"It's Not My Data Anymore": Exploring Non-Users' Privacy Perceptions of Medical Data Donation Apps

Sarah Abdelwahab Gaballah
Ruhr University Bochum
sarah.gaballah@rub.de

Lamya Abdullah
Technical University of Darmstadt
abdullah@tk.tu-darmstadt.de

Ephraim Zimmer
Technical University of Darmstadt
zimmer@privacy-trust.tu-darmstadt.de

Sascha Fahl
CISPA Helmholtz Center for
Information Security
sascha.fahl@cispa.de

Max Mühlhäuser
Technical University of Darmstadt
max@tk.tu-darmstadt.de

Karola Marky
Ruhr University Bochum
karola.marky@rub.de

Abstract

This paper contributes an in-depth investigation (N=24) of privacy perceptions in the context of medical data donation apps. *Medical data donation* refers to the act of voluntarily sharing medical data with research institutions, which plays a crucial role in advancing healthcare research and personalized medicine. To design effective medical data donation apps, we need to understand how privacy expectations affect people's willingness to use such apps. We focus on non-users—those who have no experience with medical data donation apps—because gaining a deeper understanding of their perceptions is essential for fostering the adoption of these apps. Our findings highlight the importance of trust, transparency, and anonymity as driving factors. Participants expressed a willingness to share highly sensitive medical data with the apps if they were assured of complete anonymity, yet criticism regarding the risks of de-anonymization was also raised. Based on our results, we identify privacy awareness issues, especially concerning data sensitivity. Additionally, we explain the differences between participants' privacy expectations and preferences and what existing medical data donation apps offer. Finally, we provide guidance for the development of future user-centric medical data donation apps.

Keywords

privacy, anonymity, mental models, medical data donation apps, trust, transparency, control, data sensitivity

1 Introduction

Medical data donation is the voluntary act of providing health-related information for research initiatives and medical databases [9]. Health information includes any personal data related to an individual's past, current, or future physical or mental health [20], and may also cover fitness data (e.g., heart rate, respiratory rate, blood oxygen levels) that can reveal health issues [63]. Donating such data enhances early disease detection and understanding of diseases, ultimately leading to better treatments [22]. Several research institutes have created apps to gather health data from people for various research purposes. One notable example is the Corona-Data-Donation-App (CDA) in Germany, which was launched during the COVID-19 pandemic. Although more than 500,000 users downloaded the CDA app, indicating a willingness among individuals to share their data for medical research [52], studies revealed significant resistance among many to share their data, primarily due to security and privacy concerns [14, 29, 60, 67].

Protecting the privacy of medical data donors is not only an ethical imperative but also ensures their trust and willingness to share their data [18, 64]. Given that donated data is typically highly personal and can reveal significant sensitive information about the donors' health, linking donors to their data could lead to risks such as discrimination. Therefore, it is essential in this context to extend privacy protections to guarantee anonymity [61]. Since the goal of data donation is to identify broad patterns rather than diagnose individuals, concealing the identities of data donors should not affect the research analysis and results.

Several studies investigated privacy concerns and perceptions related to sharing medical data (cf. [4, 13, 32, 37, 51, 65, 74]). However, inconsistencies exist among these studies regarding the level of trust in researchers, the understanding of potential privacy risks and protection methods, and the effectiveness of the privacy measures in encouraging data sharing. Although research highlights the importance of anonymity for facilitating medical data sharing [13, 64, 65], there has been limited exploration into how people understand anonymity and misconceptions they might have. Furthermore, there is a lack of understanding about how people's misconceptions or lack of awareness regarding privacy and anonymity might influence their willingness to use medical data donation apps.

To address these gaps in existing research, we conducted a study focusing on perceptions and understanding of privacy in the context of medical data donation apps, with an in-depth investigation of anonymity. Given the limited use of these apps, we explored the perceptions and expectations of non-users—those who never used a medical data donation app. By gaining insight into their perceptions and identifying potential obstacles or misunderstandings, we can design apps that address these issues, making them more appealing and accessible, thereby promoting greater usage. Our research specifically considers the following main research question:

RQ: What are privacy perceptions and expectations of non-users in the context of medical data donation apps?

For this, we conducted semi-structured interviews (N=24) where we asked participants to participate in a drawing exercise. Based on the drawings, we discussed the participants' wishes, expectations, perceptions, and *speculative* mental models, i.e., how non-users imagine the usage of a data donation app. From our results, we learned that most participants struggled to illustrate a detailed mental model of their expected data donation infrastructure. We also found that participants trust data donation apps that are not driven by commercial gains and are provided by research institutes. The participants expressed a desire to control what they share and how it is used but feared that this control might burden them with technical and medical complexities. Further, they were concerned about data breaches, misuse, and discrimination, but had limited understanding of how these risks could occur. They wanted strong privacy guarantees from medical data donation apps and insisted on anonymity, expressing unwillingness to use these apps if their data could be linked to their identities or locations. However, they had awareness issues about the sensitivity of the data that these apps could collect and the methods that can ensure the privacy and anonymity of data donors. When we compared participants' speculative mental models with two existing medical data donation apps, we found significant gaps that might hinder user adoption. To develop user-friendly medical data donation apps that align with privacy and anonymity expectations, we propose several key design recommendations.

Research contributions: In the course of this paper, we make the following contributions:

(1) **First mental model investigation of medical data donation apps:** We present the first investigation of perceptions of medical data donation apps. We specifically investigated the speculative privacy mental models of 24 participants (non-users) through semi-structured interviews and a drawing exercise.

(2) **Analysis of perceived expectations, risks & misconceptions:** Among our results, we show expectations regarding data collection, storage, and access in medical data donation apps and highlight perceived risks, protection measures, and misconceptions regarding privacy and anonymity.

(3) **Comparison of users' mental models to existing apps:** We compare participants' expectations regarding privacy and anonymity guarantees to the protection measures implemented by two well-known existing data donation apps proposed during the COVID-19 pandemic.

(4) **Overall recommendations for human-centered medical data donation apps:** We conclude with recommendations for research institutes regarding how to design usable medical data donation apps that meet users' needs and expectations.

2 Background & Related Work

This section provides an overview of privacy, anonymity, medical data donation apps, mental models, and a summary of related work.

2.1 Privacy & Anonymity

Privacy grants individuals the ability to prevent involuntary disclosure by affording them the right to protect personal information

across various contexts. It is commonly understood as both an individual mechanism for revealing and concealing aspects of oneself and a contextual norm governing information flows, e.g., who has access to what information [43]. Early research on Privacy Enhancing Technologies (PETs) categorized privacy into four main areas: 'freedom from intrusion', 'negotiating the public/private divide', 'identity management', and 'surveillance' [50].

Anonymity, as a means for enhancing privacy, is linked to control over identity management and surveillance. It is typically regarded not only as a way to protect identity information but also to withhold it entirely [44]. Achieving anonymity involves concealing multiple dimensions of identity knowledge, including legal name, location, behavior patterns, and personal characteristics [42].

2.2 Privacy & Anonymity Techniques for Medical Data Donation

Several techniques have been proposed in the literature to protect users' privacy and anonymity when sharing their medical data. One common technique is pseudonymization, which involves removing personal identifiers from data and replacing them with placeholder values (i.e., pseudonyms). However, this technique proves ineffective when it comes to protecting data against a wide range of re-identification threats [34]. For example, if an individual's record is unique based on information like age, job, sex, or ZIP code, an attacker with this information can directly link the record to its owner [61]. There are other techniques that provide better protection by altering personally identifiable information (PII), both direct and indirect, within data to prevent the linkage of individuals to specific data points. Common examples of such techniques include generalization, suppression, k-anonymity, and differential privacy [15, 24]. Generalization reduces data granularity by replacing specific values with more generalized ones, while suppression selectively removes sensitive information to preserve privacy. k-anonymity ensures that each record in a dataset is indistinguishable from at least k-1 others. Differential privacy (DP) protects privacy by adding noise to data; this noise can be introduced by users (local DP) or by a data aggregator. Additional techniques for privacy-preserving medical data donation extend protection beyond data to also prevent linking users to their shared data based on communication metadata, e.g., IP addresses. Such techniques are often based on secure multi-party computation or secret-sharing-based methods [26].

2.3 Medical Data Donation Apps

Over the past few years, many apps for donating medical data have been introduced. These apps can collect similar data to fitness apps, yet differ from them in their data collection purposes, user consent, and methods. Data donation apps gather data for research purposes, with users willingly and fully informed about that. In contrast, fitness apps collect data for user health monitoring, but users may be unaware that their data is shared with other third parties, e.g., for advertising and marketing purposes [28]. Additionally, data in fitness apps is typically collected only through trackers, whereas medical data donation apps might combine trackers and questionnaires.

Medical data donation apps vary in the types of data they collect, but they can have similar infrastructure and privacy protections across contexts (COVID or non-COVID). However, motivation to participate tends to be stronger during critical situations like the COVID pandemic, as individuals are often more driven to contribute to efforts aimed at managing the crisis [14]. It is important to note that medical data donation apps differ from COVID contact-tracing apps in their goals and methods. Data donation apps focus on collecting health information to create datasets for research to improve disease understanding and treatment. Conversely, contact-tracing apps are designed for real-time contact tracing, using technologies like Bluetooth or GPS to alert users about potential COVID-19 exposure and help prevent the virus's spread [1].

In our study, we presented the Corona-Data-Donation (CDA) and SafeVac apps as examples of data donation apps because they are the most recognized German apps in this field, given that the study was conducted in Germany.

Corona-Data-Donation-App (CDA). The Corona-Data-Donation-App was developed by the Robert Koch Institute (RKI) with the aim of collecting data from users for purposes, such as detecting COVID-19 symptoms and constructing fever maps [52]. Users are required to provide health-related information, including symptoms experienced, vaccination status, and any pre-existing medical conditions. Also, users are requested to link their wearable devices to the app to allow for the collection of data such as temperature, heart rate, sleep patterns, and activity levels. Additionally, the app requires demographic information (e.g., age, gender, and location) and contact details (e.g., email address or phone number). Pseudonymization is employed to safeguard user privacy. However, this may not be sufficient to protect sensitive data in the event of a breach, particularly considering the identifiable information collected, such as location, email address, and phone number. Further, since users send their data directly to the RKI, the institute can potentially link users to their donated data through IP addresses. Moreover, the app is susceptible to several other privacy and security risks, as discussed by Tschirsich et al. [41].

SafeVac. The Paul-Ehrlich Institute (PEI) developed this app to study the effects of COVID-19 vaccines [17]. The app collects demographic data (age, weight, height, and gender), vaccination details, and health status (e.g., pre-existing medical conditions and current medications). Additionally, it prompts users to complete questionnaires at specific intervals after vaccination. These questionnaires are used to track adverse events by recording symptoms and their impact, as well as to gather feedback on the vaccination experience and follow-up actions. Pseudonymization is also implemented in this app to protect user privacy [2]. Unlike CDA, where data is sent directly to the RKI, SafeVac uses a government server as an intermediary between the PEI and users [2]. This server receives data from users and forwards it to the PEI, thereby preventing the PEI from identifying which user sent the data. While SafeVac offers greater privacy compared to CDA, it still relies on pseudonymization, rendering it susceptible to re-identification attacks [48].

2.4 User Attitudes and Awareness of Sharing Health-Related Information

According to many studies [5, 54, 66, 74], users' willingness to share their fitness data is influenced by their understanding of how the data is used and the benefits, if any, they receive from sharing it. Another common finding is that users are more likely to share their fitness data when they believe the benefits outweigh the potential risks and the recipient (e.g., service providers, third-party apps, and individuals) will use the shared data positively [4, 51, 60, 67], referring to the privacy calculus model [19]. Regarding donating medical data for research, several studies [4, 13, 14, 27, 51, 58–60, 67] found that participants generally have a positive attitude, and their beliefs in the benefits of medical research strongly motivate the willingness to donate data. Research shows that people mainly donate their data for altruistic reasons, such as supporting research and advancing healthcare for the benefit of society and future patients [7, 13, 18, 25, 45, 55, 59], though some are also driven by monetary incentives [37, 55, 59, 65]. Brown et al. [13] found that participants did not identify health-related stigma as a barrier to sharing their personal health data. Additionally, in a study by Seltzer [58], the majority of participants expressed interest in receiving the results of analyses conducted on their shared data, with half of them expressing a desire for their healthcare provider to be informed about the results as well.

Valdez and Ziefle's study [65] found that people were hesitant to share data about mental health but were more comfortable sharing information about physical health. Brown et al. [13] also found that participants were very cautious about sharing information related to sexuality. Additionally, in a study by Belen-Saglam et al. [8], participants exhibited a significant resistance to disclose information they considered irrelevant or out of context. Garrison et al. [27] discovered that individuals who were concerned about privacy and confidentiality were less likely to share their data. Further, both Garrison et al. [27] and Buhr et al. [14] observed less willingness to share data among minority groups. People generally preferred sharing their data with academic researchers rather than with businesses [27, 51, 65], government databases, or pharmaceutical companies [27].

Regarding privacy awareness, Zufferey et al. [74] found that many users of wearable activity trackers were aware of the privacy implications of sharing their fitness data, contrasting with Alqhatani et al.'s [5] findings where most users were unaware of privacy risks. Richter et al. [51] reported that about 70% of their study participants trusted medical researchers to handle their data responsibly. However, in other studies, including by Voigt et al. [67], Sleigh et al. [60], and Aitken et al. [4], participants expressed concerns about researchers potentially misusing the data. Aitken et al. [4] identified worries regarding confidentiality and users' control over their data, along with low awareness of users about current data privacy practices.

There are few studies that have investigated the impact of privacy and anonymity on individuals' decisions regarding medical data sharing. Kacsmar et al. [32] examined the impact of five privacy and anonymity techniques on user acceptability. However, the results indicated that participants had a very low level of understanding regarding these techniques. Also, Kühtreiber et al.[35] found similar

results, as participants were not able to fully comprehend differential privacy. Valdez and Ziefle [65] studied two anonymization techniques, k-anonymity, and differential privacy, and found that anonymity was the most important factor for participants in their study, regardless of the used anonymization technique. Brown et al.[13] discovered that offering the option to remain anonymous encourages individuals to share health data within online communities. Contrarily, Belen-Saglam et al. [8] did not find anonymity to be of significant importance. Their study revealed that, with the exception of data related to sex lives, participants generally did not prioritize anonymity when sharing health data. The researchers suggested that this lack of emphasis on anonymity may be due to the fact that their participants were all from the UK, where there is considerable trust in the national health service.

2.5 Mental Models

Mental models are internal representations humans derive from the real world to use a technical system [30]. This can have various levels of details that differ between humans [11, 30, 33, 68, 70]. Overall, there are two main types of mental models: functional and structural models [46]. Users with functional models know how to use a system but do not understand how it works in detail. Users with structural models have a thorough understanding of how the system works. Consequently, having a mental model requires some interaction with a system. This paper investigates *speculative* mental models, which are the users' internal representations of a system they have not used yet.

Misconceptions in mental models may lead users to engage in behaviors that do not always reflect their true needs. Thus, the mental models must be sound enough for users to interact with technology effectively [36].

Privacy concerns and misconceptions have repeatedly been shown to impact the usage intention of IoT devices [3, 62, 71, 73], or digital health records [6, 49]. The majority of related work focused on privacy as a rather generic concept in the context of mental models. The solutions proposed in most papers centered on raising awareness [71, 73], enabling control [62], or education [3]. This, however, comes with several challenges considering digitization as a whole because we likely do not have the capacity to educate individuals in-depth about each and every aspect of each system to create structural mental models. Inspired by these related studies on mental models and their findings, we decided to further investigate the medical data donation domain as a special use case where data must not be linked to the identities of individuals, consequently demanding the highest level of privacy— which is anonymity.

Summary. Related work suggests that people generally have a positive attitude toward data sharing for medical research. However, there are inconsistencies among the findings of existing studies regarding awareness levels of privacy risks and protections. There is also limited knowledge about how non-users perceive data donation apps and the sensitive nature of the data they collect. Additionally, there has been insufficient exploration of how perceptions of anonymity influence the willingness to use and engage with medical data donation apps. To increase the adoption of these underused apps, we address these gaps and contribute to the literature by examining the privacy expectations, understanding, and speculative

mental models of non-users, with a particular focus on anonymity. To achieve this, we gathered users' perceptions using both drawings and conversation-based interviews. In contrast, relevant studies have primarily relied on either surveys or conversation-based interviews to capture users' perceptions.

3 Methodology

We conducted an interview study with 24 participants to answer our research question. Our study consisted of two parts: 1) a drawing exercise where participants were asked to explain and sketch their mental models; and 2) a semi-structured interview to delve deeper into their understanding. We chose to include drawing exercises because they are effective in capturing users' mental models of specific systems or technologies [31]. We used semi-structured interviews due to their balance of structure and flexibility in exploring participants' perceptions in depth [47].

Participants and Recruitment. We did our study in Germany, where the adoption of medical data donation is very low. We recruited 24 participants who were non-users of data donation apps. We utilized various methods, such as mailing lists, flyers, poster advertisements, social networks, and word-of-mouth, to reach out to potential volunteers. All participants were required to be at least 18 years of age. Thirteen participants identified as men, ten as women, and one as non-binary. The average age of all participants was 29.37 years (SD=10.68, Min=19, Max=65). The distribution of the participants' ages reveals the following frequency counts within 10-year ranges: 1 individual [10-19], 14 [20-29], 6 [30-39], 1 [40-49], 1 [50-59], 1 [60-69].

Eleven of the participants attended school or university. Ten were employed full-time, and one participant was retired. Two participants identified themselves as housewives. There was a variation in the educational levels: nine individuals had a high school diploma, and five had a bachelor's degree. Ten had advanced degrees: one participant held a PhD, and nine had a master's degree. An overview of our sample is presented in Table 1. We used the ATI scale [23], which ranges from 1 to 6, to determine participants' affinity for technology. A higher score indicates a greater affinity for technology. Our sample had an average ATI score of 4.09 (minimum = 3, maximum = 5.11, SD = 0.66), which suggests that the participants have a high affinity for technology [23]. To assess participants' privacy perception, we considered the 10-item IUIPC questionnaire [38]. Overall, they rated their desire for control at a mean of 5.81, their awareness of privacy practices at a mean of 6.30, and the perceived ratio between collection and benefits at a mean of 5.96. That indicates that participants were more concerned about their privacy as they had high scores on the IUIPC scale. For more detailed values, see Table 1.

Study Procedure. The session we had with each participant consisted of five main parts. All the questions asked to participants are included in the Appendix A.3. The sessions were audio-recorded, with the drawing process being video-recorded as well. Each session lasted about an hour in total. The detailed procedure is as follows:

1) Consent & Demographics. Participants were first informed of their rights, and the collected data, and that they could end the study at any time without any negative consequences. Additionally,

ID	Age	Gender	Education	Job	Study Field	ATI Scale	IUIPC Control	IUIPC Awareness	Collection	Spec. Mental Model
P1	31	Woman	PhD	Researcher	Psychology	5.11	7	7	6.5	Intermediate understanding
P2	23	Woman	B.Sc.	M.Sc. Student	Industrial eng.	5	6.67	7	6	Advanced understanding
P3	23	Male	High school	B.Sc. Student	Informatics	4.89	5.67	5.33	3.75	Advanced understanding
P4	19	Woman	High school	B.Sc. Student	Informatics	4.67	5.67	5.34	5	Advanced understanding
P5	23	Man	High school	B.Sc. Student	Informatics	3.44	7	6.67	6	Advanced understanding
P6	28	Man	M.Sc.	Research Associate	Informatics	4.78	5.67	5.67	4	Advanced understanding
P7	22	Man	High school	B.Sc. Student	Informatics	4.11	5.67	6	6.75	Intermediate understanding
P8	20	Man	High school	B.A. Student	Cognitive science	4.11	4	6.33	7	Intermediate understanding
P9	30	Woman	High school	B.Sc. Student	Informatics	3	6	6.67	6.5	Intermediate understanding
P10	31	Woman	M.Sc.	Software engineer	Informatics	3	3	5.33	7	Misconception-based understanding
P11	29	Man	B.A.	M.A. Student	Psychology	4	7	5.33	7	Intermediate understanding
P12	25	Man	M.Sc.	Software engineer	IT-Security	4.78	6	7	7	Advanced understanding
P13	30	Man	High school	B.Sc. Student	Cognitive science	4.56	6.33	6.67	6	Advanced understanding
P14	42	Woman	M.Sc.	Engineer	Electronics eng.	3.89	5.33	7	7	Intermediate understanding
P15	32	Woman	M.Sc.	Engineer	Architectural eng.	4	6.33	6.67	5.25	Misconception-based understanding
P16	21	Man	High school	B.Sc. Student	Civil eng.	4.56	4.67	6.33	5.25	Intermediate understanding
P17	30	Man	M.A.	Admin. Specialist	Management	4.11	6.67	7	6	Misconception-based understanding
P18	26	Man	B.Sc.	Software engineer	Informatics	3.44	6	6	5	Misconception-based understanding
P19	21	Man	High school	B.A. Student	Cognitive science	4.78	5.33	7	6	Intermediate understanding
P20	26	Nonbinary	M.Sc.	Research Associate	Bio-medical	3.22	5.33	5.67	7	Intermediate understanding
P21	65	Man	M.Sc.	Retired	Physics	3.11	6.33	6.33	5.75	Intermediate understanding
P22	54	Woman	B.A.	Housewife	Arts	4.22	5.67	6.67	6	Intermediate understanding
P23	25	Woman	B.A.	Housewife	Law	3.78	5.33	5.33	5.25	Misconception-based understanding
P24	29	Woman	M.A.	Research Associate	Economics	3.67	6.67	6.67	6	Intermediate understanding

Table 1: Participants' demographics, education, occupation, ATI scale, IUIPC scores, and mental models.

they were informed that the interview was audio-recorded and that the drawing exercise was filmed without their faces being captured. This, along with additional information about data and privacy protection ensured for participants, was provided to them in an information sheet, which also included a consent form. Participants were asked to read and sign the consent form. Following this, each participant provided demographic information, such as age, gender, education, and occupation. They also completed the questionnaires of the ATI scale [23] and the IUIPC scale [38].

2) Warm-Up & Anchoring. We asked the participants about their understanding of medical data donation and whether they had already heard about it. To make sure all participants have a common understanding, we explained our definition of medical data donation. Following that, we asked warm-up questions, such as whether they had ever donated their medical data, if they had any experience with medical data donation apps, and what situations or settings would encourage them to donate their data. We then introduced the following scenario: there is an app that lets users donate medical data to a research institute. The collected data is used by researchers in this research institute to better understand diseases and improve public health. To make the scenario more tangible for our participants, we provided them with two examples of data donation apps—specifically, the two apps explained in Section 2.3. In terms of the information shared with participants about these apps, we provided only the app's name, the research institute that developed it, its purpose, and the method used to collect data from users (SafeVac uses a questionnaire to gather data, while CDA retrieves data from users' fitness trackers). Then, we asked them if they had experience with any of the mentioned apps. For details regarding the participants' familiarity and prior experience with data donation and its apps, please see Table 3 in Appendix A.2.

3) Drawing Exercise. After that, the participants were requested to conduct the drawing exercise. We asked them to sketch their expectations of how a medical data donation app works. This involved

illustrating the data flow and connections between various components. It's important to note that, while SafeVac and CDA were provided as examples, participants were not restricted to depicting the infrastructure of either one.

We provided the participants with paper in DIN A3 size and pens in different colors as recommended by related work [39, 72]. Research indicates that free-hand drawing without support can require excessive cognitive effort, which can be reduced by using cutout figures to make the drawing task easier [53, 56]. Therefore, we provided a wide range of printed cut-outs of several components, such as a user, a researcher, a research institute, a smartphone, a smartwatch, smart glasses, a questionnaire, the Internet, a server, a printer, a scanner, message, email, and router. Moreover, we explained that they do not need to use all the provided icons and they should only pick the ones that they wish to use in their sketch. During the drawing exercise, we encouraged participants to think aloud [10] and comment on what they were drawing, so that we could understand their thinking process. Several previous studies demonstrated the effectiveness of this combination [39, 71, 72]. We also asked the participants follow-up questions after finishing their drawing to ensure all drawn parts were explained in detail.

4) Semi-Structured Interview. We used the sketch from the previous part as the basis for the interview. We asked questions about storage, access, control, trust, privacy, and anonymity. To refine the interview script, we first conducted pilot interviews to identify and address issues with clarity of questions, structure, flow, timing, and participant comfort. Researchers took notes during these interviews, analyzed the feedback, and made necessary adjustments to the questions and procedures.

5) End & Reimbursement. Following the interview, we gave participants the opportunity to ask questions and provide additional feedback. Finally, each participant received ten euros as compensation, which was not subject to tax. This amount was consistent with Germany's minimum wage requirements: given that the minimum wage at the time of our study was €12.00 per hour pre-tax [57], ten

euros would be equivalent to or more than the after-tax earnings for one hour of work.

Ethical Considerations. In conducting our study, we adhered to the guidelines set by the ethics committees in the institutions of the authors. At our institutions, user studies must restrict the gathering of personal data to safeguard the participants' privacy. Every participant was given a random identifier. Before the interview, each participant signed a consent form, which was stored separately from all other information to ensure that it could not be linked to their identities. Prior to the study, we received approval from the ERB of the Technical University of Darmstadt.

Data Analysis. We ensured that all collected data was anonymized prior to the analysis. The audio transcripts were converted to text and personal information was replaced with neutral markers. To prevent participants from being identified through their handwriting in the drawn models, machine-generated text by a picture editing tool was used to conceal the handwriting. The sketches and interview transcripts were then analyzed in two parts using thematic analysis [12].

First, we analyzed the mental models expressed in the sketches. We ordered the sketches from undetailed to very detailed. Then, we used an open-coding approach with two authors serving as coders. By reviewing all sketches the two coders generated and agreed on a final codebook. The codebook consisted of four codes for the expressed level of detail. They then coded each sketch independently. The results were discussed, and the final code allocations for each drawing were decided. We considered the audio recordings throughout the analysis to supplement the information expressed in the sketches in cases where parts of the drawing were unclear.

Second, we analyzed the interview transcripts to capture the participants' mental models. We conducted open coding by assigning codes to meaningful and relevant concepts related to our research questions. One researcher, who conducted the interviews and was familiar with the data, proposed an initial codebook. A second researcher who was present during some of the interviews and had also reviewed the transcripts agreed on the final codebook in discussion with the first researcher. The final codebook consists of eight final categories of codes and 72 codes (see also Appendix A.1). One researcher followed the methodology guidelines for conducting thematic analysis and coded all statements using the codebook. The second researcher verified this, and any disagreements were resolved. It should be noted that thematic analysis guidelines advise against using double or multiple independent codings and relying on inter-rater reliability to demonstrate reliability [16]. This is because qualitative research acknowledges the researcher's influence on the process.

Limitations. Those are the limitations of our study: First, due to the qualitative nature of our study, we cannot make any quantitative conclusions. Also, our study relies on self-reported data and assessments, which might be biased due to social desirability, availability bias, and wrong recalls or self-assessments. As a result, our data only reflects the highly subjective perspectives of our participants. Additionally, we captured the speculative mental models of non-users, those might alter when using a data donation app. Further, we analyzed participants' mental models in relation

to the CDA and SafeVac apps, given their relevance to the study location (Germany), which might also restrict the generalizability of the findings. Moreover, our results may be influenced by cultural bias. Therefore, the findings might reflect a perspective shaped by German cultural attitudes toward privacy, which are unique due to the country's history and privacy laws [69]. Research [14, 64] also shows that people in Germany generally have a positive attitude toward data donation for research purposes, which may differ from attitudes in other countries.

Finally, despite our efforts to recruit a diverse sample, our study may lack representativeness, as all participants had college-level education or higher, which typically indicates good knowledge, awareness, and cognitive skills. As a result, our findings might not generalize well to less-educated individuals. Nevertheless, our exploratory study provided an initial step in examining the speculative mental models of non-users of medical data donation apps. Future work should investigate a more representative sample.

4 Results

This section outlines the outcomes of our study. We begin by explaining the sketches. Then, we delve into the thematic analysis results, organized by themes. We offer quantifiers of mentions to give the reader an impression of how often a certain aspect was brought up, yet this is not an attempt to quantify our findings.

It's important to note that there is no singular, definitive infrastructure for medical data donation apps, which makes it challenging to establish a single ground truth. In our study, we chose to use CDA and SafeVac—the most notable medical data donation apps in Germany, where the study was conducted—as baselines and compared participants' expectations with these apps.

4.1 Level of Detail

Based on the sketches and the interviews, we found three types of speculative mental models about medical data donation apps, each with a different level of detail. Our classification process involved the evaluation of various factors: data flow, connections among different entities, the presence of key entities (such as data sources, storage, and data recipients like researchers or research institutes), the complexity of depicted entities, and incorporation of any security-related measures. To see each participant's mental model type, refer to Table 1.

1) Misconception-based Understanding. Five participants illustrated an infrastructure for medical data donation that represents a misconception-based understanding. This model would form a functional mental model, given the provided details are limited [46]. The sketches created by these participants either portrayed an abstract data flow or failed to resemble that of a realistic data donation app. Some participants depicted only a few entities of the infrastructure (see Figure 1a). Connections between entities often deviated from real-world scenarios. Also, participants expressed difficulty in determining which entities should be connected to the internet. For instance, P23 connected the researcher directly to the user's smartphone and smartwatch, with no internet connection between the researcher and these devices, while an internet connection was established between the questionnaire and the smartphone (see Figure 1b).

(a) Sketch of P18 (Misconception-based Understanding)

(b) Sketch of P23 (Misconception-based Understanding)

(c) Sketch of P8 (Intermediate Understanding)

(d) Sketch of P19 (Intermediate Understanding)

(e) Sketch of P2 (Advanced Understanding)
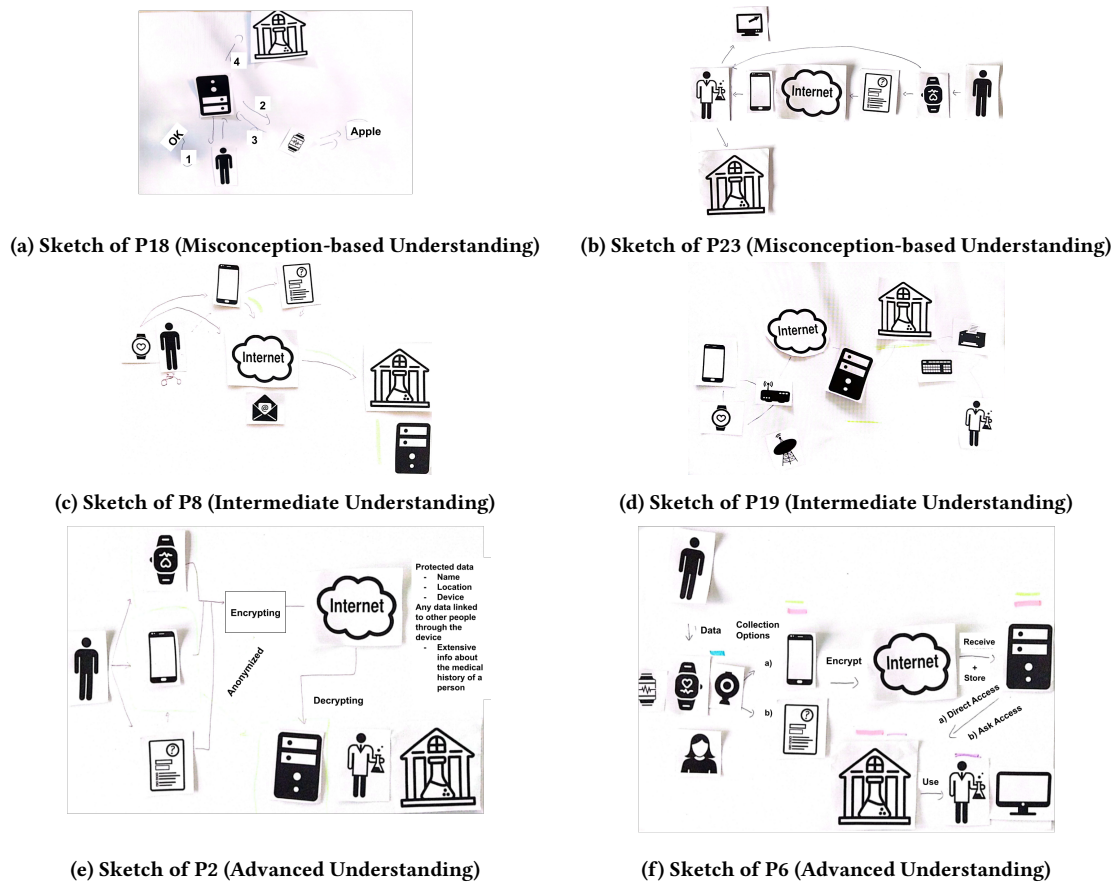

(f) Sketch of P6 (Advanced Understanding)

Figure 1: Examples of Participants' Sketches

All participants with this mental model were neither familiar with the term *data donation* nor had they experience with medical data donation or data donation apps. According to their scores on the ATI scale, three participants had medium technology affinity, while two had high technology affinity. As for their scores on the IUIPC scale, all participants scored 5 or higher in three categories: control, awareness, and collection, with the exception of one participant who scored 3 for control. These high scores indicate significant privacy concerns, a sentiment expressed also during the interviews. This shows that even individuals with basic knowledge can have strong worries or needs about privacy.

2) Intermediate Understanding. Twelve participants had this model type. They sketched the main components of a realistic infrastructure and connected these components in a realistic manner. Some of them also included multiple data sources, routers, and icons representing data transmission, processing, searching, and printing in their sketches. For examples, see the sketch of P8 (Figure 1c) and the sketch of P19 (Figure 1d). The level of understanding in this model, with its attention to detail and anticipation of data flow and connections, represents a structural model [46].

Most participants with this model showed a high affinity for technology interaction based on their scores on the ATI scale. Additionally, their IUIPC scores were high, indicating significant privacy

concerns and awareness. Although participants with this model demonstrated a more sophisticated understanding compared to those with the basic understanding model, they were unfamiliar with the *data donation* term and lacked experience with related apps, except for two who mentioned some familiarity but hadn't used the CDA or SafeVac app.

3) Advanced Understanding. This model was demonstrated by seven participants. They provided very detailed sketches that demonstrated deep understanding. The illustrated components, their connections, and the depicted data flow reflect what could be realized in a realistic infrastructure. Unlike participants with other models, all participants with this model included security-related entities or measures in their sketches. For example, all of them used encryption to protect transmissions, and some participants also sketched other security safeguards, such as anonymity or access control. Refer to the sketch of P6 (Figure 1f), where the representation of data flow was comprehensive, integrating security measures like encryption and access control. Similarly, P2's sketch (Figure 1e) emphasized the necessity of encrypting data before transmission over the internet, followed by decryption at the server end. P2 also highlighted the importance of anonymizing communication between the smartphone and the server to hide the user's location. This model is a sophisticated structural model [46].

Most participants with this mental model had previously donated their medical data, but only in offline settings. Among them, only two had heard about CDA or SafeVac, and none had experience using these apps or medical data donation apps in general. During the interviews, these participants demonstrated a high understanding of privacy, anonymity, potential threats, and protection measures.

4.2 Motivations & Expectations

Participants were questioned about factors that could motivate their use of medical data donation apps and their expectations from such apps.

Diverse Motivations & Privacy-Related Barriers. We found that participants' motivations varied, which aligns with research on other domains like blood donation that demonstrates the multifaceted nature of prosocial motivation [21]. Ten participants expressed that their motivation to use medical data donation apps would be to contribute to the advancement of medical research, a sentiment consistent with findings from other studies on different data sharing scenarios [7, 18, 55, 59]. Another mentioned motivation was the desire to aid humanity and those in need, particularly during times of crisis (N=5), similar to findings in [7, 13, 18, 45, 59]. Sample comments from the participants include P2 stating, *"To help the research basically, because I think the more people participate, the better the results of the research will be"* and P1 commenting, *"If I see a benefit for society or for people in general not just for a company"*. Other motivations were getting a reward, personal benefits, or financial gain (N=3), similar to related work on the privacy calculus [19] and the studies in [55, 59, 65]. Also, three participants mentioned that if they were suffering from specific or rare diseases requiring further study, they would be motivated to use a data donation app that collects data on these conditions.

However, there are participants (N=2) who stated that nothing could persuade them to donate their data: they either had strict privacy needs or did not trust the researchers as they believed the researchers could misuse the data and not protect the participants' privacy as they promised. For example, P15 said: *"Nothing could encourage me because I'm not someone who would agree to share medical data. I'd rather keep my data private to safeguard my privacy. Even if I were assured of data protection, I would still decline to share my data"*.

Desired Data Control, Yet Perceived Infeasible. In line with other studies on data sharing [7, 14, 45, 64], participants (N=15) voiced expectations related to transparency. Their expectations included understanding how data would be used, collected, stored, and secured, as well as knowing who would have access to the data and when it would be deleted. Interestingly, we found that some participants (N=3) also expressed a desire to track the usage of their data. For example, P10 said, *"I need to know if the data will be used by only one organization or if it will be shared among several organizations or research institutions. I would like to be able to track where my data is stored and who has access to it"*. Similarly, P24 mentioned, *"I have concerns about how the researchers may use my data. They may use it for bad things, rather than for the benefit of society as they claim. If researchers give me complete information about their research and goals, as well as the ability to track what*

they do with my data and the results of their research, I will be willing to donate my data".

The transparency mentioned above enables participants to exert control. We found that participants (N=10) demanded control over the shared data, and also wanted to choose the studies that benefit from the data, e.g., P4: *"Because, it's my data, so I should have the right to choose which parts of it I want to share. Even if I can't immediately say which data I want to keep private, there might be still something I don't want to share. In that case, having the ability to decide not to donate that specific data would be important to me"*.

Even though the participants clearly wanted control, they had misconceptions about it. Some (N=3) believed that controlling what to donate and which studies could use it was infeasible. They saw only a binary choice: agree to all the terms and provide all requested data or decline participation entirely. Additionally, others (N=2) believed that if control options were provided, they would be unable to utilize them because it would require medical or technical expertise, which they, as normal users, lack. For example, P22 said: *"No, I do not need to have control because I am not a medical expert"*.

Research-Exclusive Use is Favored over Private Companies. Participants emphasized the importance of having confidence in the institute responsible for overseeing the data donation app. Factors contributing to trustworthiness include the institute's renown, positive reputation, official (governmental) status, or exclusive research focus. Similar to the findings of Buhr et al. [14], the majority of participants (N=19) in our study voiced limited trust in private institutes or companies, expressing negative opinions about sharing their medical data with them, e.g., P19 mentioned: *"I think we hear more scandals and problems about the private companies which have these data breaches or something, so surely I would not trust a data donation app from a private company"*. Participants were divided regarding data donation to apps operated by government institutes versus those managed by research institutes. While some (N=3) perceived government institutes to possess superior data protection capabilities, others (N=4) asserted that research institutes could offer better protection, with more commitment to data integrity and non-misuse.

Similar to findings in [13, 27, 51, 65], which show that participants are unwilling to share their data with businesses or for commercial purposes. Many participants (N=6) in our study indicated they would only use data donation apps if they were assured their data would be used exclusively for research purposes. They stressed that their data should not be used for marketing or financial gain, but rather to further knowledge and provide new research perspectives. For instance, P8 said: *"I would share my data only when I'm sure that this data will be collected just for research and not for any other reason"*.

Privacy and Anonymity Guarantees Impact Sharing. Participants expressed their willingness to donate their data if the researchers possessed a high level of expertise in privacy and security, along with a commitment to anonymous data collection. Additionally, they emphasized that their confidence in sharing data would increase if the research institute operating the app were based in a country with strong data privacy laws and protections. Moreover, one participant (P2) emphasized the significance of protecting privacy by ensuring data collection from a large and diverse

population: *"If the study involves a small number of participants, I wouldn't agree to share my data. This is because I think that with a limited group of individuals, the risk of being identified increases. Thus, it is important for me that the study then has a diverse and large number of participants"*.

> **Key Findings:** In summary, participants highly prioritized the recipient of donated data, favoring research institutes over companies. They desired control over the collected data and its recipient but expressed concerns that exercising such control could introduce complexity on their end, and they may not be able to benefit from this option due to their lack of expertise. Additionally, they desired anonymity for their data while also wanting the ability to track their data within the system to monitor its use. However, these two conflicting desires may create tension, as once records are anonymized, they can no longer be linked to individuals, thus disabling control after anonymization.

4.3 Perceptions of Data Storage & Access

We explored participants' views on the storage location of collected data and their beliefs regarding entities that might have access to it.

Perceived Storage Locations Vary. The participants had various ideas regarding the storage location of the collected data. The most common expectation was that the data would be stored on a server. However, there were different views on where this server would be located. Most participants (N=14) believed that the server would be within the research institute that offers the app, such as P3, who said, *"It will be stored on a server on the researchers' side, and this server will be connected to a person's smartwatch"*. Some (N=3) mentioned that the server might be located within the software company responsible for the app's technical development. Other expectations included storage on a government server (N=1) or on cloud servers (N=5). One participant (P7) demonstrated advanced technical knowledge by proposing a distributed system, envisioning a network of interconnected servers where data could be stored. This participant also suggested that donated data could be divided across different parts of the network based on its nature, though not necessarily on all servers.

Six participants believed that data should only be stored on the user's smartphone or smart device (e.g., fitness tracker). Others (N=4) suggested a combination, with data stored on both the user's device and a server owned by the research institute. P1 and P11 proposed that data could be stored across all components and devices integrated into the infrastructure.

Two participants highlighted the importance of storing data in the same country where the users of the app live. Others (N=2) noted that the choice of storage location is influenced and governed by state data protection regulations and laws, e.g., P5 stated: *"If we're discussing this scenario in the context of Germany, I think that, in compliance with data protection laws, the data should be stored within the country"*.

Researchers Have Access, But Others Might Too. When the participants were asked specifically regarding their perceptions of who might have access to the donated raw data, we again found

different perceptions. First, almost all participants identified researchers as the primary group with access. According to three participants, after the app has collected the data, those who donated it, the users, will undoubtedly still have access to it. Their rationale behind this was that since the users own the data they contribute, they inherently possess the right to access it at any time.

Perhaps interestingly, two participants (P16 & P19) held the perspective that data contributed through the app might be also accessible to official authorities, such as the health ministry or other health and social care agencies. They expected that these authorities could have an interest in examining the collected data to gain insights into the citizens.

Five participants (P7, P9, P11, P17 & P18) believed that individuals, such as app developers, system administrators, or service providers like internet service providers or cloud service providers might have the capability to access the data contributed through the app, e.g., P7 said: *"I guess there are always some kind of administration people who are not really interested in the data itself but in organizing all the structure of the network, and I guess they could also get some kind of access. But I guess they would not need to use the data for their job"*. Additionally, they emphasized that the companies responsible for manufacturing the smartwatches worn by users could also potentially access the users' data.

Finally, two participants mentioned that anyone or any entity within the infrastructure could potentially access the data. One of the two specified that this access should only be permitted with the user's consent.

Lack of Awareness About Access Control. Most participants were unaware of the possibility of having different levels and types of access to their data. Only three participants recognized the importance of access control and stressed that only individuals with proper permissions should be allowed to access the data. However, participants struggled describing who grants these permissions; some pointed to servers, developers, or researchers as responsible entities, whereas others argued that the users themselves are responsible. A sample quote by P11: *"I would say the person who gave the data, who is like the origin of the data, has the most right to determine who can get access to it"* or P10: *"I would like to think that only limited people have access to my data. And, it's important to clarify that granting access to an institution doesn't automatically grant access to all employees within that institution"*.

> **Key Findings:** In sum, the participants expressed a variety of possible locations where the data would be stored and expected that while researchers would have access to their donated data, other parties, including themselves, would also have access. Additionally, findings show that participants had limited knowledge about access control.

4.4 Perceptions of Privacy & Anonymity

Our aim was to acquire a deeper understanding of participants' speculative mental models regarding privacy in data donation apps. We initiated our exploration by questioning which data should be protected to maintain users' privacy. This also involved pinpointing potential threats or sources that this data should be protected from, determining the party accountable for ensuring this protection, and exploring different methods for achieving user privacy. After

gathering the participants' opinions on data privacy protection, we introduced the concept of anonymity and asked about their understanding of it.

Obvious PII is Considered Sensitive, Medical Data Not. All participants provided a range of examples of medical and demographic data that a medical data donation app could gather. Most examples of medical data provided by participants include information that can be collected by fitness trackers or wearable devices, such as heart rate, blood pressure, temperature, steps, and sleep patterns. They also mentioned self-reported data, including personal and family health history, symptoms, blood type, and current medications.

When participants were asked about the data they perceived as sensitive, the majority of them highlighted information related to PII and demographics, such as name (N=10) and address/location (N=9), with phone numbers coming closely after. Subsequent mentions included national or social IDs, gender, age, birth date, bank account details, occupation, religion, national or social identifiers, education level, phone number, race, and workplace location. Examples of participants' statements include P1 mentioning, *"I think it's probably the name and probably also the location because it's easier to identify a certain person from the location"*, and P14 stating, *"The address, the last name, and the phone number are the most important things to be protected. The other information is not that sensitive"*. All participants emphasized the necessity of protecting demographic data due to its capability of identifying users or disclosing their true identities, demonstrating a strong understanding of the sensitivity associated with demographics.

Only six participants recognized the sensitivity of medical data and the need for its protection. A sample comment is given by P22: *"The person's medical history or the medical history of the person's family should be protected because if this type of information is revealed, it can be used against the person to ruin his or her life, for example, the person may lose his or her work or reputation"*.

Two participants held the viewpoint that all data collected by the app from users was sensitive and should be protected, e.g., P18: *"I think all the data that is provided by the users to the app should be protected to ensure the unlinkability between people and their data"*.

One participant (P3) expressed a willingness to share all of their data with researchers, stating that they did not possess any specific information they deemed sensitive: *"There is no personal medical information about me that I consider myself sensitive. I'm not only talking about medical data but also personal information. Even my name, I don't think that is very sensitive for me"*.

Lack of Awareness About Metadata. We found that most participants displayed limited awareness regarding the collection and sensitivity of metadata. Two participants even believed their location-based data would not be gathered. For example, P1 said, *"I think they will respect my privacy, so they will not ask for too much demographic or collect location-based data"*.

Only P2 and P5 mentioned examples of sensitive metadata and recognized that it could be used to link app users to their donated data, e.g., P5: *"The app could collect some metadata that can be traced back, such as the server storing the path from which the data originates. You can then trace the specific points where the data has traveled and ultimately trace it back to the person"*. Examples of

metadata provided by these two participants included location metadata (e.g., IP addresses), device metadata (e.g., MAC addresses), the duration taken to complete the questionnaire, and the time when the questionnaire was completed. These participants emphasized the importance of protecting metadata, with P2 suggesting that if metadata is absolutely necessary, it should be stored separately and deleted after a certain period.

Researchers Lead Protection, with Room for Shared Responsibility. When we asked about who is responsible for privacy and anonymity protection, most of the participants (N=17) pointed to the research institute that provides the app as the main entity responsible for leading data protection efforts. However, interestingly, one participant (P5) believed that anonymity preservation should not be the responsibility of researchers but rather of an external party: *"Anonymity should not be done by researchers because I think that's a conflict of interest. You need an external company to maintain anonymity software running in the cloud or on a server. If it's open source, well, people can check the code, but still, someone has to still maintain it. You could also have a new company every five years or something. Like to switch out so you don't have the same partner for a long time. That's what I'd recommend"*.

Many participants (N=9) emphasized that technical companies involved in developing data donation apps or manufacturing smart devices used by users to generate medical data should also oversee data protection. Some participants (N=5) stated that the responsibility should also lie with the government by implementing laws and regulations to ensure user protection. One participant (P24) mentioned that users should protect their data when donating it to the app by using security tools (e.g., antivirus software) on their devices. Only one participant (P12) mentioned that users are responsible for preserving their privacy by choosing what to disclose: *"The users themselves just have to look at what kind of data is collected and think about, okay, could I possibly imagine any of these data to identify myself? If I were given this data, could I identify someone with it? And then, if it looks good, the user can participate"*.

Perceived Threats. We asked participants about the potential threats that data should be protected against. Table 2 provides a summary of the threats they mentioned. When we explored who might have the potential to violate user privacy or break their anonymity, the responses of participants included either researchers or an external attacker who gains control over users' devices or servers. For example, P5 said: *"I'm assuming the institution or the group responsible for creating the system is a good actor. The bad actors who want to compromise anonymity are coming from the outside"*.

Interdependent Privacy Overlooked. Only one participant (P8) raised concerns about the privacy implications of gathering sensitive data on individuals who are not app users themselves, such as the friends and family of the app user. P8 argued that while information on family health histories, like parental cancer risks, could be valuable for understanding health conditions, collecting such data is problematic and raises serious privacy and ethical issues, particularly because it involves individuals who are not using the app and have not given explicit consent for their data to be collected.

Threat	Description
Unauthorized Access & Data Breaches	Concerns about donated data being stolen or accessed by those who do not have permission.
Data Misuse & Discrimination	Concerns about unethical use of donated data and the potential negative consequences. For example, P22 was concerned that researchers might disclose sensitive information about a donor, especially regarding stigmatized conditions, potentially damaging the donor's reputation and leading to discrimination. P16 feared that researchers might reveal the donated data to insurance companies which could lead to higher premiums.
Phishing Attacks	Concerns about attempts to obtain sensitive data through fake data donation apps, with P5 highlighting the need for protection against such attacks.

Table 2: Perceived Threats

Anonymity as Prerequisite. The majority of participants demonstrated a general familiarity with anonymity. For instance, 20 out of 24 confirmed they had heard the anonymity term before. Most participants defined anonymity as data that couldn't be traced back to the individual who provided it. Also, they understood the implications of breaking anonymity. For example, P1 mentioned, *"That the health data is connected to the person, the name, the birth date or maybe also the location"*. Similarly, P21 stated, *"It means discovering the identity of the person who gave data, which means he no longer has privacy"*.

We asked participants whether they would be willing to use a medical data donation app if they were aware that this app could link their donated data to their name, mobile phone number, or location. Fourteen participants completely declined to donate their data under such conditions. For instance, P2 said: *"If I knew that the data could be linked or thought that the data could be linked back, this is the point where I would say no so that I wouldn't participate"*. Also, P23 stated: *"No, because it is very difficult to trust researchers in this case"*. Only four participants were open to donate their data even if anonymity was not assured. Six participants expressed that ensuring anonymous data donation is very important to them, but they might agree under certain circumstances to denote their data when anonymity is not ensured. These conditions include perceiving the donated data as non-sensitive, having trust in the research institute's commitment to user protection, and believing that the country in which the research institute is located would enforce privacy protection. A sample comment by P10: *"I would be more specific about which data to donate. So for example I know there are medical information or medical data that I would not mind being linked to me personally because they are maybe more common. For example headaches, flu, and some illnesses that you have that everyone has. But as soon as it comes to very specific things like very specific illnesses or very specific cases then I would like to keep that to myself. If I can be linked to these, that could be used against me in some way"*.

Perceived Privacy & Anonymity Preservation Methods. When we asked participants how medical data donation apps could maintain privacy, we received a wide range of responses. Participants described various methods that align with common protection techniques, although most were not familiar with the specific names of techniques. The mentioned ones included encryption (N=4), data aggregation (N=4), access control (N=3), anonymity (N=3), and

pseudonymization (N=2). One participant stressed the importance of raising awareness, suggesting that countries should educate their residents about data significance, handling, and self-protection. Yet, another participant emphasized data protection can be achieved through state laws and official data protection regulations. One participant proposed protecting data by making users donate only outdated data, as according to their understanding, this data would no longer relate to the same individual; P11: *"If they only have data from the past, it's not actually about me. It's about past me that's quite different from me now"*.

When we asked specifically about how the anonymity of users can be ensured in medical data donation apps, some participants (N=5) mentioned traditional security techniques like encryption. Other participants (N=7) suggested that the app should obscure or eliminate PII from the donated data, while two participants stated that no PII should be collected at all. Additionally, some participants (N=2) proposed that ensuring anonymity could involve deliberately supplying inaccurate data to researchers. For example, P14 said, *"A person can give wrong or fake answers like saying he is female although he is male. But this sure will affect the research results"*.

Furthermore, some participants described methods that align with the following techniques: pseudonymization (N=3), data aggregation (N=2), suppression (N=1), and generalization (N=1). For instance, P6 referred to generalization, stating, *"Well, I know some techniques to anonymize data. For example, if the user inserted the exact age, like 28, then it should be converted to an age range. So we will end up with 20 to 30 or 25 to 30, things like that"*.

Data shuffling (N=2) was also brought up; e.g., P2 mentioned, *"I hope they also change the order in which the data was collected. So you cannot say, okay, this particular data came at this particular point in time, or it came from this location"*. Another participant held the perspective that anonymity could be established by opting for data collection through paper-based questionnaires rather than relying on apps or fitness trackers.

Privacy & Anonymity Awareness Issues. Many participants exhibited limited or inaccurate knowledge regarding the protection measures, such as P9: *"I don't know how the data could be protected because actually, I don't know how the data safety in Germany or Europe works. I heard of that already often, but I don't know. Especially about the medical data, I don't have any idea"*. Also, several had awareness issues about how their privacy or anonymity could be compromised and who might be motivated to do so. For example,

nine participants believed that breaking user anonymity was not feasible as long as the data they provided did not contain explicit personal identifiers, which was proven wrong in [61]. E.g., P14: *As long as the data does not include a last name, email address, or location, no one could know the person's identity or link data to him or her".* or P18: *If the data is just medical data and no birthdate or birthplace, I think it is very difficult to compromise anonymity in this case".*

Some participants (N=3) thought that the demographic information could not be used to break the anonymity of individuals who provided their data via a data donation app. This was rooted in a missing distinction between the app's user base and the larger population in a nation or worldwide: *'Because there are millions of people in the world who have the same age, weight, height, and other characteristics"* (P23). Also, the sensitivity of other data types, such as donated medical data, was rarely mentioned even though it is possible to identify individuals based on such data [48].

Only three participants (P2, P7 & P21) recognized the importance of unique demographics, medical data, or distinct data patterns in comparison to other app users as potential factors contributing to de-anonymization. E.g., P2: *'But if I know that the study is not very large, then my nationality or my race could be traced back to me. I previously took part in a study where I was the only participant of my nationality. I had concerns that if they included a statement from a participant of my nationality in their report, it could easily be traced back to me".*

Generally, we found that the participants who had knowledge gaps regarding privacy and anonymity tended to be more resistant to considering the use of medical data donation apps.

> **Key Findings:** The participants highly valued anonymity. They did not want to be identified in any way through the shared data. While they recognized that obvious data, like demographics and PII, could identify them, they did not consider that medical data (e.g., a specific disease) might also reveal their identity. Additionally, participants feared discrimination as a potential negative consequence of de-anonymization.

4.5 Comparison to Existing Apps

Most participants, across all types of speculative mental models, perceived that the donated data is sent directly from users to researchers or research institutes, aligning with the infrastructure of CDA. An exception was participant P5, who anticipated the presence of an anonymizer entity between users and the research institute, similar to SafeVac. However, this participant had higher expectations regarding the role of the anonymity entity. They anticipated the anonymizer not only routing data to the research institute, as in SafeVac, but also anonymizing the data before transmission: *"Here, we will basically have a sort of anonymizer. So all the data goes first to it to be anonymized, then the anonymizer directly sends the data in the anonymized format to the institute".*

Pseudonymization, as the approach used in CDA and SafeVac to protect users' privacy, matched the protection method expected by three participants (note that they did not name the approach but instead described what aligns with how it works). However, the majority of participants, including these three, strongly expressed a desire for anonymity when donating their medical data

via apps. They wanted their data to be untraceable to them or their locations. Pseudonymization alone cannot ensure this level of protection [26, 61]. This suggests that CDA and SafeVac do not provide the protection guarantees that participants need. Interestingly, many participants, including P2 & P12, who claimed familiarity with the apps, anticipated that the apps were employing much stronger measures, such as data removal, shuffling, aggregation, and generalization. For instance, P2 believed that data would be anonymized locally, possibly aggregated with data from other users before being received by the research institute. Furthermore, some participants (N=3) expected the apps to implement access control measures to restrict access to the donated data and allow only authorized individuals to have access. However, neither CDA nor SafeVac provided any information about whether they implement such measures.

While most participants drew infrastructures close to that of CDA, very few considered the app's ability in this case to link users to their data through the IP address. Given that only four participants agreed to donate their data if the app could link it to their location/address, it suggests a disparity between participants' understanding of privacy and anonymity within the data donation apps and their actual privacy and anonymity needs. Moreover, this highlights a gap between the functionalities of existing apps and participants' preferences, as most participants want an app that ensures strong anonymity, including hiding the origin of the data, i.e., concealing location or IP address.

5 Discussion

Our findings from interviewing participants suggest the importance of user-friendly data donation app designs to encourage people to use these apps. While privacy was considered in many debates on data donation apps (cf. [64]), our study focus was more on anonymity as it is crucial in the domain of data donation where donors have to be certain that the sensitive data that could be linked to them, is correctly anonymized.

Perspective of Medical Researchers. In addition to investigating the perspectives of potential future users of medical data donation apps, we also explored the viewpoints of medical researchers. Before beginning our study, we held discussions with members of our medical faculty and leading researchers from the Paul Ehrlich Institute, which provides the SafeVac app. From these meetings, it was clear that medical data donation is highly valued by researchers. While there was strong support for protecting donor privacy and complying with regulations like GDPR, researchers also expressed a need for data to be flexible enough for various types of analysis. They discussed the trade-off between privacy and utility, expressing concerns that excessive anonymization might reduce the data's usefulness for research. For instance, they mentioned that anonymizing data might involve removing outliers to mitigate re-identification risks, but these outliers can sometimes be crucial for analyses. They emphasized the importance of anonymizing data in a way that preserves its utility. Additionally, researchers from the Paul Ehrlich Institute highlighted the need to connect data points that come from the same donor, even if the data is anonymized and the researchers do not know the donor's real identity.

Recommendations. Based on our findings and collected insights, we discuss key design recommendations for creating a user-centric medical data donation app:

1) Make data donation research exclusive. Our participants have expressed that data donation apps should be exclusively for research purposes and have indicated that they would refuse to share their medical data if it could be used for commercial gain. Based on that, we recommend that the data donation apps should not be driven by any profit motives and limited to research only or the participants can opt-out of commercial studies by companies.

2) Make data recipients, studies & results transparent. When it comes to collecting medical data, transparency is crucial for establishing user trust. Similar to other privacy-sensitive domains, medical data donation apps should clearly outline what information they gather and why. Additionally, they should explain the types of studies that may use the collected data, how the data is stored, who can access it, and when it will be deleted. As stated above, users do not receive a direct personal benefit, yet might be driven by the benefit of society as a whole. Based on that, we recommend notifying app users about research results, new treatments, or similar where their data was used. This could make users proud of data donation yet needs further investigation in future work.

3) Make sharing highly customizable. Even though privacy and anonymity have different levels of control, we argue that control should not be completely taken away from individuals. Some of our participants were only willing to donate specific data or wanted to select which data about them to donate. This aligns with findings from other privacy-sensitive domains, such as IoT [40, 62], and studies on medical data sharing [13, 67]. Instead of providing only all-or-nothing settings, the app should enable users to easily and conveniently choose which data to donate and specify which studies can use it.

4) Data minimization. Some participants were concerned about sharing information beyond what was necessary or relevant to the research. They worried that such data might not be used for research purposes or could be misused. Hence, we recommend that data donation apps refrain from gathering any non-essential demographic or medical information from their users, and not collecting any data that could explicitly identify an individual. We observed strong rejection among participants regarding sharing their addresses or locations. Hence, we highly advise against collecting this kind of information. However, we are aware that explorative research endeavors might collect data that later on proves to be not useful. Such cases must be clearly communicated to their users.

5) Ensure anonymity by default. Some participants would donate any kind of data if they were assured that the app maintains unlinkability between users and their donated data. Our participants valued their anonymity and were hesitant to donate if their information could be traced back to them. However, they also expressed difficulty in judging what data can be used to track them. Therefore, to gain user trust and willingness to share data, we recommend that apps deploy strong anonymity measures by default that safeguard against de-anonymization risks in both data and communication. The users should be taken out of the loop by allowing to only donate anonymous data. Moreover, medical data donation apps should clearly communicate the level of security they provide in a way

that is easy for the average user to understand. Tracing apps in the COVID-19 pandemic showed that this is not an easy task [64]. Several countries used different app infrastructures offering various privacy levels. Further, it might be possible to de-anonymize datasets in the future with novel algorithms. Future work should investigate techniques for a) robust anonymity that lasts long-term and b) means communicating this to users in a verifiable way allowing users to verify the anonymization of their donated data.

6) Balancing Privacy and Data Utility. All the medical researchers we spoke with emphasized the need to balance privacy protection with the requirement for high-quality, useful data that can support a variety of research analyses. As known, anonymization techniques vary in their utility and privacy capabilities, and no single technique is universally applicable. Therefore, before designing a data donation app, we recommend consulting with potential data recipients (researchers who will use the data) to identify scenarios where the data might be anonymized in a way that renders it not useful for their analyses. This will help in selecting the most appropriate anonymization technique that maximizes privacy while meeting researchers' utility requirements.

Future Work. In future research, it would be interesting to explore how privacy mental models differ between users and non-users of medical data donation apps. Also, it would be worthwhile to investigate the differences between the privacy mental models of participants from different cultures, as the cultural factor has been shown by many studies to have a significant impact on participants' perceptions and awareness of privacy.

6 Conclusion

This paper examines the perceptions and expectations of non-users of medical data donation apps. Our findings reveal that participants had difficulty understanding how these apps work, highlighting the need for clearer information. Trust, transparency, strong security, and full anonymity were essential for their participation. Although participants understood what data could be collected, they lacked awareness about data sensitivity and protection methods, including anonymity. Privacy concerns, such as data breaches and discrimination, were noted, but understanding of these risks was limited. Those with less knowledge about privacy protections were less willing to donate data. We compared participants' expectations with two existing apps and identified gaps between the apps' protections and user needs, offering design recommendations to better align with privacy and anonymity expectations.

Acknowledgments

References

[1] 2020. COVID-19: Implications for the EU and its economy. https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/652711/IPOL_BRI(2020)652711_EN.pdf.

[2] 2020. SafeVac FAQ. https://www.pei.de/DE/service/faq/coronavirus/faq-coronavirus-safevac-app-tabelle.html.

[3] Noura Abdi, Kopo M. Ramokapane, and Jose M. Such. 2019. More than Smart Speakers: Security and Privacy Perceptions of Smart Home Personal Assistants. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS '19).* USENIX Association, Berkeley, CA, USA, 1–16.

[4] Mhairi Aitken, Jenna de St. Jorre, Claudia Pagliari, Ruth Jepson, and Sarah Cunningham-Burley. 2016. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC medical ethics* 17 (2016), 1–24.

[5] Abdulmajeed Alqhatani and Heather Richter Lipford. 2019. " There is nothing that I need to keep secret": Sharing Practices and Concerns of Wearable Fitness Data.. In *SOUPS@ USENIX Security Symposium*.

[6] Norm Archer and Mihail Cocosila. 2014. Canadian patient perceptions of electronic personal health records: An empirical investigation. *Communications of the Association for Information Systems* 34, 1 (2014), 20.

[7] Khadija Baig, Reham Mohamed, Anna-Lena Theus, and Sonia Chiasson. 2020. " I'm hoping they're an ethical company that won't do anything that I'll regret" Users Perceptions of At-home DNA Testing Companies. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[8] Rahime Belen-Saglam, Jason RC Nurse, and Duncan Hodges. 2022. An investigation into the sensitivity of personal information and implications for disclosure: A UK perspective. *Frontiers in Computer Science* 4 (2022), 908245.

[9] Matthew Bietz, Kevin Patrick, and Cinnamon Bloss. 2019. Data donation as a model for citizen science health research. *Citizen Science: Theory and Practice* 4, 1 (2019).

[10] Ted Boren and Judith Ramey. 2000. Thinking aloud: Reconciling theory and practice. *IEEE transactions on professional communication* 43, 3 (2000), 261–278.

[11] Christine L Borgman. 1999. The user's mental model of an information retrieval system: an experiment on a prototype online catalog. *International journal of human-computer studies* 51, 2 (1999), 435–452.

[12] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.

[13] Richard Brown, Elizabeth Sillence, Lynne Coventry, Emma Simpson, Jo Gibbs, Shema Tariq, Abigail C. Durrant, and Karen Lloyd. 2022. Understanding the attitudes and experiences of people living with potentially stigmatised long-term health conditions with respect to collecting and sharing health and lifestyle data. *Digital health* 8 (2022), 20552076221089798.

[14] Lorina Buhr, Silke Schicktanz, Eike Nordmeyer, et al. 2022. Attitudes toward mobile apps for pandemic research among smartphone users in Germany: national survey. *JMIR mHealth and uHealth* 10, 1 (2022), e31857.

[15] Tânia Carvalho, Nuno Moniz, Pedro Faria, and Luís Antunes. 2023. Survey on Privacy-Preserving Techniques for Microdata Publication. *ACM Comput. Surv.* 55, 14s, Article 309 (jul 2023), 42 pages.

[16] Victoria Clarke and Virginia Braun. 2013. Successful qualitative research: A practical guide for beginners. *Successful qualitative research* (2013), 1–400.

[17] D. Mentzer D. Oberle and G. Weber. 2020. Befragung zur Verträglichkeit der Impfstoffe gegen das neue Coronavirus (SARS-CoV-2) mittels Smartphone-App SafeVac 2.0. https://www.pei.de/SharedDocs/Downloads/EN/newsroom-en/pharmacovigilance-bulletin/single-articles/2020-safevac-app-en.pdf?__blob=publicationFile&v=3.

[18] Daniel Diethei, Jasmin Niess, Carolin Stellmacher, Evropi Stefanidi, and Johannes Schöning. 2021. Sharing heartbeats: motivations of citizen scientists in times of crises. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[19] Tamara Dinev and Paul Hart. 2006. An extended privacy calculus model for e-commerce transactions. *Information systems research* 17, 1 (2006), 61–80.

[20] European Union. 2011. General Data Protection Regulation (GDPR) - Recital 35. https://gdpr-info.eu/recitals/no-35/.

[21] Roy Evans and Eamonn Ferguson. 2014. Defining and measuring blood donor altruism: a theoretical approach from biology, economics and psychology. *Vox sanguinis* 106, 2 (2014), 118–126.

[22] Keolu Fox. 2020. The illusion of inclusion—The "All of Us" research program and indigenous peoples' DNA. *New England Journal of Medicine* 383, 5 (2020), 411–413.

[23] Thomas Franke, Christiane Attig, and Daniel Wessel. 2019. A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. *International Journal of Human–Computer Interaction* 35, 6 (2019), 456–467.

[24] Benjamin CM Fung, Ke Wang, Rui Chen, and Philip S Yu. 2010. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (Csur)* 42, 4 (2010), 1–53.

[25] Marco Furini, Silvia Mirri, Manuela Montangero, and Catia Prandi. 2020. Can IoT wearable devices feed frugal innovation?. In *Proceedings of the 1st Workshop on Experiences with the Design and Implementation of Frugal Smart Objects*. 1–6.

[26] Sarah Abdelwahab Gaballah, Lamya Abdullah, Mina Alishahi, Thanh Hoang Long Nguyen, Ephraim Zimmer, Max Mühlhäuser, and Karola Marky. 2024. Anonify: Decentralized Dual-level Anonymity for Medical Data Donation. *Proceedings on Privacy Enhancing Technologies* 3 (2024), 1–15.

[27] Nanibaa'A Garrison, Nila A Sathe, Armand H Matheny Antommaria, Ingrid A Holm, Saskia C Sanderson, Maureen E Smith, Melissa L McPheeters, and Ellen W Clayton. 2016. A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. *Genetics in Medicine* 18, 7 (2016), 663–671.

[28] Kit Huckvale, John Torous, and Mark E Larsen. 2019. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. *JAMA network open* 2, 4 (2019), e192542–e192542.

[29] Maximilian Häring, Eva Gerlitz, Christian Tiefenau, Matthew Smith, Dominik Wermke, Sascha Fahl, and Yasemin Acar. 2021. Never ever or no matter what: Investigating Adoption Intentions and Misconceptions about the Corona-Warn-App in Germany. In *Seventeenth Symposium on Usable Privacy and Security, SOUPS 2021, August 8-10, 2021*. USENIX Association, 77–98. https://www.usenix.org/conference/soups2021/presentation/acar

[30] Philip Nicholas Johnson-Laird. 1983. *Mental models: Towards a cognitive science of language, inference, and consciousness.* Number 6. Harvard University Press.

[31] David Jonassen and Young Hoan Cho. 2008. Externalizing mental models with mindtools. *Understanding models for learning and instruction* (2008), 145–159.

[32] Bailey Kacsmar, Kyle Tilbury, Miti Mazmudar, and Florian Kerschbaum. 2022. Caring about Sharing: User Perceptions of Multiparty Data Sharing. In *31st USENIX Security Symposium (USENIX Security 22)*. 899–916.

[33] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. {"My} Data Just Goes {Everywhere:"} User Mental Models of the Internet and Implications for Privacy and Security. In *Eleventh symposium on usable privacy and security (SOUPS 2015)*. 39–52.

[34] Florian Kohlmayer, Ronald Lautenschläger, and Fabian Prasser. 2019. Pseudonymization for research data collection: is the juice worth the squeeze? *BMC medical informatics and decision making* 19 (2019), 1–7.

[35] Patrick Kühtreiber, Viktoriya Pak, and Delphine Reinhardt. 2022. Replication: The Effect of Differential Privacy Communication on German Users' Comprehension and Data Sharing Attitudes. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. 117–134.

[36] Todd Kulesza, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng-Keen Wong. 2013. Too much, too little, or just right? Ways explanations impact end users' mental models. In *2013 IEEE Symposium on visual languages and human centric computing*. IEEE, 3–10.

[37] Hyunsoo Lee, Soowon Kang, and Uichin Lee. 2022. Understanding privacy risks and perceived benefits in open dataset collection for mobile affective computing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–26.

[38] Naresh K Malhotra, Sung S Kim, and James Agarwal. 2004. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information systems research* 15, 4 (2004), 336–355.

[39] Karola Marky, Sarah Prange, Max Mühlhäuser, and Florian Alt. 2021. Roles matter! Understanding differences in the privacy mental models of smart home visitors and residents. In *Proceedings of the 20th International Conference on Mobile and Ubiquitous Multimedia*. 108–122.

[40] Karola Marky, Alexandra Voit, Alina Stöver, Kai Kunze, Svenja Schröder, and Max Mühlhäuser. 2020. "I Don't Know How to Protect Myself": Understanding Privacy Perceptions Resulting from the Presence of Bystanders in Smart Environments. In *Proceedings of the Nordic Conference on Human-Computer Interaction* (Tallinn, Estonia). ACM, New York, NY, USA, Article 4, 11 pages. https://doi.org/10.1145/3419249.3420164

[41] Patrick Jäger Martin Tschirsich and André Zilch. 2020. Blackbox-Sicherheitsbetrachtung Corona-Datenspende-App des RKI. https://www.ccc.de/system/uploads/297/original/CCC_Analyse_Datenspende.pdf.

[42] Gary T Marx. 1999. What's in a Name? Some Reflections on the Sociology of Anonymity. *The information society* 15, 2 (1999), 99–112.

[43] Nora McDonald and Andrea Forte. 2020. The politics of privacy theories: Moving from norms to vulnerabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[44] Nora McDonald, Rachel Greenstadt, and Andrea Forte. 2023. Intersectional thinking about PETs: A study of library privacy. *Proceedings on Privacy Enhancing Technologies* (2023).

[45] Roisin McNaney, Catherine Morgan, Pranav Kulkarni, Julio Vega, Farnoosh Heidarivincheh, Ryan McConville, Alan Whone, Mickey Kim, Reuben Kirkham, and Ian Craddock. 2022. Exploring Perceptions of Cross-Sectoral Data Sharing with People with Parkinson's. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–14.

[46] Donald A Norman. 2014. Some observations on mental models. In *Mental models*. Psychology Press, 15–22.

[47] Briony J Oates, Marie Griffiths, and Rachel McLean. 2022. *Researching information systems and computing.* Sage.

[48] Iyiola E Olatunji, Jens Rauch, Matthias Katzensteiner, and Megha Khosla. 2022. A review of anonymization for healthcare data. *Big data* (2022).

[49] Rebecca Panskus, Max Ninow, Sascha Fahl, and Karola Marky. 2023. Privacy Mental Models of Electronic Health Records: A German Case Study. In *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)*. 525–542.

[50] David J Phillips. 2004. Privacy policy and PETs: The influence of policy regimes on the development and social implications of privacy enhancing technologies. *New Media & Society* 6, 6 (2004), 691–706.

[51] Gesine Richter, Christoph Borzikowsky, Bimba Franziska Hoyer, Matthias Laudes, and Michael Krawczak. 2021. Secondary research use of personal medical data:

patient attitudes towards data donation. *BMC medical ethics* 22, 1 (2021), 1–10.

[52] RKI. 2019. Corona Data Donation Project. https://corona-datenspende.de/science/en/.

[53] Mike Scaife and Yvonne Rogers. 1996. External cognition: how do graphical representations work? *International journal of human-computer studies* 45, 2 (1996), 185–213.

[54] Stefan Schneegass, Romina Poguntke, and Tonja Machulla. 2019. Understanding the impact of information representation on willingness to share information. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[55] Eva-Maria Schomakers, Chantal Lidynia, and Martina Ziefle. 2020. All of me? Users' preferences for privacy-preserving data markets and the importance of anonymity. *Electronic Markets* 30, 3 (2020), 649–665.

[56] Annett Schwamborn, Richard E Mayer, Hubertina Thillmann, Claudia Leopold, and Detlev Leutner. 2010. Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology* 102, 4 (2010), 872.

[57] Christina Schön, Roland Speidel, and Sabine Welsch. 2023. Minimum Wage and Mini-Job Threshold will rise on January 1, 2024. https://www.bdo.de/en-gb/insights/updates/tax-legal/minimum-wage-and-mini-job-threshold

[58] Emily Seltzer, Jesse Goldshear, Sharath Chandra Guntuku, Dave Grande, David A Asch, Elissa V Klinger, and Raina M Merchant. 2019. Patients' willingness to share digital health and non-health data for research: a cross-sectional study. *BMC medical informatics and decision making* 19 (2019), 1–8.

[59] Anya Skatova and James Goulding. 2019. Psychology of personal data donation. *PloS one* 14, 11 (2019), e0224240.

[60] Joanna Sleigh. 2018. Experiences of donating personal data to mental health research: an explorative anthropological study. *Biomedical Informatics Insights* 10 (2018), 1178222618785131.

[61] Latanya Sweeney. 2002. k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems* 10, 05 (2002), 557–570.

[62] Madiha Tabassum, Tomasz Kosinski, and Heather Richter Lipford. 2019. " I don't own the data". End User Perceptions of Smart Home Device Data Practices and Risks. In *Fifteenth symposium on usable privacy and security (SOUPS 2019)*. 435–450.

[63] Nora Tophof and Maximilian Tischer. 2024. Data Donation: Better Health and Quality of Life for All. https://www.data4life.care/en/library/journal/data-donation-in-medicine/.

[64] Christine Utz, Steffen Becker, Theodor Schnitzler, Florian M Farke, Franziska Herbert, Leonie Schaewitz, Martin Degeling, and Markus Dürmuth. 2021. Apps against the spread: Privacy implications and user acceptance of COVID-19-related smartphone apps on three continents. In *Proceedings of the 2021 chi conference on human factors in computing systems*. 1–22.

[65] André Calero Valdez and Martina Ziefle. 2019. The users' perspective on the privacy-utility trade-offs in health recommender systems. *International Journal of Human-Computer Studies* 121 (2019), 108–121.

[66] Lev Velykoivanenko, Kavous Salehzadeh Niksirat, Noé Zufferey, Mathias Humbert, Kévin Huguenin, and Mauro Cherubini. 2021. Are those steps worth your privacy? Fitness-tracker users' perceptions of privacy and utility. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–41.

[67] Torsten H Voigt, Verena Holtz, Emilia Niemiec, Heidi C Howard, Anna Middleton, and Barbara Prainsack. 2020. Willingness to donate genomic and other medical data: results from Germany. *European Journal of Human Genetics* 28, 8 (2020), 1000–1009.

[68] Rick Wash. 2010. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*. 1–16.

[69] James Q Whitman. 2003. The two western cultures of privacy: Dignity versus liberty. *Yale LJ* 113 (2003), 1151.

[70] Yaxing Yao, Davide Lo Re, and Yang Wang. 2017. Folk models of online behavioral advertising. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1957–1969.

[71] Eric Zeng, Shrirang Mare, and Franziska Roesner. 2017. End user security and privacy concerns with smart homes. In *thirteenth symposium on usable privacy and security (SOUPS 2017)*. 65–80.

[72] Verena Zimmermann, Merve Bennighof, Miriam Edel, Oliver Hofmann, Judith Jung, and Melina von Wick. 2018. 'Home, smart home'–exploring End users' mental models of smart homes. *Mensch und Computer 2018-Workshopband* (2018).

[73] Verena Zimmermann, Paul Gerber, Karola Marky, Leon Böck, and Florian Kirchbuchner. 2019. Assessing Users' Privacy and Security Concerns of Smart Home Technologies. *i-com* 18, 3 (2019), 197–216. https://doi.org/10.1515/icom-2019-0015

[74] Noé Zufferey, Kavous Salehzadeh Niksirat, Mathias Humbert, and Kévin Huguenin. 2023. "Revoked just now!" Users' Behaviors toward Fitness-Data Sharing with Third-Party Applications. *Proceedings on Privacy Enhancing Technologies* 1 (2023), 1–21.

A Appendix

A.1 Codebook

The bullet points represent the categories of our coding tree. The frequency is given in brackets.

- **Motivation**
  - helping humanity (N=5)
  - improving research (N=10)
  - when data really needed (N=3)
  - no motivation (N=2)
  - getting a reward (N=3)
- **Expectations**
  - having control (N=10)
  - protecting data (N=3)
  - trust & reputation (N=13)
  - transparency (N=15)
  - data collected only for research (N=6)
  - diverse & large population (N=1)
- **Data Collection**
  - collected data (N=24)
  - not collected data (N=1)
  - sensitive data (N=22)
  - non-sensitive data (N=2)
- **Data Storage**
  - server (N=15)
  - user side (N=6)
  - research institute (N=14)
  - cloud (N=5)
  - database (N=1)
  - depending on laws & regulations (N=2)
  - everywhere in the infrastructure (N=2)
  - no idea (N=1)
- **Data Access**
  - access control (N=4)
  - data accessed by users (N=3)
  - data accessed by researchers (N=20)
  - data accessed by state (N=2)
  - data accessed by technicians (N=5)
  - data accessed by anyone (N=2)
- **Data Protection**
  - examples data needs protection (N=24)
  - protection by users (N=1)
  - protection by researchers (N=17)
  - protection by state (N=5)
  - protection by technicians (N=9)
  - protection by everyone (N=1)
  - protection against data breaches (N=3)
  - protection against data misuse (N=4)
  - protection against discrimination (N=3)
  - protection against phishing attacks (N=1)
  - protection using encryption (N=4)
  - protection using aggregation (N=4)
  - protection using anonymity (N=3)
  - protection using laws (N=1)
  - protection using awareness (N=1)
  - protection using access control (N=3)
  - protection using pseudonymization (N=2)
  - protection using a donation of old data (N=1)
- **Anonymity**
  - anonymization definition (N=20)
  - anonymity using pseudonymization (N=3)
  - anonymity using encryption (N=5)
  - anonymity using laws & consents (N=3)
  - anonymity using shuffling (N=2)
  - anonymity using aggregation (N=2)
  - anonymity using generalization (N=1)
  - anonymity using suppression (N=1)
  - anonymity using no PII collection (N=2)
  - anonymity using paper-based donation (N=1)
  - anonymity using data removal (N=7)
  - anonymity using incorrect data donation (N=2)
  - anonymity by research institute (N=3)
  - anonymity by an external company (N=3)
- **De-anonymization**
  - de-anonymization definition (N=22)
  - de-anonymization by an external attacker (N=2)
  - de-anonymization by researchers (N=2)
  - de-anonymization using location (N=5)
  - de-anonymization using demographics (N=4)
  - de-anonymization using medical data (N=1)
  - de-anonymization possible without PII in data (N=15)
  - de-anonymization difficult without PII in data (N=9)
  - accept non-anonymous donation (N=4)
  - refuse non-anonymous donation (N=14)
  - conditional accept non-anonymous donation (N=6)

ID	Familiarity with the anonymity term	Familiarity with data donation term	Experience with medical data donation	Usage Experience with data donation apps
P1	No	Yes	Yes (online)	No, but mentioned having prior information about them
P2	Yes	Yes	Yes (offline)	No
P3	Yes	No	Yes (offline)	No, but mentioned having prior information about them
P4	Yes	No	No	No
P5	Yes	Yes	No	No
P6	Yes	Yes	Yes (offline)	No
P7	Yes	No	No	No
P8	No	No	No	No
P9	Yes	No	No	No
P10	Yes	No	No	No
P11	Yes	Yes	Yes (offline)	No
P12	Yes	Yes	Yes (offline)	No, but mentioned having prior information about them
P13	Yes	No	Yes (offline)	No
P14	Yes	No	No	No
P15	No	No	No	No
P16	Yes	No	No	No
P17	Yes	No	No	No
P18	Yes	No	No	No
P19	Yes	No	No	No
P20	Yes	No	No	No
P21	Yes	No	No	No
P22	Yes	No	No	No
P23	Yes	No	No	No
P24	No	No	No	No

Table 3: The participants' familiarity with anonymity and data donation concepts, as well as their prior experience in data donation and its apps.

A.2 Details about Participants

Table 3 offers insights into participants' familiarity with anonymity and data donation concepts, along with their previous experiences related to medical data donation and associated applications.

A.3 Interview Questions

- *Have you ever heard about data donation?* If yes: *In which context?*
  - Let me explain to you what data donation in the scope of health data is: Data donation is a concept that aims to improve scientific research by giving citizens the opportunity to provide data concerning their health to researchers.
- *Have you ever donated your medical data, e.g., by participating in a questionnaire?*
  - If yes: *which context? What was the topic of the study? How was the data donated: paper-based or online? What types of data have you provided in this study? What encouraged you to participate in this study?*
  - If no, *why not?*
- *What situations or settings would encourage you to donate your medical data?*
- *What kind of information about data donation is important for you when you make your decision?*
- *Is it necessary for you to be able to choose which data to donate and which medical studies this data can be used in? Why?*
- Let's consider a specific scenario: There is an app that users can use to donate their medical data to research institutes. The collected data will be used by researchers to better understand diseases and improve public health. The Corona-Datenspende-App (Corona data donation app) by the Robert Koch institute (RKI) and the SafeVac app by the Paul Ehrlich institute are two examples of such app. In the Corona data donation app, the donated data is collected from users' fitness trackers like an Apple watch. In the SafeVac app, the donated data is collected via a questionnaire. *Have you heard about any of these two apps?*
- Drawing Exercise: *Can you please draw on these papers how you think a medical data donation app like the Corona data donation app or the SafeVac app works, including the data flow?* When you draw, please keep in mind how things work in this app behind the scenes. Also, please think aloud while you draw your sketch so that I can understand what you are drawing and why you are drawing it. Another important remark: keep in mind that there are no correct answers to the questions—just answer them based on your own knowledge and experiences.
- Considering the infrastructure that you have just drawn: *what kind of data can the medical data donation app know about the data donor? Where is the data stored?*
- Now, use different colors to mark *which entity stores data about you, and which entity has data that can be linked to your person. Please also list the specific data that is stored, for instance, medical data, and demographic data.* Be as specific as possible.
- Please mark *what information should be protected about the data donors? From what should the data donors be protected?*
- *Do you have an opinion regarding who is responsible for providing this protection?*
- Depending on what the participant drew: *Which medical or personal information do you consider so sensitive that you would refuse to share it with a medical data donation app?*
- According to your understanding, *who can access your donated data?*

- *What information about the data donors can the researchers obtain?*
- *Would you still agree to donate data if you knew the medical data donation app could link your donated data to your name, mobile phone number, or location?*
  - *If yes, why?*
  - *If not, will you change your mind if you know the application is run by an official authority? Why?*
- *Have you heard about anonymity? What does it mean to you in the context of medical data donation?*
- *How, to your understanding, does a medical data collection app like the Corona data donation app or the SafeVac app protect the anonymity of the data donors?*
- *Which entity or entities are responsible for ensuring anonymity?*
- *What does breaking the anonymity of data donors mean, in your opinion?*
- *How, in your opinion, can the anonymity of data donors be broken?*
- *Is it still possible to compromise the anonymity of data donors if the donated data does not contain information that explicitly identifies an individual, such as a name, social security number, phone number, address, or driver's license? Why?*