



OPEN ACCESS

EDITED BY

Peter David Tolmie,
University of Siegen, Germany

REVIEWED BY

Claudia Müller-Birn,
Free University of Berlin, Germany
Arne Berger,
Anhalt University of Applied Sciences,
Germany

*CORRESPONDENCE

Alejandra Gómez Ortega
✉ alejandra@dsv.su.se

RECEIVED 21 March 2024

ACCEPTED 18 April 2025

PUBLISHED 30 May 2025

CITATION

Gómez Ortega A (2025) Sensitive data
donation in practice: unforeseen challenges
and lessons learned.

Front. Hum. Dyn. 7:1404855.

doi: 10.3389/fhumd.2025.1404855

COPYRIGHT

© 2025 Gómez Ortega. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Sensitive data donation in practice: unforeseen challenges and lessons learned

Alejandra Gómez Ortega*

Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden

Data donation is increasingly being used to collect personal data for scientific research. Through it, researchers directly interact with individuals and communities as they invite them to contribute to a specific research project by donating their data. This practice is facilitated by the ubiquity of products and services collecting data in people's daily lives, which means data is already available and could potentially be re-used. Additionally, it is enabled by recent directives such as the General Data Protection Regulation in Europe, which gives individuals the right to access their data. From 2020 to 2024, I investigated and applied data donation following a Research through Design process across three case studies, each focusing on different data. In this paper, I briefly introduce the case studies, reflect on my experience, and discuss the unforeseen challenges I faced and the practical considerations I learned from these. I hope these can aid researchers and practitioners in applying and further developing data donation research.

KEYWORDS

data donation, personal data, sensitive data, privacy, intimate data

1 Introduction

Data donation is an approach to personal data¹ collection that is increasingly being discussed and used across various domains, including philosophy, psychology, health, social sciences and communication, and Human-Computer Interaction (HCI). In practice, data donations constitute a voluntary transaction of personal data that is initiated by a request from a researcher or research institution (Skatova and Goulding, 2019). It has been facilitated by the ubiquity of connected products and services that continually collect personal data and with which people interact on a daily basis (e.g., digital apps, fitness trackers, internet browsers). In this way, existing data collected by a wearable device or a connected appliance can be used to investigate critical research questions. Furthermore, it has been enabled by the implementation in Europe of the General Data Protection Regulation (GDPR),² particularly the rights of access and data portability (European Parliament, Directorate General for Parliamentary Research, 2022, Art. 13 and 14) which allow people to receive their data from data controllers in a machine-readable format and (re)use it.

1 Personal data is defined in the GDPR as data through which a person can be directly or indirectly identified (European Parliament, Directorate General for Parliamentary Research, 2022, Art. 4).

2 The GDPR applies to the population of the European Union. Yet, in practice, the rights to access and data portability are available worldwide since international companies rarely limit them by geography (Bowyer et al., 2022).

Researchers rely on one of three approaches to enable people to donate their data. First, digital apps or APIs that require the donor's permission to access their data; such as the Corona-Datenspende-App through which people could donate their health data (e.g., heart rate, sleep duration, step count) from wearable devices to monitor the COVID-19 pandemic in Germany (Wiedermann et al., 2022). Second, applications to which donors give permission to scrape data using their personal account(s); such as a browser plugin that donors could install to donate information about public posts from their personal Facebook feeds (e.g., post title, number of comments) to a research project about media use (Breuer et al., 2023). Third, the most commonly used approach (e.g., Van Driel et al., 2022; Boeschoten et al., 2022; Razi et al., 2022; Zannettou et al., 2024), research repositories or digital platforms to which donors upload a copy of their personal data previously requested from a data controller.³ For instance, Van Driel et al. (2022) supported donors in obtaining a copy of their Instagram logs (e.g., posts, comments) and uploading it to a research repository to investigate the psychological consequences of social media use in adolescents. Similarly, Razi et al. (2022) developed a web-based data donation system where adolescents could upload their private Instagram messages and annotate them based on perceived safety.

Data donation practices, although emerging, are believed to be increasingly applied in the near future—especially due to recent restrictions of API access (Ohme et al., 2021; Breuer et al., 2023). Thus, it is important to reflect not only on conceptual and methodological aspects but also on practice. Practical aspects often occur “behind the scenes” and are hidden beneath the lines in publications. In this paper, I shed light on these aspects and considerations as I examine: **what practical and procedural challenges emerge when applying data donation?** To do so, I reflect on my experience developing and applying data donation across three case studies. I contribute with a set of unforeseen practical challenges that might emerge when applying data donation and a list of practical recommendations for whether and how to approach sensitive data donation in practice. I hope these can aid in furthering data donation research and practice.

2 Background

2.1 Personal data

Disclosing personal data is essential for data donation. In most cases, these data originate from people's interactions with connected products and services. Thus, personal data—potentially available through data donation—have multiple forms, types, and formats (Wiese et al., 2017); including (1) digital communications (e.g., DMs on Tinder), (2) entertainment consumption (e.g., Netflix logs), (3) finances (e.g., credit card purchases), (4) physical activity (e.g., daily steps), and (5) physiological signals (e.g., heart rate), among many others. These data are often considered “sensitive data,” defined in the GDPR as a special category of personal data that includes racial or ethnic origin, political opinions, religious

or philosophical beliefs; health-related data; and data concerning a person's sex life or sexual orientation, among others (European Parliament. Directorate General for Parliamentary Research, 2022, Art. 9).

Therefore, ethical considerations around personal data are critical in data donation. The main consideration is informed consent (Hummel et al., 2019; Jones, 2019; Ohme and Araujo, 2022), and how donors can exert their autonomy (Prainsack, 2019; Jones, 2019) and preferences (e.g., deciding whether/ what/ and to whom to donate Strotbaum et al., 2019). Especially when digital-trace data is opaque—and thus, it is difficult for donors to know what is in their data—and potentially sensitive or invasive (Hummel et al., 2019; Jones, 2019). Further, data donation concerns ethical aspects related to the relationship between donors and recipients. These include aspects such as data minimization (Ohme and Araujo, 2022; Ohme et al., 2024; Boeschoten et al., 2022), mitigating harm (Prainsack, 2019; Boeschoten et al., 2022), and uncertainties around future data use (Nickel, 2019; Hummel et al., 2019). Additionally, these require researchers to be transparent, provide sufficient information, and honor donors' contributions (Prainsack et al., 2022; Hummel et al., 2019). In this paper, I explore these ethical considerations from the perspective of an academic researcher developing digital tools and methods for individuals to donate their data to (design) research. I emphasize the practical and procedural aspects of my experience.

2.2 Data donation

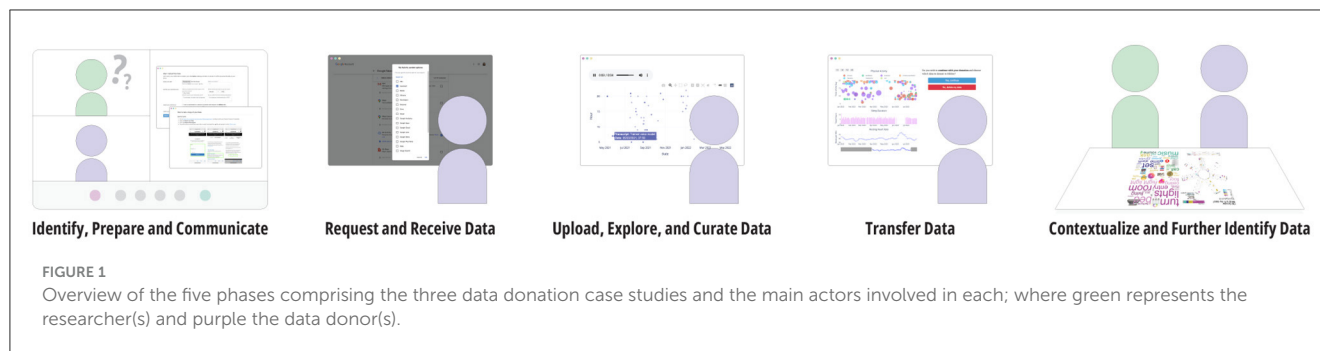
Data donation is an approach to collecting personal data for research purposes. Similar to the donation of blood, the donation of data is a voluntary transaction of personal data from an individual who “has”⁴ data to another individual or entity who needs it (e.g., a healthcare center or research institution). When a person donates her personal data she is actively consenting to provide it for research purposes (Skatova and Goulding, 2019). Recently, data donation has shown promise as an alternative to platform-centric approaches to personal data collection, such as Application Programming Interfaces (APIs) (Ohme and Araujo, 2022; Breuer et al., 2023; Van Driel et al., 2022). Especially considering the current landscape, in which access to data through the APIs of certain platforms (e.g., Facebook, Twitter, Reddit) is increasingly restricted (Breuer et al., 2023). As a result, data donation is considered an individual-centered (Breuer et al., 2023) or user-centered approach (Ohme and Araujo, 2022); where researchers are able to access data *directly* from individuals—who get to opt-in or consent to their participation.⁵

One of the main strengths of data donation is the nature of the data that it allows access to Van Driel et al. (2022),

³ A data controller is an entity (e.g., private company or public authority) that collects personal data. They are required by the GDPR to provide a copy of that data in a machine-readable format.

⁴ The term is in quotation marks as legal scholars have extensively argued about the limits of ownership—as exclusive use—in the context of personal data (Prainsack, 2019; Hummel et al., 2019).

⁵ Platform-centric approaches don't require individual participants to consent to the use of their data. In some cases, it is assumed they consented to secondary uses of their data through the Terms and Conditions of a specific product or service.



Breuer et al. (2023), Gómez Ortega et al. (2022), and Razi et al. (2022). In contrast to the public data available through platform-centric approaches (e.g., all public posts from an individual on Reddit), data available through data donation are individual-level private data all public posts from an individual, and the direct messages they have changed with others on Reddit). The individual-level private data offers researchers access to new types of insights and open the way to investigate new research questions across sensitive domains. For instance, Razi et al. (2022) collected private Instagram conversations from teens to identify online risks; including nudity and porn, sexual messages or solicitations, harassment, and violence, among others. Yet, requesting and accessing private data at the individual level also poses new challenges (See Section 2.1). In this paper, I describe the process of designing, developing, and applying data donation across three case studies, each with different types of data: (1) menstrual tracking logs, (2) speech records, and (3) physical activity logs. These case studies focus on intimate contexts, such as the body and the home, where the sensitive nature of the data raises critical privacy considerations and pragmatic decisions that introduce tensions.

2.3 Research through design

Research through Design (RtD) is an approach where design actions and design activities play a formative role in the generation of knowledge (Stappers and Giaccardi, 2014). A research topic is investigated through an iterative process that involves creating an artifact and reflecting on the creation process and the resulting artifact—which enables an iterative (re)framing of the research topic (Zimmerman et al., 2007). Thus, the creation and development of artifact(s) play a central role in the knowledge-generating process, for instance, by giving form to an alternative future state and seeing whether and how it works (Stappers and Giaccardi, 2014; Bardzell et al., 2012). Artifacts are the concrete embodiment of a specific concept or idea and are a way to share and communicate it with others (Zimmerman et al., 2007; Stappers and Giaccardi, 2014). They are shaped by design decisions made to represent said concept or idea (Stappers and Giaccardi, 2014). Further, they are shaped by the technical opportunities and constraints around “making” them (Stappers and Giaccardi, 2014). Thus, artifacts allow researchers to derive knowledge from “making” them as well as sharing them with others. In this paper, I describe an iterative RtD process where the artifacts

correspond to various instances of a Sensitive Data Donation approach (Gómez Ortega et al., 2024b) embodied by a digital data donation platform (Section 3). These create a possibility for people to engage in a specific form of data donation that was not possible before the design and that becomes experienceable through the design.

3 Method

This paper describes an iterative Research through Design (RtD) approach⁶ of designing, developing, and applying specific instances of Sensitive Data Donation (Gómez Ortega et al., 2024b)—embodied by a digital data donation platform—and deploying them in what (Koskinen et al., 2011) refer to as the “field,” a specific case study where people can engage with and experience a form of data donation. This process comprises three iterations of the Sensitive Data Donation approach, each applied in a given context. Along this process, I derive knowledge in two ways. First, through “making” or developing the data donation process and platform. This involves engaging with specific theories and concepts (e.g., privacy as boundary management) and facing practical decisions (e.g., How to encourage donors to set boundaries around their data?), opportunities (e.g., I can create a tool for them to explore their data), and constraints (e.g., What if they don’t feel like using the tool?). I generate insights by documenting and reflecting on my own struggles and successes (Stappers and Giaccardi, 2014). In this paper, I report on these insights—in the form of lessons learned around the practical and procedural challenges that emerged from applying data donation. Second, through sharing a specific data donation instance with others. More specifically, by inviting others to partake in or experience a specific form of data donation in a given context and reflect on their experience. I generate insights by learning from individual data donation experiences as people go through them. These insights are reported in publications specific to each case study (Gómez Ortega et al., 2022, 2023, 2024a) and are out of the scope of this paper.

In what follows, I briefly introduce the Sensitive Data Donation method and its application across the three Case Studies. I reflect on my experiences through these to examine the practical and procedural challenges that emerged when applying data donation.

⁶ This work was carried out from 2020 to 2024 at Delft University of Technology.

3.1 Sensitive data donation

Sensitive Data Donation (Gómez Ortega et al., 2024b) is a data donation approach informed by feminist perspectives on data. It comprises five principles:

- (P1) **Balanced value:** it calls for recognizing and honoring donors' contributions and efforts by intentionally integrating activities into the data donation process that allows them to derive value (e.g., dedicated learning activities for donors).
- (P2) **Sensitive data:** it calls for recognizing the sensitive nature of the data, and supporting donors in knowing their data and drawing better-informed boundaries around them.
- (P3) **Multiple knowledge(s):** it calls for valuing multiple knowledge(s) in data donation and not only the digital-trace data; by involving donors in interpreting and contextualizing their data and prioritizing their embodied and situated knowledge.
- (P4) **Ongoing consent:** it calls for embedding informed consent as an ongoing incremental process that accounts for donors' preferences regarding participation and disclosure of sensitive information over time.
- (P5) **Shared goals:** it calls for supporting different degrees of participation in data donation (i.e., contributors, collaborators, and co-creators Shirk et al., 2012) and inviting interested potential donors to relate to and shape the research project and goals from the start.

These principles are integrated into a five-phase approach (Figure 1), where donors upload a copy of their data to a digital platform; in this case, a web-based data donation platform developed using the open-source Python web framework Django. In the following, I summarize each phase of the Sensitive Data Donation approach and describe how I applied it throughout the three case studies.

- **Identify, prepare, and communicate:** researchers and potential donors co-create and scope the research questions and goals and co-define the value-gain strategy (i.e., how donors will derive value from their participation), data needs, and how these can be flexible enough to suit individual preferences. Additionally, researchers invite potential donors to participate in the research. When applying this step in the Case Studies, I identify what data is relevant to investigate a research question; how potential donors can access their data; and how potential donors can gain value through their participation—often in collaboration with them. At this point, I submit our project for review to the institution's Human Research Ethics Committee and Privacy team. Additionally, I prepare the data donation platform to parse and visualize the data. This process entails requesting, obtaining, and understanding the Takeout files returned by the specific data controller. In the data donation platform, I provide information about the project, and I share this information with existing communities of potential donors as I invite them to participate in the research.

- **Request and receive data:** potential donors respond to researchers' invitation by following the steps to request and obtain a copy of their data from data controllers, with assistance from researchers if necessary. When applying this step in the Case Studies, I become familiar with the practicalities of the data download process and provide detailed instructions for potential donors to access their data. I remain available to support potential donors when necessary. Those interested in participating in the research follow the instructions, request, and obtain a copy of their personal data. This process normally entails direct contact with the data controller(s).
- **Upload, explore, and curate data:** researchers support potential donors to autonomously explore and draw boundaries around their data and the information they wish to disclose before they disclose it. When applying this step in the Case Studies, I develop tools aimed at facilitating potential donors' interaction with their data even before they become research participants by making the (opaque) data visible, inspectable, and easy to understand and manipulate. Once they upload their data to the platform, they are invited to explore it so they can better understand what information it contains and curate their data so they explicitly decide whether and what data to donate. This is done through interactive data visualizations in the data donation platform (Figure 2).
- **Transfer data:** donors transfer (a part of) their data to researchers—after having explored and defined clear boundaries around it—and consent to their participation in the research; they can (re)evaluate their consent from this point onward. When applying this step in the Case Studies, I create an explicit moment for donors to decide whether to continue with their donation (i.e., confirm consent) or delete their data (i.e., revoke consent) once they have explored and curated their data. If potential donors decide to continue with their donation, they become donors, and the research team gains access to their data until they decide (re)assess their participation and remove it from the platform or until the end of the project.
- **Contextualize and further identify data:** donors are invited to interpret and contextualize their data with the researchers and (re)evaluate their participation (i.e., confirm or revoke consent). This requires researchers to prepare and represent the data as a tool to elicit and invite multiple forms of knowledge. When applying this step in the Case Studies, I conduct a voluntary session with donors to interpret and contextualize their data. During the session, I introduce the data through interactive data visualizations.

3.2 Data analysis

This paper describes an iterative Research through Design (RtD) process across three case studies (Table 1). Each case study represents an application of the Sensitive Data Donation method

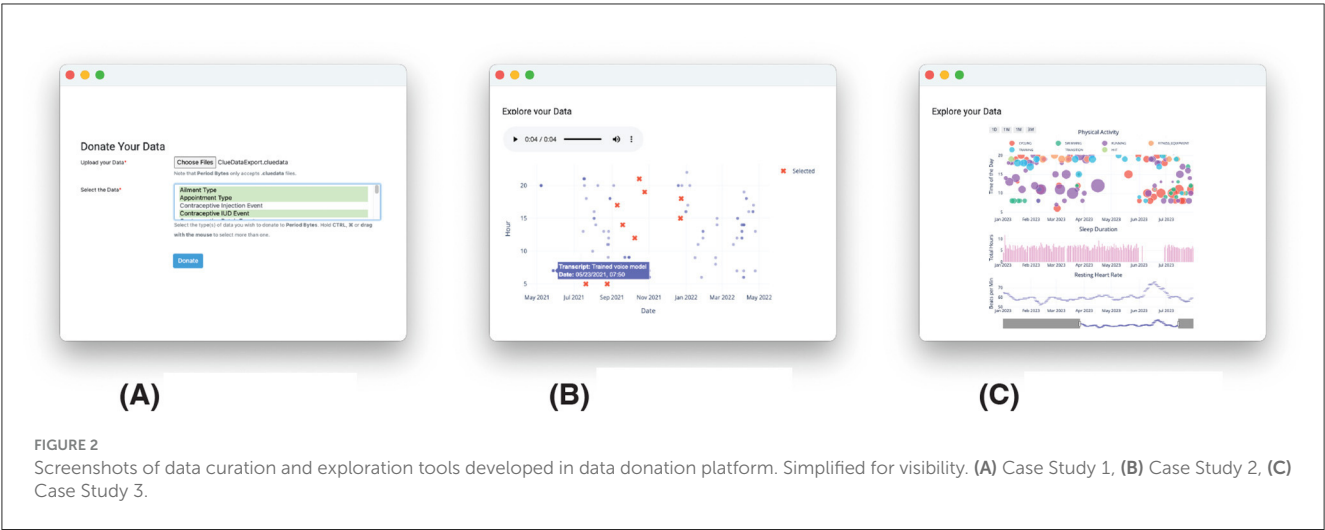


TABLE 1 Overview of the actors involved in each case study, the type(s) of data, and the practicalities of donating data to the research team.

Case study	1	2	3
Recipient(s)	Research team	Research team	Research team
Donor(s)	Adults (18+) Active clue users.	Adults (18+) Active Google Assistant users.	Adults (18+) Garmin or Apple Health users.
Data source	Clue	Google Assistant	Garmin and Apple Health
Data type	Menstrual logs	Google Assistant	Garmin and Apple Health
Request data	Through clue app. Available in minutes.	Through dashboard. Available in up to 30 days.	Garmin: through dashboard. Available in up to 30 days. Apple: Through Apple Health app. Available in minutes.
Receive data	.JSON	Zip folder: .mp3 and .JSON	Garmin: Zip folder, .JSON Apple: .XML
Explore data	After donation.	Before donation.	Before donation.
Curate data	Which types to donate.	Which points to donate.	Which types and points to donate.

(Gómez Ortega et al., 2024b); in a given context and with specific design decisions and practical considerations. The iterative and cumulative process means Case Studies influence each other, and each one builds on and expands the prior design decisions and considerations made. In line with the RtD methodology, this paper contributes to (behind-the-scenes) knowledge from this iterative process of “making”—met with practical decisions and constraints.

During each Case Study, I systematically kept a personal research journal (See Appendix), where I documented and reflected on what happened across every step of the data donation process: from the implementation of the data donation platform to interacting with potential donors. When the three Case Studies concluded, I gathered all the journal entries and clustered them by theme and phase of the Sensitive Data Donation method with a focus on practical and procedural challenges. I report on these below.

4 Case studies

4.1 Case study 1: menstrual tracking logs

The goal of this case study, described in depth in Gómez Ortega et al. (2022), was to explore how to better support people who track their menstrual cycle in interacting with the data they produce. Therefore, it was essential to gain access to the data (i.e., menstrual tracking logs) and situate and contextualize it within individual tracking practices and experiences. I collaborated with a group of potential donors (i.e., users of menstrual tracking technologies) to define a value-gain strategy (i.e., a concrete way for donors to derive value from their participation). It comprised a dedicated workshop with an expert in reproductive health and sexuality. I decided to focus on all data collected via the menstrual tracking app Clue,⁷ as at the time of the research, it was relatively

7 Clue menstrual tracking app: <https://hellocue.com/>.

easy for Clue users to obtain a copy of their data—they could download a takeout file directly from the app—in contrast with other menstrual tracking apps. The self-reported data corresponds to up to 31 different types of data Clue users can log, arranged in three groups: (1) single choice categorical data (e.g., sleep duration: 0–3 h, 3–6 h, 6–9 h, and 9 h or more); (2) multiple choice categorical data (e.g., menstrual pain: cramps, headache, ovulation pain, tender breasts); and (3) manual input (e.g., weight: numeric value).

In the platform, I aimed to support donors in exercising their autonomy and data sovereignty by inviting them to explicitly decide what data to donate. Thus, they were invited to select which types of data to donate out of the 31 different types collected by Clue through a form in the data donation platform (Figure 2A); there were no “select all” or pre-checked options. Having selected the types of data, donors could proceed with their donation and see a visualization with an overview of their data. From this point onward, they could remove it from the platform. From mid-October to mid-November 2021 I used convenience and snowball sampling to reach out to worldwide Clue users online, by posting on social media and reaching out to existing online communities. Thirty-five people volunteered to donate their data of which 27 people decided to contextualize them. To contextualize the data, I designed a timeline visualization of all the donated data (Figure 3A), following the recommendations by Pins et al. (2022) and Tolmie et al. (2016), and conducted individual sessions with donors. Seventeen donors donated all the types of data while 18 donated only a subset. This resulted in diverse datasets, shaped by how donors interacted with Clue and their specific choices. Data was retrospective and vast, from when a donor started using Clue to the date of donation; in some cases, this time period was more than 3 years.

4.2 Case study 2: speech records

The goal of this case study, described in depth in Gómez Ortega et al. (2023), was to investigate people's perceptions of their speech records when faced with a comprehensive view as opposed to individual data points. In this case, situating and contextualizing the data was a way for donors to better understand the different types of information they capture; how it relates to their lived experiences; and reflect on its implications. The hypothesis was that donors could derive value from increasing their understanding of their data, thus the value-gain strategy was aligned with the contextualization activities. I decided to focus on speech records collected by Google Assistant, as Google has an extensive pool of users and it offers a relatively simple and quick process to obtain a Takeout of the data and well-structured metadata. The data corresponds to speech records. These are logs of every interaction between potential donors and their Google Assistant, each including a timestamp, a transcript, and an audio recording.

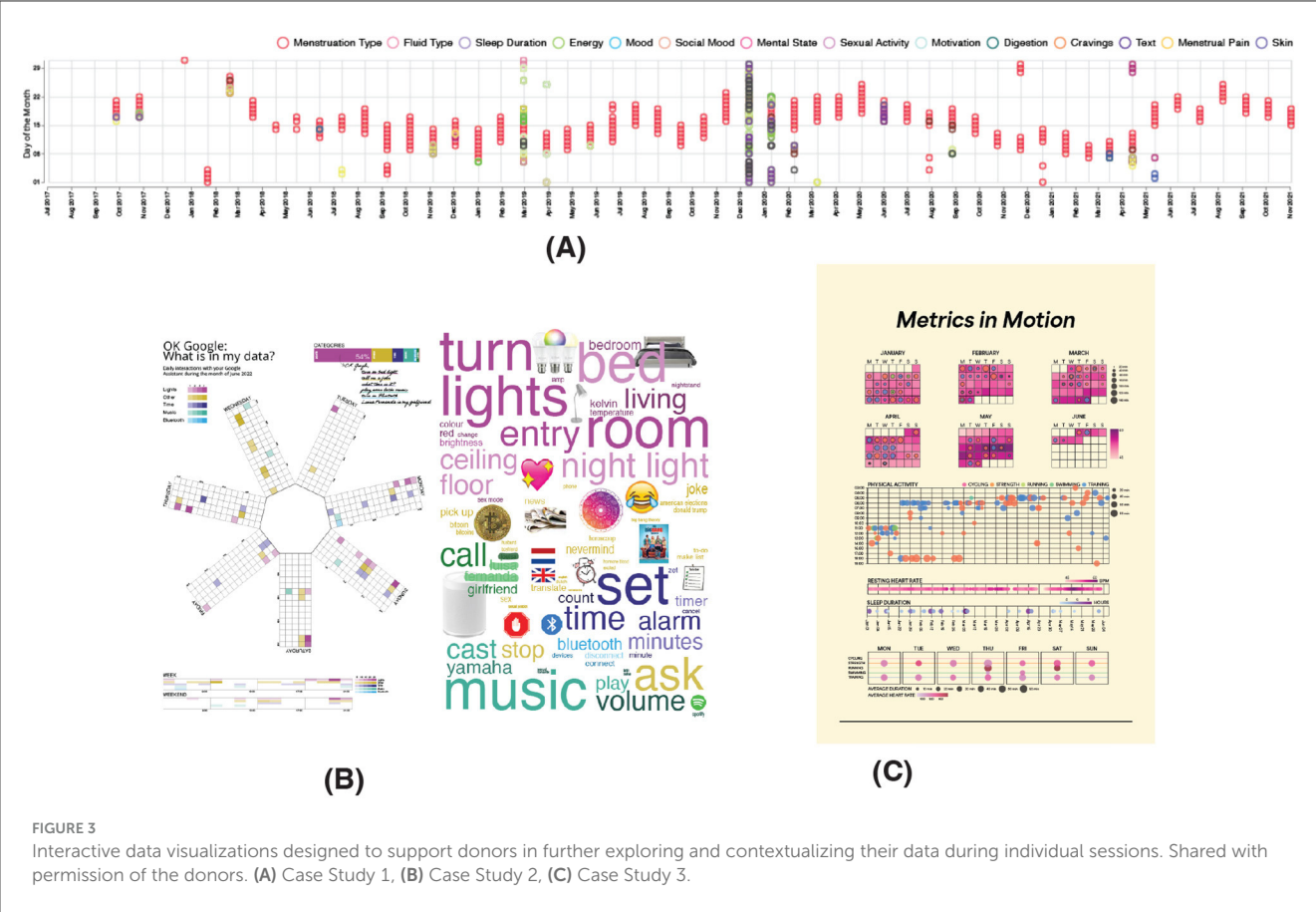
In the platform, I aimed to support donors in the autonomous exploration of their data *before* they decide whether to donate it. This was one of the shortcomings of Case Study 1. At the time of

donation donors often “*don't know what they don't know*” Jones (2019) about their data, and the files in confusing formats returned by Google—and similar data controllers—don't promote their understanding (Bowyer et al., 2022; Alizadeh et al., 2019). Thus, I developed an interactive graph (Figure 2B) where donors could see and listen to their data when hovering at a point representing a speech record. Through this graph, donors could also decide which specific points to donate. Having selected these, donors could proceed with their donation. From early April to mid-June 2022 I used snowball sampling to reach out to worldwide Google Assistant users online, by posting on social media and mailing lists. Twenty-two people volunteered to donate their data; of which 17 also decided to contextualize them. To contextualize the data, I designed a series of visualizations highlighting several aspects of the data, including the temporal patterns and the most frequent keywords (Figure 3B). Most of the donors (21/22) decided not to remove any points from their donation. Due to changes in Google settings in 2018, speech records comprising the audio recordings were not saved by default, and for some users, data contained only speech records stored up to that point (See Figure 3C).

4.3 Case study 3: physical activity logs

The goal of this case study, described in depth in Gómez Ortega et al. (2024a), was to investigate how athletes perceive the impact of the menstrual cycle in sports. In this case, situating and contextualizing the data was a way for donors to reflect on their experience over time and identify interesting correlations between different factors and data. I collaborated with a small group of potential donors (i.e., athletes) and a sports gynecologist to determine a value-gain strategy (i.e., a concrete way for donors to derive value from their participation) and identify the relevant types of data. The value-gain strategy comprised communicating partial results to donors and giving each donor a poster representing their data; as athletes enjoy displaying their achievements and mementos which are captured by their data. The data corresponded to (1) menstrual cycle, (2) sleep, (3) heart rate, and (4) physical activity logs. Unlike the previous Case Studies, I decided to focus on data from Garmin devices and the Apple Health ecosystem to lower the participation threshold and support participation by athletes who use more than one specific device.

Building upon the considerations for Case Study 2, through the data donation platform, I aimed to support donors in the local and autonomous exploration of their data *before* they decide whether to donate it. Moreover, I decided to explicitly limit the time frame of the data by inviting donors to decide whether they wanted to share data from the last 3 or 6 months; unlike Case Studies 1 and 2 where I accessed all available retrospective data. Thus, similar to Case Study 2, donors could explore and curate their data through an interactive data visualization (Figure 2C); where they could decide which types of data to donate and for how long. From early June to mid-July 2023 I used purposive and snowball sampling to reach out worldwide to professional, semi-professional, and amateur athletes using a Garmin device or a device synchronized with the Apple Health ecosystem. Specifically, I posted flyers on social media, local sports competitions, and sports



associations. 20 athletes volunteered to donate their data; of which 16 also decided to contextualize them. To contextualize the data, I designed a poster highlighting the physical activity over time in a calendar view and a timeline and integrating the sleep and heart rate data (Figure 3C). Most of the donors decided to donate all types of data (18/20) from the last 6 months (19/20).

5 Unforeseen challenges and considerations when applying data donation

In this section, I describe some of the unforeseen challenges I faced through the three case studies and the consequent practical considerations in the form of reflective questions. They come from my personal experience approaching data donation; which I noted as entries in a research journal over the past 3 years.

5.1 Ethical approval: dealing with unusual requests

A significant challenge in all three Case Studies, but especially in the first one, was getting the research reviewed and approved by the Human Research Ethics Committee and Privacy Team at the institution where this research was conducted. Their main concern was that data donation is a different way of collecting personal data

that they (and the general public) were not familiar with. They raised important questions about (1) operationalizing (dynamic) informed consent through a digital platform (e.g., How can donors understand the purpose of the research? How can donors ensure all their data has been removed?); (2) the sensitive nature of the data (e.g., How to mitigate privacy concerns? What additional measures may be needed?); (3) secure data processing and storage (e.g., Where is data stored? How is the platform GDPR compliant?); and (4) limited data use (e.g., How can donors ensure their data is not duplicated for further use? How can donors ensure data is not accessed beyond the research team?).

Together, we carefully examined these questions along with every step in the data donation process; we agreed on essential considerations including identifying and documenting all the personal data processing activities on the platform and the roles of responsibilities of the research team and limiting access to the data to the research team for the duration of the project. However, for some of these, implementation is limited and, in many ways, depends on trust. For example, I technically cannot guarantee that data is not duplicated for future use, although ethically, I commit to doing so. Further limitations come in the form of questions to which I don't have an answer, such as: how to deal with a donor uploading someone else's data?

Considerations How can you best align with your institution's Human Research Ethics Committee? How can you best communicate your approach and the ethical considerations it entails? Note that these might not be fully aligned with what is expected in more "traditional" research processes. Consider

walking them through your proposed data donation process and detailing how you plan to address critical aspects such as informed consent and the relationship and interactions with participants (and their data). Reach out to them a few months before the start of your study so you have enough time to carefully examine and revise the critical aspects. What commitments have you made to donors? How can you remain accountable to them throughout the process? Consider communicating effectively with donors about the status of these, for instance, by reaching out to them when the project is finished and their data is permanently removed.

5.2 Parsing the data: changing structures

Parsing and processing the data in the data donation platform is essential for receiving donations. To do so, I first request my data, or that of a dummy account, from the specific data controller (i.e., Clue, Google, Garmin, or Apple Health) and inspect the structure and the location of the data I need (e.g., individual speech records in the second Case Study). Once I'm familiar with this I write a script to process it in the platform. I then test the script with my own files, files from dummy accounts, and files from colleagues who volunteer to help. Here is where I have realized that context matters. At the time of writing and conducting this research, I am based in the Netherlands, and I am surrounded by international friends and colleagues. Our shared language is English, and we tend to configure our devices in English, which is also the default language for many data-driven products and services. Thus, in the second Case Study, the language of all the files I tested with was English. In this context, it was difficult for me to anticipate that accounts in other parts of the world or configured in a different language would result in Takeout files with different languages. Nonetheless, when I started reaching out to potential donors worldwide, I started receiving emails reporting issues uploading their files to the data donation platform. The reason behind these issues: the language of the multiple folders and files was different (Figure 4). As a result, I had to modify the script in the middle of the data donation campaign to account for this difference. It introduced delays for donors whose Google Assistant was not configured in English, some of whom attempted to upload their files to the platform multiple times while others gave up.

Having learned from this, in the third Case Study, I set up Garmin and Apple Health accounts in different languages (i.e. English, Spanish, Italian, Dutch) and requested the data each time. The files from Garmin were all in English regardless of the language of configuration, while the name of the files from Apple Health changed depending on the language of configuration. Although files from Garmin and Apple Health contained similar types of data (i.e., sleep, heart rate, and physical activity), their format and structure were different (Figure 5). This is an important limitation and raises a critical practical consideration: receiving data from multiple devices means a larger pool of potential donors; at the same time, it means more labor and opportunity for unforeseen challenges. For instance, in the case of Garmin, the structure of the files inside the Takeout folder changed from early June 2023 to mid-June 2023; changing the location of the file containing the heart rate data (Figure 6), Van Driel et al. (2022) reported a similar

phenomenon with the Instagram Takeout files. Thus, some of the donations I received in mid-June were incomplete; lacking the heart rate data. I again had to modify the script in the middle of the data donation campaign. This illustrates the attention needed to ensure that the data donation platform keeps working, which largely constitutes invisible labor. In addition, I reached out to donors whose donations were incomplete, requesting that they upload their files again. Most donors were highly motivated and agreed; however, some expressed frustration with this process.

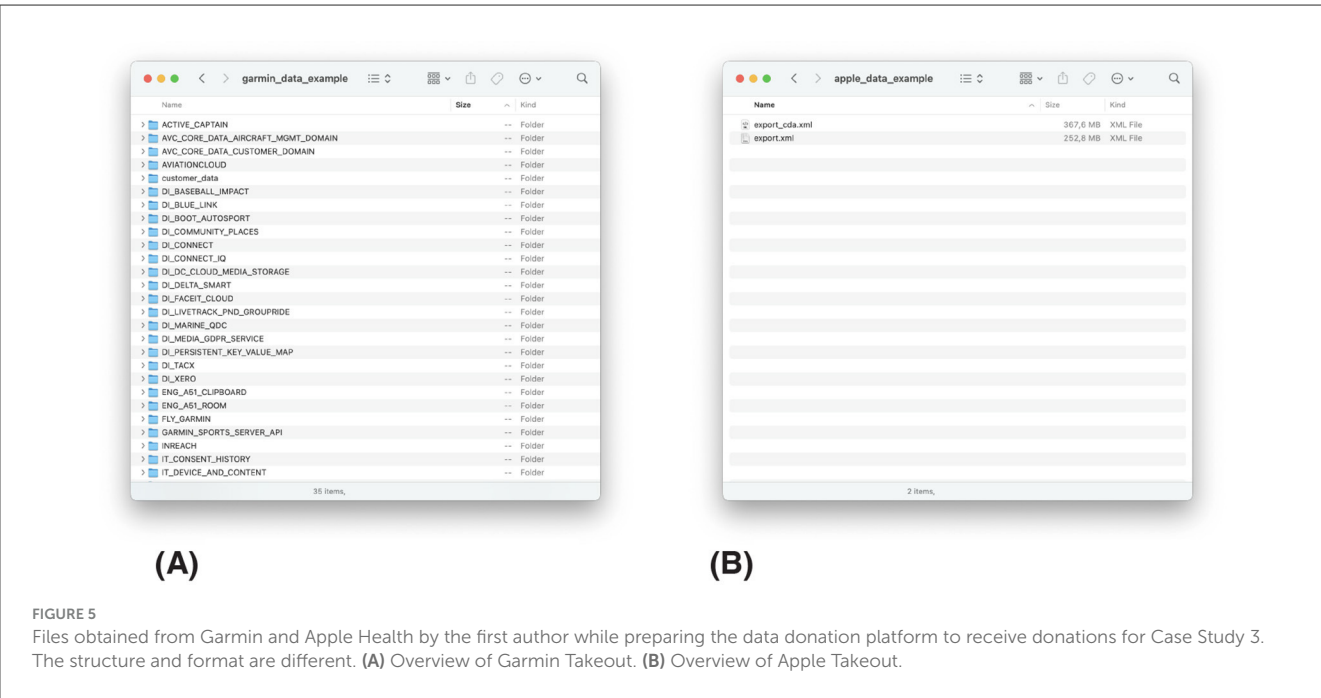
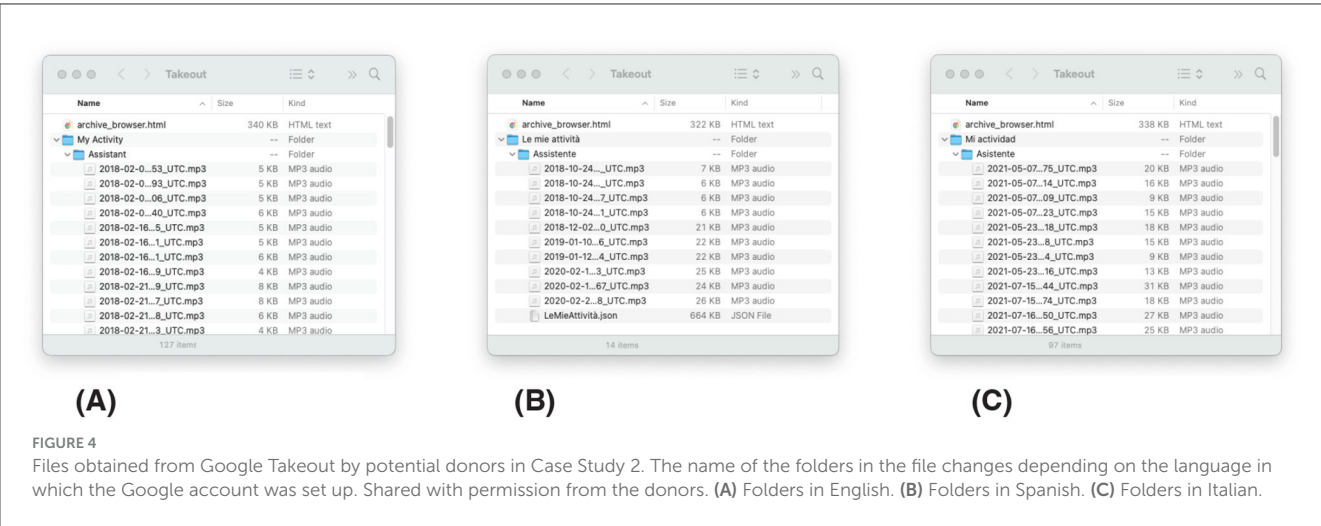
Considerations What is the most appropriate source for the data? If it is an existing type of product(s) or service(s) consider the trade-off between the availability of potential donors and the labor and complexity going into the data donation platform or system. What is the context in which you and your potential participants are immersed? How could this context introduce variability to the data? Consider the possibility of having changing data structures over time and based on different languages and device configurations. How can this introduce messiness and variations to the data? Consider these contexts as you become familiar with the way data is returned by the specific data controller(s). Monitor the file(s) structure(s) over time so you can respond to unexpected changes in a timely manner.

5.3 Requesting the data: unexpected difficulties and unpredictable time-frames

When requesting a copy of their data potential donors might face unforeseen difficulties such as downloading the wrong files, files in the wrong format, forgetting their login credentials, or not completing the data download process within the time frame provided by the data controllers. Sometimes the complexity of this transaction discourages potential donors from completing the data donation process.

For instance, in the second Case Study, some donors obtained empty files or files with less data than expected due to the introduction of the GDPR in 2018. If they did not opt-in to voice data collection after Google's policy change in 2018, their data only contained speech records before that date, when some donors weren't even Google Assistant users. When the files were empty they could not upload them to the platform, which resulted in them expressing frustration with the whole process and giving up. While in the third Case Study, when the data was ready for download Garmin sent an email with a link that expired in 3 days. Some donors forgot to download their data within that time frame while others only opened the email after 3 days. To continue with their donation they would have to request their data again, and hopefully remember to download it or open the email within 3 days. In both cases, I attempted to schedule emails through the data donation platform reminding potential donors to complete the process, yet the unpredictability of the processing time of each request made these emails often obsolete—generally, these requests can take from a few hours to up to 30 days⁸. Therefore, it is critical to remain

⁸ Once people complete the Google Takeout process they see the following message: "Google is creating a copy of files from My Activity. This process can take a long time (possibly hours or days) to complete".



accessible and available to support potential donors in navigating these procedures.

Considerations Making a data portability request from a data controller might be a new and confusing process for data donors; support them as much as possible. Are the instructions you provide clear enough? Can others follow the process only with the instructions? Do donors know how to contact you if necessary? Become familiar with the process and anticipate potential challenges, such as a donor forgetting

their login credentials, by reminding them to have their login credentials ready. Anticipate the potential challenges donors might face when requesting their data from different devices, different configurations, and various geographical locations, among others.

5.4 Exploring and curating the data: sensitive datasets and best practices

In Case Studies 1 and 2 the Takeout processes enabled donors to retrieve only their menstrual cycle logs and speech records respectively. In Case Study 3 donors, specifically Garmin users (Figure 5A), obtained a large folder including different types of

Once people complete the Garmin Takeout process they see the following message: "It takes ~ 48 h to prepare files, but depending on the number of requests being processed and the amount of data associated with your profile, this could take up to 30 days."

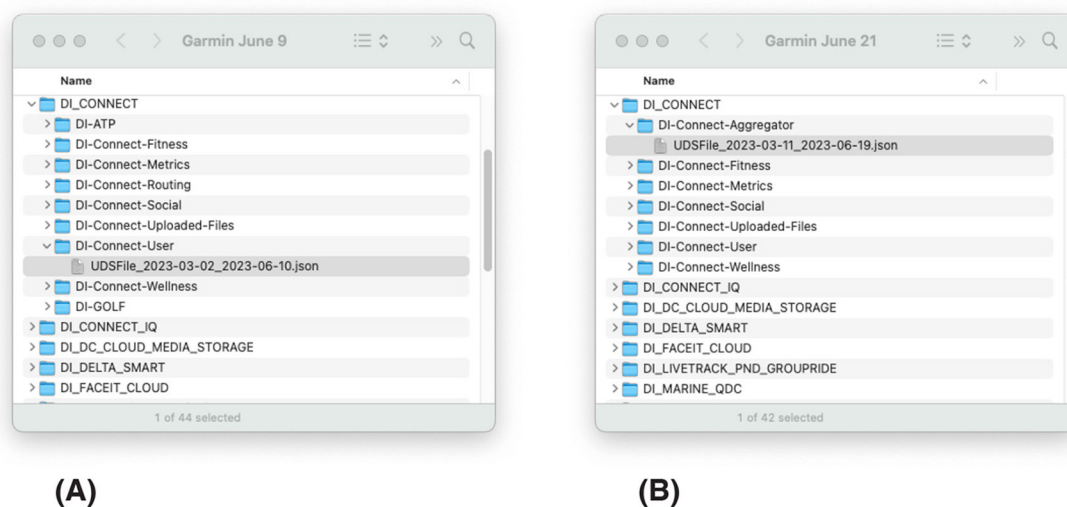


FIGURE 6

Files obtained from Garmin by the first author while receiving donations for Case Study 3. The structure of the folders in the file changes, changing the location of the JSON file titled UDSFile, containing the heart rate data. (A) Garmin Takeout Structure on June 9, 2023, (B) Garmin Takeout Structure on June 21, 2023.

personal information (e.g., photos, followers, interactions with followers). These folders were processed locally on the devices of potential donors; still, the vast and sensitive information available could potentially be extracted without donors' knowledge or permission and misused. This is where research(ers) ethics, and principles such as data minimization, play a fundamental role. It is critical to be transparent with the information being requested and collected and to honor donor's preferences and choices. In addition to adhering to the principle of data minimization, the data donation journey in all three Case Studies aimed to encourage donors to exercise their data sovereignty and make meaningful choices about what data to donate. These choices varied incrementally: (1) What types of data? With almost half of the donors excluding specific types in Case Study 1; (2) What data points? With the majority of the donors donating all their data in Case Study 2; (3) What types and points of data? For how long? With the majority of the donors donating all their data within the last 6 months in Case Study 3. This poses two open questions: What choices are meaningful choices? How to best encourage and support donors in exploring and curating their data? How to balance meaningful choices with obtaining meaningful and usable research data? Additionally, it reinforces the importance of research(ers) ethics and data minimization when donors are highly motivated to contribute to scientific research.

Considerations What data is necessary to answer your research question? How is this communicated to donors? What choices do they have? How do their choices reflect on the data you might and might not have? Recognize the sensitive nature of data and support donors to define clear boundaries about what they do and do not want to share over time. Respect these boundaries.

5.5 Contextualizing the data: discovering and introducing the unexpected

To support donors in exploring and contextualizing their data I conducted individual sessions in-person and online. This format allowed for high flexibility and adaptability to donors' preferences and specific situations such as the COVID-19 pandemic. During these sessions, I relied on highly personal interactive data visualizations (Figure 3) that I crafted for each individual donor. One important limitation is that crafting these visualizations is time-consuming and scales poorly. In Case Study 1, I was not able to carry out all the sessions because I did not have enough time.

Additionally, in all three Case Studies, when exploring their visualizations donors expressed feelings of discomfort from discovering something unexpected in their data or seeing it through other lenses. I reiterated to donors that they could withdraw their donations at any time; yet, none did. Among the unexpected, in Case Study 2 donors found the voices of their friends, family, and neighbors as part of their data. For them, this felt like a violation of their friends, family, and neighbors' privacy. It poses questions on the relational aspect of the data and informed consent. If my mom uses my voice assistant once, should she consent to me donating my voice assistant data—which includes her one recording—too? Should I remove her recording from the data that I decide to donate? Moreover, the process of exploring and conceptualizing the data was also insightful and exciting for donors. So much so, that in Case Study 3 donors had expectations about what they wanted to learn from their data and even reached out to me inquiring about the possibility of including data collected through different means (e.g., personal spreadsheet, training diaries). The exploration and contextualization promote better relationships between donors and their data and thus I consider it important to be flexible and

support donors in gaining as much as possible from this process. In this case, I adapted the visualization (Figure 3C) I used during the session to include the additional data provided by the donor. However, this required additional labor on my side in an already time-consuming process.

Considerations What is important that donors learn from their data? How can you best support them to explore and interpret their data? How can this process empower them to reflect upon their decisions and reassess their participation if necessary? Consider designing simple visualizations and leveraging the temporal dimension of the data to anchor it with the donors' lived and embodied experiences. Visualizing or representing the data is time-consuming, take the time. This process might result in valuable questions and even assumptions. Note them and bring them up for discussion during the session.

6 Practical recommendations: a checklist

In this section, I summarize the challenges and considerations described above into a checklist of practical aspects to consider when applying data donation—these are mapped into the five phases of the Sensitive Data Donation Method.

1. Phase 1: identify, prepare, and communicate

- (a) Identify what data is relevant to answer your research question. How can you and others access it?
- (b) Familiarize with the process of requesting and obtaining a copy of the data and the way the data is structured. Make sure you account for potential contextual variations in the data.
- (c) Prepare the platform or system to parse and process the data. Test, test, test! Can people follow and understand the different steps?
- (d) Reach out to your institution's Human Research Ethics Committee and walk them through the process if possible. Create a step-by-step slideshow of the process. (How) Does it fit their forms? What questions arise? How can you best address them? Take the necessary action.
- (e) Clearly state your research goals and how you will use and process the data.
- (f) Reach out to existing communities who might be interested in contributing to your research.

2. Phase 2: request and receive data

- (a) Provide visual instructions where each step is visually demonstrated. Can other people follow them?
- (b) Test all the steps with different devices and browsers. Does everything still work?
- (c) Support potential donors if necessary. Do they know how to reach out to you? Take into account the time between the download request and receiving the data.

3. Phase 3: upload, explore, and curate data

- (a) Outline the process. Will this be asynchronous or do you need to be available for potential donors? If it is asynchronous, provide guiding questions to better support them in the exploration of their data.

- (b) Remind donors they can always choose what information to share and how to participate. How easy is it for them to choose?

4. Phase 4: transfer data

- (a) Verify if data has been adequately processed once donors transfer their data. Something went wrong? Be sure to reach out to donors if necessary.
- (b) Review the structure of the files that donors would request and receive. Can you see any variations? How can you best address them? Inform donors if necessary.

5. Phase 5: contextualize and further identify data

- (a) Represent the data in a way that can be relatable and easy to interpret. Harness the temporal dimension. What questions do you have about the data?
- (b) Let donors lead and come up with their own questions and interpretations about their data. Remind donors they can always choose to withdraw.

7 Limitations

In this section, I acknowledge several limitations of the approach described and applied in this paper. First, most of this research was conducted in The Netherlands and is subject to Human Research Ethics procedures specific to Dutch institutions. Thus, some of the challenges I encountered might not translate to other contexts or countries. Second, although I received data donations from donors primarily residing in the European Union and Latin America, the GDPR applies only to the population residing in the European Union. It is not a guarantee that data donation, as described in this paper, can be applied in other countries or contexts. Similarly, participation in the three Case Studies described is limited by several factors. For instance, owning a specific device, having used the device for a given time, having sufficient digital abilities to request and transfer the data, and having sufficient trust in the research, among others. Third, given the time and resource constraints of this research, which took part from 2020 to 2024, the data donation processes typically operated on short timescales. Finally, although I strived to explore different types of digital-trace data from diverse sources, the challenges that might emerge from other data types and characteristics remain unexplored.

8 Conclusion

In this paper, I presented a series of case studies of how I approached and applied sensitive data donation over the past 3 years. I briefly introduced the three case studies—each focused on a different type of data: (1) menstrual tracking logs, (2) speech records, and (3) physical activity logs, reflected on my experience, and discussed the unforeseen challenges I faced and the practical considerations I learned from these. Further, I proposed a set of recommendations that I hope can aid and inspire others seeking to approach and apply data donation.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Human Research Ethics Committee, at the Delft University of Technology. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

AGO: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

References

- Alizadeh, F., Jakobi, T., Boldt, J., and Stevens, G. (2019). "GDPR-reality check on the right to access data: claiming and investigating personally identifiable data from companies," in *Proceedings of Mensch und Computer 2019* (Hamburg Germany: ACM), 811–814. doi: 10.1145/3340764.3344913
- Bardzell, S., Bardzell, J., Forlizzi, J., Zimmerman, J., and Antanitis, J. (2012). *Critical Design and Critical Theory: The Challenge of Designing for Provocation*. doi: 10.1145/2317956.2318001
- Boeschoten, L., Mendrik, A., Van Der Veen, E., Vloothuis, J., Hu, H., Voorvaart, R., et al. (2022). Privacy-preserving local analysis of digital trace data: a proof-of-concept. *Patterns* 3:100444. doi: 10.1016/j.patter.2022.100444
- Bowyer, A., Holt, J., Go Jefferies, J., Wilson, R., Kirk, D., and David Smeddinck, J. (2022). "Human-GDPR interaction: practical experiences of accessing personal data," in *CHI Conference on Human Factors in Computing Systems*, (New Orleans, LA, USA: ACM), 1–19. doi: 10.1145/3491102.3501947
- Breuer, J., Kmetty, Z., Haim, M., and Stier, S. (2023). User-centric approaches for collecting Facebook data in the 'post-API age': experiences from two studies and recommendations for future research. *Inf. Commun. Soc.* 26, 2649–2668. doi: 10.1080/1369118X.2022.2097015
- European Parliament. Directorate General for Parliamentary Research Services. (2022). *Governing Data and Artificial Intelligence for All: Models for Sustainable and Just Data Governance*. Publications Office, LU.
- Gómez Ortega, A., Bourgeois, J., Hutiri, W. T., and Kortuem, G. (2023). Beyond data transactions: a framework for meaningfully informed data donation. *AI Soc.* doi: 10.1007/s00146-023-01755-5
- Gómez Ortega, A., Bourgeois, J., and Kortuem, G. (2022). "Reconstructing intimate contexts through data donation: a case study in menstrual tracking technologies," in *Nordic Human-Computer Interaction Conference* (Aarhus Denmark: ACM), 1–12. doi: 10.1145/3546155.3546646
- Gómez Ortega, A., Bourgeois, J., and Kortuem, G. (2024a). Participation in data donation: co-creative, collaborative, and contributory engagements with athletes and their intimate data," in *Proceedings of the 2024 ACM Designing Interactive Systems Conference, DIS '24* (New York, NY, USA: Association for Computing Machinery), 2388–2402. doi: 10.1145/3643834.3661503
- Gómez Ortega, A., Bourgeois, J., and Kortuem, G. (2024b). "Sensitive data donation: a feminist reframing of data practices for intimate research contexts," in *Proceedings of the 2024 ACM Designing Interactive Systems Conference, DIS '24* (New York, NY, USA: Association for Computing Machinery), 2420–2434. doi: 10.1145/3643834.3661524
- Hummel, P., Braun, M., and Dabrock, P. (2019). "Data donations as exercises of sovereignty," in: *The Ethics of Medical Data Donation* (Cham Publisher: Springer International Publishing, New York). doi: 10.1007/978-3-030-04363-6_3
- Jones, K. H. (2019). "Incongruities and dilemmas in data donation: juggling our 1s and 0s," in *The Ethics of Medical Data Donation*, vol 137, eds. J. Krutzinna and L. Floridi (Springer International Publishing: Cham). doi: 10.1007/978-3-030-04363-6_5
- Koskinen, I. K., Zimmerman, J., Binder, T., Redstr, J., and Wensveen, S. (2011). *Design Research Through Practice: From the Lab, Field, and Showroom*. Morgan Kaufmann Publishers, Inc.: Waltham. doi: 10.1016/B978-0-12-385502-2.00006-7
- Nickel, P. (2019). "The ethics of uncertainty for data subjects," in *The Ethics of Medical Data Donation*, eds. P. Dabrock, M. Braun, and P. Hummel (Springer Verlag: New York), 55–74. doi: 10.1007/978-3-030-04363-6_4
- Ohme, J., and Araujo, T. (2022). Digital data donations: a quest for best practices. *Patterns* 3:100467. doi: 10.1016/j.patter.2022.100467
- Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., et al. (2024). Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. *Commun. Methods Meas.* 18, 124–141. doi: 10.1080/19312458.2023.2181319
- Ohme, J., Araujo, T., De Vreese, C. H., and Piotrowski, J. T. (2021). Mobile data donations: assessing self-report accuracy and sample biases with the iOS screen time function. *Mob. Media Commun.* 9, 293–313. doi: 10.1177/2050157920959106
- Pins, D., Jakobi, T., Stevens, G., Alizadeh, F., and Krer, J. (2022). Finding, getting and understanding: the user journey for the GDPR's right to access. *Behav. Inf. Technol.* 41, 2174–2200. doi: 10.1080/0144929X.2022.2074894
- Prainsack, B. (2019). Logged out: ownership, exclusion and public value in the digital data and information commons. *Big Data Soc.* 6:205395171982977. doi: 10.1177/2053951719829773
- Prainsack, B., El-Sayed, S., Forgó, N., Szoszkiewicz, U., and Baumer, P. (2022). Data solidarity: a blueprint for governing health futures. *Lancet Digital Health* 4, e773–e774. doi: 10.1016/S2589-7500(22)00189-3
- Razi, A., Alsoubai, A., Kim, S., Naher, N., Ali, S., Stringhini, G., et al. (2022). "Instagram data donation: a case study on collecting ecologically valid social media data for the purpose of adolescent online risk detection," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, (New Orleans, LA, USA: ACM), 1–9. doi: 10.1145/3491101.3503569
- Shirk, J. L., Ballard, H. L., Wilderman, C. C., Phillips, T., Wiggins, A., Jordan, R., et al. (2012). Public participation in scientific research: a framework for deliberate design. *Ecol. Soc.* 17:art29. doi: 10.5751/ES-04705-170229

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Skatova, A., and Goulding, J. (2019). Psychology of personal data donation. *PLoS ONE*. 14:e0224240. doi: 10.1371/journal.pone.0224240
- Stappers, P. J., and Giaccardi, E. (2014). *Research Through Design*. Interaction Design Foundation-IxDF. Available online at: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/research-through-design?srltid=AfmBOorkcGiLAXfpYdwMkOGis8_gZW90KB7xcYyrrawfbBxeRjxVCU1T
- Strotbaum, V., Pobiruchin, M., Schreiweis, B., Wiesner, M., and Strahwald, B. (2019). Your data is gold—data donation for better healthcare? *Inf. Technol.* 61, 219–229. doi: 10.1515/itit-2019-0024
- Tolmie, P., Crabtree, A., Rodden, T., Colley, J., and Luger, E. (2016). “This has to be the cats: personal data legibility in networked sensing systems,” in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing*, (San Francisco, CA, USA: ACM), 491–502. doi: 10.1145/2818048.2819992
- Van Driel, I. I., Giachanou, A., Pouwels, J. L., Boeschoten, L., Beyens, I., and Valkenburg, P. M. (2022). Promises and pitfalls of social media data donations. *Commun. Methods Meas.* 16, 266–282. doi: 10.1080/19312458.2022.2109608
- Wiedermann, M., Rose, A. H., Maier, B. F., Kolb, J. J., Hinrichs, D., and Brockmann, D. (2022). Evidence for positive long- and short-term effects of vaccinations against COVID-19 in wearable sensor metrics—Insights from the German Corona Data Donation Project. *arXiv:2204.02846*. doi: 10.1093/pnasnexus/pgad223
- Wiese, J., Das, S., Hong, J. I., and Zimmerman, J. (2017). Evolving the ecosystem of personal behavioral data. *Hum. Comput. Interact.* 32, 447–510. doi: 10.1080/07370024.2017.1295857
- Zannettou, S., Nemeth, O.-N., Ayalon, O., Goetzen, A., Gummadi, K. P., Redmiles, E. M., et al. (2024). Analyzing user engagement with tiktok’s short format video recommendations using data donations. *arXiv:2301.04945*. doi: 10.1145/3613904.3642433
- Zimmerman, J., Forlizzi, J., and Evenson, S. (2007). “Research through design as a method for interaction design research in HCI,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (San Jose, CA, USA: ACM), 493–502. doi: 10.1145/1240624.1240704

Appendix

Appendix: personal research journal template

Research journal entry template

To be filled by the principal researcher while conducting a case study, including the data donation platform development and participant recruitment.

***Day, Month, Year* — What did we do? What happened? Reflections on process and/or outcomes.**

Example

13/06/2023 — A potential data donor forgot to download her data within the 3 days given by Garmin and the download link expired. She had to restart the process and request her data again. This introduced an unexpected delay to her data donation process that we failed to account for when providing the instructions. In addition, illustrates a dependency on the data controllers.