security* 148 (Jan. 2025), 104025. https://doi.org/10.1016/j.cose.2024.104025

[38] Michelle N. Meyer. 2018. Practical Tips for Ethical Data Sharing. *Advances in Methods and Practices in Psychological Science* 1, 1 (March 2018), 131–144. https://doi.org/10.1177/2515245917747656

[39] William K. Michener. 2015. Ten Simple Rules for Creating a Good Data Management Plan. *PLOS Computational Biology* 11, 10 (Oct. 2015), e1004525. https://doi.org/10.1371/journal.pcbi.1004525

[40] Arvind Narayanan and Vitaly Shmatikov. 2008. Robust De-Anonymization of Large Sparse Datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*. IEEE, Oakland, CA, USA, 111–125. https://doi.org/10.1109/SP.2008.33 ISSN: 1081-6011.

[41] Michael Nii Laryeafio and Omoruyi Courage Ogbewe. 2023. Ethical consideration dilemma: systematic review of ethics in qualitative data collection through interviews. *Journal of Ethics in Entrepreneurship and Technology* 3, 2 (Dec. 2023), 94–110. https://doi.org/10.1108/JEET-09-2022-0014

[42] Linda Nordling. 2023. Falling behind: Postdocs in their thirties tire of putting life on hold. *Nature* 622, 7984 (Oct. 2023), 881–883. https://doi.org/10.1038/d41586-023-03296-9

[43] B. A. Nosek, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, and T. Yarkoni. 2015. Promoting an open research culture. *Science* 348, 6242 (June 2015), 1422–1425. https://doi.org/10.1126/science.aab2374

[44] State of Connecticut. 2022. Senate Bill 6: An Act Concerning Personal Data Privacy and Online Monitoring (The Connecticut Data Privacy Act). https://www.cga.ct.gov/2022/ACT/PA/PDF/2022PA-00015-R00SB-00006-PA.PDF

[45] Code of Virginia. 2021. Chapter 53. Consumer Data Protection Act. https://law.lis.virginia.gov/vacodefull/title59.1/chapter53/

[46] Anna-Marie Ortloff, Matthias Fassl, Alexander Ponticello, Florin Martius, Anne Mertens, Katharina Krombholz, and Matthew Smith. 2023. Different Researchers, Different Results? Analyzing the Influence of Researcher Experience and Data Type During Qualitative Analysis of an Interview and Survey Study on Security Advice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–21. https://doi.org/10.1145/3544548.3580766

[47] Sachin R Pendse, Amit Sharma, Aditya Vashistha, Munmun De Choudhury, and Neha Kumar. 2021. "Can I Not Be Suicidal on a Sunday?": Understanding Technology-Mediated Pathways to Mental Health Support. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16.

[48] Katharina Pfeffer, Alexandra Mai, Edgar Weippl, Emilee Rader, and Katharina Krombholz. 2022. Replication: Stories as Informal Lessons about Security. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, 1–18. https://www.usenix.org/conference/soups2022/presentation/pfeffer

[49] Andreas Pfitzmann and Marit Hansen. 2010. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf

[50] Emilee Rader and Rick Wash. 2015. Identifying patterns in informal sources of security information. *Journal of Cybersecurity* 1, 1 (12 2015), 121–144. https://doi.org/10.1093/cybsec/tyv008 arXiv:https://academic.oup.com/cybersecurity/article-pdf/1/1/121/7002665/tyv008.pdf

[51] Emilee Rader, Rick Wash, and Brandon Brooks. 2012. Stories as informal lessons about security. In *Proceedings of the Eighth Symposium on Usable Privacy and Security* (Washington, D.C.) *(SOUPS '12)*. Association for Computing Machinery, New York, NY, USA, Article 6, 17 pages. https://doi.org/10.1145/2335356.2335364

[52] Joel R. Reidenberg, Travis Breaux, Lorrie Faith Cranor, Bran Fench, Amanda Grannis, James T. Graves, Fei Liu, Aleecia McDonald, Thomas B. Norton, Rohan Ramanath, N. Cameron Russel, Norman Sadeh, and Florian Schaub. 2015. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Journal of International Law* 1, 30 (2015), 39–88.

[53] Arianna Rossi and Gabriele Lenzini. 2020. Transparency by design in data-informed research: A collection of information design patterns. *Computer Law & Security Review* 37 (July 2020), 105402. https://doi.org/10.1016/j.clsr.2020.105402

[54] Antti Mikael Rousi, Reid Isaac Boehm, and Yan Wang. 2024. Data stewardship: case studies from North American, Dutch and Finnish universities. *Journal of Documentation* 80, 7 (Sept. 2024), 306–324. https://doi.org/10.1108/JD-12-2023-0264

[55] Rebekah Rousi, Joni-Roy Piispanen, and Jani Boutellier. 2024. I trust you Dr. Researcher, but not the company that handles my data - trust in the data economy. In *57th Hawaii International Conference on System Sciences (HICCS), Waikiki Beach Resort, January 3-6, 2024*. University of Hawaii at Manoa, Honolulu, 4632–4641. OCLC: 1427461703.

[56] Kavous Salehzadeh Niksirat, Lahari Goswami, Pooja S. B. Rao, James Tyler, Alessandro Silacci, Sadiq Aliyu, Annika Aebli, Chat Wacharamanotham, and Mauro Cherubini. 2023. Changes in Research Ethics, Openness, and Transparency in Empirical Studies between CHI 2017 and CHI 2022. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–23. https://doi.org/10.1145/3544548.3580848

[57] Shruti Sannon and Andrea Forte. 2022. Privacy Research with Marginalized Groups: What We Know, What's Needed, and What's Next. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (Nov. 2022), 455:1–455:33. https://doi.org/10.1145/3555556

[58] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2018. A Design Space for Effective Privacy Notices*. In *The Cambridge Handbook of Consumer Privacy* (1 ed.), Evan Selinger, Jules Polonetsky, and Omer Tene (Eds.). Cambridge University Press, Canada, 365–393. https://doi.org/10.1017/9781316831960.021

[59] James Scheibner, Marcello Ienca, Sotiria Kechagia, Juan Ramon Troncoso-Pastoriza, Jean Louis Raisaro, Jean-Pierre Hubaux, Jacques Fellay, and Effy Vayena. 2020. Data protection and ethics requirements for multisite research with health data: a comparative examination of legislative governance frameworks and the role of data protection technologies. *Journal of Law and the Biosciences* 7, 1 (July 2020), lsaa010. https://doi.org/10.1093/jlb/lsaa010

[60] Christopher Smith, Maliha Tabassum, Soumya Chowdary Daruru, Gaurav Kulhare, Arvin Wang, Ethan L. Miller, and Erez Zadok. 2024. Secure Archival is Hard... Really Hard. In *Proceedings of the 16th ACM Workshop on Hot Topics in Storage and File Systems*. ACM, Santa Clara CA USA, 38–46. https://doi.org/10.1145/3655038.3666093

[61] Elizabeth Stowell, Mercedes C Lyson, Herman Saksono, Reneé C Wurth, Holly Jimison, Misha Pavel, and Andrea G Parker. 2018. Designing and Evaluating mHealth Interventions for Vulnerable Populations: A Systematic Review. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–17.

[62] Nguyen Thanh Tam, Nguyen Tien Huy, Le Thi Bich Thoa, Nguyen Phuoc Long, Nguyen Thi Huyen Trang, Kenji Hirayama, and Juntra Karbwang. 2015. Participants' understanding of informed consent in clinical trials over three decades: systematic review and meta-analysis. *Bulletin of the World Health Organization* 93, 3 (March 2015), 186–198H. https://doi.org/10.2471/BLT.14.141390 Publisher: World Health Organization.

[63] The European Parliament And The Council Of The European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation). , 88 pages. https://eur-lex.europa.eu/eli/reg/2016/679/oj

[64] Evert-Ben van Veen. 2018. Observational Health Research in Europe: Understanding the General Data Protection Regulation and Underlying Debate. *European Journal of Cancer* 104 (Nov. 2018), 70–80. https://doi.org/10.1016/j.ejca.2018.09.032

[65] Ashley Marie Walker, Yaxing Yao, Christine Geeng, Roberto Hoyle, and Pamela Wisniewski. 2019. Moving Beyond 'One Size Fits All' Research Considerations for Working with Vulnerable Populations. *Interactions* 26, 6 (2019), 34–39.

[66] Jenny Waycott, Cosmin Munteanu, Hilary Davis, Anja Thieme, Stacy Branham, Wendy Moncur, Roisin McNaney, and John Vines. 2017. Ethical Encounters in

HCI: Implications for Research in Sensitive Settings. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 518–525. https://doi.org/10.1145/3027063.3027089

[67] Travis Weller and Amalia Monroe-Gulick. 2014. Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech* 32, 3 (Sept. 2014), 467–482. https://doi.org/10.1108/LHT-02-2014-0021

[68] Christian Wendelborn, Michael Anger, and Christoph Schickhardt. 2023. What is data stewardship? Towards a comprehensive understanding. *Journal of Biomedical Informatics* 140 (April 2023), 104337. https://doi.org/10.1016/j.jbi.2023.104337

[69] Arne Freya Zillich, Daniela Schlütz, Eva-Maria Roehse, Wiebke Möhring, and Elena Link. 2024. Forschungsethische Prinzipien und methodische Güte in der Umfrageforschung. *Publizistik* 69, 3 (July 2024), 237–266. https://doi.org/10.1007/s11616-024-00845-8

A Interview Guides

A.1 Researcher Interview Guide

- **Warm-up:** What kind of empirical studies have you conducted in recent years?
- What kind of personal data have you collected, particularly identifiable data?
- How do you handle personal data during the course of the study?
- What processes are in place to ensure compliance?

A.2 DPO Interview Guide 1st Verstion

- **Warm-up:** In studies, such as interviews or surveys, personal data is often collected, which requires special handling. How does it work with publicly accessible information, such as your email address? To what extent does it count as personal data?
- What would the correct process for a researcher look like? *(We followed up until all of the following points have been addressed.)*
  - What is the role of informed consent?
  - Who is responsible for ensuring data is deleted?
  - Which data must be deleted or anonymized?
  - Are there any data that must be retained? In what form?
  - Are there guidelines or training available for researchers?
  - Are there processes enforcing compliance?
  - When is the study considered complete? What are the timelines?
  - How is this handled in international collaborations? Are there differences between countries, within and outside the EU?
- What is the role of the DPO in this process? Does the DPO have oversight over what happens?
- What is PII/personal data? What are considered identifiers (IP addresses, survey panel IDs, email addresses, demographics)?
- Can you provide us with any documents?

A.3 DPO Interview Guide 2nd version

- **Warm-up:** In studies, such as interviews or surveys, personal data is often collected, which requires special handling.
- What is personal data in empirical data? What are considered identifiers (e.g., IP addresses, survey panel IDs, email addresses, demographics)? Does it make a difference regarding data protection if data are publicly available?
- What would the correct process for a researcher look like? *(We showed the figure of a typical research process. We followed up until all of the following points have been addressed.)*
  - What is the role of informed consent?
  - Who is responsible for ensuring data is deleted?
  - Are there guidelines or training available for researchers?
  - Which data must be deleted or anonymized?
  - Are there any data that must be retained? In what form?
  - When is the study considered complete? What are the timelines?
  - How is this handled in international collaborations? Are there differences between countries, within and outside the EU?
- What is the role of the DPO in this process? Does the DPO have oversight over what happens?
- Can you provide us with any documents?